illumina®

Comprehensive Analysis of *LRRK2* Mutation Carriers in Parkinson's Disease with the TruSeq™ Neurodegeneration Panel

Contributed by Alan Pittman, PhD, Emmeline Brown, PhD, Deborah Hughes, PhD*

Introduction

Parkinson's disease (PD) is a complex multifactorial neurodegenerative disease, affecting an estimated 7–10 million people worldwide.[1] PD is caused by an interplay of genetic, environmental, and epigenetic factors. Known genetic factors associated with PD include 18 chromosomal locations, termed *PARK* (to denote their putative link to PD) 1-18,[2] and 28 common variants across 24 loci that act as risk factors for sporadic disease.[3]

The most common single mutation linked to PD is a Gly2019Ser (c.6055G->A) missense mutation in the leucine-rich repeat kinase (*LRRK2*) gene with worldwide prevalence estimated at 1–4% of PD cases.[4] Prevalence varies greatly between different ethnic populations: North Africans (34–41%), Ashkenazi Jewish (10–25%), and European (1–2%). G2019S exhibits age-dependent incomplete penetrance and variable phenotype severity. Recent studies have estimated penetrance at 25–42.5% at age 80,[5] suggesting a significant contribution from other genetic and environmental modifiers. The heritability in age at onset (AAO) for *LRRK*2 G2019S PD has not been established, although several regions have been implicated but not independently replicated.

The TruSeq Neurodegeneration Panel is a targeted next-generation sequencing (NGS) panel that covers 118 risk genes associated with common neurodegenerative diseases, including Alzheimer's disease, PD, and others. It features a highly optimized probe set that covers 8.7 Mb of genomic content, including exons, introns, untranslated regions (UTRs), and promoter regions of candidate genes.

This application note presents results using the TruSeq Neurodegeneration Panel in AAO genetic modifier discovery. Relevant findings include the characterization of the G2019S *LRRK2* mutation haplotype in different ethnic groups and identification of potential genetic modifiers in the *LRRK2* trans haplotype or novel loci.

Methods

Library Preparation

DNA samples were obtained from 41 individuals with PD and 7 unaffected relatives of North African, Ashkenazi Jewish, and European Caucasian origin that were *LRRK2* G2019S mutation carriers. Sequencing libraries were prepared using the TruSeq Neurodegeneration Panel, following the assay protocol.

Sequencing

Prepared libraries were sequenced using paired end 2 × 150 bp reads on the HiSeq™ 3000 System.

Data Analysis

NGS data processing was based on recommendations of best practice from Illumina and the Genome Analysis Toolkit (GATK v3.7) for identifying germline single nucleotide polymorphisms (SNPs) and insertions/deletions (indels). After sequencing, the resulting data underwent rigorous quality control (QC) and genetic analysis.

Sample Gender Determination

Incongruous gender determination between genetic and clinical data indicates that an error has occurred in database management or laboratory processing. Alternatively, clinical missexing of the patient could have occurred or the patient may not have a determinate sex. In this study, sample gender was determined using the Y count markers designed in the TruSeq Neurodegeneration Panel. Samples with a Y count ≥ 40 were determined as male. Samples with a Y count < 40 were categorized as female. All samples used in this study had an expected gender.

*University College London, Department of Molecular Neuroscience, Institute of Neurology, London, UK.

LASER analysis

Principal components of ancestry (PCA) was performed to use as covariates in the association analyses (to correct for systematic ancestry differences) and as an additional QC measure to check ancestry. LASER (Locating Ancestry from SEquence Reads) software was also used to assess PCA.[6–8] LASER overcomes the issue of nongenome wide coverage by analysing off-target sequence reads that were produced as a by-product in the sequencing process. Early-stage BAM files (including off-target information) were converted using Samtools (invoked by LASER) into Pileup files. A python script converted Pileup files to a SEQ file containing all study samples, LASER then used a GENO file of references (Worldwide (imputed HGDP) reference panel was used in this project) and extracts PCAs for the study samples.

Genetic Analysis

Logistic regression, Kaplan Meier survival curve analysis, and Cox proportional hazards modeling were used in assessing AAO modification of *LRRK2* and for discovery of additional modifiers. Logistic regression was performed in PLINK on genome-wide data from the samples. One covariate for clinical sex and three principal components extracted from LASER were initially added as covariates to overcome issues of population stratification that can impact association studies.

Results

Using the TruSeq Neurodegeneration Panel, all PD samples were found to be *LRRK2* G2019S carriers and the trans *LRRK2* haplotype was defined in all cases. Variants identified with the TruSeq Neurodegeneration Panel were then searched for rare and common genetic modifiers that may influence AAO penetrance and phenotype severity.

Ancestry Analysis

PCA analysis with LASER software for ancestry determination showed that project samples cluster separately from reference samples (Figure 1A). Magnification of these sample clusters revealed distinct clustering of European Caucasion and Ashkenazi Jewish samples (Figure 1B).

Genetic Analysis

The phenotype (outcome variable) in logistic regression analysis was specified as either early-onset PD (EOPD) or late-onset PD (LOPD). EOPD was defined as < 56 years; LOPD was defined as ≥ 57 years. These categorizations were based on visual inspection of a histogram for AAO for G2019S carriers, which showed a bimodal distribution with a minimum at ~56. The Manhattan plot for logistic regression on 6282 genome-wide SNPs shows no significant associations; it did not reach the commonly accepted genome-wide significance criterion (Figure 2). The cost effectiveness of the TruSeq Neurodegeneration Panel should enable analysis of larger sample sizes needed for variant identification.
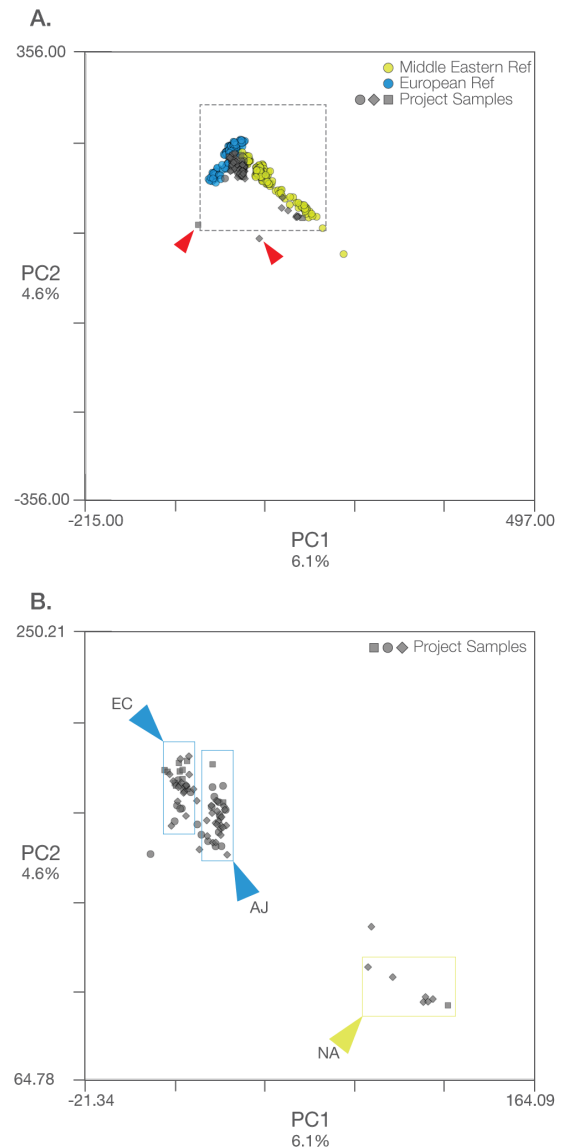
Figure 1: Principal components of ancestry plots from LASER—(A) Plot overview with reference samples shown (yellow = Middle Eastern reference; Blue = European reference; Grey = Project samples. Outliers are noted with red triangles. (B) Magnification of the project samples, gray dashed box in (A), show that North African (NA, yellow triangle) samples lie in the Middle Eastern component of the PCA. European samples appear to form two clusters (European Caucasian (EC) and Ashkenazi Jewish (AJ) (blue arrows).
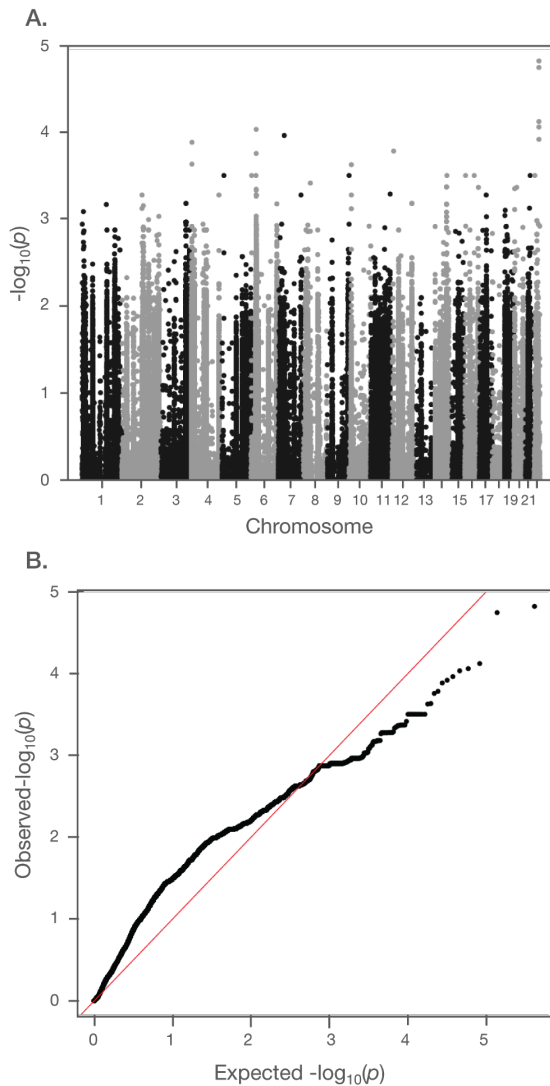
Figure 2: Logistic Regression Using 4 Covariates—(A) A Manhattan plot displays genomic coordinates on the X-axis plotted against the negative logarithm of the association p-value ($-\log_{10}(p)$) for 6282 SNPs (gray and black dots). SNPs with stronger associations have higher $-\log_{10}(p)$ values. (B) A Q-Q plot of observed versus expected $-\log_{10}(p)$ values. The linearity of the data points suggests a normal distribution.

Haplotype Analysis

All haplotypes found in this project were consistent with haplotype being the same origin. All LRRK2 G2019S haplotypes characterized in this study were the same, which is consistent with there being a founder effect, ie, a loss of genetic variation when a new population is established by a small number of individuals from a larger population.

After calling variants, LRRK2 trans haplotype data for all G2019S carriers was constructed using 24 biallelic tag SNPs, characterized during haplotype analysis. SNPs that did not deviate from the reference among the individuals included in analysis were removed. SNPs that were missing for certain individuals were imputed to the

mean. Three participants of ancestry and one principal component of sex were added. Seven SNPs yielded significant p values, however after correction for linkage disequilibrium, no single SNP remained significant.

Discussion

This project had two main aims: 1) to characterize the G2019S pathogenic haplotype in a new multiethnic cohort, 2) to find additional genetic modifiers of AAO in G2019S Parkinson's disease. While the first aim was successful, the sample sizes used in this study were not large enough to find novel genetic modifiers of AAO.

Overall, accruing larger sample sizes will be essential in addressing the question of genetic modifiers of LRRK2 G2019S disease. The cost effectiveness of the TruSeq Neurodegeneration Panel will enable analysis of the larger sample sizes needed. The current data could also be assessed in other ways: for instance, using additional techniques to analyze affected sib pairs, or in assessing intronic variation and adding in additional samples into the study as they are collected and enrolled for study in the clinic.

The TruSeq Neurodegeneration Panel offers a quick and effective means for not only pinpointing the genetic basis for many common monogenic neurodegenerative diseases, but also as an effective research tool for studying the underlying genetic architecture of more complex neurodegenerative diseases.

References

1. Tysnes OB, Storstein A. Epidemiology of Parkinson's disease. *J Neural Transm (Vienna)*. 2017;124(8):901–905.

2. Klein C, Westenberger A. Genetics of Parkinson's Disease. *Cold Spring Harb Perspect Med*. 2012;2(1):a008888.

3. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet*. 2014;46(9):989–993.

4. Li JQ, Tan L, Yu JT. The role of the LRRK2 gene in Parkinsonism. *Mol Neurodegener*. 2014;9:47.

5. Lee AJ, Wang Y, Alcalay RN, et al. Penetrance estimate of LRRK2 p.G2019S mutation in individuals of non-Ashkenazi Jewish ancestry. *Mov Disord*. 2017;32(10):1432–1438.

6. Wang C, Zhan X, Bragg-Gresham J, et al. Ancestry estimation and control of population stratification for sequence-based association studies. *Nat Genet*. 2014;46(4):409–415.

7. Wang C, Zhan X, Liang L, Abercasis GR, Lin X. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. *Am J Hum Genet*. 2015;96(6):926–937.

8. Taliun D, Chothani SP, Schonherr S, et al. LASER server: ancestry tracing with genotypes or sequence reads. *Bioinformatics*. 2017;33(13):2056–2058.

For Research Use Only. Not for use in diagnostic procedures.

illumina®