Human loss-of-function variants suggest that partial LRRK2 inhibition is a safe therapeutic strategy for Parkinson's disease

Nicola Whiffin[1,2,3]*, Irina M. Armean[3,4]*, Aaron Kleinman[5]*, Jamie L. Marshall[3], Eric V. Minikel[3], Konrad J. Karczewski[3,4], Beryl B. Cummings[3,4,6], Laurent Francioli[3,4], Kristen Laricchia[3,4], Qingbo Wang[3,4,7], Anna Guan[5], Babak Alipanahi[5†], Peter Morrison[5], Marco A.S. Baptista[8], Kalpana M. Merchant[8], Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, James S. Ware[1,2,3], Aki S. Havulinna[9,10], Bozenna Iliadou[11], Jung-Jin Lee[12], Girish N. Nadkarni[13,14], Cole Whiteman[15], the 23andMe Research Team, Mark Daly[3,4,16], Tõnu Esko[3,17], Christina Hultman[11,14], Ruth J.F. Loos[13,18], Lili Milani[17], Aarno Palotie[4,9,16], Carlos Pato[15], Michele Pato[15], Danish Saleheen[12,19,20], Patrick F. Sullivan[11,21], Jessica Alföldi[3,4], Paul Cannon[5]**, Daniel G. MacArthur[3,4]**

[1]National Heart & Lung Institute and MRC London Institute of Medical Sciences, Imperial College London, London UK
[2]Cardiovascular Research Centre, Royal Brompton & Harefield Hospitals NHS Trust, London UK
[3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA
[4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA
[5]23andMe, Inc.
[6]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, 02115, USA
[7]Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA 02115, USA
[8]Michael J. Fox Foundation
[9]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland
[10]National Institute for Health and Welfare, 00271, Helsinki, Finland.
[11]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[12]Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA
[13]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY
[14]Department of Medicine, Icahn School of Medicine at Mount Sinai
[15]Department of Psychiatry and the Behavioral Sciences, State University of New York, Downstate Medical Center, New York, New York.
[16]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA
[17]Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia
[18]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY
[19]Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA
[20]Center for Non-Communicable Diseases, Karachi, Pakistan

[21]Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA

*These authors contributed equally to this work
**These authors contributed equally to this work
†Current affiliation: Google, Inc.

Correspondence should be addressed to Daniel G MacArthur danmac@broadinstitute.org or Paul Cannon pcannon@23andme.com

Abstract

Human genetic variants causing loss-of-function (LoF) of protein-coding genes provide natural *in vivo* models of gene inactivation, which are powerful indicators of gene function and the potential toxicity of therapeutic inhibitors targeting these genes[1,2]. Gain-of-kinase-function variants in *LRRK2* are known to significantly increase the risk of Parkinson's disease[3,4], suggesting that inhibition of LRRK2 kinase activity is a promising therapeutic strategy. Whilst preclinical studies in model organisms have raised some on-target toxicity concerns[5–8], the biological consequences of LRRK2 inhibition have not been well characterized in humans. Here we systematically analyse LoF variants in *LRRK2* observed across 141,456 individuals sequenced in the Genome Aggregation Database (gnomAD)[9] and over 4 million participants in the 23andMe genotyped dataset, to assess their impact at a molecular and phenotypic level. After thorough variant curation, we identify 1,358 individuals with high-confidence predicted LoF variants in *LRRK2*, several with experimental validation. We show that heterozygous LoF of *LRRK2* reduces LRRK2 protein level by ~50% but is not associated with reduced life expectancy, or with any specific phenotype or disease state. These data suggest that therapeutics that downregulate LRRK2 levels or kinase activity by up to 50% are unlikely to have major on-target safety liabilities. Our results demonstrate the value of large scale genomic databases and phenotyping of human LoF carriers for target validation in drug discovery.

Main text

Novel therapeutic strategies are desperately needed in Parkinson's disease (PD), one of the most common age-related neurological diseases, which affects about 1% of people over 60[10,11]. Around 30% of familial and 3-5% of sporadic PD cases have been linked to a genetic cause[12]. *LRRK2* variants account for a significant fraction of cases, including high-penetrance variants[13], moderately penetrant variants such as G2019S[14], and risk factors identified in genome-wide association studies[15]. There is emerging consensus that G2019S as well as other disease-causing missense variants in *LRRK2* result in increased kinase activity[3,4,16,17]. Conversely, predicted loss-of-function (pLoF) variants in *LRRK2* are not associated with PD risk[18]. LRRK2 has therefore become a prominent drug target, with multiple LRRK2 kinase inhibitors and suppressors[19] in development as disease-modifying treatments for PD[20–22].

Despite these promising indications, concerns remain about the potential toxicity of LRRK2 inhibitors. These mainly arise from preclinical studies, where homozygous knockouts of *LRRK2* in mice, and high dose toxicology studies of LRRK2 kinase inhibitors in rats and primates, have shown abnormal phenotypes in lung, kidney and liver[5–8]. While model organisms have been invaluable in understanding the function of LRRK2, they also have important limitations as exemplified by the inconsistencies in the phenotypic consequences of reduced LRRK2 activity seen among yeast, fruit flies, worm, mice, rats and non-human primates[23]. Complementary data from natural human knockouts are critical for understanding both gene function and the potential consequences of a long-term reduction in LRRK2 kinase activity in humans.

Large-scale human genetics is an increasingly powerful source of data for the discovery and validation of therapeutic targets in humans[1]. pLoF variants, predicted to largely or entirely abolish the function of affected alleles, are a particularly informative class of genetic variation. Such variants are natural models for life-long organism-wide inhibition of the target gene, and can provide information about both the efficacy and safety of a candidate target[2,24–27]. However, pLoF variants tend to be rare in human populations[28], and are also enriched for both sequencing and annotation artefacts[29]. As such, leveraging pLoF variation in drug target assessment typically requires very large collections of genetically and phenotypically characterized individuals, combined with deep curation of the target gene and candidate variants[30].

We identified *LRRK2* pLoF variants, and characterized the associated phenotypic changes, in two large cohorts of genetically characterized individuals. Firstly, we annotated *LRRK2* pLoF variants in the gnomAD 2.1 data set, which contains 125,748 exomes and 15,708 genomes from unrelated individuals sequenced as part of various disease-specific and population genetic studies, and is described in detail in an accompanying publication[9]. We identified 633 individuals in gnomAD with 123 unique candidate *LRRK2* LoF variants, a combined carrier frequency of 0.45%. All variants were

observed only in the heterozygous state. Compared to the spectrum observed across all genes, *LRRK2* is not significantly depleted for LoF variants (loss-of-function observed/expected upper bound fraction (LOEUF)[9]=0.64).

To ensure that the 123 identified variants are indeed expected to cause LoF, we manually curated them to remove those of low-quality or with annotation errors (Figure 1a; Supplementary Table 1). We removed twelve variants identified either as low-confidence by the Loss-of-function Transcript Effect Estimator (LOFTEE; 6 variants in 285 individuals)[9], or manually curated as low-quality or unlikely to cause true LoF (6 variants in 93 individuals; see methods). This included one variant (12-40626187-T-C), found in 77 gnomAD carriers, that was identified in an individual with RNA-seq data in the Genotype-Tissue Expression (GTEx) project[31]. Assessing the read distribution of a linked heterozygous variant in this individual showed convincingly that the variant has no discernible effect on transcript splicing (Supplementary Figure 1).

This analysis reaffirms the importance of thorough curation of candidate LoF variants[30], and results in a final gnomAD dataset of 255 individuals with 111 unique high-quality pLoF variants (Figure 1a), with an overall carrier frequency of 0.19%. A subset of 25 samples with 19 unique *LRRK2* LoF variants with DNA available were all successfully validated by Sanger sequencing (Supplementary Table 2).

Secondly, we examined *LRRK2* pLoF variants in over four million consented research participants from the personal genetics company 23andMe, Inc. Participants have been genotyped on a variety of platforms, and imputed against a reference panel comprising 56.5 million variants from 1000 Genomes phase 3[32], and UK10K[33]. Eight putative *LRRK2* LoF variants were identified in the 23andMe cohort (LOFTEE high-confidence). One variant was the same splice variant that was excluded from gnomAD after inspection of GTEx data (Supplementary Figure 1) and was similarly excluded here. Carrier status of the remaining seven variants was confirmed using Sanger sequencing. The resulting cohort consisted of 749 individuals, each of whom is a Sanger-confirmed carrier for one of three LoF variants (Supplementary Table 3). Subsequent expansion of the 23andMe dataset identified 354 additional putative carriers of rs183902574. Across the two most frequent pLoF alleles we observed an extremely small number (<5) of sequence-confirmed homozygotes; however, given the very small number of observations, we can make no robust inference except that homozygous inactivation of LRRK2 appears to be compatible with life. For the remainder of this manuscript we focus on heterozygous pLoF carriers.

The combined datasets from gnomAD and 23andMe describe a total of 1,358 carrier individuals with 111 unique *LRRK2* pLoF variants. These variants are found across all major continental populations (Figure 1b; Supplementary Figure 2) and show neither any obvious clustering along the

length of the LRRK2 protein, nor specific enrichment or depletion in any of the known annotated protein domains (Chi-square *P*=0.23; Figure 1c), consistent with signatures of true LoF[30].

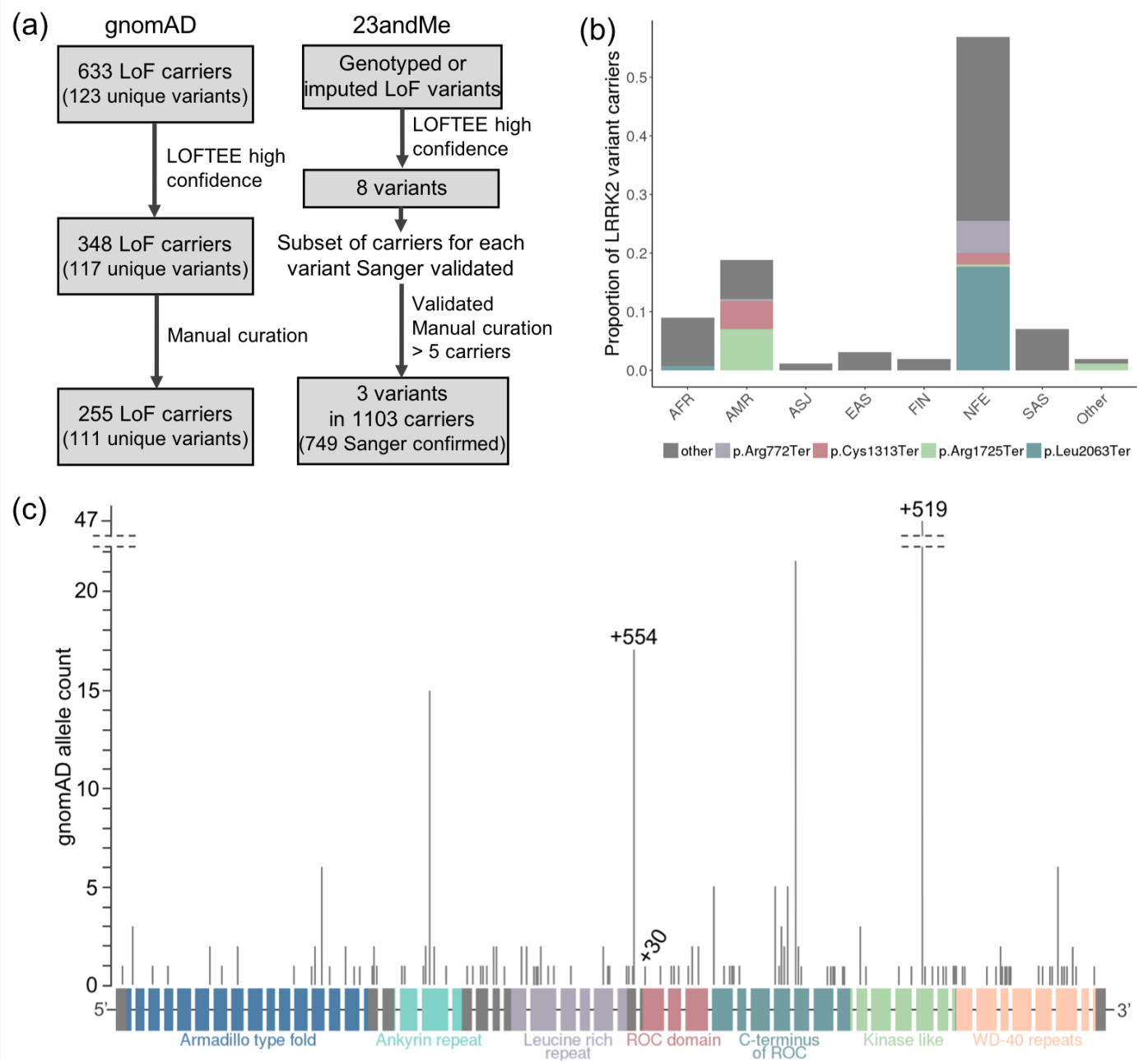


Figure 1: Annotation and curation of candidate *LRRK2* LoF variants. a) Flow chart showing the variant filtering and curation of candidate *LRRK2* LoF variants in the gnomAD and 23andMe cohorts. Of the 1,103 carriers identified in 23andMe, 749 were confirmed by Sanger sequencing with the remainder untested. b) The ancestry distribution of *LRRK2* pLoF variant carriers in gnomAD (AFR - African/African American; AMR - American/Latino; ASJ - Ashkenazi Jewish; EAS - East Asian; FIN - Finnish; NFE - Non-Finnish European; SAS - South Asian). pLoF variants seen more than 10 times appear in colour with remaining variants in grey. *LRRK2* pLoF variants are mostly individually extremely rare (less than

1/10,000 carrier frequency), with the exception of two nonsense variants almost exclusively restricted to the admixed American/Latino population (Cys1313Ter and Arg1725Ter), and two largely Non-Finnish European-specific variants (Leu2063Ter and Arg772Ter). All variant protein descriptions are with respect to ENSP00000298910. c) Schematic of the LRRK2 gene with pLoF variants marked by position, with the height of the marker corresponding to allele count in gnomAD. The 51 exons are shown as rectangles coloured by protein domain, with the remaining exons in grey. The three variants genotyped in the 23andMe cohort are annotated with their sample count in black text.

To confirm that *LRRK2* pLoF variants lead to a reduction in LRRK2 protein levels, we analysed total protein lysates from cell lines with three unique pLoF variants by immunoblotting. We obtained lymphoblastoid cell lines (LCLs) from two families with naturally occuring heterozygous LoF variants, and a third variant was CRISPR/Cas9 engineered into embryonic stem cells (Supplementary Figure 3), which were differentiated into cardiomyocytes. In all instances, LRRK2 protein levels were visibly reduced compared to non-carrier controls (Figure 2).

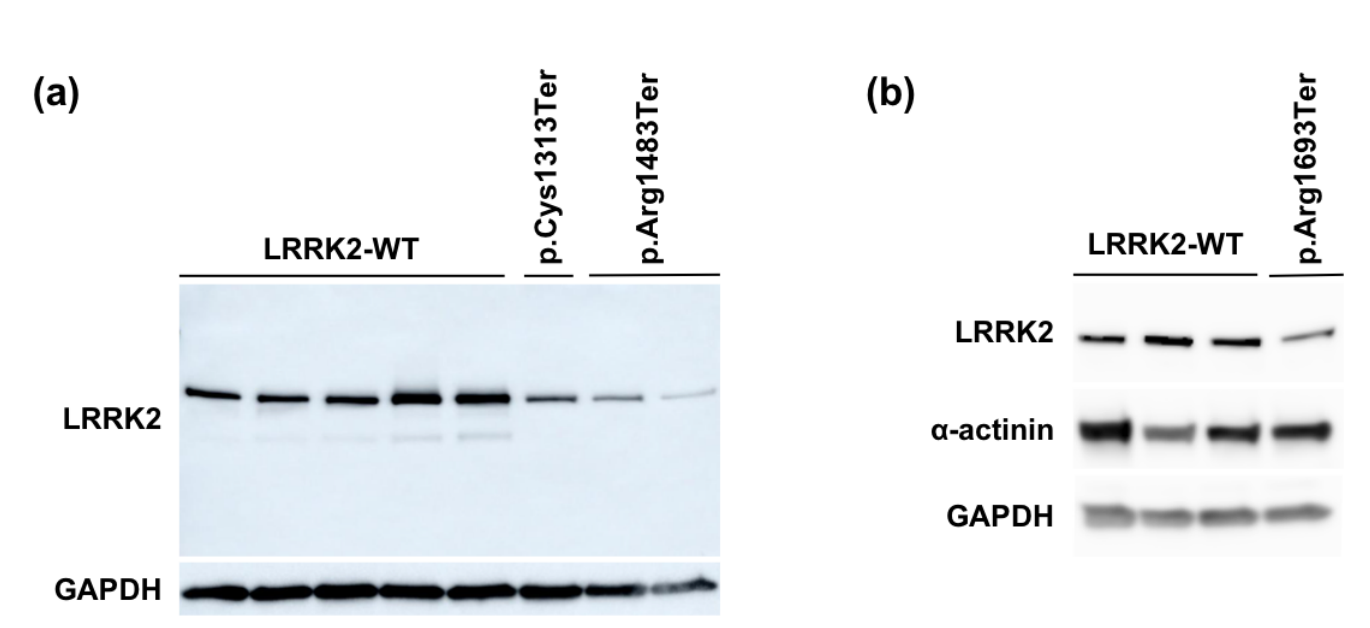


Figure 2: *LRRK2* LoF heterozygotes have reduced LRRK2 protein compared to cells harboring no LoF variants. a) Immunoblot of LRRK2 and loading control GAPDH on lymphoblastoid cell lines from five individuals harboring no pLoF variants (LRRK2-WT) and three individuals harboring a heterozygous LoF variant (Cys1313Ter - 12-40699748-T-A; Arg1483Ter - 12-40704362-C-T). b) Immunoblot of LRRK2, α-actinin (specific to muscle), and GAPDH on three control lines and one CRISPR/Cas9 engineered LRRK2 heterozygous line of cardiomyocytes differentiated from ESCs (Arg1693Ter - 12-40714897-C-T). All variant protein descriptions are with respect to ENSP00000298910.

We next sought to determine whether lifelong lowering of LRRK2 protein levels through LoF results in an apparent reduction in life-span. We found no significant difference between the age distribution of *LRRK2* pLoF variant carriers and non-carriers in either the gnomAD or 23andMe datasets (Kolmogorov-Smirnov *P*=0.085 and 0.46 respectively; Figure 3a), suggesting no major impact on lifespan.

For a subset of studies within gnomAD, phenotype data is available from study or national biobank questionnaires or from linked electronic health records (for full details see online methods). We manually reviewed these records for all 60 of the 255 gnomAD *LRRK2* LoF carriers for whom data were available, and recorded any phenotypes affecting the lung, liver, kidney, cardiovascular system, nervous system, immunity, and cancer (Supplementary Table 4). We did not find an over-representation of any particular phenotype or phenotype category in *LRRK2* LoF carriers. In addition, for the 12 *LRRK2* carriers from the Pakistan Risk of Myocardial Infarction Study (PROMIS) we observed no significant association between *LRRK2* carrier status and 77 blood biomarkers, including creatinine, a widely-used marker of kidney function (Figure 3b; Supplementary Figure 4; Supplementary Table 5).

The 23andMe dataset includes self-reported data for thousands of phenotypes across a diverse range of categories. We performed a phenome-wide association study on *LRRK2* LoF carriers compared to non-carriers and found no significant association between any individual phenotype and carrier status (Figure 3c). In particular, we found no significant associations with any lung, liver or kidney phenotypes (Supplementary Table 6).

Our results indicate that approximately one in every 550 humans is heterozygous for a pLoF variant in *LRRK2*, resulting in a systemic lifelong decrease in LRRK2 protein levels to approximately 50% of wild-type levels, and that this partial inhibition has no discernible effect on survival or health. These results suggest that drugs that reduce LRRK2 kinase activity are unlikely to show toxicity due to on-target effects. This observation is particularly important given that individuals at risk of PD will need to take therapeutics chronically, likely soon after clinical diagnosis and extending for decades. While initial phase I studies of the safety of LRRK2 kinase inhibitors have shown a promising safety profile[22,34], these are not yet able to address the question of long-term, on-target pharmacology-related safety profiles.

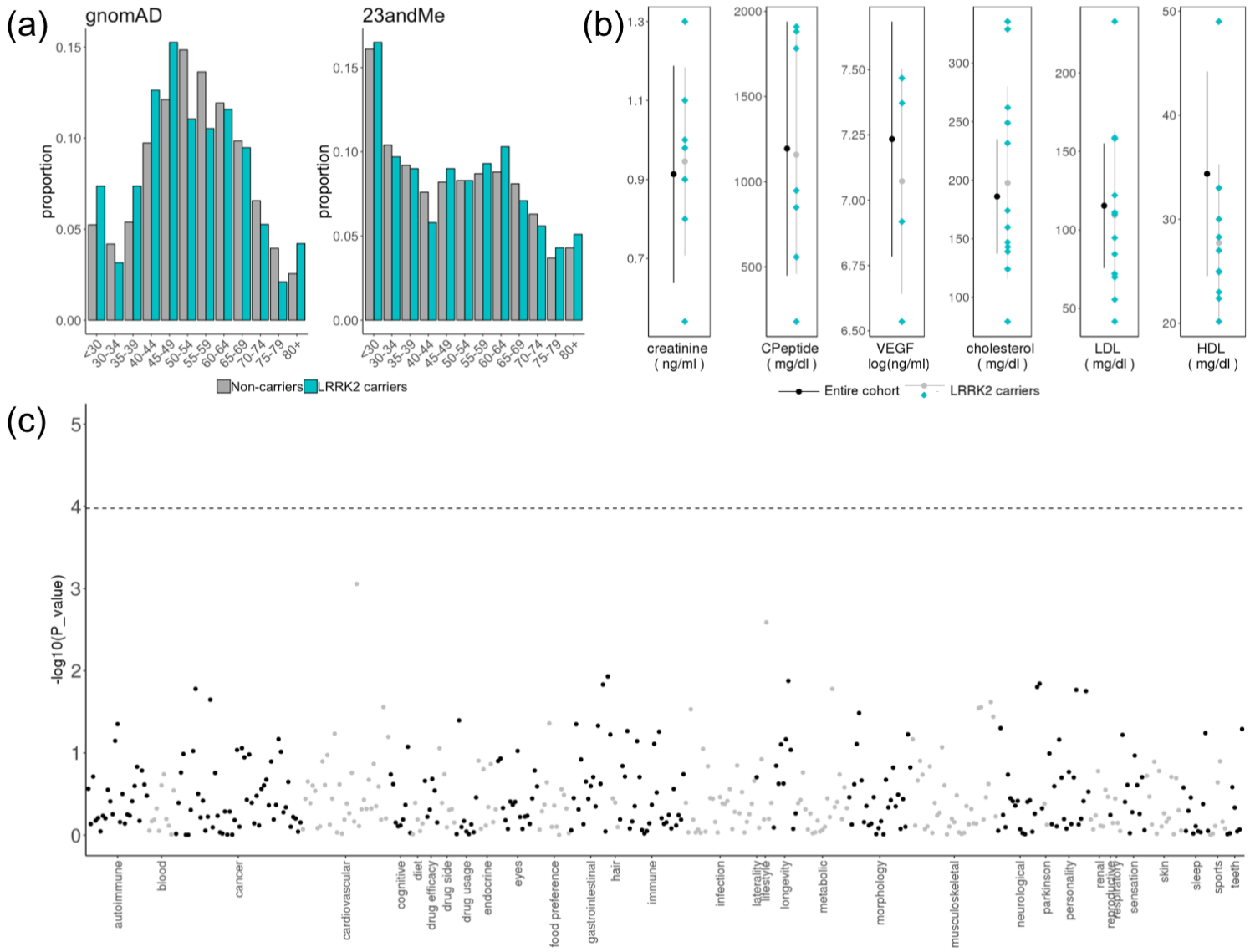


Figure 3: *LRRK2* LoF variants are not associated with reduced lifespan, are not outliers for blood biomarkers and are not associated with any adverse phenotypes. (a) The age distributions of *LRRK2* LoF carriers are not significantly different from those of non-carriers in both gnomAD and 23andMe. (b) Six candidate blood serum biomarkers in the pLoF carriers compared to the measurements in the 10,503 PROMIS parent cohort: cohort mean and standard deviations are shown in black (mean as a circle with +/- 1 SD as black lines) and the *LRRK2* LoF carriers as blue diamonds with the mean and standard deviations shown in grey. Biomarkers: serum creatinine ([mg/dl]) used as a biomarker for kidney failure; Vascular endothelial growth factor (VEGF) log [ng/ml] used as a cancer biomarker; C-Peptide [mg/dl], High Density Lipoprotein (HDL) [mg/dl], Cholesterol [mg/dl] and Low Density Lipoprotein [mg/dl] levels used as markers of cardiovascular disease. Plots for the remaining assayed biomarkers are shown in Supplementary Figure 4. (c) Manhattan plot of phenome-wide association study results for carriers of three *LRRK2* pLoF variants against non-carriers in the 23andMe cohort. Each point represents a distinct phenotype with these grouped into related categories (delineated by alternating black and grey points).

We acknowledge multiple limitations to our study. Firstly, we relied on heterogeneous phenotype data mostly derived from self-reported questionnaires. Secondly, both 23andMe and gnomAD record only age at recruitment, and participants are of a relatively young age compared to the typical age of onset for PD. Thirdly, a subset of the 23andMe individuals treated as non-carriers could be carriers of *LRRK2* pLoF variants not genotyped or imputed in this dataset. Finally, the low-frequency of naturally occuring pLoF variants in *LRRK2* results in a relatively small number of carriers which could be assessed for each biomarker and phenotype. Additionally, the rarity of pLoF variants in *LRRK2* prevents us from directly assessing whether *LRRK2* inhibition may reduce the incidence of PD, at current sample sizes.

Despite this, our findings illustrate the value of large genetically and phenotypically characterized cohorts, combined with thorough variant curation, in exploring the safety profile of candidate drug targets. Over the coming years, the availability of complete exome or genome sequence data for hundreds of thousands of individuals who are deeply phenotyped and/or available for genotype-based recontact studies, combined with deep curation and experimental validation of candidate pLoF variants, will provide powerful resources for therapeutic target validation as well as broader studies of the biology of human genes.

Acknowledgments

We thank the research participants and employees of 23andMe, Inc., and the research participants in gnomAD, for making this work possible. N.W. is supported by a Rosetrees and Stoneygate Imperial College Research Fellowship. E.V.M. is supported by NIH F31 AI22592. KJK was supported by NIGMS F32 GM115208. This work was supported by the Michael J. Fox Foundation for Parkinson's Research grant 12868, NIDDK U54DK105566, NIGMS R01GM104371, Wellcome Trust [107469/Z/15/Z]; Medical Research Council (UK); NIHR Royal Brompton Biomedical Research Unit; NIHR Imperial Biomedical Research Centre. T.E is supported by Estonian Research Council grant PUT1660. L.M. is supported by Estonian Research Council grant PRG184.

Conflicts of interest

A.K., B.A., A.G., P.M., P.C., and members of the 23andMe Research Team are current or former employees of 23andMe, Inc., and hold stock or stock options in 23andMe. D.G.M. is a founder with equity in Goldfinch Bio.

References

1. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
2. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* **12**, 581–594 (2013).
3. Greggio, E. *et al.* Kinase activity is required for the toxic effects of mutant LRRK2/dardarin. *Neurobiol. Dis.* **23**, 329–341 (2006).
4. West, A. B. *et al.* Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 16842–16847 (2005).
5. Andersen, M. A. *et al.* PFE-360-induced LRRK2 inhibition induces reversible, non-adverse renal changes in rats. *Toxicology* **395**, 15–22 (2018).
6. Fuji, R. N. *et al.* Effect of selective LRRK2 kinase inhibition on nonhuman primate lung. *Sci. Transl. Med.* **7**, 273ra15 (2015).
7. Baptista, M. A. S. *et al.* Loss of leucine-rich repeat kinase 2 (LRRK2) in rats leads to progressive abnormal phenotypes in peripheral organs. *PLoS One* **8**, e80705 (2013).
8. Hinkle, K. M. *et al.* LRRK2 knockout mice have an intact dopaminergic system but display alterations in exploratory and motor co-ordination behaviors. *Mol. Neurodegener.* **7**, 25 (2012).
9. Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210 (2019). doi:10.1101/531210
10. de Lau, L. M. L. & Breteler, M. M. B. Epidemiology of Parkinson's disease. *Lancet Neurol.* **5**, 525–535 (2006).

11. Polymeropoulos, M. H. *et al.* Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science* **274**, 1197–1199 (1996).

12. Klein, C. & Westenberger, A. Genetics of Parkinson's disease. *Cold Spring Harb. Perspect. Med.* **2**, a008888 (2012).

13. Zimprich, A. *et al.* Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **44**, 601–607 (2004).

14. Goldwurm, S. *et al.* Evaluation of LRRK2 G2019S penetrance: relevance for genetic counseling in Parkinson disease. *Neurology* **68**, 1141–1143 (2007).

15. Do, C. B. *et al.* Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet.* **7**, e1002141 (2011).

16. West, A. B. Achieving neuroprotection with LRRK2 kinase inhibitors in Parkinson disease. *Exp. Neurol.* **298**, 236–245 (2017).

17. Smith, W. W. *et al.* Kinase activity of mutant LRRK2 mediates neuronal toxicity. *Nat. Neurosci.* **9**, 1231–1233 (2006).

18. Blauwendraat, C. *et al.* Frequency of Loss of Function Variants in LRRK2 in Parkinson Disease. *JAMA Neurol.* **75**, 1416–1422 (2018).

19. Zhao, H. T. *et al.* LRRK2 Antisense Oligonucleotides Ameliorate α-Synuclein Inclusion Formation in a Parkinson's Disease Mouse Model. *Mol. Ther. Nucleic Acids* **8**, 508–519 (2017).

20. Di Maio, R. *et al.* LRRK2 activation in idiopathic Parkinson's disease. *Sci. Transl. Med.* **10**, (2018).

21. Chen, Z.-C. *et al.* Phosphorylation of amyloid precursor protein by mutant LRRK2 promotes AICD activity and neurotoxicity in Parkinson's disease. *Sci. Signal.* **10**, (2017).

22. Chen, J., Chen, Y. & Pu, J. Leucine-Rich Repeat Kinase 2 in Parkinson's Disease: Updated from

Pathogenesis to Potential Therapeutic Target. *Eur. Neurol.* **79**, 256–265 (2018).

23. Daniel, G. & Moore, D. J. Modeling LRRK2 Pathobiology in Parkinson's Disease: From Yeast to Rodents. in *Behavioral Neurobiology of Huntington's Disease and Parkinson's Disease* (eds. Nguyen, H. H. P. & Cenci, M. A.) 331–368 (Springer Berlin Heidelberg, 2015).
24. Cohen, J. C., Boerwinkle, E., Mosley, T. H., Jr & Hobbs, H. H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
25. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute *et al.* Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N. Engl. J. Med.* **371**, 22–31 (2014).
26. Myocardial Infarction Genetics Consortium Investigators *et al.* Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N. Engl. J. Med.* **371**, 2072–2082 (2014).
27. Minikel, E. V. *et al.* Quantifying prion disease penetrance using large population control cohorts. *Sci. Transl. Med.* **8**, 322ra9 (2016).
28. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
29. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
30. Minikel, E. V. *et al.* Evaluating potential drug targets through human loss-of-function genetic variation. *bioRxiv* 530881 (2019). doi:10.1101/530881
31. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
32. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

33. UK10K Consortium *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).

34. Troyer, M. Safety, Tolerability and Target Engagement Demonstrated in Phase 1 Study of LRRK2 Inhibitor DNL201 in Healthy Young and Elderly Adults. in *https://denalitherapeutics.com/uploads/documents/events/DNL201_MJFF_PDTx_Presentation.pdf* (MJFF Parkinson's Disease Therapeutics Conference, 2018).

35. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

36. Pato, M. T. *et al.* The genomic psychiatry cohort: partners in discovery. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **162B**, 306–312 (2013).

37. Saleheen, D. *et al.* Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **544**, 235–239 (2017).

38. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).

39. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address: douglas.ruderfer@vanderbilt.edu & Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**, 1705–1715.e16 (2018).

40. Genovese, G. *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).

41. Borodulin, K. *et al.* Cohort Profile: The National FINRISK Study. *Int. J. Epidemiol.* (2017). doi:10.1093/ije/dyx239

42. Leitsalu, L. *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).

43. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

44. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).

45. Wang, Q. & Ui-Tei, K. Computational Prediction of CRISPR/Cas9 Target Sites Reveals Potential Off-Target Risks in Human and Mouse. in *Genome Editing in Animals: Methods and Protocols* (ed. Hatada, I.) 43–53 (Springer New York, 2017).

46. Dale, R. K., Pedersen, B. S. & Quinlan, A. R. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).

47. Naito, Y., Hino, K., Bono, H. & Ui-Tei, K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* **31**, 1120–1123 (2015).

Supplementary Tables and Figures

Supplementary Table 1: Curated *LRRK2* LoF variants identified in gnomAD. All positions are with respect to GRCh37.

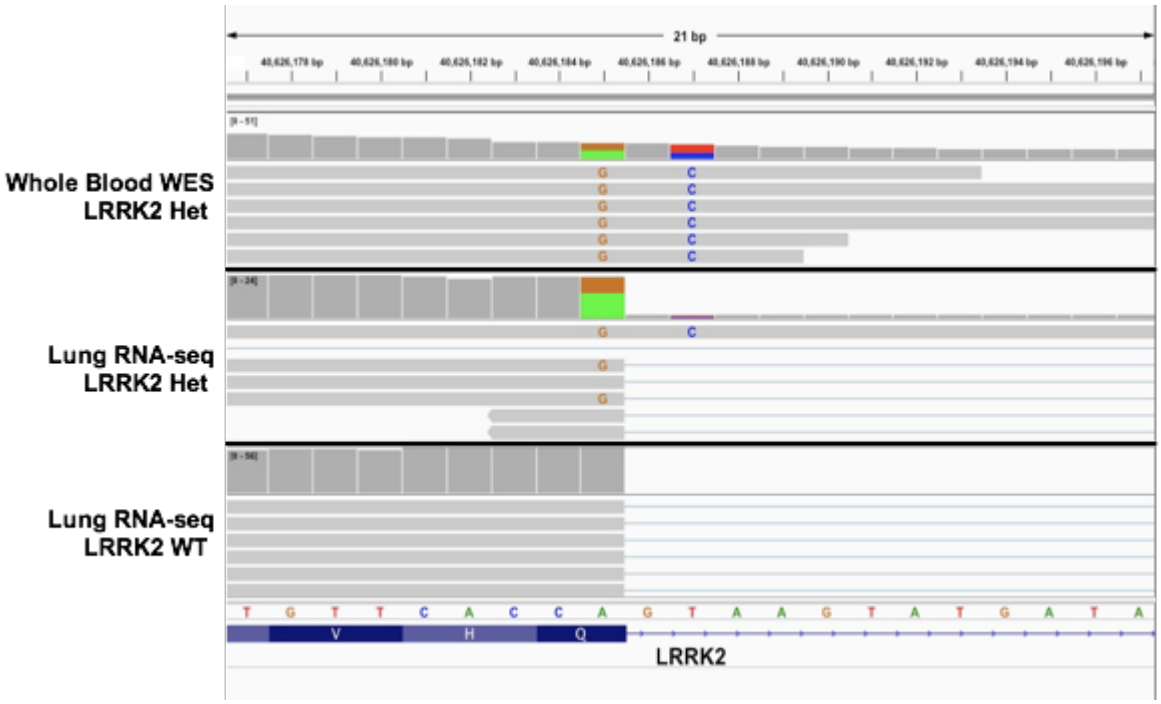
Supplementary Table 2: Details of variants and primers for Sanger validation of *LRRK2* LoF carriers. All positions are with respect to GRCh37.

Supplementary Table 3. Details of LRRK2 variants identified in the 23andMe cohort. All positions are with respect to GRCh37. All variant protein descriptions are annotated on ENSP00000298910.

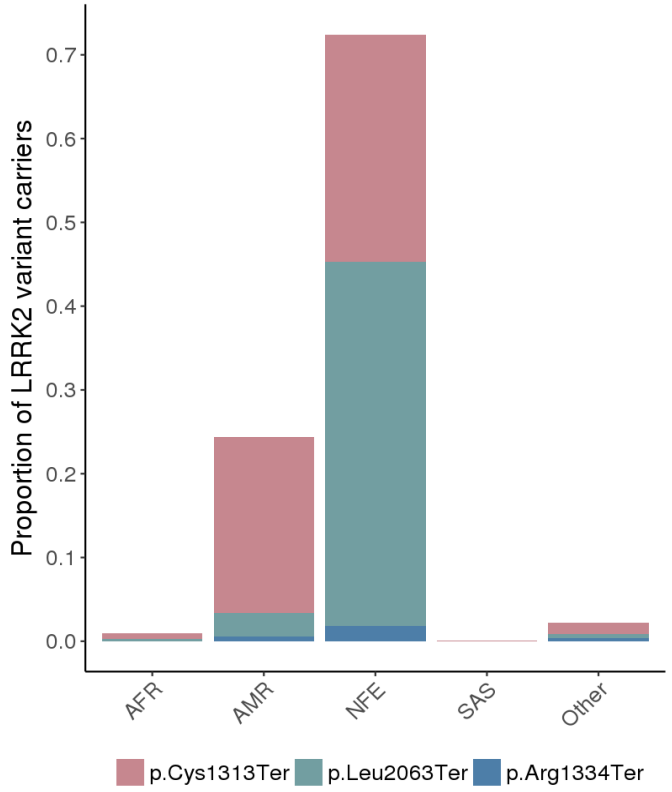
Supplementary Table 4. Summary of phenotype data from gnomAD *LRRK2* LoF carriers. All positions are with respect to GRCh37.

Supplementary Table 5: Results from a student's T-test comparing *LRRK2* pLoF carriers to non-carriers for 77 biomarkers (with data for >1 carrier) in the PROMIS study.

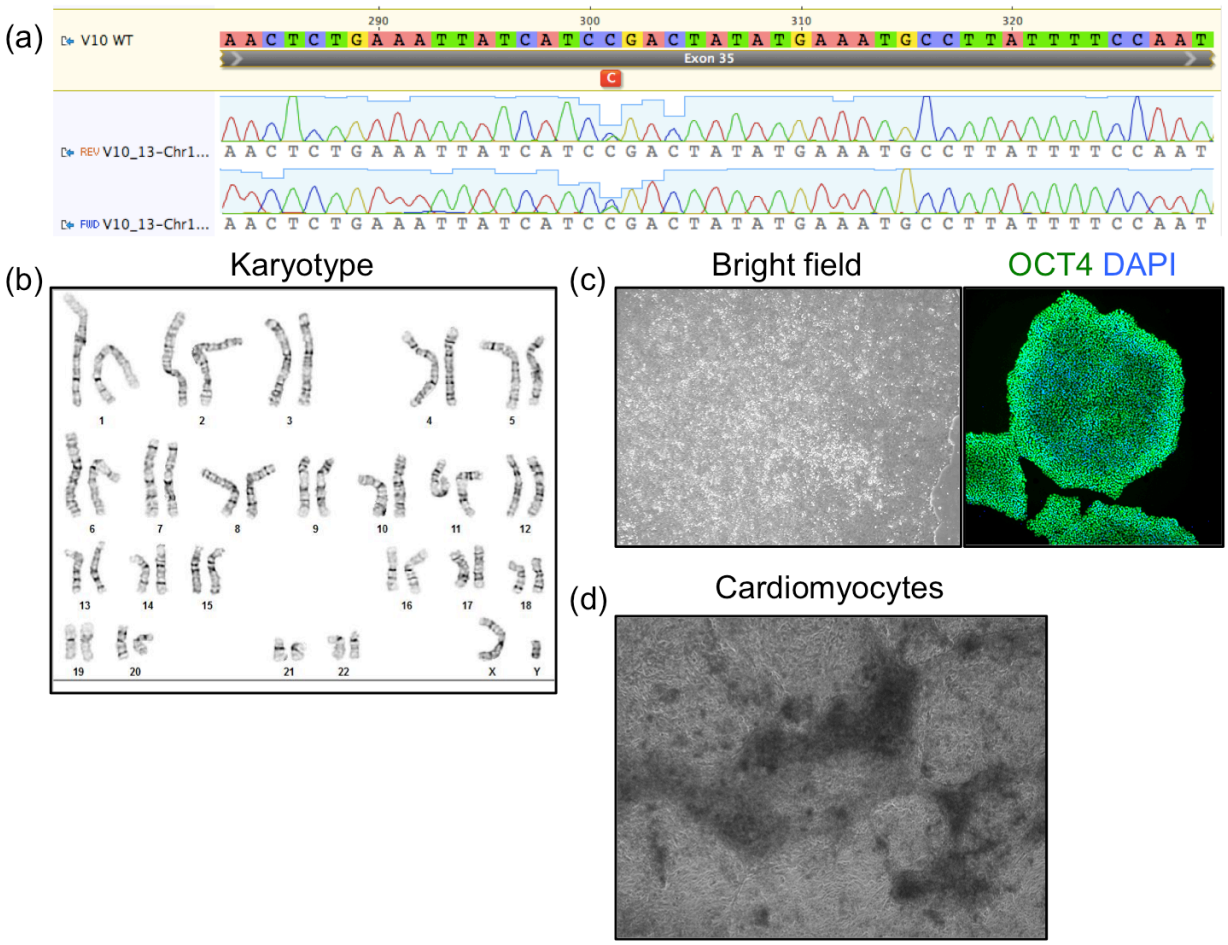
Supplementary Table 6: Association statistics for phenotypes specific to lung, kidney and liver (also plotted in Figure 3c). For privacy reasons, counts of less than five are rounded. None of the phenotypes associated with carrier status, even without correction ($p < 0.05$).



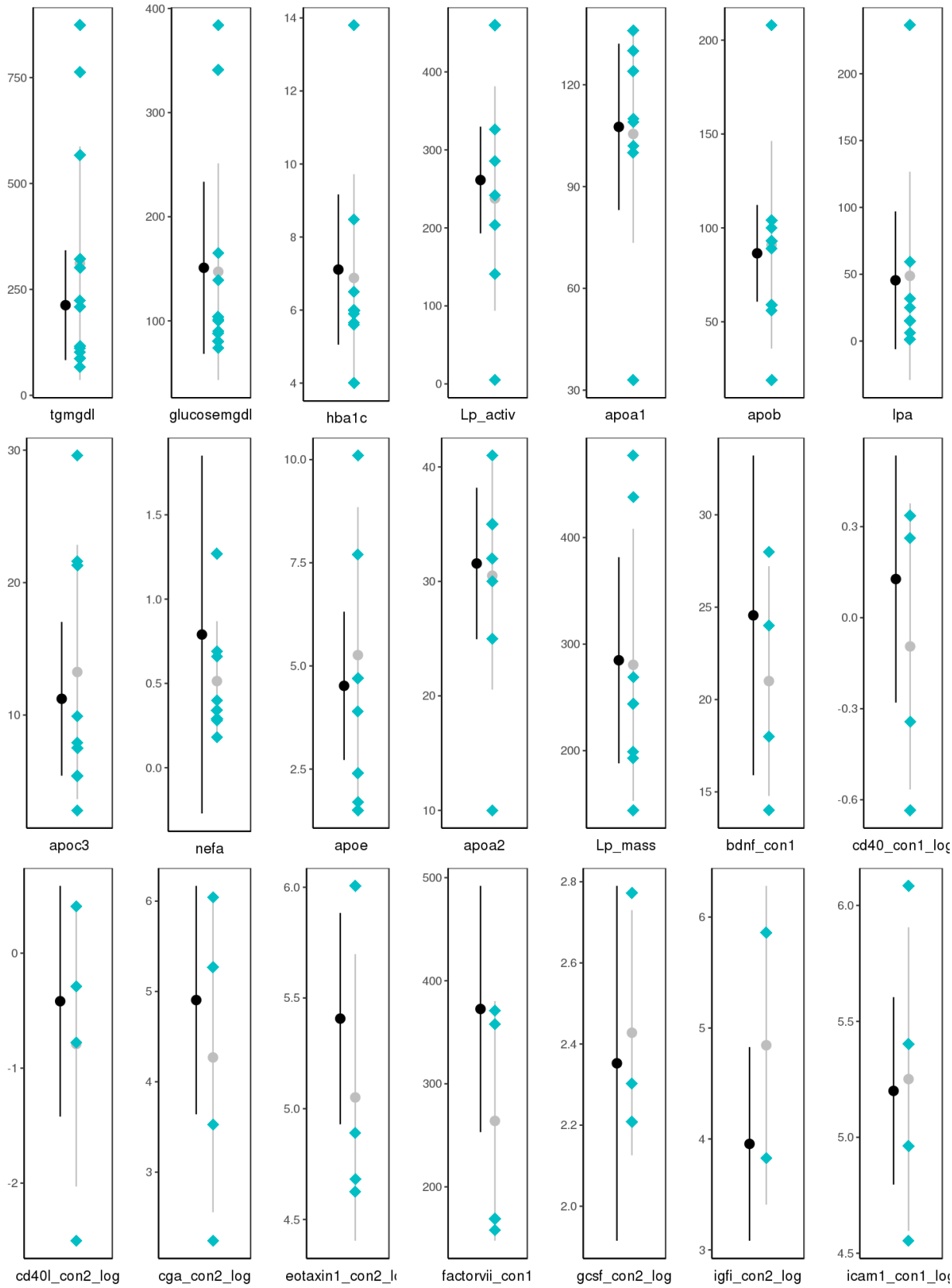
Supplementary Figure 1. IGV visualization of the splice donor variant GRCh37:12-40626187-T-C in the GTEx *LRRK2* pLoF carrier exome sequencing data and lung tissue RNA-seq data compared to a control GTEx lung RNA-seq sample. The pLoF variant is observed on reads containing an anchoring missense variant, GRCh37:12-40626185-A-G (A green and G orange), and these reads are presenting normal splicing as seen in the control RNA-seq sample.

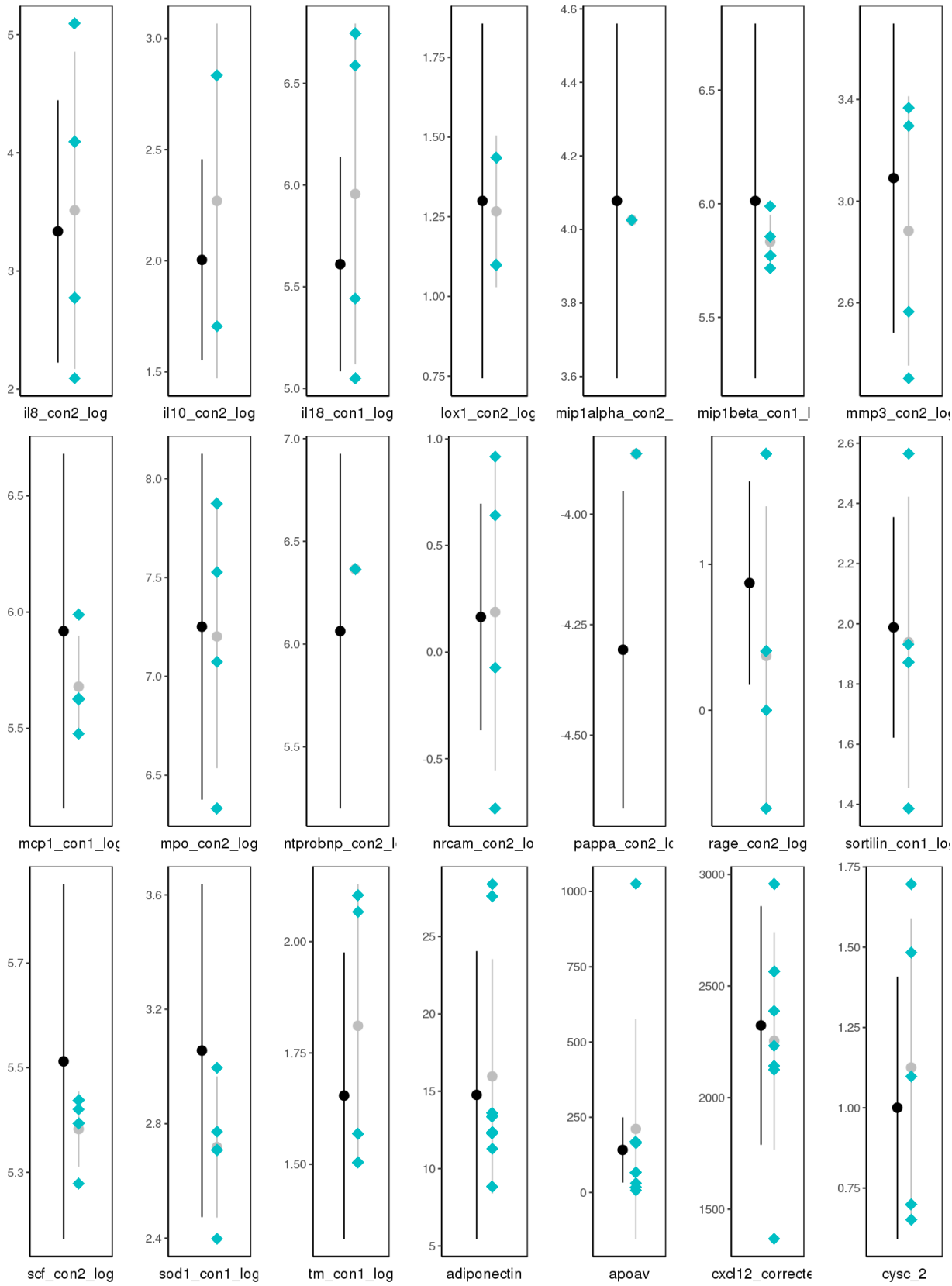


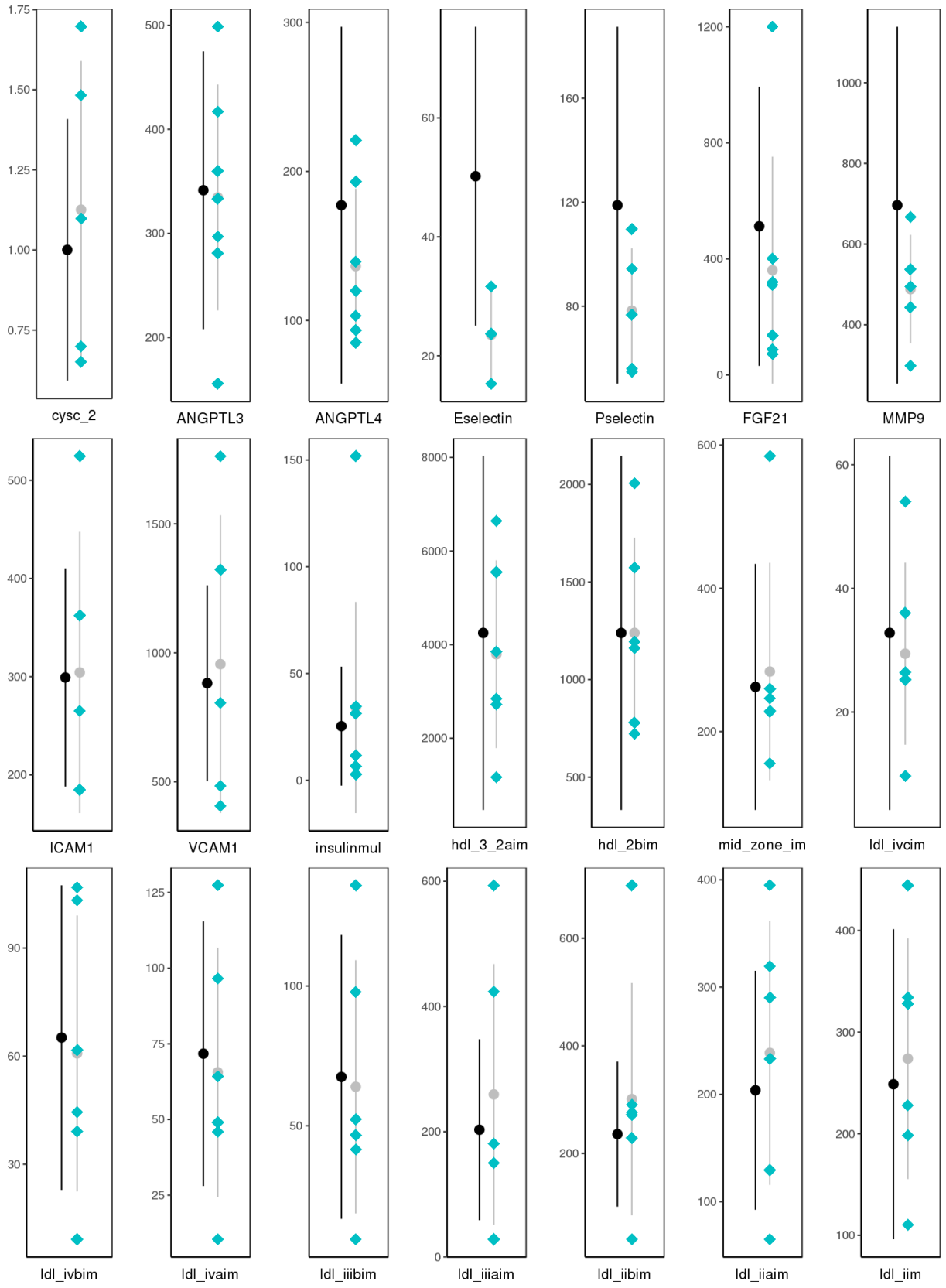
Supplementary Figure 2. Ethnicity distribution of LRRK2 LoF carriers in the 23andMe cohort.

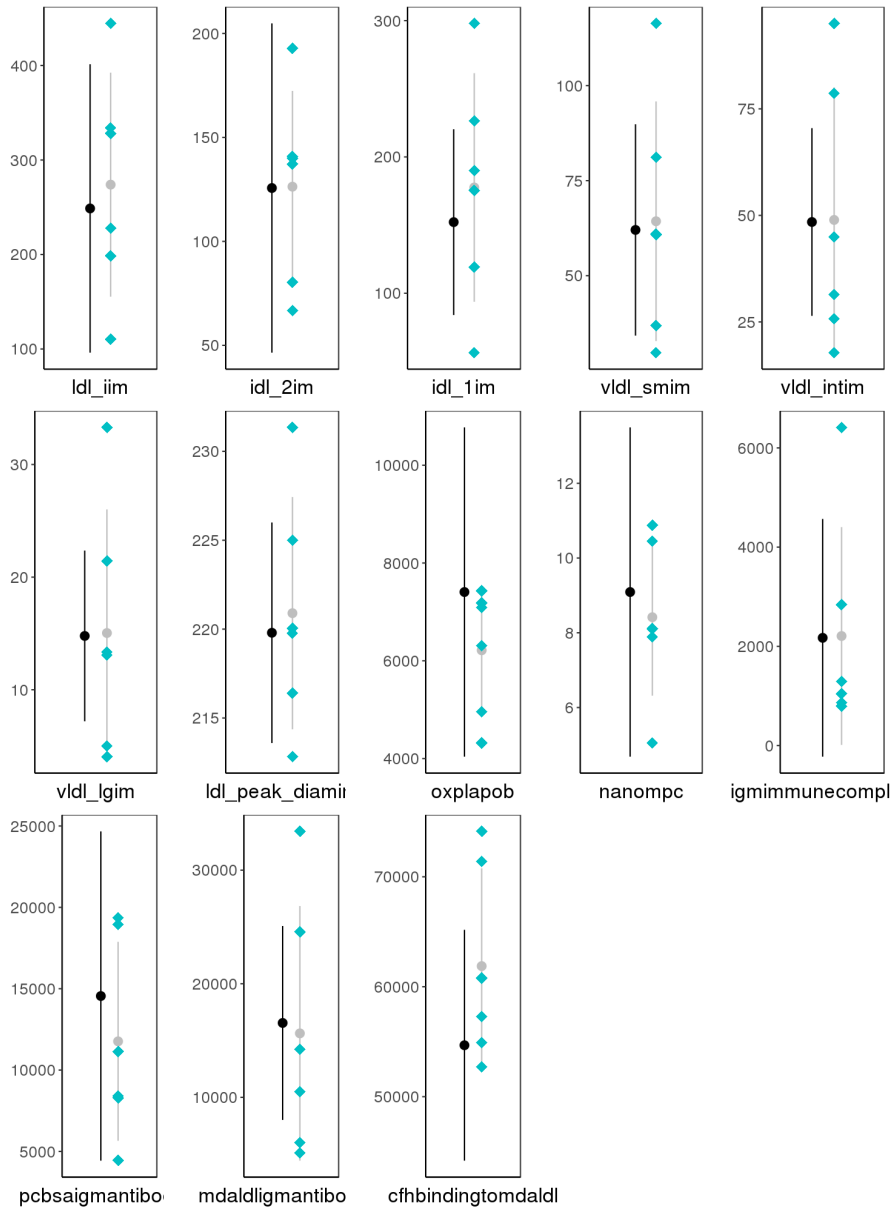


Supplementary Figure 3. a) Sanger sequence of isogenic hESC engineered colony for heterozygous LRRK2 variant 10, clone 13 (GRCh37:12-40714897-C-T). The engineered cell line maintains b) a normal karyotype, c) normal colony morphology and expression of OCT4, and d) differentiates into the cardiomyocyte lineage. The bright field image of cardiomyocytes was captured at day 17 of the differentiation protocol.


tgmgdl
glucosemgdl
hba1c
Lp_activ
apoa1
apob
lpa
apoc3
nefa
apoe
apoa2
Lp_mass
bdnf_con1
cd40_con1_log
cd40l_con2_log
cga_con2_log
eotaxin1_con2_l
factorvii_con1
gcsf_con2_log
igfi_con2_log
icam1_con1_log


il8_con2_log
il10_con2_log
il18_con1_log
lox1_con2_log
mip1alpha_con2_
mip1beta_con1_l
mmp3_con2_log
mcp1_con1_log
mpo_con2_log
ntprobnp_con2_l
nrcam_con2_lo
pappa_con2_lo
rage_con2_log
sortilin_con1_log
scf_con2_log
sod1_con1_log
tm_con1_log
adiponectin
apoav
cxcl12_correcte
cysc_2


cysc_2
ANGPTL3
ANGPTL4
Eselectin
Pselectin
FGF21
MMP9
ICAM1
VCAM1
insulinmul
hdl_3_2aim
hdl_2bim
mid_zone_im
ldl_ivcim
ldl_ivbim
ldl_ivaim
ldl_iiibim
ldl_iiiaim
ldl_iibim
ldl_iiaim
ldl_iim



Supplementary Figure 4. Blood serum biomarkers in *LRRK2* LoF carriers compared to the measurements in the 10,503 PROMIS parent cohort: cohort mean and standard deviations are shown in black (mean as a circle with +/- 1 SD as black lines) and the *LRRK2* LoF carriers as blue diamonds with the mean and standard deviations shown in grey.

Online Methods

gnomAD variant annotation and curation

The gnomAD resource, including both sample and variant quality control (including sample ancestry assignment), is fully described in our companion paper[9]. gnomAD version 2.1 was used for this analysis. Putative LoF variants were defined as stop gained, frameshift or essential splice site (splice donor or splice acceptor) as annotated by the Ensembl Variant Effect Predictor[35]. Variants were included if they were annotated as LoF on any transcript expressed in the lung, liver or kidney. Variants were filtered out if they were flagged as low-confidence by the Loss-of-function Transcript Effect Estimator (LOFTEE)[9]. For the remaining variants, a manual curation was performed, including inspection of variant quality metrics, read distribution and the presence of nearby variants using the Integrative genome viewer (IGV) and splice site prediction algorithms using Alamut. For a single splice-site variant found in an individual in the GTEx project, RNA-seq reads were manually inspected to look for any effect on splicing. All available tissues were assessed with reads from lung tissue shown in Supplementary Figure 1.

We have complied with all relevant ethical regulations. This study was overseen by the Broad Institute's Office of Research Subject Protection and the Partners Human Research Committee. Informed consent was obtained from all participants.

Sanger validation of gnomAD variant carriers

Sanger validation was performed on genomic DNA derived from peripheral blood under the following PCR conditions: 98°C 2min; 30 cycles 20sec 98°C, 20sec 54°C, 1min 72°C; 3min 72°C using the Herculase II Fusion DNA polymerase (Agilent, 600679). PCR products (5ul) were analyzed on a 2% agarose gel and the remaining product was purified with the Qiagen PCR Purification kit. Sequence analysis was performed with both PCR primers at Quintarabio. Details of variants and PCR primers used for each can be found in Supplementary Table 2.

gnomAD phenotype curation and cohort descriptions

The below described studies with *LRRK2* pLoF carriers had available phenotype data. For each study, all available records were manually reviewed to identify any reports of health problems which were categorised into the following classes: lung, liver, kidney, cardiovascular, nervous system, immune and cancer.

The Genomic Psychiatry Cohort (GPC) project

The GPC cohort is a longitudinal resource with the aim of making population-based data available through the National Institute of Mental Health (NIMH). The NIMH repository contains whole-genome sequencing data and detailed clinical and demographic data particularly focussed on schizophrenia and bipolar disorders. A large fraction of participants, 88%, have consented for re-contact[36].

The screening questionnaire consisted of 32 yes/no questions about mental health issues and 23 yes/no questions covering other medical problems including liver, digestive and cardiovascular problems. There were no specific questions relating to lung or kidney phenotypes although participants were asked to answer yes/no to having any additional health problems. If a participant answered yes to this question, we marked the existence of lung or kidney disease as 'unknown'. One sample was excluded due to conflicting questionnaire answers.

The age of the 25 *LRRK2* carriers ranged from 19 to 67. Two carriers, aged 55 and 60, reported having had liver problems with 4 participants over 60 years reporting no liver problems.

The Pakistan Risk of Myocardial Infarction Study (PROMIS)

The PROMIS cohort comprises 10,503 individuals characterized using a phenotype questionnaire with >350 items covering demographic and dietary characteristics and over 80 blood biomarker measurements[37]. The predominant focus of the questionnaire was cardiac function and phenotype. Whist the participants were specifically asked to report suffering from asthma, no other lung, liver, kidney, nervous system or immune phenotypes were directly assayed so these were marked as 'unknown' for these individuals.

The 12 *LRRK2* LoF carriers in PROMIS did not differ in terms of age, sex and myocardial infarction status when compared to the entire cohort. In addition, for these 12 individuals, we obtained data for up to 80 blood serum biomarkers along with summary statistics (mean, median and standard deviation) for the entire cohort. A Welch Two Sample t-test was used to test for difference between *LRRK2* pLoF carriers and non-carriers for the 77 biomarkers with data for >1 carrier individual.

The Swedish Schizophrenia and Bipolar Studies

Cases with schizophrenia or bipolar disorder were identified from Swedish national hospitalization registers[38,39]. Controls were selected at random from population registers. All individuals had whole exome sequencing data[40]. All available ICD codes from inpatient hospitalizations and outpatient specialist treatment contacts were provided for each patient.

FINRISK: the National FINRISK study

The FINRISK survey has been carried out for 40 years since 1972 every five years using independent, random and representative population samples from different parts of Finland. For this work, we used sequencing and health register data from FINRISK surveys 1992—2007[41].

Full health records including ICD10 codes were reviewed by study coordinators who provided us with yes/no answers for each of our phenotype classes.

The BioMe biobank at The Charles Bronfman Institute for Personalized Medicine at Mount Sinai

The Mount Sinai BioMe Biobank, founded in September 2007, is an ongoing, broadly-consented electronic health record (EHR)-linked bio- and data-repository that enrolls participants non-selectively from the Mount Sinai Medical Center patient population (New York City). BioMe participants represent a broad racial, ethnic and socioeconomic diversity with a distinct and population-specific disease burden, characteristic of the communities served by Mount Sinai Hospital. Currently comprised of over 47,000 participants, BioMe participants are of African (AA, 24%), Hispanic/Latino (HL, 35%), European (EA, 32% of whom 40% Ashkenazi Jewish) and other/mixed ancestry.

BioMe is linked to Mount Sinai's system-wide Epic EHR, which captures a full spectrum of biomedical phenotypes, including clinical outcomes, covariate and exposure data from past, present and future health care encounters. The median number of outpatient encounters is 21 per participant, reflecting predominant enrollment of participants with common chronic conditions from primary care facilities. Clinical phenotype data has been meticulously harmonized and validated.

Genome-wide genotype data and whole exome sequencing data is available for >30,000 participants. In addition, whole genome sequencing data is available for >11,000 participants. The full electronic records of three BioMe *LRRK2* pLoF carriers were reviewed by local clinicians and we were provided with detailed summaries.

Estonian Biobank of the Estonian Genome Center, University of Tartu

The Estonian Biobank cohort is composed of volunteers from the general Estonian resident adult population[42]. The current number of participants - close to 165,000--represents 15%, of the Estonian adult population, making it ideally suited to population-based studies. Participants were recruited throughout Estonia by medical personnel, and participants receive a standardized health examination, donate blood, and fill out a 16-module questionnaire on health-related topics such as

lifestyle, diet and clinical diagnoses. A detailed phenotype summary from a health survey and linked data including ICD10 codes, clinical lab values and treatment and medication information is annually updated through linkage with national electronic health databases and registries.

23andMe variant annotation and curation

Putative *LRRK2* LoF variants were defined as those classified as high-confidence by LOFTEE. Variants were manually assessed for call rate, genotyping and imputation quality and manually curated to ensure they were expected to cause true LoF.

For each of the two genotyped *LRRK2* pLoF, we determined carrier status by manually inspecting and custom calling the probe intensity plots. For the imputed variants, carrier status was determined from the Minimac-Imputed dosage. As these calling methods might produce false positives, we confirmed the participants' genotypes through Sanger sequencing. This resulted in a cohort of 749 individuals, each of whom is a Sanger sequence-confirmed carrier for one of three pLoF variants (Supplementary Table 3).

During initial selection and sequencing, expansion of the database led to inclusion of a number of additional individuals genotyped for one of the pLoF variants, rs183902574. We performed custom-calling on these individuals and found 354 deemed high-confidence carriers (Supplementary Table 3). As these individuals were not Sanger sequenced, all subsequent analyses were performed both including and excluding these individuals.

Association testing in the 23andMe dataset

Phenotype selection

The 23andMe dataset includes self-reported phenotype data for thousands of phenotypes across a diverse range of categories. These phenotypes have different sample sizes and prevalences, so the power to detect associations varies widely. We began with a curated set of 748 disease phenotypes. We then applied a liberal filter based on our power to detect an association with carrier status. More specifically, assuming a MAF of 2E-05, we restricted to phenotypes where we had power 0.1 to detect an association effect with OR > 1.3 (for binary traits) or beta > 0.2 (for quantitative traits) at alpha=0.0001 significance. This left us with 481 binary and 15 quantitative phenotypes.

Association testing

For each phenotype, we first restricted to individuals for whom we had phenotypic data. We calculated pairwise identity by descent (IBD) over all individuals using a modified version of IBD64, and then iteratively removed individuals until we were left with a set of participants, no two of whom shared > 700cM in IBD. We then tested the association between the phenotype and carrier status, controlling for age, sex, genotyping platform, and the first ten genetic principle components. We used logistic regression for binary phenotypes, and linear regression for quantitative phenotypes.

To control for population structure we restricted our analyses to participants with > 97% European ancestry, but the results did not qualitatively change when we dropped this restriction. We also tested associations using only individuals whose carrier status was confirmed by Sanger sequencing, but this also did not result in any meaningful difference.

Analysis of LRRK2 protein levels

Cell culture

All cell lines tested negative for mycoplasma contamination on a monthly basis with the MycoAlert™ Detection kit (Lonza, LT07-118) and MycoAlert™ Assay Control Set (Lonza, LT07-518). Cells were grown at 37°C with 5% CO2.

Human embryonic stem cell culture

All pluripotent stem cells were approved by Harvard ESCRO protocol #E00052 and #E00067. Human embryonic stem cells (hESCs) were obtained from WiCell Research Institute (WA01, H1) under an MTA. Cell lines were authenticated by visual inspection of cell morphology with brightfield microscopy, staining with anti-Oct4 antibody to determine maintenance of pluripotency (Santa Cruz, sc-5279, data not shown), sent to WiCell Research Institute after 6 months of passaging or after isogenic cell line generation for karyotyping, and in some cases PCA of RNA sequencing data to confirm clustering with other pluripotent stem cell lines. Pluripotent stem cells were plated onto hESC qualified matrigel (VWR, BD354277) coated 6 well plates, mTeSR1 media was changed daily (StemCell Technologies, 85850), and cells were passaged every 5-7 days with 0.5mM EDTA.

Lymphoblastoid cell culture

Lymphoblastoid cell lines (LCLs) were obtained from Coriell Biorepository (GM18500, GM18501, GM18502, HG01345, HG01346) and approved by the Broad Institute Office of Research Subject

Protection protocol #3639. Cell lines were authenticated by visual inspection of cell morphology with brightfield microscopy and in some cases PCA of RNA sequencing data to confirm clustering with GTEx LCLs. LCL media was changed every other day with RPMI 1640 (Life Technologies), 2mM L-glutamine (Life Technologies), and 15% FBS (Sigma).

Cardiomyocyte differentiation

Cardiomyocyte differentiation of the control and engineered H1 hESC lines was performed according to the protocol by Lian et al., 2012. Briefly, 500,000 cells were plated on hESC qualified matrigel (VWR, BD354277), grown in mTeSR1 media for four days (StemCell Technologies, 85850), and switched to RPMI media (Life Technologies) with B27 supplement (Life Technologies), switching to B27 with insulin at day 7, for the remainder of the protocol. On day 0 of differentiation, 12µM CHIR99021 (Tocris) was applied for 24 hours. At day 3, cultures were treated with 5µM IWP2 (Tocris) for 24 hours. Bright field images and movies were acquired at day 17 and cells were harvested for protein/RNA extraction at day 19.

Isogenic cell line engineering

The following guide and HR template were delivered into single cell H1 hESCs via nucleofection (Lonza 4D-Nucleofector X unit) using the P3 Primary Cell Kit (V4XP-3024), pulse code CA137, and pX459 (Addgene): AATAAGGCATTTCATATAGT and ACAGGCCTGTGATAGAGCTTCCCCATTGT GAGAACTCTGAAATTATCATCTGACTATATGAAATGCCTTATTTTCCAATGGGATTTTGGTCAAGATTAA.
Cells were allowed to recover from nucleofection in mTeSR supplemented with 10µM Rock Inhibitor (Y-27632, Tocris) overnight. The following three days the cells were treated with 0.25µg/ml puromycin (VWR) in mTeSR. Cells were then cultured in mTeSR until colonies were ready to be split. Engineered cells were split into single cells and plated in matrigel coated 96-well plates at a density of 0.5 cells per well. Plates were screened for colonies 8-10 days after plating and grown until colonies were ready to be split. Colonies were then split with 0.5mM EDTA into two identical 96-well plates, one for DNA extraction/PCR/sequencing and one for freezing cells. Once colonies were ready to be split, 96-well plates were frozen in mFreSR (Stem Cell Technologies) and stored at -80°C until HR positive wells were identified. HR edited cells were then thawed and expanded for 4 generations, validated by Sanger sequencing, karyotyping, and OCT4 staining prior to proceeding with cardiomyocyte differentiation.

Off-target analysis of CRISPR/Cas9 engineering

To detect any potential off-target effects caused by CRISPR/Cas9 genome editing, whole-genome sequencing (WGS) was conducted for both the engineered and control cell lines. DNA extraction, Quality Control (QC) and 30x PCR-Free WGS were performed by the Genomics Platform at the Broad

Institute. AllPrep DNA/RNA extraction kit was used, following its protocol. Alignment, marking of duplicates, recalibration of quality scores and variant calling are all performed using GATK best practices[43].

We identified 157,230 variants in the engineered cell line that were not found in the control cell line as candidate variants. For the guide RNA (gRNA) used, we defined potential off-target regions as those with <4bps mismatch against the full 20bp gRNA sequence (334 regions), and/or <2bp mismatch against the seed (=PAM proximal) 12bp of the gRNA sequence (5,780 regions), each followed by the NGG PAM. We looked for any candidate variant that fall into the potential off-target region, resulting in detection of only one variant (chr8-65084564-A-AT) that falls onto a region with 1 mismatch against the seed 12bp of gRNA sequence (chr8:65084560-65084575). No variants with <4bp mismatch against the full 20bp gRNA sequence or perfect match at the seed region were detected. Since a mismatch at the seed region decreases the likelihood of off-target variants[44,45], and also because the single variant we detected is a known variant (rs1161563412) observed in the population without apparent phenotypic association, we concluded no major off-target effect exists at the level of violating the main steps of our research. All the analysis for the detection of potential off-target were conducted using pybedtools[46] and CRISPRdirect[47] software.

Western blot analysis

Cell pellets were snap-frozen in liquid nitrogen and stored at -80°C. Cells were dounce homogenized in ice-cold radioimmunoprecipitation assay buffer (RIPA) (89901; Thermo Scientific) containing protease inhibitors (HaltTM Protease Inhibitor, Thermo Scientific). Homogenates were rotated at 4°C for 30min, followed by centrifugation at 15,000g for 20min at 4°C. Equal amounts of protein (50µg) were electrophoresed on 4-20% SDS-PAGE (Biorad) and transferred to nitrocellulose membranes. The following antibodies were used for immunoblotting: LRRK2 (75-253, UC Davis/NIH NeuroMab Facility), α-actinin (A7811, Sigma), GAPDH (sc-25778, Santa Cruz), anti-rabbit IgG HRP (7074, Cell Signaling), and anti-mouse IgG HRP (7076, Cell Signaling). Immunoblots were developed using enhanced chemiluminescence (SuperSignal West Pico Chemiluminescent Substrate, Thermo Scientific) on an Amersham Imager 600.

GROUP AUTHORS

Genome Aggregation Database Production Team: Jessica Alföldi[1,2], Irina M. Armean[1,2,3], Eric Banks[4], Louis Bergelson[4], Kristian Cibulskis[4], Ryan L Collins[1,5,6], Kristen M. Connolly[7], Miguel Covarrubias[4], Beryl Cummings[1,2,8], Mark J. Daly[1,2,9], Stacey Donnelly[1], Yossi Farjoun[4], Steven Ferriera[10], Laurent Francioli[1,2], Stacey Gabriel[10], Laura D. Gauthier[4], Jeff Gentry[4], Namrata Gupta[10,1], Thibault Jeandet[4], Diane Kaplan[4], Konrad J. Karczewski[1,2], Kristen M. Laricchia[1,2], Christopher Llanwarne[4], Eric V. Minikel[1], Ruchi Munshi[4], Benjamin M Neale[1,2], Sam Novod[4], Anne H. O'Donnell-Luria[1,11,12], Nikelle Petrillo[4], Timothy Poterba[9,2,1], David Roazen[4], Valentin Ruano-Rubio[4], Andrea Saltzman[1], Kaitlin E. Samocha[13], Molly Schleicher[1], Cotton Seed[9,2], Matthew Solomonson[1,2], Jose Soto[4], Grace Tiao[1,2], Kathleen Tibbetts[4], Charlotte Tolonen[4], Christopher Vittal[9,2], Gordon Wade[4], Arcturus Wang[9,2,1], Qingbo Wang[1,2,6], James S Ware[14,15,1], Nicholas A Watts[1,2], Ben Weisburd[4], Nicola Whiffin[14,15,1]

1. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA
2. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA
3. European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom
4. Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA
5. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
6. Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA 02115, USA
7. Genomics Platform, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA
8. Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, 02115, USA
9. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA
10. Broad Genomics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA
11. Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts 02115, USA
12. Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA
13. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK
14. National Heart & Lung Institute and MRC London Institute of Medical Sciences, Imperial College London, London UK
15. Cardiovascular Research Centre, Royal Brompton & Harefield Hospitals NHS Trust, London UK

Genome Aggregation Database Consortium: Carlos A Aguilar Salinas[1], Tariq Ahmad[2], Christine M. Albert[3,4], Diego Ardissino[5], Gil Atzmon[6,7], John Barnard[8], Laurent Beaugerie[9], Emelia J. Benjamin[10,11,12], Michael Boehnke[13], Lori L. Bonnycastle[14], Erwin P. Bottinger[15], Donald W Bowden[16,17,18], Matthew J Bown[19,20], John C Chambers[21,22,23], Juliana C. Chan[24], Daniel Chasman[3,25], Judy Cho[15], Mina K. Chung[26], Bruce Cohen[27,25], Adolfo Correa[28], Dana Dabelea[29], Mark J. Daly[30,31,32], Dawood Darbar[33], Ravindranath Duggirala[34], Josée Dupuis[35,36], Patrick T. Ellinor[30,37], Roberto Elosua[38,39,40], Jeanette Erdmann[41,42,43], Tõnu Esko[30,44], Martti Färkkilä[45], Jose Florez[46], Andre Franke[47], Gad Getz[48,49,25], Benjamin Glaser[50], Stephen J. Glatt[51], David Goldstein[52,53], Clicerio Gonzalez[54], Leif Groop[55,56], Christopher Haiman[57], Craig Hanis[58], Matthew Harms[59,60], Mikko Hiltunen[61], Matti M. Holi[62], Christina M. Hultman[63,64], Mikko Kallela[65], Jaakko Kaprio[56,66], Sekar Kathiresan[67,68,25], Bong-Jo Kim[69], Young Jin Kim[69], George Kirov[70], Jaspal Kooner[23,22,71], Seppo Koskinen[72], Harlan M. Krumholz[73], Subra Kugathasan[74], Soo Heon Kwak[75], Markku Laakso[76,77], Terho

Lehtimäki[78], Ruth J.F. Loos[15,79], Steven A. Lubitz[30,37], Ronald C.W. Ma[24,80,81], Daniel G. MacArthur[31,30], Jaume Marrugat[82,39], Kari M. Mattila[78], Steven McCarroll[32,83], Mark I McCarthy[84,85,86], Dermot McGovern[87], Ruth McPherson[88], James B. Meigs[89,25,90], Olle Melander[91], Andres Metspalu[44], Benjamin M Neale[30,31], Peter M Nilsson[92], Michael C O'Donovan[70], Dost Ongur[27,25], Lorena Orozco[93], Michael J Owen[70], Colin N.A. Palmer[94], Aarno Palotie[56,32,31], Kyong Soo Park[75,95], Carlos Pato[96], Ann E. Pulver[97], Nazneen Rahman[98], Anne M. Remes[99], John D. Rioux[100,101], Samuli Ripatti[56,66,102], Dan M. Roden[103,104], Danish Saleheen[105,106,107], Veikko Salomaa[108], Nilesh J. Samani[19,20], Jeremiah Scharf[30,32,67], Heribert Schunkert[109,110], Moore B. Shoemaker[111], Pamela Sklar*[112,113,114], Hilkka Soininen[115], Harry Sokol[9], Tim Spector[116], Patrick F. Sullivan[63,117], Jaana Suvisaari[108], E Shyong Tai[118,119,120], Yik Ying Teo[118,121,122], Tuomi Tiinamaija[56,123,124], Ming Tsuang[125,126], Dan Turner[127], Teresa Tusie-Luna[128,129], Erkki Vartiainen[66], James S Ware[130,131,30], Hugh Watkins[132], Rinse K Weersma[133], Maija Wessman[123,56], James G. Wilson[134], Ramnik J. Xavier[135,136]

1. Unidad de Investigacion de Enfermedades Metabolicas. Instituto Nacional de Ciencias Medicas y Nutricion. Mexico City
2. Peninsula College of Medicine and Dentistry, Exeter, UK
3. Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA.
4. Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.
5. Department of Cardiology, University Hospital, 43100 Parma, Italy
6. Department of Biology, Faculty of Natural Sciences, University of Haifa, Haifa, Israel
7. Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY, USA, 10461
8. Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44122, USA
9. Sorbonne Université, APHP, Gastroenterology Department, Saint Antoine Hospital, Paris, France
10. NHLBI and Boston University's Framingham Heart Study, Framingham, Massachusetts, USA.
11. Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA.
12. Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA.
13. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109
14. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
15. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY
16. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA
17. Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA
18. Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA
19. Department of Cardiovascular Sciences, University of Leicester, Leicester, UK
20. NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK
21. Department of Epidemiology and Biostatistics, Imperial College London, London, UK
22. Department of Cardiology, Ealing Hospital NHS Trust, Southall, UK
23. Imperial College Healthcare NHS Trust, Imperial College London, London, UK
24. Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China.
25. Department of Medicine, Harvard Medical School, Boston, MA
26. Departments of Cardiovascular Medicine, Cellular and Molecular Medicine, Molecular Cardiology, and Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio, USA.
27. McLean Hospital, Belmont, MA
28. Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, USA
29. Department of Epidemiology, Colorado School of Public Health, Aurora, Colorado, USA.
30. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
31. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA
32. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
33. Department of Medicine and Pharmacology, University of Illinois at Chicago
34. Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA

35. Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA
36. National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA 01702, USA
37. Cardiac Arrhythmia Service and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA
38. Cardiovascular Epidemiology and Genetics, Hospital del Mar Medical Research Institute (IMIM). Barcelona, Catalonia, Spain
39. CIBER CV, Barcelona, Catalonia, Spain
40. Departament of Medicine, Medical School, University of Vic-Central University of Catalonia. Vic, Catalonia, Spain
41. Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany
42. 1. DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, 23562 Lübeck, Germany
43. University Heart Center Lübeck, 23562 Lübeck, Germany
44. Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia
45. Helsinki University and Helsinki University Hospital, Clinic of Gastroenterology, Helsinki, Finland.
46. Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital; Programs in Metabolism and Medical & Population Genetics, Broad Institute; Department of Medicine, Harvard Medical School
47. Institute of Clinical Molecular Biology (IKMB), Christian-Albrechts-University of Kiel, Kiel, Germany
48. Bioinformatics Program, MGH Cancer Center and Department of Pathology
49. Cancer Genome Computational Analysis, Broad Institute.
50. Endocrinology and Metabolism Department, Hadassah-Hebrew University Medical Center, Jerusalem, Israel
51. Department of Psychiatry and Behavioral Sciences; SUNY Upstate Medical University
52. Institute for Genomic Medicine, Columbia University Medical Center, Hammer Health Sciences, 1408, 701 West 168th Street, New York, New York 10032, USA.
53. Department of Genetics & Development, Columbia University Medical Center, Hammer Health Sciences, 1602, 701 West 168th Street, New York, New York 10032, USA.
54. Centro de Investigacion en Salud Poblacional. Instituto Nacional de Salud Publica MEXICO
55. Lund University, Sweden
56. Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland
57. Lund University Diabetes Centre
58. Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX 77030
59. Department of Neurology, Columbia University
60. Institute of Genomic Medicine, Columbia University
61. Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland
62. Department of Psychiatry, PL 320, Helsinki University Central Hospital, Lapinlahdentie, 00 180 Helsinki, Finland
63. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
64. Icahn School of Medicine at Mount Sinai, New York, NY, USA
65. Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland.
66. Department of Public Health, Faculty of Medicine, University of Helsinki, Finland
67. Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA
68. Cardiovascular Disease Initiative and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA
69. Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do, Republic of Korea.
70. MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University School of Medicine, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ
71. National Heart and Lung Institute, Cardiovascular Sciences, Hammersmith Campus, Imperial College London, London, UK.
72. Department of Health, THL-National Institute for Health and Welfare, 00271 Helsinki, Finland.
73. Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut3Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, Connecticut.
74. Division of Pediatric Gastroenterology, Emory University School of Medicine, Atlanta, Georgia, USA.
75. Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea
76. The University of Eastern Finland, Institute of Clinical Medicine, Kuopio, Finland
77. Kuopio University Hospital, Kuopio, Finland
78. Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Health Technology, Tampere University, Finland
79. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY

80. Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China.
81. Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China.
82. Cardiovascular Research REGICOR Group, Hospital del Mar Medical Research Institute (IMIM). Barcelona, Catalonia.
83. Department of Genetics, Harvard Medical School, Boston, MA, USA
84. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LJ UK
85. Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK
86. Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford OX3 9DU, UK
87. F Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.
88. Atherogenomics Laboratory, University of Ottawa Heart Institute, Ottawa, Canada
89. Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, 02114
90. Program in Population and Medical Genetics, Broad Institute, Cambridge, MA
91. Department of Clinical Sciences, University Hospital Malmo Clinical Research Center, Lund University, Malmo, Sweden.
92. Lund University, Dept. Clinical Sciences, Skane University Hospital, Malmo, Sweden
93. Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, 14610, Mexico
94. Medical Research Institute, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK.
95. Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea
96. Department of Psychiatry, Keck School of Medicine at the University of Southern California, Los Angeles, California, USA.
97. Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
98. Division of Genetics and Epidemiology, Institute of Cancer Research, London SM2 5NG
99. Medical Research Center, Oulu University Hospital, Oulu, Finland and Research Unit of Clinical Neuroscience, Neurology, University of Oulu, Oulu, Finland.
100. Research Center, Montreal Heart Institute, Montreal, Quebec, Canada, H1T 1C8
101. Department of Medicine, Faculty of Medicine, Université de Montréal, Québec, Canada
102. Broad Institute of MIT and Harvard, Cambridge MA, USA
103. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA.
104. Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA.
105. Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA
106. Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA
107. Center for Non-Communicable Diseases, Karachi, Pakistan
108. National Institute for Health and Welfare, Helsinki, Finland
109. Deutsches Herzzentrum München, Germany
110. Technische Universität München
111. Division of Cardiovascular Medicine, Nashville VA Medical Center and Vanderbilt University, School of Medicine, Nashville, TN 37232-8802, USA.
112. Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA
113. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
114. Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
115. Institute of Clinical Medicine, neurology, University of Eastern Finad, Kuopio, Finland
116. Department of Twin Research and Genetic Epidemiology, King's College London, London UK
117. Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA
118. Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore
119. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
120. Duke-NUS Graduate Medical School, Singapore
121. Life Sciences Institute, National University of Singapore, Singapore.
122. Department of Statistics and Applied Probability, National University of Singapore, Singapore.
123. Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland
124. HUCH Abdominal Center, Helsinki University Hospital, Helsinki, Finland

125. Center for Behavioral Genomics, Department of Psychiatry, University of California, San Diego
126. Institute of Genomic Medicine, University of California, San Diego
127. Juliet Keidan Institute of Pediatric Gastroenterology, Shaare Zedek Medical Center, The Hebrew University of Jerusalem, Israel
128. Instituto de Investigaciones Biomédicas UNAM Mexico City
129. Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán Mexico City
130. National Heart & Lung Institute & MRC London Institute of Medical Sciences, Imperial College London, London UK
131. Cardiovascular Research Centre, Royal Brompton & Harefield Hospitals NHS Trust, London UK
132. Radcliffe Department of Medicine, University of Oxford, Oxford UK
133. Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, Groningen, the Netherlands
134. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA
135. Program in Infectious Disease and Microbiome, Broad Institute of MIT and Harvard, Cambridge, MA, USA
136. Center for Computational and Integrative Biology, Massachusetts General Hospital

23andMe Research Team: Michelle Agee, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, Barry Hicks, David A. Hinds, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Keng-Han Lin, Nadia K. Litterman, Matthew H. McIntyre, Kimberly F. McManus, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, G. David Poznik, J. Fah Sathirapongsasuti, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, Xin Wang, Catherine H. Wilson

Affiliation: 23andMe, Inc.