Parkinsonism and Related Disorders 20 (2014) 93–98

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis

Association mapping of the *PARK10* region for Parkinson's disease susceptibility genes

Jia Y. Wan [a], Karen L. Edwards [a], Carolyn M. Hutter [a], Ignacio F. Mata [b,c], Ali Samii [b,c], John W. Roberts [d], Pinky Agarwal [e], Harvey Checkoway [a,f], Federico M. Farin [f], Dora Yearout [b,c], Cyrus P. Zabetian [b,c,*]

[a] *Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA*
[b] *Veterans Affairs Puget Sound Health Care System, Seattle, WA, USA*
[c] *Department of Neurology, University of Washington School of Medicine, Seattle, WA, USA*
[d] *Virginia Mason Medical Center, Seattle, WA, USA*
[e] *Booth Gardner Parkinson's Care Center, Evergreen Hospital Medical Center, Kirkland, WA, USA*
[f] *Department of Environmental and Occupational Health Sciences, School of Public Health, University of Washington, Seattle, WA, USA*

ARTICLE INFO

Article history:
Received 2 July 2013
Received in revised form
20 September 2013
Accepted 1 October 2013

Keywords:
PARK10
Parkinson's disease
Replication
GWAS

ABSTRACT

Background: Previous studies indicate that as many as six genes within the *PARK10* region (*RNF11*, *UQCRH*, *HIVEP3*, *EIF2B3*, *USP24*, *ELAVL4*) might modify susceptibility or age at onset in Parkinson's disease (PD).
Methods: We sought to identify new PD susceptibility genes and to validate previously nominated candidate genes within the *PARK10* region using a two-stage design. We used data from a large, publicly-available genome-wide association study (GWAS) in the discovery stage (n = 2000 cases and 1986 controls) and data from three independent studies for the replication stage (total n = 2113 cases and 2095 controls). Marker density was increased by imputation using HapMap 3 and 1000 Genomes reference panels, and over 40,000 single nucleotide polymorphisms (SNPs) were used in the final analysis. The association between each SNP and PD was modeled using logistic regression with an additive allele dosage effect and adjusted for sex, age, and axes of geographical variation.
Results: Although the discovery stage yielded promising findings for SNPs in several novel genes, including *DAB1*, none of the results were validated in the replication stage. Furthermore, in meta-analyses across all datasets no genes within *PARK10* reached significance after accounting for multiple testing.
Conclusion: Our results suggest that common variation in the *PARK10* region is not associated with PD risk. However, additional studies are needed to assess the role of *PARK10* in modifying age at onset and to determine whether rare variants in this region might affect PD susceptibility.
Published by Elsevier Ltd.

1. Introduction

Parkinson's disease (PD) is the second-most common neurodegenerative disorder with a lifetime risk estimated at 1.6% [1]. Linkage and association studies have implicated over 20 susceptibility loci for PD, though in many instances the specific disease gene has not been clearly identified [2,3]. In 2002, two independent studies reported evidence of linkage to PD on chromosome 1p32; one to PD risk [4] and the other to age at onset [5]. This region, which was designated *PARK10*, spans 19.2 megabases (Mb) (40.2–59.4 Mb on NCBI build 36) and contains approximately 200 genes. Several subsequent studies [6–8] nominated potential PD genes within the *PARK10* region including the: (1) ring-finger protein 11 gene (*RNF11*), (2) ubiquinol-cytochrome c reductase hinge protein gene (*UQCRH*), (3) human immunodeficiency virus enhancer-binding protein 3 gene (*HIVEP3*), (4) gamma-subunit of the translation initiation factor EIF2B gene (*EIF2B3*), (5) ubiquitin-specific protease 24 gene (*USP24*), and (6) embryonic lethal, abnormal vision, Drosophila-like 4 gene (*ELAVL4*). However, attempts to replicate these candidate genes have yielded mixed results [9–11] and no markers within the *PARK10* region have reached genome-wide significance in any of the recently published PD genome-wide association studies [3,12–19]. Thus, there is considerable uncertainty as to whether the *PARK10* region actually harbors PD susceptibility genes.

* Corresponding author. Geriatric Research Education and Clinical Center S-182, VA Puget Sound Health Care System, 1660 South Columbian Way, Seattle, WA 98108, USA.
E-mail address: zabetian@u.washington.edu (C.P. Zabetian).

1353-8020/$ – see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.parkreldis.2013.10.001

We re-examined the *PARK10* region in the publicly available NeuroGenetics Research Consortium (NGRC) Genome-Wide Association Study (GWAS) dataset enriched with additional markers imputed from the HapMap 3 and 1000 Genomes reference panels. Furthermore, we utilized two other publicly available GWAS datasets [15,18] and unpublished data from a third case-control study [20] for replication. Our primary goals were to identify new PD-associated genes that might have been previously missed due to low genotyping coverage, and to verify associations with the aforementioned candidate genes.

2. Methods

2.1. Studies

The discovery phase for novel risk variants used data from the NGRC Genome-Wide Association Study of Parkinson Disease: Genes and Environment (phs000196.v1.p1) with 2000 subjects with PD and 1986 controls that were recruited from movement disorder clinics in Oregon, Washington, Georgia, and New York [14]. All subjects were genotyped on the IlluminaHumanOmni1-Quad_v1-0_B array and data were downloaded from dbGaP. Details of the quality control processes employed in the analysis are provided (Appendix A. Supplementary methods).

The replication phase used two publicly available GWASs, the Center for Inherited Disease Research (CIDR) GWAS in Familial PD (phs000126.v1.p1) and the National Institute of Neurological Disorders and Stroke (NINDS) Genome-Wide Genotyping in PD Study (phs000089.v3.p2), and a population-based study of PD at the University of Washington and Group Health Cooperative (GHC-UW) [20]. Information on both the CIDR [15] and NINDS [18] GWAS datasets has been published elsewhere. Because controls for both GWASs came from the NINDS Human Genetics Repository (http://ccr.coriell.org/Sections/Collections/NINDS), we performed identity-by-descent (IBD)-estimation using 287,341 independent single nucleotide polymorphisms (SNPs) to elucidate any overlap in controls that showed evidence of being duplicates or were a result of cryptic relatedness (PI $\geq$ .5). We found evidence that nine controls overlapped between the NINDS and CIDR GWAS datasets, and these individuals were subsequently omitted from the NINDS dataset. The GHC-UW sample consisted of 758 non-Hispanic, white subjects (321 cases, 437 controls) who were genotyped for 1138 tagging SNPs spanning the *PARK10* region and 440 ancestry informative markers (AIMs) using custom Affymetrix GeneChip Universal 3K Tag Arrays. The AIMs were unlinked SNPs selected to distinguish intercontinental population structure [21,22] as well as European substructure [23]. Before imputation, SNPs were omitted if Hardy–Weinberg equilibrium was violated ($P < 0.001$ using an exact test) in the combined sample or if the genotyping rate was less than 90%. The project was approved by the Veterans Affairs Puget Sound Health Care System and University of Washington institutional review boards, and all subjects provided written informed consent.

Patients in the NGRC, CIDR, and NINDS studies met UK PD Society Brain Bank clinical diagnostic criteria for PD [14,15,18]. Patients in the GHC-UW study met similar clinical diagnostic criteria for PD as previously described [20]. The age at onset of PD was similar across studies with the following mean ±SD onset age: 58.3 ± 11.9 years (NGRC), 61.9 ± 10.9 years (CIDR), 58.4 ± 13.2years (NINDS). In the GHC-UW study, age at onset data were not collected, but the mean ±SD age at diagnosis was 66.2 ± 10.4 years.

2.2. Genotype enrichment and imputation

A subset of 499 subjects from the NGRC study population was previously genotyped using the same custom Affymetrix 3K Array as was used in the GHC-UW study. Prior to imputation, these genotype data were merged into the NGRC dataset using the default consensus genotype call approach in PLINK [24]. Additionally, standard data-cleaning methods (e.g. strand flipping) were performed. We then used IMPUTE2 [25] to enrich the *PARK10* region with imputed markers. To ensure that rare variants were adequately covered, we used two phased reference panels from both HapMap 3 and the 1000 Genomes pilot data, with release dates of Feb 2009 and Jun 2010, respectively, and we imputed genotypes for every 5-Mb interval in the 40–60 Mb region of chromosome 1 for each study separately. We omitted SNPs with an information metric less than .30 and noted those SNPs with information metrics between .30 and .50. After genotype imputation, the number of SNPs ($N_{SNP\text{-}STUDY}$) analyzed in the *PARK10* region for the NGRC, CIDR, NINDS, and GHC-UW studies were: $N_{SNP\text{-}NGRC} = 43{,}799$, $N_{SNP\text{-}CIDR} = 43{,}243$, $N_{SNP\text{-}NINDS} = 43{,}362$ and $N_{SNP\text{-}GHC\text{-}UW} = 40{,}351$, respectively, with 43,744 SNPs common to two or more studies.

2.3. Statistical analysis

Failure to recognize population structure can lead to confounding results. Thus, we accounted for differences in ancestry between cases and controls in our analyses using a principal components analysis (PCA) approach as implemented in the EigenStrat program [26]. For the three GWASs, we obtained a subset of uncorrelated markers by pruning the genome-wide SNPs to approximate linkage equilibrium as described in the Supplement. These markers were then used to calculate the axes of geographical variance or principal components (PCs) that describe the genetic variation. The first 28, 10, and 9 PCs were found to be significant (Tracy–Widom P-value $< .05$) for the NGRC, NINDS, and CIDR studies, respectively. However, for the GHC-UW dataset, we used the 440 AIMs to perform PCA, and although Tracy–Widom statistics may not be applicable to AIMs, we conservatively chose the top six PCs (Tracy–Widom P-value $< .05$).

We used PLINK [24] to analyze the genotyped and imputed SNPs based on the genotype probabilities with an additive dosage model in a logistic regression analysis adjusted for sex, age, and the significant PCs from EigenStrat PCA analysis. The asymptotic P-value was obtained from the Z-test assessing the significance of the association with PD. SimpleM [27] was used to infer the effective number of independent tests (Meff) after accounting for LD between SNPs with a minor allele frequency (MAF) of at least .005; Meff was 15,572 in the NGRC study. Thus, after accounting for multiple testing using a Bonferroni adjustment on the effective number of tests, we set the significance threshold for the overall study to $P < 3.2 \times 10^{-6}$ (i.e., .05/15,572). However, in selecting SNPs from the discovery phase for replication we used a less stringent threshold of $P < 5 \times 10^{-4}$ to reduce the probability of false negative results. Additionally, we performed a meta-analysis for each SNP across all studies using the classical approach of pooling effect size estimates and standard errors using a random effects model in PLINK. Cochrane's Q statistic was used to test for the presence of heterogeneity, and I^2 was calculated to examine the degree of heterogeneity present, and values of 25%, 50%, and 75% generally represent low, moderate, and high levels of heterogeneity, respectively.

Power was calculated using Quanto (http://hydra.usc.edu/gxe) and assuming a 1.5% prevalence of PD, a significance of 5×10^{-4} and a rare variant under a log-additive genetic model with a minor allele frequency (MAF) of .05 and .25.

3. Results

In the NGRC dataset, SNPs in five genes (*DAB1*, *SLFNL1*, *OMA1*, *SSBP3*, and *AGBL4*) met the predefined, exploratory significance