Parkinsonism and Related Disorders 19 (2013) 595–599

Contents lists available at SciVerse ScienceDirect

Parkinsonism and Related Disorders

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

Genetic susceptibility loci, environmental exposures, and Parkinson's disease: A case–control study of gene–environment interactions

Sun Ju Chung [a,b], Sebastian M. Armasu [c], Kari J. Anderson [c], Joanna M. Biernacka [c], Timothy G. Lesnick [c], David N. Rider [c], Julie M. Cunningham [d], J. Eric Ahlskog [b], Roberta Frigerio [e], Demetrius M. Maraganore [f,*]

[a] Department of Neurology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[b] Department of Neurology, Mayo Clinic, Rochester, MN, USA
[c] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA
[d] Department of Laboratory Medicine, Mayo Clinic, Rochester, MN, USA
[e] Research Institute, NorthShore University HealthSystem, Evanston, IL, USA
[f] Department of Neurology, Neurological Institute, NorthShore University HealthSystem, 2650 Ridge Avenue, Evanston, IL 60201, USA

ARTICLE INFO

Article history:
Received 9 September 2012
Received in revised form
16 January 2013
Accepted 15 February 2013

Keywords:
Parkinson's disease
Genes
Environment
Interactions

ABSTRACT

Background: Prior studies causally linked mutations in *SNCA*, *MAPT*, and *LRRK2* genes with familial Parkinsonism. Genome-wide association studies have demonstrated association of single nucleotide polymorphisms (SNPs) in those three genes with sporadic Parkinson's disease (PD) susceptibility worldwide. Here we investigated the interactions between SNPs in those three susceptibility genes and environmental exposures (pesticides application, tobacco smoking, coffee drinking, and alcohol drinking) also associated with PD susceptibility.
Methods: Pairwise interactions between environmental exposures and 18 variants (16 SNPs and two variable number tandem repeats, or "VNTRs") in *SNCA*, *MAPT* and *LRRK2*, were investigated using data from 1098 PD cases from the upper Midwest, USA and 1098 matched controls. Environmental exposures were assessed using a validated telephone interview script.
Results: Five pairwise interactions had uncorrected *P*-values < 0.05. These included pairings of pesticides $\times$ *SNCA* rs3775423 or *MAPT* rs4792891, coffee drinking $\times$ *MAPT* H1/H2 haplotype or *MAPT* rs16940806, and alcohol drinking $\times$ *MAPT* rs2435211. None of these interactions remained significant after Bonferroni correction. Secondary analyses in strata defined by type of control (sibling or unrelated), sex, or age at onset of the case also did not identify significant interactions after Bonferroni correction.
Conclusions: This study documented limited pairwise interactions between established genetic and environmental risk factors for PD; however, the associations were not significant after correction for multiple testing.

1. Introduction

The causes of Parkinson's disease (PD) are largely unknown. Both genetic and environmental factors have been implicated. Genetic loci that have been causally linked with familial Parkinsonism and reproducibly associated with PD susceptibility worldwide include α–synuclein (*SNCA*), microtubule associated protein tau (*MAPT*), and leucine rich repeat kinase 2 (*LRRK2*) [1–7]. Immunohistochemical studies demonstrated that Lewy bodies, the neuropathological hallmark of PD, contain not only α–synuclein, but also MAPT [8] and LRRK2 [9] proteins. However, as individual factors or combined, these genetic susceptibility loci account for only a small fraction of PD. Environmental exposures that have been reproducibly associated with PD include pesticides application ("pesticides"), tobacco smoking ("smoking"), coffee drinking ("coffee"), and in some studies alcohol drinking ("alcohol") [10–12]. However, as individual factors or combined, these environmental exposures also account for only a small fraction of PD.

Since genetic susceptibility loci and environmental exposures independently account for only a small fraction of PD, it has been postulated that interactions between genetic and environmental factors may be associated with much greater risk. For example, pesticides accelerate the rate of α–synuclein fibril formation in vitro and exacerbate the pathology associated with causal *SNCA* mutations in transgenic mice [13,14]. We previously reported that *SNCA* genotypes and herbicides had independent effects on PD risk,

* Corresponding author. Tel.: +1 847 570 1678; fax: +1 847 733 5565.
E-mail addresses: dmaraganore@northshore.org, dmaraganore@me.com (D.M. Maraganore).

1353-8020/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.parkreldis.2013.02.008

with no significant pairwise interactions [15]. However, in another study of interactions, while analyses of interactions were limited by small sample sizes, risk due to *SNCA* variations seemed to vary with pesticide exposure and smoking, especially in younger onset cases, suggesting an age-of-onset effect [16]. Here we expand the scope of our previous studies of genetic susceptibility loci (main effects and gene–gene interactions analyses) [17,18], to also include environmental factors (gene–environment interaction analyses), focusing on the genetic and environmental factors that have been reproducibly associated with PD.

2. Methods

2.1. Study subjects

All subjects were recruited as part of a National Institutes of Health funded study of the molecular epidemiology of PD (2R01ES10751). The enrollment of matched cases and controls has been previously described [15,17]. PD cases were referred sequentially to the Department of Neurology of Mayo Clinic in Rochester, MN, from June 1, 1996 through June 30, 2007. Controls consisted of unaffected siblings of PD cases or matched unrelated controls. Cases were matched to a single participating sibling first by sex (when possible) and then by closest age. Cases without an available sibling were matched to unrelated controls of the same sex, age (year of birth ± 2 years), and residential region (Minnesota, Wisconsin, Iowa, or North and South Dakota pooled together). Unrelated controls of ages 65 and older were randomly selected from the Centers for Medicare and Medicaid Services (CMS) lists. Unrelated controls younger than 65 years were selected using random digit dialing, according to standard techniques [19]. Initially 1103 cases and 1103 matched controls were enrolled in the study [15,17]. Genomic DNA was collected, extracted, and stored as previously described [15]. Five cases were excluded subsequently because of indeterminate diagnosis. Thus 1098 cases and 1098 matched controls were used in subsequent analyses. The Institutional Review Board of the Mayo Clinic approved the study, and all 2196 subjects provided written informed consent.

2.2. Genotyping

Single nucleotide polymorphisms (SNPs) in species-conserved regions and tag SNPs for the *SNCA*, *MAPT*, and *LRRK2* loci were selected for genotyping as previously described [17,18]. In total, 19 SNPs in *SNCA*, 35 in *MAPT*, and 65 in *LRRK2* were successfully genotyped using a bead array platform (Illumina GoldenGate). In addition, two variable number tandem repeats (VNTRs; *SNCA* REP1 and *MAPT* H1/H2 haplotype) that have been shown to be associated with PD worldwide via regularly updated meta-analysis (www.pdgene.org), were genotyped using a sequencing platform (Applied Biosystems). In total, 121 variants in the three susceptibility gene loci were successfully genotyped.

2.3. Selection of SNPs for gene–environment interaction analyses

Variants with minor allele frequency <0.05 or showing departures from Hardy–Weinberg equilibrium ($P < 0.001$) were excluded from the analyses. We limited the gene–environment interaction analyses to SNPs with at least marginal evidence of association with PD ($P < 0.1$ in a univariate test of SNP main effect under the assumption of log-additive allele effects). We further applied a tag-SNP selection strategy to the resulting SNP list, using the pairwise Tagger algorithm with $r^2 = 0.9$ implemented in Haploview 4.2 [20]. This procedure resulted in the selection of 8 SNPs in *SNCA*, 6 SNPs in *MAPT*, and 2 SNPs in *LRRK2* that had *P*-values < 0.1 in a trend test for association with PD. In addition, *SNCA* REP1 (coded based on the number of 259 bp alleles, the number of 263 bp alleles, or as the previously described REP1 score [15]) and the *MAPT* VNTR that distinguishes the H1/H2 haplotype were included in the analyses, as these variables also showed marginal evidence of association at the $P < 0.1$ level. The 18 SNPs and VNTRs used for the interaction analysis are listed in Table 1 [18].

2.4. Ascertainment of environmental exposures

All exposures were ascertained via telephone by direct or proxy (for subjects who had died or incapacitated subjects) interviews using a structured risk factors questionnaire administered by specifically trained study assistants. We previously described the ascertainment of pesticides exposures, the coding of types of pesticides (herbicides, insecticides, or fungicides), and the reliability of the exposures ascertainment [15,21]. Subjects were also asked to provide information on smoking (ever or never) and number of cigarettes per day and number of years (with details about periods of life with different amounts). Subjects were also queried about coffee drinking (ever or never), number of cups (quantified as "small cups", i.e., regular coffee cup of 6 ounces, or one shot of espresso; or as "medium or large cups", i.e., coffee mug of 12 ounces, or two shots of espresso) per day and number of years (with details about periods of life with different amounts) [22]. Subjects were also asked to provide information about alcohol drinking (ever or never), the number of drinks per day and number of years (with details about periods of life with different amounts) [23]. For this study, pesticide, alcohol, and tobacco use were coded as yes (ever) and no (never). Coffee consumption was coded as low (0–3 cups/day) and high (4+ cups/day) [10]. In addition, tobacco use was quantified as pack years.

Table 1
Genetic variants included in the interaction analyses.

Gene	Variants	SNP location or known function	Minor allele frequency	*P*-value[c]
SNCA	rs1372520	Intron	0.1848	0.0056
	rs2572324	Intron; associated with the extent of neurofibrillary pathology	0.3226	0.0090
	rs2583959	Intron	0.3013	0.0506
	rs2736990	Intron	0.4902	0.0017
	rs356186	Intron	0.1686	0.0119
	rs356218	3′ downstream	0.3545	0.0419
	rs3775423	Intron	0.0903	0.0090
	rs3775439	Intron	0.1290	0.0716
	REP1-259[a]	Regulates *SNCA* gene expression	0.2413	0.0345
	REP1-263[b]	Regulates *SNCA* gene expression	0.0749	0.0465
	REP1 score	Regulates *SNCA* gene expression	n.a.	0.0118
MAPT	rs16940758	Intron; regulates *MAPT* gene expression	0.1745	0.0844
	rs16940806	3′ UTR	0.1800	0.0059
	rs2435200	Intron	0.3968	0.0001
	rs2435211	Intron; regulates *MAPT* gene expression	0.3012	0.0856
	rs4792891	Intron	0.3016	0.0036
	rs8079215	Intron	0.2945	0.0660
	MAPT H1/H2	Regulates *MAPT* gene expression	0.1793	0.0042
LRRK2	rs17484286	Intron	0.0927	0.0128
	rs2404835	Intron	0.3255	0.0872

[a] REP1-259 denotes the *SNCA* VNTR REP1 coded as the number of 259 bp alleles.
[b] REP1-263 denotes the *SNCA* VNTR REP1 coded as the number of 263 bp alleles.
[c] *P*-value for trend test for association with PD.

2.5. Statistical analysis

Pairwise interactions between variants in three genes (*SNCA*, *MAPT*, and *LRRK2*) and environmental exposures (smoking, coffee, pesticides, and alcohol) were assessed using conditional logistic regression analyses. For SNPs, a log-additive genotype coding scheme was used, while for *SNCA* REP1 genotypes were converted to scores ranging from 0 to 4 as previously described [15]. REP1 genotypes coded as the number of 259 alleles and the number of 263 alleles were also used in the interaction analyses. A total of 20 pairwise interactions were tested for each environmental factor (primarily defined as ever/never for pesticides use, cigarette smoking, coffee drinking, or alcohol drinking). All analyses were adjusted for age at study and sex. For each genetic variant and environmental exposure we calculated the odds ratio (OR), 95% confidence interval (CI), and *P* value for the univariate effect in the conditional logistic regression model. In addition, the coefficient for the multiplicative interaction term and the associated *P* value were calculated.

We performed similar analyses of gene–environment interactions in subgroups, restricting either to case–unaffected sibling or case–unrelated control pairs, to men–men or women–women pairs, or to younger or older pairs as defined by median age at onset in the cases.

Although the expected effect sizes of the interaction terms were unknown, we used the assumed/observed marginal effects to estimate power for a range of values of the interaction term. The power calculations were performed using the software Quanto (http://hydra.usc.edu/GxE) [24], with an alpha level of 0.003, which takes into account the multiple genetic variants that were tested for interaction with each environmental factor. As an example, we estimated power to detect an interaction between coffee consumption and the genetic variants of primary interest (*SNCA* REP1 and *MAPT* H1/H2 variation). The observed allele frequencies were used in the power calculations, and dominant effects of genetic variants were assumed. We considered models with marginal effect sizes consistent with those observed in our data (with both the 259 REP1 allele and the *MAPT* H2 haplotype having protective effects, with marginal odds ratios of approximately 0.80 and 0.75, respectively, and