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 Steigert, Jennifer Schymick, Michael S Okun, Ronald J Mandel, Hubert H Fernandez, Kelly D Foote, Ramón I 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. 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. 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, 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 variants throughout the genome.

The completion of stages I and II of 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 Sciences and the National Institute on Aging (NIA). The NINDS Genetics Unit (H-C Fung MD, M Matarin PhD, J Simón-Sánchez BS, B N Gur A Singleton PhD) and the Laboratory of Neurogenetics (J A Hardy PhD, J Schymick BS), Molecular Genetics Unit (S Scholz MD, M Matarin PhD, J Simón-Sánchez BS), D Hernandez MS, A Britton MS, A Singleton PhD), and Computational Biology Core (J R Gibbs BS), National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; Section on Biostatistics, Department of Public Health Sciences, Wake Forest University Health Sciences, Winston-Salem, NC, USA; C Langefeld PhD, M Steigert MS; Movement Disorders Center, University of Florida, Departments of Neurology, Neuroscience, and Neurosurgery, Gainesville, FL, USA (M Okun MD, R J Mandel PhD, H H Fernandez MD, K D Foote MD, R I Rodríguez MD; and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA, R E Peddada DO, K Gwinn-Hardy MD) Unitat de Genètica Molecular, Departamento de Genòmica y Proteómica, Instituto de Biomedicina de Valencia-CSIC, Valencia, Spain (J Simón-Sánchez); Department of Neurology, Chang Gung Memorial Hospital and College of Medicine, Chang Gung University, Taipei, Taiwan (H-C Fung); and Rita Lila Weston Institute of Neurological Studies, University College London, London, London, UK (H-C Fung).

Correspondence to: Andrew Singleton, Molecular Genetics Unit, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA singleta@mail.nih.gov

Published Online September 27, 2006 DOI:10.1016/S0140-6736(06)69278-6
See Reflection and Reaction page R96

Laboratory of Neurogenetics (H-C Fung MD, F W De Vrieze PhD, J A Hardy PhD, J Schymick BS), Molecular Genetics Unit (S Scholz MD, M Matarin PhD, J Simón-Sánchez BS), D Hernandez MS, A Britton MS, A Singleton PhD), and Computational Biology Core (J R Gibbs BS), National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; Section on Biostatistics, Department of Public Health Sciences, Wake Forest University Health Sciences, Winston-Salem, NC, USA; C Langefeld PhD, M Steigert MS; Movement Disorders Center, University of Florida, Departments of Neurology, Neuroscience, and Neurosurgery, Gainesville, FL, USA (M Okun MD, R J Mandel PhD, H H Fernandez MD, K D Foote MD, R I Rodríguez MD; and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA, R E Peddada DO, K Gwinn-Hardy MD) Unitat de Genètica Molecular, Departamento de Genòmica y Proteómica, Instituto de Biomedicina de Valencia-CSIC, Valencia, Spain (J Simón-Sánchez); Department of Neurology, Chang Gung Memorial Hospital and College of Medicine, Chang Gung University, Taipei, Taiwan (H-C Fung); and Rita Lila Weston Institute of Neurological Studies, University College London, London, London, UK (H-C Fung).
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. 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.

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 (ND01447 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 SNPs, based on the highest number of fully genotyped samples (271 samples), show evidence of association with idiopathic PD at genome-wide significance levels. These are presented in the Table 1.

### Table 1: p values with uncorrected significance >0.0001 for SNPs that gave successful genotypes in >95% of samples

<table>
<thead>
<tr>
<th>Chromosome location</th>
<th>dbSNP ID</th>
<th>Location (genome build 36.1)</th>
<th>No of geno</th>
<th>Gene</th>
<th>Putative function</th>
<th>HWE p value</th>
<th>p value 2DF</th>
<th>Empirical p value 2DF</th>
<th>p value D/A/R</th>
<th>OR (95% CI)</th>
<th>Empirical p value D/A/R</th>
</tr>
</thead>
<tbody>
<tr>
<td>4q13.2</td>
<td>rs22423205</td>
<td>68129844</td>
<td>537</td>
<td>BRDG1</td>
<td>Docking protein acting downstream of Tec</td>
<td>0.708</td>
<td>1.7×10⁻⁴</td>
<td>1.2×10⁻⁵</td>
<td>2.9×10⁻⁴</td>
<td>0.5 (0.4-0.7)</td>
<td>&lt;1×10⁻⁴</td>
</tr>
<tr>
<td>10q11.21</td>
<td>rs1480597</td>
<td>71730166</td>
<td>537</td>
<td>BRDG1</td>
<td>As above</td>
<td>1.000</td>
<td>2.0×10⁻⁴</td>
<td>1.1×10⁻⁵</td>
<td>4.6×10⁻⁴</td>
<td>0.3 (0.3-0.7)</td>
<td>2.0×10⁻⁴</td>
</tr>
<tr>
<td>1q25</td>
<td>rs18872791</td>
<td>182176783</td>
<td>537</td>
<td>GLT2D2</td>
<td>Glycosyltransferase 25 domain containing 2</td>
<td>0.424</td>
<td>5.7×10⁻⁵</td>
<td>3.5×10⁻⁶</td>
<td>1.2×10⁻⁵</td>
<td>0.5 (0.4-0.7)</td>
<td>6.0×10⁻⁵</td>
</tr>
<tr>
<td>2q13</td>
<td>rs11090762</td>
<td>46133989</td>
<td>536</td>
<td>Intergenic</td>
<td>As above</td>
<td>0.730</td>
<td>6.3×10⁻⁴</td>
<td>4.2×10⁻⁴</td>
<td>1.2×10⁻⁴</td>
<td>0.4 (0.3-0.6)</td>
<td>8.0×10⁻⁴</td>
</tr>
</tbody>
</table>

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 webfi gure 1.

---

Articles

http://neurology.thelancet.com Vol 5 November 2006 913
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 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 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. Side-by-side comparison of the current data and the most significantly associated SNPs, published by Maraganore and colleagues, 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

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>Location (genome build 36.1)</th>
<th>Successful genotypes</th>
<th>HWE p value</th>
<th>Genotype p value</th>
<th>Dominant p value</th>
<th>Additive p value</th>
<th>Recessive p value</th>
<th>2-marker haplotype p value</th>
<th>3-marker haplotype p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs356474</td>
<td>90849924</td>
<td>rs7</td>
<td>0·852</td>
<td>0·919</td>
<td>0·711</td>
<td>0·808</td>
<td>0·960</td>
<td>0·050</td>
<td>0·289</td>
</tr>
<tr>
<td>rs11913074</td>
<td>90858538</td>
<td>rs7</td>
<td>0·411</td>
<td>0·199</td>
<td>0·123</td>
<td>0·092</td>
<td>0·243</td>
<td>0·157</td>
<td>0·179</td>
</tr>
<tr>
<td>rs6842093</td>
<td>90866888</td>
<td>rs7</td>
<td>1·000</td>
<td>0·325</td>
<td>0·154</td>
<td>0·199</td>
<td>0·996</td>
<td>0·176</td>
<td>0·391</td>
</tr>
<tr>
<td>rs2735423</td>
<td>90876514</td>
<td>rs7</td>
<td>0·421</td>
<td>0·190</td>
<td>0·123</td>
<td>0·092</td>
<td>0·243</td>
<td>0·157</td>
<td>0·179</td>
</tr>
<tr>
<td>rs356204</td>
<td>90882565</td>
<td>rs7</td>
<td>0·262</td>
<td>0·334</td>
<td>0·143</td>
<td>0·186</td>
<td>0·510</td>
<td>0·220</td>
<td>0·228</td>
</tr>
<tr>
<td>rs356168</td>
<td>90893454</td>
<td>rs7</td>
<td>0·262</td>
<td>0·334</td>
<td>0·143</td>
<td>0·186</td>
<td>0·510</td>
<td>0·220</td>
<td>0·228</td>
</tr>
<tr>
<td>rs2583985</td>
<td>90977968</td>
<td>rs7</td>
<td>0·326</td>
<td>0·694</td>
<td>0·406</td>
<td>0·576</td>
<td>0·943</td>
<td>0·889</td>
<td>0·971</td>
</tr>
<tr>
<td>rs1372519</td>
<td>90976332</td>
<td>rs7</td>
<td>0·126</td>
<td>0·270</td>
<td>0·104</td>
<td>0·066</td>
<td>0·245</td>
<td>0·277</td>
<td>0·253</td>
</tr>
<tr>
<td>rs2583978</td>
<td>90974962</td>
<td>rs7</td>
<td>0·298</td>
<td>0·297</td>
<td>0·123</td>
<td>0·166</td>
<td>0·496</td>
<td>0·563</td>
<td>0·282</td>
</tr>
<tr>
<td>rs356191</td>
<td>90970143</td>
<td>rs7</td>
<td>0·161</td>
<td>0·347</td>
<td>0·175</td>
<td>0·756</td>
<td>0·269</td>
<td>0·227</td>
<td>0·253</td>
</tr>
<tr>
<td>rs356188</td>
<td>90970560</td>
<td>rs7</td>
<td>0·202</td>
<td>0·270</td>
<td>0·104</td>
<td>0·066</td>
<td>0·245</td>
<td>0·277</td>
<td>0·253</td>
</tr>
<tr>
<td>rs3775439</td>
<td>90928764</td>
<td>rs7</td>
<td>0·124</td>
<td>0·116</td>
<td>0·057</td>
<td>0·041</td>
<td>0·248</td>
<td>0·274</td>
<td>0·271</td>
</tr>
<tr>
<td>rs2197120</td>
<td>90948625</td>
<td>rs7</td>
<td>0·161</td>
<td>0·347</td>
<td>0·175</td>
<td>0·756</td>
<td>0·269</td>
<td>0·227</td>
<td>0·253</td>
</tr>
<tr>
<td>rs2119095</td>
<td>90955540</td>
<td>rs7</td>
<td>0·508</td>
<td>0·655</td>
<td>0·672</td>
<td>0·452</td>
<td>0·364</td>
<td>0·372</td>
<td>0·274</td>
</tr>
<tr>
<td>rs1812923</td>
<td>90958562</td>
<td>rs7</td>
<td>0·252</td>
<td>0·804</td>
<td>0·738</td>
<td>0·993</td>
<td>0·678</td>
<td>0·380</td>
<td>0·544</td>
</tr>
<tr>
<td>rs10005233</td>
<td>90962354</td>
<td>rs7</td>
<td>0·491</td>
<td>0·646</td>
<td>0·784</td>
<td>0·457</td>
<td>0·350</td>
<td>0·800</td>
<td>0·909</td>
</tr>
<tr>
<td>rs2583978</td>
<td>90969349</td>
<td>rs7</td>
<td>0·262</td>
<td>0·469</td>
<td>0·473</td>
<td>0·698</td>
<td>0·419</td>
<td>0·950</td>
<td>0·758</td>
</tr>
<tr>
<td>rs1471483</td>
<td>90973315</td>
<td>rs7</td>
<td>0·489</td>
<td>0·650</td>
<td>0·842</td>
<td>0·492</td>
<td>0·356</td>
<td>0·814</td>
<td>0·948</td>
</tr>
<tr>
<td>rs2583985</td>
<td>90974962</td>
<td>rs7</td>
<td>1·000</td>
<td>0·895</td>
<td>0·639</td>
<td>0·685</td>
<td>0·916</td>
<td>0·873</td>
<td>0·661</td>
</tr>
<tr>
<td>rs1372519</td>
<td>90976332</td>
<td>rs7</td>
<td>0·126</td>
<td>0·693</td>
<td>0·617</td>
<td>0·781</td>
<td>0·569</td>
<td>0·885</td>
<td>0·970</td>
</tr>
<tr>
<td>rs301134</td>
<td>9097968</td>
<td>rs7</td>
<td>0·343</td>
<td>0·694</td>
<td>0·460</td>
<td>0·576</td>
<td>0·943</td>
<td>0·889</td>
<td>0·971</td>
</tr>
<tr>
<td>rs6787945</td>
<td>90981722</td>
<td>rs7</td>
<td>0·343</td>
<td>0·766</td>
<td>0·467</td>
<td>0·576</td>
<td>0·862</td>
<td>0·889</td>
<td>0·955</td>
</tr>
<tr>
<td>rs1372522</td>
<td>90984309</td>
<td>rs7</td>
<td>0·435</td>
<td>0·702</td>
<td>0·426</td>
<td>0·621</td>
<td>0·993</td>
<td>0·834</td>
<td>-</td>
</tr>
<tr>
<td>rs2736988</td>
<td>90995368</td>
<td>rs7</td>
<td>1·000</td>
<td>0·861</td>
<td>0·638</td>
<td>0·586</td>
<td>0·683</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
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 1.2×10−7 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.

Contributors
H-CF, SS, MM, and JS-S all contributed equally to this study. H-CF, SS, MM, JS-S, DH, AB, FWDV, and JS all participated in the laboratory-based genotyping and data analysis and critical revision of the manuscript. JRG did data manipulation and statistical analysis and critical revision of the manuscript. CL and MLS did statistical analysis. MSO, RIM, HHF, KDF, RLR, EP, and KG-H participated in the collection and characterisation of patients in addition to critical revision of the manuscript. JAH undertook critical revision of the manuscript and was involved in the design of the study. AS drafted the manuscript and designed and supervised the study.

Conflicts of interest
We have no conflicts of interest.

Acknowledgments
We thank the participants and the submitters for depositing samples at the NINDS Neurogenetics repository. The samples for this study are derived from the NINDS Neurogenetics repository at Coriell Cell Repositories. Access to the samples and to these data are available from the website. This study used the high-performance computational capabilities of the Biowulf PC/Linux cluster at the National Institutes of Health, Bethesda, MD, USA. This work was supported by the intramural programmes of the National Institute on Aging and the National Institute on Neurological Disorders and Stroke (NINDS) as well as by an extramural NINDS contract funding the Coriell Repository.

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

Webfigure 1: Representative p value plots generated from the output of Snpgwa
The Y axes show log p values and X axes are SNPs across a chromosome or chromosomal region. Log p values for are plotted across chromosome 4 (upper panel) and across a significant region (lower panel). A high degree of linkage disequilibrium can be seen across the significant region, which explains the significant association of many neighbouring SNPs. LPV=log p value. PV=p value. 2DF=two degree of freedom genotype test. Add=additive model. Rec=recessive model. Dom=dominant model. Allelic=allelic association. LRS 2 marker=two marker haplotype association (two contiguous markers). LRS 3 marker=three marker haplotype association (three contiguous markers). HWE=p value indicating deviation from Hardy Weinberg equilibrium. HW Comb=for cases and controls. HW cases=for cases. HW controls=controls. D Prime and R squared are both measures of linkage disequilibrium.

Correspondence to: Andrew Singleton, Molecular Genetics Unit, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA singleta@mail.nih.gov
Webfigure 2: Bar and triangle plots from STRUCTURE using 267 random autosomal SNPs
A) Bar plot for K=2, sorted by putative population where population 1 consists of 271 white controls and population 2 consists of 267 patients with sporadic PD. B) Bar plot for K=2, sorted by putative population using the same set of 267 SNPs where population 1 consists of 536 whites (sporadic PD case/control series) and population 2 consists of 173 non-white participants. C) Triangle plot with same putative populations as bar plot A) but with K=4, where blue dots are population 1 (controls) and red dots are population 2 (PD). D) Triangle plot with same putative populations as bar plot B) but with K=4, where blue dots are population 1 (white sporadic PD patients and controls) and red dots are population 2 (non-white participants). The non-white population are self-identified African American subjects from the NIA sponsored study Healthy Aging in Neighborhoods of Diversity across the Life Span (http://handls.nih.gov/).

Correspondence to:
Andrew Singleton, Molecular Genetics Unit, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA
singleta@mail.nih.gov
### Genome-wide genotyping in Parkinson’s disease and neurologically normal controls: first stage analysis and public release of data

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>dbSNP ID</th>
<th>Location bp (genome build 36-1)</th>
<th>No.</th>
<th>Geno</th>
<th>HWE p value</th>
<th>p value</th>
<th>Dominant model</th>
<th>p value</th>
<th>Additive model</th>
<th>p value</th>
<th>Recessive model</th>
<th>p value</th>
<th>2 marker haplo</th>
<th>p value</th>
<th>3 marker haplo</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p value 2DF</td>
<td></td>
<td>p value OR (95% CI)</td>
<td></td>
<td>p value OR (95% CI)</td>
<td></td>
<td>p value OR (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xq28</td>
<td>rs787823</td>
<td>10505701</td>
<td>536</td>
<td>0.836</td>
<td>0.008</td>
<td>0.094</td>
<td>0.7 (0.47-1.96)</td>
<td>0.304</td>
<td>0.87 (0.69-1.08)</td>
<td>0.922</td>
<td>0.78 (0.62-1.34)</td>
<td>0.290</td>
<td>0.40 (0.24-0.67)</td>
<td>0.182</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs6615259</td>
<td>138882861</td>
<td>536</td>
<td>0.770</td>
<td>0.281</td>
<td>0.326</td>
<td>0.83 (0.58-1.2)</td>
<td>0.595</td>
<td>1.09 (0.81-1.48)</td>
<td>0.346</td>
<td>1.58 (0.64-4.15)</td>
<td>0.026</td>
<td>0.06 (0.02-2.08)</td>
<td>0.192</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs11736972</td>
<td>125300823</td>
<td>536</td>
<td>0.840</td>
<td>0.380</td>
<td>0.218</td>
<td>0.78 (0.53-1.15)</td>
<td>0.400</td>
<td>0.27 (0.14-0.52)</td>
<td>0.031</td>
<td>0.7 (0.47-1.26)</td>
<td>0.457</td>
<td>0.19 (0.06-0.57)</td>
<td>0.192</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs16251318</td>
<td>125300823</td>
<td>567</td>
<td>0.095</td>
<td>0.864</td>
<td>0.668</td>
<td>1.11 (0.63-1.97)</td>
<td>0.617</td>
<td>1.11 (0.73-1.68)</td>
<td>0.665</td>
<td>1.39 (0.31-6.28)</td>
<td>0.811</td>
<td>0.19 (0.06-0.59)</td>
<td>0.192</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Only two SNPs overlapped between the sets (rs2313982 and rs7520966), so in every other instance the closest flanking SNPs are shown. SNPs described as significant by Maraganore and colleagues are shown in bold italics. 2 marker haplotype or 3 marker haplotype p value of disease association to a two marker or three marker haplotype where the interrogated SNP is the first of two or three contiguous SNPs placed into a haplotype. Chrom=chromosome. Haplo=haplotype.

**Webtable:** Comparison of significant loci identified by Maraganore et al (Am J Hum Genet 2005; 77: 685–93), with the current data set

---

Correspondence to:
Andrew Singleton, Molecular Genetics Unit, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA
singletas@mail.nih.gov

http://neurology.thelancet.com Published online September 27, 2006 Webtable