Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data

Hon-Chung Fung, Sonja Scholz, Mar Matarin, Javier Simón-Sánchez, Dena Hernandez, Angela Britton, J Raphael Gibbs, Carl Langefeld, Matt L Stiegert, Jennifer Schymick, Michael S Okun, Ronald J Mandel, Hubert H Fernandez, Kelly D Foote, Ramón L Rodríguez, Elizabeth Peckham, Fabienne Wavrant De Vrieze, Katrina Gwinn-Hardy, John A Hardy, Andrew Singleton

Summary

Background Several genes underlying rare monogenic forms of Parkinson's disease have been identified over the past decade. Despite evidence for a role for genetics in sporadic Parkinson's disease, few common genetic variants have been unequivocally linked to this disorder. We sought to identify any common genetic variability exerting a large effect in risk for Parkinson's disease in a population cohort and to produce publicly available genome-wide genotype data that can be openly mined by interested researchers and readily augmented by genotyping of additional repository subjects.

Methods We did genome-wide, single-nucleotide-polymorphism (SNP) genotyping of publicly available samples from a cohort of Parkinson's disease patients (n=267) and neurologically normal controls (n=270). More than 408 000 unique SNPs were used from the Illumina Infinium I and HumanHap300 assays.

Findings We have produced around 220 million genotypes in 537 participants. This raw genotype data has been publicly posted and as such is the first publicly accessible high-density SNP data outside of the International HapMap Project. We also provide here the results of genotype and allele association tests.

Interpretation We generated publicly available genotype data for Parkinson's disease patients and controls so that these data can be mined and augmented by other researchers to identify common genetic variability that results in minor and moderate risk for disease.

Introduction

Parkinson's disease is a chronic neurodegenerative disease with a cumulative prevalence of greater than one per thousand people.[1] The estimated sibling risk ratio for Parkinson's disease is around 1·7 (70% increased risk for Parkinson's disease if a sibling has the disease) for all ages, and increases by more than seven times for those younger than 66 years.[2] These data are consistent with a significant genetic contribution to disease risk.

Although attempts to define the underlying lesions in monogenic forms of Parkinson's disease have been successful,[3–8] traditional testing of candidate-gene associations has been less successful. Few common variants have shown repeatable association with risk for Parkinson's disease, the notable exception being common variation in *SNCA*, a gene originally implicated by results from family-based studies.[9]

The completion of stages I and II of the International Haplotype Map project,[10,11] in concert with the arrival of efficient, affordable, high-density typing methods of single-nucleotide polymorphisms (SNPs), promises to provide an approach with which to define the role of common genetic variation in risk for disease. This approach, much like traditional linkage methods, provides researchers with the ability to test variation in the genome in a relatively unbiased global manner, and thus does not rely on a-priori hypotheses of mechanistic underpinnings of disease. The International Haplotype Map project has provided a resource with which to calculate a minimum set of SNPs, often called tagging SNPs (tSNPs), which act as proxy markers for neighbouring genetic variation. Thus, a well-chosen set of several hundred thousand tSNPs will provide information about several million common genetic variants throughout the genome.

To begin to address the role of common genetic variation in idiopathic Parkinson's disease we did genome-wide SNP typing using more than 408 000 unique SNPs across the genome. By using Illumina Infinium I and HumanHap300 assays, we undertook a genome-wide association study in 276 patients with Parkinson's disease and 276 neurologically normal controls. The samples used for this study were derived from the National Institute of Neurological Disorders and Stroke (NINDS) funded Neurogenetics repository, which includes samples from patients with Parkinson's disease, cerebrovascular disease, epilepsy, and amyotrophic lateral sclerosis, and from neurologically normal controls.

Methods

Participants

Samples were derived from the NINDS Neurogenetics repository hosted by the Coriell Institute for Medical

research (NJ, USA). All patients gave written informed consent to participate in the study. Six precompiled panels each consisting of 92 cases or controls were selected for the analysis. The panels that contained samples from patients with Parkinson's disease were NDPT001, NDPT005, and NDPT007; these included DNA from 273 unique participants and three replicate samples. The panels that contained samples from neurologically normal controls were NDPT002, NDPT006, and NDPT008; these comprised DNA from 275 unique participants and one replicate sample. For the control population used in these experiments, blood samples were drawn from neurologically normal, unrelated, white individuals at many different sites within the USA. Each participant underwent a detailed medical history interview. None had a history of Alzheimer's disease, amyotrophic lateral sclerosis, ataxia, autism, bipolar disorder, brain aneurysm, dementia, dystonia, or Parkinson's disease. Folstein mini-mental state examination scores ranged from 26–30. All participants were interviewed for family history in detail. None had any first-degree relative with a known primary neurological disorder, including amyotrophic lateral sclerosis, ataxia, autism, brain aneurysm, dystonia, Parkinson's disease, and schizophrenia. The mean age of participants at sample collection was 68 years (range 55–88).

For the Parkinson's disease cohort, blood was obtained from unique and unrelated white individuals with idiopathic Parkinson's disease. The age of patients at onset of the disease ranged from 55 years to 84 years. Disease onset was defined as the time when symptoms of the disease were first noted, including at least one of the following: resting tremor, rigidity, bradykinesia, gait disorder, postural instability. All patients were queried about family history of parkinsonism, dementia, tremor, gait disorders, and other neurological dysfunction. Both those with and without a reported family history of Parkinson's disease were included on this panel. None were included who had three or more relatives with parkinsonism, nor with apparent Mendelian inheritance of Parkinson's disease.

Procedures

DNA for the genotyping experiments was extracted using a salting out procedure from Epstein-Barr virus immortalised lymphocyte cell lines (LCLs). The average passage number for each line was five (range five to seven). Epstein-Barr virus immortalisation was undertaken as previously described.[12] At the same time, DNA was extracted from 0·5 mL of blood from all participants for subsequent quality control steps in the cell-banking process.

All samples were assayed with the Illumina Infinium I and Infinium HumanHap300 SNP chips (Illumina, San Diego, CA, USA). These products assay 109 365 gene-centric SNPs (Infinium I) and 317 511 haplotype tagging SNPs derived from phase I of the International HapMap project (HumanHap300). There are 18 073 SNPs in common between the two arrays; thus the assays combined provide data for 408 803 unique SNPs. Any samples with a call rate below 95% were repeated on a fresh DNA aliquot and if the call rate persisted below this level the sample was excluded from the analysis. Low-quality genotyping led us to repeat 11 individual samples, of which seven were ultimately excluded from the analysis.

Statistical analysis

For each SNP we computed a series of estimates and tests using a program developed at Wake Forest University called Snpgwa. Each SNP was tested for departures from Hardy-Weinberg equilibrium. Five tests of genotypic association were computed: two degrees of freedom overall test for 2×3 tables, dominant model, additive model (Cochran-Armitage trend test), recessive model, and lack of fit to an additive model. We calculated odds ratios (ORs), 95% CIs, and p values for each of the association models. We used the program Dandelion, which ran within Snpgwa, to do two-marker and three-marker moving-window haplotype-association analysis for those SNPs that were consistent with Hardy-Weinberg equilibrium in controls. For all p values with an uncorrected significance of less than 0·05 we did permutation tests within Snpgwa using a variable number of permutations based on the p value of the test. For each permutation, Snpgwa permutes the affection status (case or control) of the entire sample represented in the input file while preserving the total number of cases and total number of controls in each permutation. The permutation is done using a Wichman-Hill random number generator.

In an attempt to detect the presence of significant population substructure or ethnically mismatched individuals we selected 267 random, unlinked SNPs from throughout the genome (available on request) and ran the program STRUCTURE on these data from all genotyped individuals.[13]

Role of the funding source

The study sponsors had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

We genotyped 276 samples from patients with Parkinson's disease and 276 from unrelated population controls. In the Parkinson's disease cohort there were 273 unique individuals and in the control cohort there were 275 unique individuals. Genotyping of the four replicate samples with the Infinium I assay gave genotype concordance rates of greater than 99·99%. Analysis of the 18 073 SNPs that overlap between the Infinium I and

HumanHap300 products revealed genotype concordance rates of 99·94% between the assays across 537 samples. Four samples were dropped from the control cohort due to low-quality genotyping; further analysis revealed that two of these samples (ND01630 and ND01666) were contaminated and the other two samples (ND03447 and ND03704) did not meet the genotype quality threshold (95% call rate) after repeated assay. Thus, the total number of fully genotyped samples in the control cohort was 271. Six samples were dropped from the Parkinson's disease cohort, this included three young-onset samples that were erroneously included in panel NDPT007 (ND05074, ND05416, and ND05841). Samples ND01500, ND04424, and ND04744 were excluded from analysis because of genotype call rates below 95% after being assayed twice.

For the 408 803 SNPs studied, the genotype call rate was greater than 99% for each of 395 275 SNPs (96·6%) and greater than 95% for 406 312 SNPs (99·4%). The Hardy-Weinberg equilibrium p value was higher than 0·001 for 395 493 SNPs and higher than 0·05 for 375 527 SNPs. The average minor allele frequency in autosomes was 26·47%. A total of 219 577 497 unique genotype calls were made and the average call rate across all samples was 99·6%.

Statistical analysis of association was done for all genotypes, irrespective of Hardy-Weinberg disequilibrium or minor allele frequency. The most significantly associated

Chromosome location	dbSNP ID	Location (genome build 36.1)	No of geno	Gene	Putative function	HWE p value	p value 2DF	Empirical p value 2DF	p value D/A/R	OR (95% CI)	Empirical p value D/A/R
11q14	rs10501570	84095494	536	DLG2	A member of the membrane-associated guanylate kinase family, may interact at postsynaptic sites	0·396	$7·3 \times 10^{-6}$	$2·0 \times 10^{-6}$	$5·3 \times 10^{-4}$R	0·2 (0·0–0·5)	$4·9 \times 10^{-4}$R
17p11.2	rs281357	19683106	537	ULK2	Similar to a serine/threonine kinase in *C elegans* which is involved in axonal elongation	0·852	$9·8 \times 10^{-6}$	$4·0 \times 10^{-6}$	0·0002R	0·4 (0·2–0·6)	$1·5 \times 10^{-5}$R
4q13.2	rs2242330§	68129844	537	BRDG1	Docking protein acting downstream of Tec tyrosine kinase in B cell antigen receptor signaling	0·708	$1·7 \times 10^{-6}$	$1·2 \times 10^{-5}$	$2·9 \times 10^{-6}$A	0·5 (0·4–0·7)	$<1 \times 10^{-6}$A
10q11.21	rs1480597*	44481115	525	Intergenic		1·000	$1·9 \times 10^{-6}$	$7·0 \times 10^{-6}$	$3·2 \times 10^{-6}$D	0·4 (0·3–0·6)	$2·0 \times 10^{-6}$D
4q13.2	rs6826751§	68116450	536	BRDG1	As above	0·024	$2·0 \times 10^{-6}$	$1·8 \times 10^{-5}$	$3·5 \times 10^{-6}$A	0·6 (0·4–0·7)	$5·0 \times 10^{-6}$A
16q23.1	rs4888984	78066835	537	Intergenic		1·000	$2·7 \times 10^{-5}$	$1·1 \times 10^{-5}$	$4·6 \times 10^{-6}$A	0·5 (0·3–0·7)	$3·0 \times 10^{-6}$A
4q35.2	rs4862792	188438344	511	Intergenic		0·358	$3·5 \times 10^{-5}$	$8·0 \times 10^{-6}$	$6·8 \times 10^{-6}$D	2·9 (1·8–4·6)	$7·0 \times 10^{-6}$D
4q13.2	rs3775866§	68126775	537	BRDG1	As above	0·911	$4·6 \times 10^{-5}$	$3·3 \times 10^{-5}$	$7·8 \times 10^{-6}$A	0·5 (0·4–0·7)	$8·0 \times 10^{-6}$A
20q13.13	rs2235617‡	47988384	530	ZNF313	Metal ion binding, protein binding, zinc ion binding, involved in cell differentiation and spermatogenesis	0·034	$4·7 \times 10^{-5}$	$4·7 \times 10^{-5}$	$8·8 \times 10^{-6}$D	0·4 (0·3–0·6)	$1·2 \times 10^{-5}$D
1p31	rs988421	72322424	536	NEGR1	Neuronal growth regulator	0·667	$4·9 \times 10^{-5}$	$4·3 \times 10^{-5}$	$7·0 \times 10^{-4}$R	2·0 (1·3–3·0)	$8·2 \times 10^{-4}$R
10q11.21	rs7097094*	44530696	537	Intergenic		0·294	$5·0 \times 10^{-5}$	$2·7 \times 10^{-5}$	$8·9 \times 10^{-6}$D	0·5 (0·3–0·7)	$8·0 \times 10^{-6}$D
10q11.21	rs999473*	44502322	537	Intergenic		0·294	$5·0 \times 10^{-5}$	$3·8 \times 10^{-5}$	$8·9 \times 10^{-6}$D	2·2 (1·5–3·1)	$8·0 \times 10^{-6}$D
11q11	rs1912373	56240441	537	Intergenic		0·375	$5·6 \times 10^{-5}$	$6·1 \times 10^{-5}$	$9·7 \times 10^{-6}$D	2·2 (1·6–3·2)	$1·2 \times 10^{-5}$D
1q25	rs1887279†	182176783	537	GLT25D2	Glycosyltransferase 25 domain containing 2	0·424	$5·7 \times 10^{-5}$	$3·5 \times 10^{-5}$	$1·2 \times 10^{-5}$A	0·5 (0·4–0·7)	$6·0 \times 10^{-6}$A
1q25	rs2986574†	182173237	536	GLT25D2	Glycosyltransferase 25 domain containing 2	0·350	$6·3 \times 10^{-5}$	$2·4 \times 10^{-5}$	$1·3 \times 10^{-5}$A	2·0 (1·4–2·7)	$6·0 \times 10^{-6}$A
22q13	rs11090762	46133989	536	Intergenic		0·730	$6·3 \times 10^{-5}$	$4·2 \times 10^{-5}$	$1·2 \times 10^{-5}$D	0·4 (0·3–0·6)	$8·0 \times 10^{-6}$D
20q13.13	rs6125829‡	48002336	509	ZNF313	Metal ion binding, protein binding, zinc ion binding, involved in cell differentiation and spermatogenesis	0·004	$6·6 \times 10^{-5}$	$7·2 \times 10^{-5}$	$1·4 \times 10^{-5}$D	2·2 (1·6–3·2)	$1·8 \times 10^{-5}$D
7p12	rs7796855	49627992	537	Intergenic		0·931	$6·6 \times 10^{-5}$	$7·2 \times 10^{-5}$	$1·3 \times 10^{-5}$D	0·4 (0·3–0·6)	$1·2 \times 10^{-5}$D
4q13.2	rs355477§	68079120	533	BRDG1	As above	0·207	$7·9 \times 10^{-5}$	$7·4 \times 10^{-5}$	$1·5 \times 10^{-5}$A	0·6 (0·5–0·8)	$1·7 \times 10^{-5}$A
1q25	rs3010040†	182174845	537	GLT25D2	Glycosyltransferase 25 domain containing 2	0·421	$8·0 \times 10^{-5}$	$6·2 \times 10^{-5}$	$1·6 \times 10^{-5}$A	0·5 (0·4–0·7)	$1·2 \times 10^{-5}$A
1q25	rs2296713†	182176340	537	GLT25D2	Glycosyltransferase 25 domain containing 2	0·421	$8·0 \times 10^{-5}$	$6·2 \times 10^{-5}$	$1·6 \times 10^{-5}$A	2·0 (1·4–2·7)	$1·2 \times 10^{-5}$A
4q13.2	rs355461§	68063319	537	BRDG1	As above	0·150	$8·3 \times 10^{-5}$	$6·0 \times 10^{-5}$	$1·6 \times 10^{-5}$A	1·7 (1·3–2·2)	$1·9 \times 10^{-5}$A
4q13.2	rs355506§	68068677	537	BRDG1	As above	0·150	$8·3 \times 10^{-5}$	$6·0 \times 10^{-5}$	$1·6 \times 10^{-5}$A	1·7 (1·3–2·2)	$1·9 \times 10^{-5}$A
4q13.2	rs355464§	68061719	531	BRDG1	As above	0·086	$8·9 \times 10^{-5}$	$9·3 \times 10^{-5}$	$1·7 \times 10^{-5}$A	1·7 (1·3–2·2)	$2·1 \times 10^{-5}$A
4q13.2	rs1497430§	68040409	535	BRDG1	As above	0·150	$9·7 \times 10^{-5}$	$8·0 \times 10^{-5}$	$1·8 \times 10^{-5}$A	1·7 (1·3–2·2)	$1·9 \times 10^{-5}$A
4q13.2	rs11946612§	68018566	535	BRDG1	As above	0·150	$9·7 \times 10^{-5}$	$8·5 \times 10^{-5}$	$1·8 \times 10^{-5}$A	0·6 (0·5–0·8)	$2·1 \times 10^{-5}$A

Although the SNPs outlined here are candidates, an appropriate replication or joint-analysis follow-up is needed and would include genotyping of loci that are significant down to a less stringent p value. HWE=Hardy Weinberg equilibrium. D=dominant. R=recessive. A=additive. No of geno=number of successful genotypes generated. *,†,‡,§=Closely associated SNPs. §This region is also shown in webfigure 1.

Table 1: p values with uncorrected significance >0·0001 for SNPs that gave successful genotypes in >95% of samples

dbSNP	Location (genome build 36.1)	Successful genotypes	HWE p value	Genotype p value	Dominant p value	Additive p value	Recessive p value	2-marker haplotype p value	3-marker haplotype p value
rs356174	90849924	537	0·852	0·919	0·711	0·808	0·960	0·050	0·289
rs11931074	90858538	537	0·411	0·199	0·123	0·092	0·243	0·157	0·179
rs6842093	90866888	537	1·000	0·335	0·154	0·199	0·996	0·176	0·391
rs3775423	90876514	537	0·411	0·230	0·152	0·114	0·243	0·242	0·242
rs356204	90882565	537	0·262	0·334	0·143	0·186	0·510	0·220	0·228
rs356168	90893454	537	0·262	0·334	0·143	0·186	0·510	0·228	0·502
rs2736990	90897564	534	0·298	0·297	0·123	0·166	0·496	0·563	0·282
rs356191	90907143	537	0·161	0·334	0·473	0·756	0·269	0·227	0·253
rs356188	90910560	529	0·202	0·270	0·404	0·693	0·245	0·277	0·253
rs3775439	90928764	537	0·124	0·116	0·057	0·041	0·248	0·274	0·271
rs2197120	90948625	537	0·161	0·334	0·473	0·756	0·269	0·618	0·565
rs3822095	90955540	536	0·508	0·655	0·672	0·452	0·364	0·372	0·274
rs1812923	90958562	535	0·252	0·804	0·738	0·993	0·678	0·380	0·544
rs10005233	90962354	537	0·491	0·646	0·784	0·457	0·350	0·800	0·909
rs2583978	90969349	537	0·262	0·469	0·473	0·698	0·419	0·950	0·758
rs1471483	90973315	536	0·489	0·650	0·842	0·492	0·356	0·814	0·948
rs2583985	90974962	537	1·000	0·895	0·639	0·685	0·916	0·873	0·661
rs1372519	90976332	536	0·126	0·693	0·617	0·781	0·569	0·885	0·970
rs2301134	90977968	537	0·343	0·694	0·406	0·576	0·943	0·889	0·971
rs7687945	90983722	537	0·343	0·766	0·467	0·576	0·862	0·889	0·955
rs1372522	90984303	532	0·435	0·702	0·426	0·621	0·993	0·834	..
rs2736988	90995368	537	1·000	0·861	0·638	0·586	0·683	..	..

The SNCA gene begins at 90 977 156 base pairs and ends at 90 865 728 base pairs. Variation of the REP1 polymorphism approximately 10 kb 5′ to the gene has been most strongly associated with risk for disease.[9] HWE=Hardy-Weinberg equilibrium.

Table 2: p values across the *SNCA* locus, previously associated with risk for Parkinson's disease

SNPs are shown in table 1 (see also webfigure 1); however, the raw p values for all loci and under each model are available at the Coriell website.

Analysis with STRUCTURE[13] showed that there is no discernible difference in the population substructure between cases and controls (webfigure 2). Furthermore, comparison of the cases and controls pooled together versus genotypes from a cohort of 173 non-white participants showed clear separation of the Parkinson's disease and control group from the non-white group, with the exception of a single patient from the former cohort, who, based on these analyses, had significant non-white genetic background. This individual was removed from the association analysis.

Discussion

We present here the generation, analysis, and public release of genome-wide SNP data in a cohort of Parkinson's disease patients and controls derived from the NINDS-funded open-access Neurogenetics repository at Coriell Cell Repositories. Our aim was to generate publicly available genotype data for Parkinson's disease patients and controls so that these data could be mined and augmented by other researchers, and also to undertake a preliminary analysis in an attempt to localise common genetic variation exerting a large effect on risk for Parkinson's disease in a cohort of white North Americans. These are the first genome-wide SNP genotype data, outside of the International HapMap Project, to be made publicly available.

A genome-wide association analysis of Parkinson's disease has been done with a two-tiered design[14] with slightly fewer than 200 000 SNPs. Although this study used fewer than half of the SNPs used in our study, the multistage design added substantial power and sensitivity to the results. The authors of these experiments suggested that their data revealed 13 SNPs associated with risk for Parkinson's disease. We, and others, have not been able to confirm these findings in independent cohorts.[15–19] Side-by-side comparison of the current data and the most significantly associated SNPs, published by Maraganore and colleagues,[14] did not show a replication of any of these published associations (webtable). Attempts at replication of these and other potential loci revealed by this study have been hindered because to date the authors have not released raw genotype data. When these data are released, our calculations show that at least 32 127 SNPs are shared between the two studies, making pooled analysis possible. Furthermore, one plausible approach is to combine or compare odds ratios of physically close SNPs, although data compared between studies and across platforms should be viewed with appropriate caution.

Our data provides 80% power to detect an allelic association with an odds ratio of more than 2·09 and less

than 0·40 at an uncorrected significance level of p=0·000001. This calculation is based on the average observed minor allele frequency of 26% and assumes that either the causal variant is typed or that there is complete and efficient tagging of common variation by the genotyped tSNPs. Although the sample size here is of limited power there is precedence for the use of small cohorts to identify genes of large effect by gene-wide association studies; the analysis of around 100 000 SNPs in only 96 cases with age-related macular degeneration and 50 controls led to the identification of variability within the gene encoding complement factor H as a risk factor for disease.[20] These data on macular degeneration draw attention to the use of genome-wide association studies in localisation of common genetic variability associated with disease, although the size of effect in that study was much higher than would generally be expected in most complex diseases (in macular degeneration the OR for homozygous carriers was 7·4). Illustrating the size of effects expected in complex disorders, the locus most robustly associated with risk of Parkinson's disease is the SNCA gene. We did not identify a significant association at this locus (table 2); however, given that the OR associated with this locus is estimated at 1·4 it is not surprising that we were unable to identify an association.

Analysis of our data showed 26 loci with a two-degree of freedom p value less than 0·0001 (table 1), with ORs ranging from 0·2 (95% CI 0·04–0·5) to 0·6 (0·5–0·8) and from 1·7 (1·3–2·2) to 2·2 (1·6–3·2). A stringent Bonferroni correction based on 408 803 independent tests means that a precorrection p value of less than $1·2 \times 10^{-7}$ would be needed to provide a corrected significant p value of less than 0·05. Thus, none of the values listed were significant after correction. Although speculation on the plausibility and biological significance of these candidate loci is tempting, we regard these data as hypothesis generating. Furthermore, given the inevitably high false-positive rate of genome-wide association studies, the next step in these analyses should involve genotyping in additional sample series. In the first instance, this work should be done in a cohort comprising patients and controls of similar demographic characteristics to reduce the confounds of allelic and genetic heterogeneity between ethnic groups. This approach would involve continued whole-genome SNP genotyping in the additional Parkinson's disease cases and controls available from the Coriell Neurogenetics repository; however, a more cost-effective measure would be to do follow-up genotyping of several thousand of the most significantly associated SNPs in additional cases and controls. The release of genotype data and not just allele frequency data means that genotype data from additional samples can be added easily to the current set allowing investigators to undertake joint analysis rather than replication-based analysis. The former approach is more powerful than the latter in identifying common genetic risk factors.[21] The control samples in the current study have been specifically obtained so that they can be used for other neurological disorders, including but not restricted to stroke and amyotrophic lateral sclerosis, so these data will also be of use to other researchers outside of the Parkinson's disease specialty.

Our data suggest that there are no common genetic variants that exert an effect of greater than an OR of 4 in Parkinson's disease. From the standpoint of experimental design this information is very useful; however, there are important drawbacks to this interpretation. First, these results can strictly only be applied to the current population. Second, analysis of young-onset Parkinson's disease cases, where a genetic effect is thought to be stronger, could reveal genetic variants with an effect of this size.[22] Third, this statement is reliant on either genotyping the causal variant or efficient and complete tagging of the causal variant.

In summary, we present here the generation and release of genotype data derived from publicly available Parkinson's disease and neurologically normal control samples. All DNA samples, raw genotype data, and significance test results are publicly available. These data suggest that there is no common genetic variant that exerts a large genetic risk for late-onset Parkinson's disease in white North Americans. These data are now available for future mining and augmentation to identify common genetic variability that results in minor and moderate risk for disease.

References
1   Kuopio AM, Marttila RJ, Helenius H, Rinne UK. Changing epidemiology of Parkinson's disease in southwestern Finland. *Neurology* 1999; **52:** 302–08.
2   de Rijk MC, Breteler MM, Graveland GA, et al. Prevalence of Parkinson's disease in the elderly: the Rotterdam Study. *Neurology* 1995; **45:** 2143–46.
3   Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* 1997; **276:** 2045–47.

4   Kitada T, Asakawa S, Hattori N, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* 1998; **392:** 605–08.

5   Bonifati V, Rizzu P, van Baren MJ, et al. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science* 2003; **299:** 256–59.

6   Valente EM, Abou-Sleiman PM, Caputo V, et al. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science* 2004; **304:** 1158–60.

7   Paisán-Ruiz C, Jain S, Evans EW, et al. Cloning of the gene containing mutations that cause PARK8 linked Parkinson disease. *Neuron* 2004; **44:** 595–600.

8   Zimprich A, Biskup S, Leitner P, et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* 2004; **44:** 601–07.

9   Maraganore DM, de Andrade M, Elbaz A, et al. Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease. *JAMA* 2006; **296:** 661–70.

10   The International HapMap Project. *Nature* 2003; **426:** 789–96.

11   A haplotype map of the human genome. *Nature* 2005; **437:** 1299–320.

12   Miller G, Shope T, Lisco H, Stitt D, Lipman M. Epstein-Barr virus: transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. *Proc Natl Acad Sci USA* 1972; **69:** 383–87.

13   Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003; **164:** 1567–87.

14   Maraganore DM, de Andrade M, Lesnick TG, et al. High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet* 2005; **77:** 685–93.

15   Clarimon J, Scholz S, Fung HC, et al. Conflicting results regarding the semaphorin gene (SEMA5A) and the risk for Parkinson disease. *Am J Hum Genet* 2006; **78:** 1082–84.

16   Farrer MJ, Haugarvoll K, Ross OA, et al. Genomewide association, Parkinson disease, and PARK10. *Am J Hum Genet* 2006; **78:** 1084–88.

17   Goris A, Williams-Gray CH, Foltynie T, Compston DA, Barker RA, Sawcer SJ. No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. *Am J Hum Genet* 2006; **78:** 1088–90.

18   Li Y, Rowland C, Schrodi S, et al. A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan. *Am J Hum Genet* 2006; **78:** 1090–92.

19   Elbaz A, Nelson LM, Payami H, et al. Lack of replication of thirteen single-nucleotide polymorphisms implicated in Parkinson's disease: a large-scale international study. *Lancet Neurol* 2006; published online Sept 27. DOI: 10.1016/S1474-4422(06)70579-8.

20   Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005; **308:** 385–89.

21   Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; **38:** 209–13.

22   Payami H, Zareparsi S, James D, Nutt J. Familial aggregation of Parkinson disease: a comparative study of early-onset and late-onset disease. *Arch Neurol* 2002; **59:** 848–50.