Published in final edited form as:

Hum Mutat. 2012 July ; 33(7): 1087–1098. doi:10.1002/humu.22075.

Deep sequencing of the *LRRK2* gene in 14,002 individuals reveals evidence of purifying selection and independent origin of the p.Arg1628Pro mutation in Europe

Justin P. Rubio¹, Simon Topp^{1,*}, Liling Warren^{1,*}, Pamela L. St Jean¹, Daniel Wegmann², Darren Kessner², John Novembre², Judong Shen¹, Dana Fraser¹, Jennifer Aponte¹, Keith Nangle¹, Lon R. Cardon¹, Margaret G. Ehm¹, Stephanie L. Chissoe¹, John C. Whittaker¹, Matthew R. Nelson¹, and Vincent E. Mooser¹

¹Quantitative Sciences, GlaxoSmithKline R&D: Stevenage, UK; Research Triangle Park, NC, USA; Upper Merion, PA, USA

²Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA

Abstract

Genetic variation in *LRRK2* predisposes to Parkinson disease (PD), which underpins its development as a therapeutic target. Here, we aimed to identify novel genotype-phenotype associations that might support developing *LRRK2* therapies for other conditions. We sequenced the 51 exons of *LRRK2* in cases comprising 12 common diseases (n = 9,582), and in 4,420 population controls. We identified 739 single nucleotide variants (SNVs), 62% of which were observed in only one person, including 316 novel exonic variants. We found evidence of purifying selection for the *LRRK2* gene and a trend suggesting that this is more pronounced in the central (ROC-COR-kinase) core protein domains of *LRRK2* than the flanking domains. Population genetic analyses revealed that *LRRK2* is not especially polymorphic or differentiated in comparison to 201 other drug target genes. Amongst Europeans, we identified 17 carriers (0.13%) of pathogenic *LRRK2* mutations that were not significantly enriched within any disease or in those reporting a family history of PD. Analysis of pathogenic mutations within Europe reveals that the p.Arg1628Pro (c4883G>C) mutation arose independently in Europe and Asia. Taken together, these findings demonstrate how targeted deep sequencing can help to reveal fundamental characteristics of clinically important loci.

Keywords

LRRK2; Deep sequencing; novel variants; evolution; population genetics; genotype-phenotype associations

INTRODUCTION

Mutations within the *LRRK2* gene (MIM# 609007) predispose to Parkinson Disease (PD; MIM# 168600), and of 127 DNA sequence variations recorded currently in the PD mutation database (Nuytemans, et al., 2010) (<http://www.molgen.ua.ac.be/PDmutDB/>), 81 are amino

Corresponding author: Justin P. Rubio (Ph.D.) GlaxoSmithKline R&D Building 1S - Room 101 Gunnels Wood Road, Stevenage, Hertfordshire, England SG1 2NY Tel: +44-(0)1438 766 622 Fax: +44-(0)1438 762798 Mob: +44-(0)7909 00 2363 justin.p.rubio@gsk.com.

*These authors contributed equally to this work

Supporting Information for this preprint is available from the *Human Mutation* editorial office upon request (humu@wiley.com)

acid changing, or non-synonymous (NS) mutations. However, only for a relatively small number (c.4309A>C (p.Asn1437His), c.4321C>G (p.Arg1441Gly), c.4321C>T (p.Arg1441Cys), c.4322G>A (p.Arg1441His), c.4883G>C (p.Arg1628Pro), c.5096A>G (p.Tyr1699Cys), c.6055G>A (p.Gly2019Ser), c.6059T>C (p.Ile2020Thr), c.7153G>A (Gly2385Arg)) is their pathogenicity supported by co-segregation with disease in families and for some, functional studies (Asly, et al., 2010; 2010, Kahn, et al., 2005; Lewis, et al., 2007; Paisan-Ruiz, et al., 2004; Tan, et al., 2007; Ross, et al., 2008; West, et al., 2005; Zimprich, et al., 2004).

LRRK2 is a multi-domain protein with both GTPase and kinase functionality, but in general its function is poorly understood (Cookson, 2010). Pathogenic mutations in *LRRK2* cluster in these enzymatic domains, as is the case for p.Gly2019Ser, the most common known PD-causing mutation, where the pathogenic mechanism is thought to be a toxic gain-of-function of kinase activity (West, et al., 2005). As a consequence of understanding the mechanistic effect of p.Gly2019Ser on LRRK2 function, drug discovery efforts across pharmaceutical companies are aimed at developing LRRK2 inhibitors for the treatment of PD.

Of particular relevance to the development of this class of drugs, variation in *LRRK2* may have pleiotropic effects. Evidence of this comes from the genetic association of common *LRRK2* variation with susceptibility to Crohn's disease (Barrett, et al., 2008), ankylosing spondylitis (Danoy, et al., 2010) and, leprosy infection (Zhang, et al., 2009). For Crohn's disease and leprosy (albeit suggestive), these associations appear to fit well with a role for LRRK2 in innate immunity, autophagy (Alegre-Abarrategui, et al., 2009) and, host response to pathogens (Gardet, et al., 2010). There is also evidence that chromosomal amplification, leading to over-expression of *LRRK2*, promotes tumour cell growth and survival in renal and thyroid carcinomas through cooperation with MET (Looyenga, et al., 2011). Taken together, these findings suggest a broader role for LRRK2 in human disease, and that pharmacological inhibition may be beneficial for treating other conditions.

LRRK2 is a large gene, and this has clearly been a hindrance to sequencing efforts using conventional methods. Therefore, published whole-gene sequencing studies to date have generally been restricted to hundreds (Paisan-Ruiz, et al., 2008; Nuytemans, et al., 2009), rather than thousands of individuals and have often focussed only on exons encoding the functional domains that harbour pathogenic mutations. With the advent of next generation sequencing comes the opportunity to fully characterize both common and rare genetic variation for the entire human genome (Durbin, et al., 2010), and for *LRRK2* in particular.

In the present study, we have sequenced all 51 *LRRK2* exons in 14,002 individuals comprised of independent case collections of neurological, neuropsychiatric, inflammatory, respiratory, metabolic and cardiovascular conditions, and population-based controls to enable genotype-phenotype analyses. Our investigations have identified a large number of unreported variants, some with potential effects on LRRK2 function, and provide insight into some fundamental characteristics of this clinically important gene.

METHODS

Subjects

We sequenced whole blood genomic DNA samples from 14,002 participants comprising 12 different diseases, and two population control collections (Nelson, et al., submitted). All subjects were consented for the study of common diseases and medically-relevant traits. Table 1 includes basic demographic, ethnic and phenotypic characteristics of study participants.

DNA sequencing and quality control

DNA sequencing was conducted at the Beijing Genome Institute, Shenzhen, China. Detailed descriptions of DNA sequencing methods and analysis of collective data obtained for 202 drug target genes, including *LRRK2*, are outlined by Nelson et al. (Nelson, et al., submitted). Briefly, genomic DNAs (3μg) were fragmented, uniquely tagged with an 8 bp index sequence and combined into pools of 48 samples. The entire *LRRK2* gene sequence (transcript, NM_198578.3), including all 51 exons, 50 bp of flanking sequence as well as 5' and 3' untranslated regions (UTRs) (NCBI Build 36.3) were selected for capture using a custom Roche Nimblegen (Madison, Wisconsin, USA) HD2.1M sequence capture array. Paired end sequencing was conducted for each 48-sample indexed pool on a single Illumina (San Diego, CA, USA) Genome Analyzer 2x lane. Paired-end short reads were aligned with SOAP (Li, et al., 2008b) and variants were called using SOAPsnp (Li, et al., 2009). Genotypes were only called when there was a minimum depth of seven DNA sequence reads and quality consensus score ≥ 20. The quality of variant calls was assessed using several approaches for the entire set of 202 genes, including *LRRK2* (Nelson, et al., submitted). One approach showed that variant calls were of high quality with 99.1% concordance observed for heterozygous genotypes between 130 sample duplicates. Further, 240 of 245 singleton variant calls (98.0%) were validated by Sanger capillary sequencing.

Prediction of variant functionality

Two widely-used prediction programs; Sorting Intolerant From Tolerant (SIFT) (Ng and Henikoff, 2003) and Polymorphism Phenotyping (PolyPhen2) (Ramensky, et al., 2002) were used to assess the potential impact of variants identified in this study on *LRRK2* function.

Genetic diversity analysis

To assess whether patterns of genetic diversity at *LRRK2* are similar to other loci, we compared patterns of variation to that for 201 other genes sequenced in the same samples (Nelson, et al., submitted). Aberrant levels of diversity from background levels in the other 201 genes may indicate a unique history of natural selection at the *LRRK2* locus (Kelley and Swanson, 2008; Nielsen, et al., 2007). We calculated several common population genetics summary statistics for each of the 202 genes, and examined where *LRRK2* appears in the distribution of each statistic. In addition, we examined the distribution of F_{ST} values (higher F_{ST} implying a greater degree of differentiation among populations) for variant sites within *LRRK2* in Europeans. Further details regarding the calculation of the statistics may be found in the Supp. Methods.

Structural predictions of LRRK2 domains

Domains and domain boundaries were identified primarily through homology models to tertiary protein structures in the Protein Data Bank (PDB) (<http://www.rcsb.org/>) using the automated and alignment modes of the SwissModel server (<http://swissmodel.expasy.org>) (Arnold, et al., 2006). The entries on which the domains were modelled in PDB are as follows: Armadillo:1ejlA; Ankyrin:2qyjA; LRR:1g9uA; ROC:2zejA; COR:3dptA; Kinase:3dkcA; WD40:3fm0A. Refer to Supp. Methods for further details.

Genetic association analyses

Only subjects of white European Caucasian origin were included in genotype-phenotype analyses. Variants and subjects with more than 30% of genotypes missing were excluded from analysis.

Control selection—For each case collection, controls were selected from among two population control samples (CoLaus and LOLIPOP) as well as from among case collections

(ethical consent permitting), similar to that performed by the Wellcome Trust Control Consortium (WTCCC, 2007). Controls were matched to cases based on their genetic background as inferred from the principal component (PC) scores obtained from genome-wide SNP genotype data if available, or from PCs imputed from country of origin and sequence data from all 202 genes if not. Euclidean distance was calculated based on the first five PC scores to select control subjects who had the shortest distance to the case subjects (best genetic matching). A range of case-control matching ratios was evaluated for each disease collection. The optimal case-control matching ratio was determined by evaluating the distribution of the genetic distance between cases and controls with the goal to select as many controls as possible while minimizing the median genetic distance between cases and controls at around 0.02. Additionally, controls were required to be free from the disease condition being evaluated based on self-reported medical history.

Case/control association analysis—Association analysis of 17 common variants (CVs - minor allele frequency (MAF) 0.5%) and aggregated rare *LRRK2* coding variants was carried out using logistic regression for each of the 12 disease phenotypes. Three rare variant (RV) aggregation tests were carried out: one including all non-synonymous (NS) variants and another restricted to NS variants predicted to be functional by either PolyPhen or SIFT, or occurring at a not-protein coding site with high evolutionary conservation (phyloP 2 from 46-way placental mammal alignment) (Siepel, et al., 2005). Collection site, age, gender and, the first five PC scores were included as covariates whenever appropriate.

Statistical power estimation—Statistical power to test differences in genotype frequencies (or aggregate RV frequencies) between cases and controls was computed for four collections (epilepsy, irritable bowel syndrome (IBS), rheumatoid arthritis (RA) and Schizophrenia) representing the range of sample sizes. Three MAFs, including 10% and 25% for CVs and 2% for the aggregate RV tests were assessed. We assumed an additive genetic model and a test-wise significance threshold of $0.05/(20 \text{ tests (17 CV} + 3 \text{ RV tests)} \times 12 \text{ disease phenotypes}) = 0.00021$. Odds ratios obtained reflect the additive allelic effect. Refer to Supp. Methods for the power curves.

PC analysis to assess geographical distribution of *LRRK2* variant carriers

Principle components (PC) analysis was performed on a total of 19,088 unrelated European individuals who were typed at 36,222 common SNPs across the whole genome. These SNPs were genotyped in all individuals, and remained after clusters of SNPs were excluded that correlated strongly with any of the first 5 PCs. The individuals were chosen from an initial set of 20,086 predominantly European individuals, which did not include Finns or other individuals found to be outliers in an initial PC analysis. For those subjects lacking genome wide SNP data, PCs were imputed using information from country of origin and sequence data from all 202 genes. PC axes were rotated to optimize the correspondence between the median country labels and the geographic position of the countries by fitting independent linear models for latitude and longitude as predicted jointly by PC1 and PC2, as in Novembre et al. (Novembre, et al., 2008). Geographic coordinates of the central geographic position of each country were used, with the exception of Norway and Sweden for which we chose geographic locations further south to account for the more probable origins of the sampled individuals.

Haplotype analysis

Haplotypes for *LRRK2* variant carriers were inferred from SNP data derived from this study and whole genome SNP data using BEAGLE (Li, et al., 2011).

RESULTS

Frequency and distribution of LRRK2 variants

We sequenced all 51 exons, including 5' and 3' UTRs, of the *LRRK2* gene (14,227 bp) in the 14,002 subjects described in Table 1. There was more than a 99% probability to identify alleles with a frequency of 0.02%, or 6 observations, in the entire sample. Overall, we found 739 unique single nucleotide variants (SNVs) in *LRRK2*, resulting in an average SNV rate (1 every 20 bp) similar to that observed for 201 other drug target genes evaluated in the same subjects (Nelson, et al., submitted). To our knowledge, 623 of the 739 (84.3%) SNVs are novel (including 316 exonic (NS, S and nonsense) variants): not in either dbSNP or the PD mutation database (<http://www.molgen.ua.ac.be/PDmutDB>), and were not genotyped in a recent large association study (Ross, et al., 2011) or reported in a 2009 *LRRK2* mutation update (Paisan-Ruiz, 2009). Amongst individuals of European origin (n=12,514), 62% of all SNVs were private to a single individual (MAF=0.004%), 32% ranged from 2 copies to <1% and, 6% of all SNVs had a MAF $\geq$ 1% (Table 2). Supp. Figure S1 provides a pictorial representation of the distribution and frequency of SNVs throughout the *LRRK2* gene in the entire sample. Detailed information on all SNVs identified in this study, including quality control measures, population frequencies, genotype counts in the different collections, sequence- and functional annotations are provided in Supp. Table S1.

Of 260 SNVs that were NS, we found that 231 were diallelic, resulting in one variant amino acid at the corresponding position, while the remaining 29 NS variants were accounted for by 13 triallelic- (two variant amino acids) and one tetra-allelic (three variant amino acids) position. Therefore, 245 amino acid positions were affected by amino acid substitutions. In addition, there were five nonsense mutations, and one of them was triallelic with both an NS (c.7361C>T (p.Ser2454Leu) and nonsense variant (c.7361C>A (p.Ser2454X) at the same position (Fig. 1). Using SIFT, which predicts whether amino acid substitutions affect protein function based on sequence homology and the physical properties of amino acids, we observed 115 of the 260 NS variants (44%) to be functional changes (SIFT scores $\geq$ 0.05). Therefore, in comparison to a study by Mort et al. (2010), *LRRK2* NS variants were twice as likely to be functional as neutral polymorphisms (MAF $\geq$ 1% in Europeans) in the UniProtKB/Swiss-Prot database (22% functional), but less likely to be functional than heritable disease-causing mutations, of which 76% were predicted to be functional. In Europeans, the average MAF of functional *LRRK2* NS variants (0.13%) was over five-fold lower than for tolerated changes (0.72%), consistent with the influence of purifying selection in keeping damaging mutations at a low frequency. Common functional NS variants in Europeans (MAF $\geq$ 0.5%) include c.1653C>G (p.Asn551Lys), c.4624C>T (p.Pro1542Ser) and c.4937T>C (p.Met1646Thr). Refer to Supp. Table S1 for data pertaining to SIFT and Polyphen annotations.

Evolutionary and population genetics of LRRK2

The 245 variant amino acid sites in the LRRK2 protein observed here were mapped to a multiple protein alignment of 14 mammalian orthologues (including human) in order to determine the degree of conservation across species and further assess their potential functional impact (Figure 2 and Supp. Figure S2). The multiple protein alignment revealed that the central catalytic region encompassing the ROC, COR and kinase domains is most conserved, with a higher proportion of invariant residues (76-88%) across mammalian species compared to the flanking protein-protein (50-58%) interaction domains (PPI); Armadillo, Ankyrin, LRR and WD40 (Supp. Figure S3). To measure selection pressure on *LRRK2* in humans, we used the approach of comparing the ratio of NS and S variants (NS:S) stratified by allele frequency (Kryukov et al., 2007). For the entire gene, we observed a trend of decreasing NS:S with increasing variant frequency, which is suggestive

of purifying selection against damaging NS variants (Table 3). To determine whether the central core of *LRRK2* (ROC-COR-Kinase) is differentially affected by purifying selection in comparison to the flanking domains, we then compared NS:S stratified by variant frequency and also by location – in the ROC-COR-Kinase domains or not. A consistent, yet not significant trend was observed, with NS:S lower (OR range 0.4-0.88) in the central domains compared to the flanking domains for every variant frequency category (Supp. Figure S3). Therefore, the pattern of coding variation in humans is consistent with the pattern of variation across mammalian species and together these converging data indicate that the central core of *LRRK2* is less tolerant of functional NS variants than the flanking domains.

Using a range of different population genetic summary statistics we evaluated the genetic diversity of *LRRK2* in European and African American subjects in comparison to 201 other drug target genes analysed in the same samples (Nelson, et al., submitted). We found that *LRRK2* was not an outlier with respect to any of the summary statistics examined (the most extreme empirical p-value was 0.06) (Supp. Figure S4). In addition, we examined the distribution of F_{ST} values, a measure of differentiation between variant sites in *LRRK2* within Europe and found that it did not differ significantly (Kolmogorov-Smirnoff test, $p=0.35$) from the distribution observed for all genes. Therefore, these data together indicate that between and within populations *LRRK2* is not unusually polymorphic or differentiated compared to other drug target genes.

Impact of NS variation on the ROC and kinase domains

Pathogenic mutations in the ROC (e.g., p.Arg1441Cys/Gly/His) and kinase (e.g., p.Gly2019Ser, p.Ile2020Thr) domains of *LRRK2* have been shown to decrease GTPase activity (Lewis, et al., 2007) and increase kinase activity (West, et al., 2005), respectively. Thus, the occurrence and frequency of NS variation at key residues in these important enzymatic domains may have implications for *LRRK2* function and drugs targeting these domains. To better understand the possible impact of the variants identified in this study on the ROC and kinase domains, we mapped them *in silico* to structural predictions, highlighting molecular changes to amino acid side chains at each position (Fig. 3).

We observed 15 variants in the ROC domain, including four private mutations located within the predicted GTP/Mg²⁺-binding site (c.4030G>A (p.Gly1344Arg), c.4256A>G (p.Tyr1419Cys), c.4468G>A (p.Ala1490Thr) and c.4471A>G (p.Thr1491Ala) (Fig. 3 A and B). Of the remaining 11 variants in the ROC domain, only p.Arg1398His is common, with a MAF of 7.1% in Europeans. In the kinase domain, 18 variants were observed overall and two of these, c.5857G>C (p.Gly1953Arg) and c.5869C>T (p.Arg1957Cys) and c.5870G>T (p.Arg1957Leu), impact residues comprising the ATP-binding sites, which are compound-facing for the majority of known *LRRK2* inhibitors (Nichols, et al., 2009) (Fig. 3 C and D). Only five subjects carried the p.Arg1957Leu variant, and both p.Arg1957Cys and p.Gly1953Arg were private. The only common NS variant in the kinase domain was c.6241A>G (p.Asn2081Asp), with a MAF of 1.6% in Europeans. The SIFT method predicted both c.4193G>A (p.Arg1398His) (SIFT score=0.12) and p.Asn2081Asp (SIFT score=0.17) to be tolerated changes.

Genotype-phenotype analyses

We tested 17 common (MAF 0.5%) NS and S variants (Table 4) for association with susceptibility to 12 disease phenotypes to determine whether natural variation within *LRRK2* might provide insight about the potential development of *LRRK2* inhibitors for human conditions beyond PD. Further, we also tested for association three groups of aggregated RVs; NS variants, a subset of NS variants predicted to be functional and

additional not-protein coding variants which are conserved across species. We restricted association analysis to subjects of white European origin, and out of 240 association tests performed (20 variant tests x 12 diseases), 19 associations reached $p < 0.05$, against 12 expected under the null hypothesis, but no association surpassed a conservative significance threshold for multiple testing correction ($0.05/240$ tests), $p=0.00021$ (Supp. Table S2).

In white European subjects, 375 SNVs were only observed in a single collection (case or control) and 358 (95.5%) of these were singletons. The remaining collection-specific variants were present in two ($n=10$), three ($n=4$), four ($n=1$), five ($n=1$) and six ($n=1$) copies, respectively. All SNVs with counts >2 were observed in the largest collections (CoLaus controls and COPD cases), as might be expected by chance, and the combined carrier rate of all collection-specific variants ranged from 1.8%-5% across the different collections (refer to Supp. Table S1).

We identified 17 carriers (0.13%) of four pathogenic *LRRK2* mutations amongst 12,623 white European subjects (Table 5). pGly2019Ser ($n=9$) was the most commonly observed mutation, with p.Arg1628Pro ($n=6$) next most common and, p.Arg1441Cys ($n=1$) and p.Gly2385Arg ($n=1$) observed once each, respectively.

A recent study by Zhao et al. showed a modest predisposing effect of p.Arg1628Pro in Chinese Asian AD patients (Zhao, et al., 2011). Here we also observed non-significant enrichment of p.Arg1628Pro in AD cases (0.44%), compared to all other European collections combined (0.048%) (Table 5). Two of the three AD carriers had relatively early age at onset (AAO): subject #1 with AAO 56 years (APOE $\epsilon 4$ negative), subject #2 with AAO 63 years (APOE $\epsilon 4$ -positive) and, subject #3 with late onset AD (78 years) was APOE $\epsilon 4$ -positive. Sequencing of the protein coding regions of APP, *PSEN1*, *PSEN2* and *PSENEN* revealed no protein coding mutations in these subjects.

We identified five carriers of nonsense mutations; three with RA (c.2056C>T (p.Gln686X), c.4000C>T (p.Arg1334X), c.7361C>A (p.Ser2454X)), one with MS (c.5977C>T (p.Arg1993X)) and another with dyslipidemia (c.5683C>T (p.Arg1895X)). Amongst three collections (AD, Epilepsy and IBS), where 1) ethical consents permitted genetic evaluations relevant to other conditions 2) information on family history of PD was available, we set out to determine whether pathogenic and other selected *LRRK2* NS variants were enriched in those reporting a family history of PD in comparison to those who did not. Overall, amongst 1279 individuals, 88 (6.9%) reported a family of PD; 69 AD cases, 14 IBS cases and 5 epilepsy cases, and of those, 11 with AD had a self-reported diagnosis of PD. None of the 88 subjects carried pathogenic *LRRK2* variants, and of those NS variants observed (p.Asn551Lys, p.Ile723Val, p.Ile1371Val, p.Arg1398His, p.Arg1514Gln, p.Ser1647Thr, p.Asn2081Asp, p.Met2397Thr), none of them showed a predisposing effect ($p<0.05$) in comparison to controls (data not shown).

Geographical distribution and haplotype analysis of *LRRK2* variants in Europeans

We used genome-wide SNP data and PCA to study the geographical distribution of European carriers of pathogenic variants and selected NS *LRRK2* variants (Fig. 4). Carriers of the most common of the pathogenic variants, p.Gly2019Ser, were distributed along a North West to South East gradient down through Europe, which, taking into account sampling density, supports previous work showing that Southern Europe and the Middle East have the highest population prevalence of this variant (Healy, et al., 2008).

Two pathogenic variants, p.Arg1628Pro and p.Gly2385Arg, were identified in our sample that have previously only been observed in Asians of Chinese origin (Lu, et al., 2008; Ross, et al., 2008). We observed that all six p.Arg1628Pro carriers originated from North-Western

Europe, and the p.Gly2385Arg carrier originated from Northern Europe, somewhere in the vicinity of Germany (DE).

The geographic localization of p.Arg1628Pro raises the question of whether this mutation arose independently in Asia and Europe. For four of six p.Arg1628Pro carriers with *LRRK2* sequencing data available (for the other two carriers, PC scores were imputed), haplotypes were inferred to determine whether the p.1628Pro allele in these European individuals occurred as part of the same extended Asian haplotype as that observed previously by Ross and colleagues (Ross, et al., 2008). While we found that all four European carriers of the p.1628Pro allele carried the same extended haplotype, spanning the entire length (144 kb) of the *LRRK2* gene (data not shown), this extended haplotype was substantially different than the Asian haplotype described previously (Table 6), indicating an independent European origin, and there was no evidence of relationship between them. Further haplotype analysis involving a single carrier of an alternate, histidine encoding minor allele at the 1628 amino acid position (c.4883G>A, p.Arg1628His), indicated that this variant is actually part of the same extended haplotype that encompasses the p.1628Pro allele in Asian PD patients. In this instance, it is most likely that the His allele arose from mutation of the wildtype (Arg) allele on the same ancestral haplotype background that encompasses the Pro allele in Asians.

DISCUSSION

Using next generation sequencing methods, we have generated DNA sequence from 14,002 individuals for the entire protein coding- and proximal regulatory regions of the *LRRK2* gene, making this the most detailed genetic characterisation of *LRRK2* to date, and the most systematic evaluation of this gene in diseases other than PD.

Here we have identified 623 novel SNVs in *LRRK2*, including 316 exonic variants. All variants identified in this study have been submitted to dbSNP to enable public access. Most of the variants we have identified are private to a single individual in this study population, and relatively few are common. Comparison with 201 other drug target genes showed that *LRRK2* has a similar rate of variation to other drug target genes (Nelson, et al., submitted), and that it is not an outlier in terms of genetic diversity, and is not especially differentiated between European populations.

LRRK2 is a large multi-domain protein with a central enzymatic region comprising the ROC and kinase domains. The biology of *LRRK2* is not well understood, but pathogenic mutations tend to cluster in these central domains and the COR region which links them. Using two complementary approaches we showed that the central enzymatic domains of *LRRK2* are more conserved across mammals than the flanking domains and propose that purifying selection is likely acting to keep potentially damaging NS variation at a low frequency in the central region in humans. Therefore, together these data support the proposal by Cookson, that *LRRK2* GTPase (ROC) and kinase activities are pathogenically important (Cookson, 2010). Further analysis of the location and frequency of amino acid substitutions in the ROC and kinase domains indicated that while naturally-occurring very RVs may impact the active site of these enzymatic domains, there are only two CVs in these regions and neither is predicted to have a functional effect. Biochemical studies will be required to directly assess the impact of these variations on *LRRK2* function and potentially also drug-target interaction.

Genetic association between *LRRK2* and immune phenotypes (Barrett, et al., 2008; Danoy, et al., 2010; Zhang, et al., 2009) suggest that *LRRK2* may have pleiotropic effects and cell-specific functions. The purpose of this study was to try to identify genetic association between *LRRK2* and non-PD common human diseases to provide support for expanding the

development of LRRK2 inhibitors for the treatment of other conditions beyond PD. Here, we did not identify any statistically robust associations that withstood correction for multiple testing. Aside from p.Met1646Thr and p.Pro1542Ser, all other CVs tested for association in this study can be well imputed using reference samples from HapMap or 1000G (Durbin, et al., 2010) data, meaning that existing genome-wide association scan (GWAS) datasets can be used to conduct genotype-phenotype analyses for the CVs evaluated here.

All 12 case-control analyses we performed had ~80% power to detect association at the multiple testing corrected threshold of significance ($p=0.00021$) for an effect size, odds ratio (OR)=2 for CVs with a MAF ≥ 0.1 (refer to Supp. Methods). For all but two of the case collections (IBS and epilepsy) power was $>80\%$ to detect CV and aggregated RV associations with a MAF of 0.02 at an OR=2. Therefore, we can be reasonably confident that we have not failed to detect any CV associations of effect OR >2 . For RVs, we attempted to maximise power by aggregating genotype counts for different variants based on three functional annotations. Better statistical approaches for the analysis of RVs in association studies are the subject of much research at present (Bansal, et al., 2010), and it is plausible that our approach may have failed to detect a modest association due to the pooling strategy for RVs, which does not discriminate between loss-of-function, gain-of-function or, variants of neutral effect.

We investigated association of *LRRK2* NS variants in relation to family history of PD in three collections ascertained for AD, epilepsy and IBS. Amongst 88 individuals reporting a family history of PD, we identified 11 AD cases with self-reported PD. Given the potential pathophysiological overlap between AD and PD, and the average age of AD cases in this study (72 years), it is perhaps not surprising that PD (or symptoms of) was more prevalent (1.6%) in this collection than the others. However, given the small number of individuals involved, it is difficult to draw any substantive conclusions from this sub-group analysis.

To our knowledge, this is the first study to identify the p.Arg1628Pro mutation in non-Asians, and the first to show that this position is triallelic. Using genome-wide SNP data, and PCs to estimate the geographical origin of *LRRK2* NS variant carriers, we showed that the p.Arg1628Pro variant is part of a different extended haplotype in Europeans than in Asian PD patients of Chinese origin (Ross, et al., 2008), leading us to propose that European carriers of this variant are likely descended from a different common founder, independent of the Asian mutation. None of our European p.Arg1628Pro carriers had PD or a family history of PD, but we observed an approximately 10-fold increase in the carrier rate of p.Arg1628Pro in AD cases in comparison to all other case and control collections combined. While this latter observation is based on only six European p.Arg1628Pro carriers, three with AD, it provides support for the results of a small genetic study showing a suggestive predisposing effect of p.Arg1628Pro in Chinese AD patients (Zhao, et al., 2011).

In conclusion, we found no conclusive genetic evidence in this study to support developing LRRK2 therapies for the conditions evaluated here; however, this work illustrates how massively parallel sequencing in humans can enable the study of evolutionary and population genetics, and how it can be used to reveal fundamental characteristics of clinically important genes, which may translate into clinically meaningful outcomes in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank study subjects for their participation, and would like to acknowledge the work of the many collaborators, including clinicians and their research teams, academic consultants and others who have contributed to the recruitment and characterization of study subjects. We thank the team who prepared samples for sequencing, with particular contributions from Jon Charnecki, Mary Ellyn Volk, Dennis Duran, David Briley and Karen King. Ermias Waldu conducted the singleton capillary sequencing work. Our gratitude also goes to Anita Nelsen, Sonal Buhta-Halburnt, Leslie Amos and Julia Forte for their work with consent review and external collaborators, and to Mathias Chiano, Brian Reck, Paul Newcombe, Xiangyang Kong and Claudio Verzilli for their assistance with statistical analysis. The sequencing work was made possible by contributions from Geng Tian, Hui Jiang, Zheng Su, Xiao Sun, Lin Yang and Xiuqing Zhang at the Beijing Genome Institute. J.N. is supported by the Searle Scholars Program. D.K. is supported by an NIH Genome Analysis training grant. GlaxoSmithKline authors of this manuscript hold company shares.

REFERENCES

- Aasly JO, Vilariño-Güell C, Dachsel JC, Webber PJ, West AB, Haugarvoll K, Johansen KK, Toft M, Nutt JG, Payami H. Novel pathogenic LRRK2 p.Asn1437His substitution in familial Parkinson's disease. *Mov Dis.* 2010; 25(13):2156–63. others.
- Alegre-Abarrategui J, Christian H, Lufino MM, Mutihac R, Venda LL, Ansorge O, Wade-Martins R. LRRK2 regulates autophagic activity and localizes to specific membrane microdomains in a novel human genomic reporter cellular model. *Hum Mol Genet.* 2009; 18(21):4022–34. [PubMed: 19640926]
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller M, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–3402. [PubMed: 9254694]
- Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics.* 2006; 22(2):195–201. [PubMed: 16301204]
- Assimes TL, Holm H, Kathiresan S, Reilly MP, Thorleifsson G, Voight BF, Erdmann J, Willenborg C, Vaidya D, Xie C, Patterson CC, Morgan TM, Burnett MS, Li M, Hlatky MA, Knowles JW, Thompson JR, Absher D, Iribarren C, Go A, Fortmann SP, Sidney S, Risch N, Tang H, Myers RM, Berger K, Stoll M, Shah SH, Thorgeirsson G, Andersen K, Havulinna AS, Herrera JE, Faraday N, Kim Y, Kral BG, Mathias RA, Ruczinski I, Sukhtitipat B, Wilson AF, Yanek LR, Becker LC, Linsel-Nitschke P, Lieb W, König IR, Hengstenberg C, Fischer M, Stark K, Reinhard W, Winogradow J, Grassl M, Grosshennig A, Preuss M, Schreiber S, Wichmann HE, Meisinger C, Yee J, Friedlander Y, Do R, Meigs JB, Williams G, Nathan DM, MacRae CA, Qu L, Wilensky RL, Matthai WH Jr, Qasim AN, Hakonarson H, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Knouff CW, Waterworth DM, Walker MC, Mooser VE, Marrugat J, Lucas G, Subirana I, Sala J, Ramos R, Martinelli N, Olivieri O, Trabetti E, Malerba G, Pignatti PF, Guiducci C, Mirel D, Parkin M, Hirschhorn JN, Asselta R, Duga S, Musunuru K, Daly MJ, Purcell S, Eifert S, Braund PS, Wright BJ, Balmforth AJ, Ball SG, Myocardial Infarction Genetics Consortium; Wellcome Trust Case Control Consortium. Cardiogenics Ouwehand WH, Deloukas P, Scholz M, Cambien F, Hage A, Scheffold T, Salomaa V, Girelli D, Granger CB, Peltonen L, McKeown PP, Altshuler D, Melander O, Devaney JM, Epstein SE, Rader DJ, Elosua R, Engert JC, Anand SS, Hall AS, Ziegler A, O'Donnell CJ, Spertus JA, Siscovick D, Schwartz SM, Becker D, Thorsteinsdottir U, Stefansson K, Schunkert H, Samani NJ, Quertermous T. Lack of association between the Trp719Arg polymorphism in kinesin-like protein-6 and coronary artery disease in 19 case-control studies. *J Am Coll Cardiol.* 2010; 56(19):1552–63. [PubMed: 20933357]
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller M, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–3402. [PubMed: 9254694]
- Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet.* 2010; 11(11):773–85. [PubMed: 20940738]
- Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, Radue EW, Lindberg RL, Uitendaa BM, Johnson MR, Angelakopoulou A, Hall L, Richardson JC, Prinjha RK, Gass A, Geurts JJ, Kragt J, Sombekke M, Vrenken H, Qualley P, Lincoln RR, Gomez R, Caillier SJ, George

- MF, Mousavi H, Guerrero R, Okuda DT, Cree BA, Green AJ, Waubant E, Goodin DS, Pelletier D, Matthews PM, Hauser SL, Kappos L, Polman CH, Oksenberg JR. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet.* 2009; 18(4): 767–78. [PubMed: 19010793]
- Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JJ, Schumm LP, Steinhart AH, Targan SR, Xavier RJ, NIDDK IBD Genetics Consortium; Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossum A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E, Belgian-French IBD Consortium; Wellcome Trust Case Control Consortium. Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghori J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield J, Jewell D, Satsangi J, Mathew CG, Parkes M, Georges M, Daly MJ. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet.* 2008; 40(8):955–62. [PubMed: 18587394]
- Cookson MR. The role of leucine-rich repeat kinase 2 (LRRK2) in Parkinson's disease. *Nat Rev Neurosci.* 2010; 11(12):791–7. [PubMed: 21088684]
- Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14:1188–1190. [PubMed: 15173120]
- Danoy P, Pryce K, Hadler J, Bradbury LA, Farrar C, Pointon J, Ward M, Weisman M, Reveille JD, Wordworth BP, Stone MA, Spondyloarthritis Research Consortium of Canada. Maksymowych WP, Rahman P, Gladman D, Inman RD, Brown MA. Association of variants at 1q32 and STAT3 with ankylosing spondylitis suggests genetic overlap with Crohn's disease. *PLoS Genet.* 2010; 6(12):e1001195. [PubMed: 21152001]
- Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467(7319): 1061–73. [PubMed: 20981092]
- Firmann M, Mayor V, Vidal PM, Bochud M, Pecoud A, Hayoz D, Paccaud F, Preisig M, Song KS, Yuan X, Danoff TM, Stirnadel HA, Waterworth D, Mooser V, Waeber G, Vollenweider P. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord.* 2008; 8:6. [PubMed: 18366642]
- Francks C, Tozzi F, Farmer A, Vincent JB, Rujescu D, St Clair D, Muglia P. Population-based linkage analysis of schizophrenia and bipolar case-control cohorts identifies a potential susceptibility locus on 19q13. *Mol Psychiatry.* 2010; 15(3):319–25. [PubMed: 18794890]
- Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics.* 1993; 133:693–709. [PubMed: 8454210]
- Gardet A, Benita Y, Li C, Sands BE, Ballester I, Stevens C, Korzenik JR, Rioux JD, Daly MJ, Xavier RJ, Podolsky DK. LRRK2 is involved in the IFN-gamma response and host response to pathogens. *J Immunol.* 2010; 185(9):5577–85. [PubMed: 20921534]
- Healy DG, Falchi M, O'Sullivan S, Bonifati V, Durr A, Bressman S, Brice A, Aasly J, Zabetian CP, Goldwurm S, Ferreira JJ, Tolosa E, Kay DM, Klein C, Williams DR, Marras C, Lang AE, Wszolek ZK, Berciano J, Schapira AH, Lynch T, Bhatia KP, Gasser T, Lees AJ, Wood NW, International LRRK2 Consortium. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol.* 2008; 7(583):590.
- Heinzen EL, Radtke RA, Urban TJ, Cavalleri GL, Depondt C, Need AC, Walley NM, Nicoletti P, Ge D, Catarino CB, Duncan JS, Kasperavicius D, Tate SK, Caboclo LO, Sander JW, Clayton L, Linney KN, Shianna KV, Gumbs CE, Smith J, Cronin KD, Maia JM, Doherty CP, Pandolfo M, Leppert D, Middleton LT, Gibson RA, Johnson MR, Matthews PM, Hosford D, Kälviäinen R, Eriksson K, Kantanen AM, Dorn T, Hansen J, Krämer G, Steinhoff BJ, Wieser HG, Zumsteg D, Ortega M, Wood NW, Huxley-Jones J, Mikati M, Gallentine WB, Husain AM, Buckley PG, Stallings RL, Podgoreanu MV, Delanty N, Sisodiya SM, Goldstein DB. Rare deletions at 16p13.11 predispose to a diverse spectrum of sporadic epilepsy syndromes. *Am J Hum Genet.* 2010; 86(5):707–18. [PubMed: 20398883]

- Kasperaviciute D, Catarino CB, Heinzen EL, Depondt C, Cavalleri GL, Caboclo LO, Tate SK, Jamnadas-Khoda J, Chinthapalli K, Clayton LM, Shianna KV, Radtke RA, Mikati MA, Gallentine WB, Husain AM, Alhusaini S, Leppert D, Middleton LT, Gibson RA, Johnson MR, Matthews PM, Hosford D, Heuser K, Amos L, Ortega M, Zumsteg D, Wieser HG, Steinhoff BJ, Krämer G, Hansen J, Dorn T, Kantanen AM, Gjerstad L, Peuralinna T, Hernandez DG, Eriksson KJ, Kälviäinen RK, Doherty CP, Wood NW, Pandolfo M, Duncan JS, Sander JW, Delanty N, Goldstein DB, Sisodiya SM. Common genetic variation and susceptibility to partial epilepsies: a genome-wide association study. *Brain*. 2010; 133(Pt 7):2136–47. [PubMed: 20522523]
- Kelley JL, Swanson WJ. Positive selection in the human genome: from genome scans to biological significance. *Annu Rev Genomics Hum Genet*. 2008; 9:143–60. [PubMed: 18505377]
- Khan NL, Jain S, Lynch JM, Pavese N, Abou-Sleiman P, Holton JL, Healy DG, Gilks WP, Sweeney MG, Ganguly M, Gibbons V, Gandhi S, Vaughan J, Eunoson LH, Katzenschlager R, Gayton J, Lennox G, Revesz T, Nicholl D, Bhatia KP, Quinn N, Brooks D, Lees AJ, Davis MB, Piccini P, Singleton AB, Wood NW. Mutations in the gene LRRK2 encoding dardarin (PARK8) cause familial Parkinson's disease: clinical, pathological, olfactory and functional imaging and genetic data. *Brain*. 2005; 128:2786–96. [PubMed: 16272164]
- Kooner JS, Chambers JC, Aguilar-Salinas CA, Hinds DA, Hyde CL, Warnes GR, Perez FJ Gomez, Frazer KA, Elliott P, Scott J, Milos PM, Cox DR, Thompson JF. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet*. 2008; 40(2):149–51. [PubMed: 18193046]
- Kraus VB, Jordan JM, Doherty M, Wilson AG, Moskowitz R, Hochberg M, Loeser R, Hooper M, Renner JB, Crane MM, Hastie P, Sundseth S, Atif U. The Genetics of Generalized Osteoarthritis (GOGO) study: study design and evaluation of osteoarthritis phenotypes. *Osteoarthritis Cartilage*. 2007; 15(2):120–7. [PubMed: 17113325]
- Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet*. 2007; 80(4):727–39. 2007. [PubMed: 17357078]
- Lewis PA, Greggio E, Beilina A, Jain S, Baker A, Cookson MR. The R1441C mutation of LRRK2 disrupts GTP hydrolysis. *Biochem Biophys Res Commun*. 2007; 357(3):668–71. [PubMed: 17442267]
- Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, Hosford D, Barnes MR, Briley JD, Borrie M. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol*. 2008a; 65(1):45–53. others. [PubMed: 17998437]
- Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics*. 2008b; 24(5):713–4. [PubMed: 18227114]
- Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K. SNP detection for massively parallel whole-genome resequencing. *Genome Res*. 2009; 19(6):1124–32. [PubMed: 19420381]
- Li L, Li Y, Browning SR, Browning BL, Slater AJ, Kong X, Aponte JL, Mooser VE, Chisoe SL, Whittaker JC, Nelson MR, Ehm MG. Performance of Genotype Imputation for Rare Variants Identified in Exons and Flanking Regions of Genes. *PLoS ONE*. 2011; 6(9):e24945. [PubMed: 21949800]
- Ling H, Waterworth DM, Stirnadel HA, Pollin TI, Barter PJ, Kesaniemi YA, Mahley RW, McPherson R, Waeber G, Bersot TP, Cohen JC, Grundy SM, Mooser VE, Mitchell BD. Genome-wide linkage and association analyses to identify genes influencing adiponectin levels: the GEMS Study. *Obesity (Silver Spring)*. 2009; 17(4):737–44. [PubMed: 19165155]
- Looyenga BD, Furge KA, Dykema KJ, Koeman J, Swiatek PJ, Giordano TJ, West AB, Resau JH, Teh BT, MacKeigan JP. Chromosomal amplification of leucine-rich repeat kinase-2 (LRRK2) is required for oncogenic MET signaling in papillary renal and thyroid carcinomas. *Proc Natl Acad Sci U S A*. 2011; 108(4):1439–44. [PubMed: 21220347]
- Lu CS, Wu-Chou YH, Van Doeselaar M, Simons EJ, Chang HC, Breedveld GJ, Di Fonzo A, Chen RS, Weng YH, Lai SC, Oostra BA, Bonifati V. The LRRK2 Arg1628Pro variant is a risk factor for Parkinson's disease in the Chinese population. *Neurogenetics*. 2008; 9:271–276. [PubMed: 18716801]
- Mort M, Evani US, Krishnan VG, Kamati KK, Baenziger PH, Bagchi A, Peters BJ, Sathyesh R, Li B, Sun Y, Xue B, Shah NH, Kann MG, Cooper DN, Radivojac P, Mooney SD. In silico functional

- profiling of human disease-associated and polymorphic amino acid substitutions. *Hum Mutat.* Mar; 2010 31(3):335–46. [PubMed: 20052762]
- Muglia P, Tozzi F, Galwey NW, Francks C, Upmanyu R, Kong XQ, Antoniadis A, Domenici E, Perry J, Rothen S, Vandeleur CL, Mooser V, Waeber G, Vollenweider P, Preisig M, Lucae S, Müller-Myhsok B, Holsboer F, Middleton LT, Roses AD. Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol Psychiatry.* 2010; 15(6):589–601. [PubMed: 19107115]
- Nelson, MR.; Wegmann, D.; Ehm, MG.; Kessner, D.; St. Jean, PL.; Verzilli, C.; Shen, J.; Tang, Z.; Bacanu, S.; Fraser, D.; Warren, L.; Aponte, J.; Zawistowski, M.; Liu, X.; Zhang, H.; Zhang, Y.; Li, J.; Li, Y.; Li, L.; Woollard, P.; Topp, S.; Hall, MD.; Nangle, K.; Wang, J.; Abecasis, G.; Cardon, LR.; Zöllner, S.; Whittaker, JC.; Chissole, SL.; Novembre, J.; Mooser, VE. An abundance of rare functional variants in 202 drug target genes sequenced in 14002 people. *Science* submitted
- Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003; 31(13):3812–4. [PubMed: 12824425]
- Nicholas KB, Nicholas HB, Deerfield DW. GeneDoc: Analysis and visualization of genetic variation. *EMBNEW NEWS.* 1997;4.
- Nichols RJ, Dzamko N, Huttie JE, Cantley LC, Deak M, Moran J, Bamborough P, Reith AD, Alessi DR. Substrate specificity and inhibitors of LRRK2, a protein kinase mutated in Parkinson's disease. *Biochem J.* 2009; 424(1):47–60. [PubMed: 19740074]
- Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, Sninsky J, Adams MD, Cargill M. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 2005; 3:e170. [PubMed: 15869325]
- Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. Recent and ongoing selection in the human genome. *Nat Rev Genet.* 2007; 8(11):857–68. [PubMed: 17943193]
- Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR. Genes mirror geography within Europe. *Nature.* 2008; 456(7218):98–101. [PubMed: 18758442]
- Nuytemans K, Meeus B, Crosiers D, Brouwers N, Goossens D, Engelborghs S, Pals P, Pickut B, Van den Broeck M, Corsmit E, Cras P, De Deyn PP, Del-Favero J, Van Broeckhoven C, Theuns J. Relative contribution of simple mutations vs. copy number variations in five Parkinson disease genes in the Belgian population. *Hum Mutat.* 2009; 30(7):1054–61. [PubMed: 19405094]
- Nuytemans K, Theuns J, Cruts M, Van Broeckhoven C. Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. *Hum Mutat.* 2010; 31(7):763–80. [PubMed: 20506312]
- Oksenberg JR, Barcellos LF, Cree BA, Baranzini SE, Bugawan TL, Khan O, Lincoln RR, Sverdlin A, Mignot E, Lin L, Goodin D, Erlich HA, Schmidt S, Thomson G, Reich DE, Pericak-Vance MA, Haines JL, Hauser SL. Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *Am J Hum Genet.* 2004; 74(1):160–7. [PubMed: 14669136]
- Paisán-Ruiz C, Jain S, Evans EW, Gilks WP, Simón J, van der Brug M, de Munain A López, Aparicio S, Gil AM, Khan N, Johnson J, Martinez JR, Nicholl D, Carrera IM, Pena AS, de Silva R, Lees A, Martí-Massó JF, Pérez-Tur J, Wood NW, Singleton AB. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron.* 2004; 44(4):595–600. [PubMed: 15541308]
- Paisán-Ruiz C, Nath P, Washecka N, Gibbs JR, Singleton AB. Comprehensive analysis of LRRK2 in publicly available Parkinson's disease cases and neurologically normal controls. *Hum Mutat.* 2008; 29(4):485–90. [PubMed: 18213618]
- Paisan-Ruiz C. LRRK2 gene variation and its contribution to Parkinson disease. *Hum Mutat.* 2009; 30(8):1153–60. [PubMed: 19472409]
- Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, Ruppert A, Carlsen KC Lødrup, Roses A, Anderson W, Rennard SI, Lomas DA, Silverman EK, Goldstein DB, ICGN Investigators. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.* 2009; 5(3):e1000421. [PubMed: 19300482]

- Preisig M, Waeber G, Vollenweider P, Bovet P, Rothen S, Vandeleur C, Guex P, Middleton L, Waterworth D, Mooser V, Tozzi F, Muglia P. The PsyCoLaus study: methodology and characteristics of the sample of a population-based survey on psychiatric disorders and their association with genetic and cardiovascular risk factors. *BMC Psychiatry*. 2009; 9:9. [PubMed: 19292899]
- Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002; 30(17):3894–900. [PubMed: 12202775]
- Ross OA, Wu YR, Lee MC, Funayama M, Chen ML, Soto AI, Mata IF, Chen GJL, Chen CM, Tang M, Zhao Y, Hattori N, Farrer MJ, Tan EK, Wu RM. Analysis of Lrrk2 R1628P as a risk factor for Parkinson's Disease. *Ann Neurol*. 2008; 64:88–96. [PubMed: 18412265]
- Ross OA, Soto-Ortolaza AI, Heckman MG, Aasly JO, Abahuni N, Annesi G, Bacon JA, Bardien S, Bozi M, Brice A, Brighina L, Van Broeckhoven C, Carr J, Chartier-Harlin MC, Dardiotis E, Dickson DW, Diehl NN, Elbaz A, Ferrarese C, Ferraris A, Fiske B, Gibson JM, Gibson R, Hadjigeorgiou GM, Hattori N, Ioannidis JP, Jasinska-Myga B, Jeon BS, Kim YJ, Klein C, Kruger R, Kyrtazi E, Lesage S, Lin CH, Lynch T, Maraganore DM, Mellick GD, Mutez E, Nilsson C, Opala G, Park SS, Puschmann A, Quattrone A, Sharma M, Silburn PA, Sohn YH, Stefanis L, Tadic V, Theuns J, Tomiyama H, Uitti RJ, Valente EM, van de Loo S, Vassilatis DK, Vilarinho-Güell C, White LR, Wirdefeldt K, Wszolek ZK, Wu RM, Farrer MJ, Genetic Epidemiology Of Parkinson's Disease (GEO-PD) Consortium. Association of LRRK2 exonic variants with susceptibility to Parkinson's disease: a case-control study. *Lancet Neurol*. 2011; 10(10):898–908. [PubMed: 21885347]
- Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*. 2003; 31:3381–3385. [PubMed: 12824332]
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005; 15(8):1034–50. [PubMed: 16024819]
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989; 123:585–595. [PubMed: 2513255]
- Tan EK, Zhao Y, Skipper L, Tan MG, Di Fonzo A, Sun L, Fook-Chong S, Tang S, Chua E, Yuen Y, Tan L, Pavanni R, Wong MC, Kolatkar P, Lu CS, Bonifati V, Liu JJ. The LRRK2 Gly2385Arg variant is associated with Parkinson's disease: genetic and functional evidence. *Hum Genet*. 2007; 120(6):857–63. [PubMed: 17019612]
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994; 22:4673–4680. [PubMed: 7984417]
- Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, Hagan G, Knobil K, Lomas DA, MacNee W, Silverman EK, Tal-Singer R, ECLIPSE investigators. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J*. 2008; 31(4): 869–73. [PubMed: 18216052]
- Vignal C, Bansal AT, Balding DJ, Binks MH, Dickson MC, Montgomery DS, Wilson AG. Genetic association of the major histocompatibility complex with rheumatoid arthritis implicates two non-DRB1 loci. *Arthritis Rheum*. 2009; 60(1):53–62. [PubMed: 19116923]
- Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 1975; 7:256–276. [PubMed: 1145509]
- West AB, Moore DJ, Biskup S, Bugayenko A, Smith WW, Ross CA, Dawson VL, Dawson TM. Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc Natl Acad Sci U S A*. 2005; 102(46):16842–7. [PubMed: 16269541]
- WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447(7145):661–78. [PubMed: 17554300]
- Wyszynski DF, Waterworth DM, Barter PJ, Cohen J, Kesaniemi YA, Mahley RW, McPherson R, Waeber G, Bersot TP, Sharma SS, Nolan V, Middleton LT, Sundseth SS, Farrer LA, Mooser V, Grundy SM. Relation between atherogenic dyslipidemia and the Adult Treatment Program-III definition of metabolic syndrome (Genetic Epidemiology of Metabolic Syndrome Project). *Am J Cardiol*. 2005; 95(2):194–8. [PubMed: 15642551]

- Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, Cui Y, Yan XX, Yang HT, Yang RD, Chu TS, Zhang C, Zhang L, Han JW, Yu GQ, Quan C, Yu YX, Zhang Z, Shi BQ, Zhang LH, Cheng H, Wang CY, Lin Y, Zheng HF, Fu XA, Zuo XB, Wang Q, Long H, Sun YP, Cheng YL, Tian HQ, Zhou FS, Liu HX, Lu WS, He SM, Du WL, Shen M, Jin QY, Wang Y, Low HQ, Erwin T, Yang NH, Li JY, Zhao X, Jiao YL, Mao LG, Yin G, Jiang ZX, Wang XD, Yu JP, Hu ZH, Gong CH, Liu YQ, Liu RY, Wang DM, Wei D, Liu JX, Cao WK, Cao HZ, Li YP, Yan WG, Wei SY, Wang KJ, Hibberd ML, Yang S, Zhang XJ, Liu JJ. Genomewide association study of leprosy. *N Engl J Med*. 2009; 361(27):2609–18. [PubMed: 20018961]
- Zhao Y, Ho P, Yih Y, Chen C, Lee WL, Tan EK. LRRK2 variant associated with Alzheimer's disease. *Neurobiol Aging*. 2011; 32(11):1990–3. 2011. [PubMed: 20018409]
- Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, Kachergus J, Hulihan M, Uitti RJ, Calne DB, Stoessl AJ, Pfeiffer RF, Patenge N, Carbajal IC, Vieregge P, Asmus F, Müller-Mylchok B, Dickson DW, Meitinger T, Strom TM, Wszolek ZK, Gasser T. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron*. 2004; 44(4):601–7. [PubMed: 15541309]

Figure 1. LRRK2 amino acid substitutions observed in 14,002 individuals

A linear schematic diagram of human LRRK2 showing amino acid changes resulting from 260 NS and five nonsense variants (denoted by an asterisk), relative to the multi-domain structure of the LRRK2 protein. All amino acid variant alleles are indicated at those positions with greater than two minor alleles (e.g., p.Gln103Lys/His (Q103KH)). Nine common amino acid variants with minor allele frequency (MAF) > 0.5% in subjects of European origin are underlined. Variants are italicised where the SIFT score (< 0.05) for the minor allele is predicted to be functionally damaging. For those variants with more than one minor allele, only damaging alleles are italicised (e.g., p.Ala419Thr/Val (A419TV)). Variants with strong evidence for a pathogenic effect on PD are indicated in bold. Refer to Supp. Methods for a detailed explanation of the approach used to model LRRK2 domains. Domain abbreviations are; leucine-rich repeats (LRR), GTPase Ras of complex (ROC) and, c-terminal of ROC (COR).

Figure 2. Summary of LRRK2 protein sequence conservation across species at amino acid residues affected by non-synonymous variation in humans

All 245 NS variant positions in human LRRK2 are depicted by “logos”, excluding nonsense mutations. Logos were created by the WebLogo tool (Crooks, et al., 2004) based on a ClustalW (Thompson, et al., 1994) multiple alignment of the human protein and 13 mammalian orthologues (chimpanzee, orangutan, baboon, marmoset, red-bellied titi, horse, panda, cow, pig, armadillo, rat, mouse and opossum: Refer to Supp. Figure S2). Only species that had full length LRRK2 sequences available in public databases were included. The human major and minor alleles are indicated below each position, e.g., C9G (p.Cys9Gly) indicates cysteine is the major, and glycine the minor allele at the 9th amino acid in the LRRK2 protein. The combined height of each position is proportional to the degree of conservation found amongst the species that were assessed. Therefore, the maximum height of 3.2 bits indicates a residue that is invariant across species. The frequency of amino acids at each position across species is relative to the size of each coloured single amino acid letter. The central catalytic region includes the ROC (amino acids (aa) 1335-1512), COR (1526-1846) and kinase (1856-2136) domains, respectively.

Figure 3. Impact of LRRK2 amino variants on structural predictions of the ROC and kinase domains

For the ROC and kinase domains, respectively, naked backbone (A and C) and ribbon format (B and D) structures are depicted, with backbone structures showing the position of amino acid variants, including the side chains for both the major (green) and minor (red) alleles. The GTP/Mg²⁺ binding site of the ROC domain (A) and the ATP-binding site of the kinase domain (C) are indicated by orange segments on the naked backbone structure. For variants with asterisks, there is functional evidence of an impact on GTPase (R1441C (p.Arg1441Cys)) and kinase (G2019S (p.Gly2019Ser)) activity, respectively. Underlined variants are common, and have a MAF > 1% in Europeans. Note, that the ATP-binding pocket of the kinase domain is compound-facing residues for the majority of known LRRK2 inhibitors (Nichols, et al., 2009). Protein structures for the ROC and kinase domains were modelled on Protein database (PDB) structures 2zej and 3dkc, respectively.

Figure 4. Geographical distribution of *LRRK2* variants in Europe

Principle components (PC) analysis was performed for 19,088 unrelated European individuals (including *LRRK2* variant carriers) genotyped at 36,222 common SNPs distributed genome-wide (see methods). Grey letters represent an individual whose ancestry can be attributed to a single European country (ISO two-letter country codes; AL=Albania, AT=Austria, BA=Bosnia and Herzegovina, BE=Belgium, BG=Bulgaria, CH=Switzerland, CZ=Czech Republic, DE=Germany, DK=Denmark, EE=Estonia, ES=Spain, FR=France, GB=United Kingdom, GR=Greece, HR=Croatia, HU=Hungary, IE=Ireland, IT=Italy, LT=Lithuania, MA=Macedonia, NL=Netherlands, NO=Norway, PL=Poland, PT=Portugal, RO=Romania, RS=Serbia and Montenegro, SE=Sweden, SK=Slovakia, TR=Turkey, UA=Ukraine) or is of Jewish (Je) ancestry. For each country with five or more individuals, circles were added at the median PC position. PCs for *LRRK2* variant carriers are represented by the symbols in the legend.

Table 1

Demographic and phenotypic details of case and control collections

Phenotype	Subjects passing QC	Mean age	Gender % males	Age range	Ethnicity	Subject information
AD	687	72	40	40-97	W Cauc	All from memory referral clinics in Canada: late onset AD (Li, et al., 2008a)
Bipolar	777	47	34	20-82	W Cauc	Canada (n=374) and UK (n=403): DSM-IV or ICD-10 bipolar I or bipolar II disorder (Francks, et al., 2010)
CAD	604	55	71	25-84	W Cauc	All from the U.S.: acute coronary syndrome (Assimes, et al., 2010)
COPD#1	780	62	60	41-75	W Cauc	All from Norway: moderate to severe COPD (GOLD criteria (Pillai, et al., 2009)
COPD#2	988	65	62	41-95	W Cauc	46 centers in 12 countries (Bulgaria, Canada, Czech Republic, Denmark, Netherlands, Norway, Slovenia, Spain, UK, and U.S.): ECLIPSE cohort GOLD stage II-IV (Vestbo, et al., 2008)
Dyslipid	1643	52	60	20-83	W Cauc	862 cases and 781 controls from Australia, Canada, Finland, Turkey, Switzerland, and U.S.: cases were ascertained for atherogenic dyslipidemia (Ling, et al., 2009; Wyszynski, et al., 2005)
Epilepsy	275	44	44	18-76	W Cauc	Finland (n=164) and Switzerland (n=111): temporal lobe epilepsy (Heinzen, et al., 2010; Kasperaviciute, et al., 2010)
IBS	317	42	18	18-78	W Cauc	Canada (n=165) and U.S. (n=152): Rome II criteria + either colonoscopy/barium enema supporting diagnosis
MS#1	670	47	23	18-67	W Cauc	U.S. (n=337), Netherlands (n=158) and Switzerland (n=175): relapsing and progressive MS subtypes (Baranzini, et al., 2009)
MS#2	591	49	32	20-86	Afr Am	U.S. cases (n=339) and controls (n=252) (Oksenberg, et al., 2004)
OA	832	67	16	30-95	W Cauc	U.S. (n=394) and UK (n=298): multiple joint OA characterized both clinically and radiographically (Kraus, et al., 2007)
RA	611	62	27	20-90	W Cauc	All from the UK : 1987 American College of Rheumatology classification for RA (Vignal, et al., 2009)
Schz	1099	42	72	18-81	W Cauc	UK (n=518), Germany (n=330) and Canada (n=254): DSM-IV or ICD-10 schizophrenia criteria (Francks, et al., 2010)
Unipolar	741	51	32	20-89	W Cauc	All from Germany: recurrent major depressive disorder (Muglia, et al., 2010)
Control #1	2059	55	47	35-75	W Cauc	Switzerland: Colaus population-based sample from Lausanne, subset of total collection with detailed psychiatric, cardiovascular, and metabolic phenotype (Firmann, et al., 2008; Preisig, et al., 2009)
Control #2	1322	48	67	23-76	W Cauc, IA, mixed	UK: LOLIPOP population sample collection recruited in West London (Kooner, et al., 2008)
QC trios	6	-	-	-	W Cauc, Afr Am	Trio families (both parents and offspring) from Nigeria (n=1) and U.S. (n=1) sequenced as part of the 1000G project (Durbin, et al., 2010)
Total	14,002					

Abbreviations: Alzheimer's disease (AD), Bipolar disorder (Bipolar), coronary artery disease (CAD), chronic obstructive pulmonary disease (COPD), Dyslipid (Dyslipidemia), irritable bowel syndrome (IBS), multiple sclerosis (MS), osteoarthritis (OA), rheumatoid arthritis (RA), Schizophrenia (Schz), Unipolar depression (Unipolar), white Caucasian (W Cauc), African American (Afr Am), Indian Asian (IA), quality control (QC), 1,000 genomes project (1000G).

The COPD#2, MS#2 and, QC trios collections were not included in genotype-phenotype analyses.

Table 2Description of DNA variation in the *LRRK2* gene

Location in <i>LRRK2</i>	Prevalence in 12,514 subjects of European Caucasian origin			
	SNVs in 14, 002 subjects	Singletons, MAF=0.004%	Doubletons, to <1%	Variants, MAF 1%
5'UTR	10 (1.4)	5 (1.4)	1 (0.5)	0
Exons	392 (53.0)	209 (56.5)	98 (51.3)	16 (44.4)
NS	260 (35.2)	140 (37.8)	60 (31.4)	8 (22.2)
Synonymous	127 (17.2)	64 (17.3)	38 (19.9)	8 (22.2)
Nonsense	5 (0.7)	5 (1.4)	0	0
Splice site	29 (3.9)	16 (4.3)	7 (3.7)	1 (2.8)
Introns	216 (29.2)	96 (25.9)	59 (30.9)	13 (36.1)
3'UTR	92 (12.4)	44 (11.9)	26 (13.6)	6 (16.7)
Total	739 (100)	370 (100)	191 (100)	36 (100)

Numbers of SNVs are indicated relative to their position and effect on the *LRRK2* gene. Numbers in parentheses indicate the percentage of variations per category relative to the total variant count at the bottom of the column. For clarity, exonic variants have been sub-stratified into NS, S and, nonsense variants. To assist with interpretation, variant frequencies (e.g., singletons (observed once only with MAF=0.004%), doubletons (present in two copies) and up to 1% MAF, and variants with MAF 1%) are shown only for the subset of 12,514 subjects of white European origin. Proportions may not add up to 100% due to rounding. Variants within 20 bp of an exon/intron boundary are referred to as "Splice site" variants. Untranslated region =UTR.

Table 3

Evidence of purifying selection at the *LRRK2* locus

	NS variants	S variants	NS:S	OR 95% CI	P-value
Singletons	140	64	2.2	-	-
No singletons	68	46	1.5	0.68 (0.41-1.12)	0.11
No singletons or doubletons	47	28	1.7	0.77 (0.43-1.40)	0.39
MAF > 0.1%	11	10	1.1	0.50 (0.18-1.40)	0.15
MAF > 0.5%	9	8	1.1	0.51 (0.17-1.61)	0.19
MAF > 5%	5	8	0.6	0.29 (0.07-1.04)	0.03

Numbers of non-synonymous (NS) and synonymous (S) variants in 12,514 white Europeans are indicated for the entire *LRRK2* gene. Ratio of NS to S variants (NS:S) were calculated for different variant frequency categories because purifying selection increases with variant frequency for functionally damaging NS variants (Kryukov et al. 2007). The “Singletons” variant category is the subset of NS and S variants observed once in 12,514 white Europeans. The “No singletons” variant category is the subset of all NS and S variants that are not singletons, and the “No singletons or doubletons” category is the subset of all variants observed more than twice. Association tests were conducted for NS:S by testing variant counts in two-by-two contingency tables for each frequency category compared to “Singletons”. MAF = minor allele frequency.

Table 4Common *LRRK2* variants tested for association with 12 different diseases

cDNA	Protein	MAF
c.457T>C	p.Leu153Leu	0.394
c.1653C>G	p.Asn551Lys	0.069
c.2167A>G	p.Ile723Val	0.068
c.2857T>C	p.Leu953Leu	0.129
c.4193G>A	p.Arg1398His	0.071
c.4269G>A	p.Lys1423Lys	0.071
c.4541G>A	p.Arg1514Gln	0.009
c.4624C>T	p.Pro1542Ser	0.032
c.4872C>A	p.Gly1624Gly	0.352
c.4911A>G	p.Lys1637Lys	0.463
c.4937T>C	p.Met1646Thr	0.015
c.4939T>A	p.Ser1647Thr	0.296
c.5457T>C	p.Gly1819Gly	0.460
c.6241A>G	p.Asn2081Asp	0.016
c.6324G>A	p.Glu2108Glu	0.314
c.7155A>G	p.Gly2385Gly	0.146
c.7190T>C	p.Met2397Thr	0.337

Association analyses included 12,514 white subjects of European origin. MAF is indicated for the entire study population. cDNA position corresponds to transcript, NM_198578.3

Table 5

LRKK2 pathogenic mutation carriers stratified by collection

cDNA	Protein	Domain	AD (687)	Bp (777)	CAD (604)	COPD (1768)	Dylip cases (862)	Dylip conts (780)	Epi (275)	IBS (317)	MS (670)	OA (832)	RA (611)	SZ (1099)	Unip (741)	CoLaus (2059)	LP (541)	Total carriers (12623)	Variant Frequency (%)
c.4321C>T	p.Arg1441Cys	ROC	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0.016
c.4883G>C	p.Arg1628Pro	COR	3	0	0	0	0	1	0	0	0	1	1	0	0	0	0	6	0.048
c.6055G>A	p.Gly2019Ser	kinase	0	1	1	0	0	1	0	0	3	2	0	0	0	1	0	9	0.071
c.7153G>A	p.Gly2385Arg	WD40	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0.008
Carriers (n)																			
			3	1	1	0	0	2	0	0	3	3	1	1	0	2	0	17	
Rate (%)																			
			0.44	0.13	0.17	0	0	0.26	0	0	0.45	0.36	0.16	0.09	0	0.10	0	0.13	

The number of variant carriers per collection is indicated in parentheses for white Caucasian subjects only, and control collections are shaded. Both COPD collections (Refer to Table 1) were pooled for this analysis. Collection abbreviations: AD=Alzheimer disease, Bp=bipolar disorder, CAD=coronary artery disease, COPD=chronic obstructive pulmonary disease, Dylip=dyslipidemia (conts=controls), Epi=epilepsy, IBS=irritable bowel syndrome, MS=multiple sclerosis, OA=osteoarthritis, RA=rheumatoid arthritis, SZ=schizophrenia, Unip=Unipolar disorder, CoLaus=CoLaus and LP=LOLIPOP (white) population controls, respectively.

Table 6
Haplotype analysis of *LRKK2* c.4883G>C (p.Arg1628Pro) and c.4883G>A (p.Arg1628His) carriers

SNP ID	Chr12 Position	cDNA	Protein	MAF	European p.Arg1628Pro haplotype			R1628H haplotype	Asian p.Arg1628Pro haplotype*		
					A	B	C		D	E	F
rs10878245	38,918,058	c.457T>C	p.Leu153Leu	0.406	T	T	T	C	T	C	-
rs7955902	38,931,524	c.1102-10C>A	-	0.373	C	C	C	A	C	A	-
rs11175964	38,989,254	c.426G>A	p.Lys1423Lys	0.066	G	G	G	G	G	G	G
rs1896252	39,000,026	c.4828-31T>C	-	0.459	T	T	T	C	T	C	C
rs1427263	39,000,101	c.4872C>A	p.Gly1624Gly	0.346	A	A	A	A	A	A	A
rs33949390	39,000,112	c.4883G>C/A	p.Arg1628Pro/His	0.0004/0.0002	C	C	C	A	C	A	C
rs11176013	39,000,140	c.4911A>G	p.Lys1637Lys	0.457	A	A	A	A	A	G	G
rs11564148	39,000,168	c.4939T>A	p.Ser1647Thr	0.292	T	T	T	A	T	A	A
rs7307276	39,001,280	c.5170+23T>A	-	0.452	T	T	T	A	T	A	-
rs10878371	39,002,527	c.5457T>C	p.Gly1819Gly	0.453	T	T	T	C	T	C	-
rs10878405	39,028,521	c.6324G>A	p.Glu2108Glu	0.306	G	G	G	A	G	A	A
rs33962975	39,043,597	c.7155A>G	p.Gly2385Gly	0.136	G	G	G	A	G	A	-

* Haplotypes were inferred using BEAGLE (Li, et al., 2011) for four (A-D) carriers of p.Arg1628Pro, and one carrier (E) of p.Arg1628His, respectively. All four p.Arg1628Pro carriers share a common haplotype, which is distinct from the Asian ancestral haplotype described by Ross et al. (Ross, et al., 2008). “-” indicates no SNP data available. SNP markers delineate a chromosome 12 genomic segment 125.5 kb in length. Haplotypes for all four R1628P carriers extend the entire length (144 kb) of the *LRKK2* gene (data not shown). Chromosome position is relative to NCBI Build 36.3 of the human genome sequence, and cDNA variant positions correspond to transcript NM_198578.3.