Genomewide Association Study for Onset Age in Parkinson Disease

Citation

Published Version

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:4621703

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Abstract

**Background:** Age at onset in Parkinson disease (PD) is a highly heritable quantitative trait for which a significant genetic influence is supported by multiple segregation analyses. Because genes associated with onset age may represent invaluable therapeutic targets to delay the disease, we sought to identify such genetic modifiers using a genomewide association study in familial PD. There have been previous genomewide association studies (GWAS) to identify genes influencing PD susceptibility, but this is the first to identify genes contributing to the variation in onset age.

**Methods:** Initial analyses were performed using genotypes generated with the Illumina HumanCNV370Duo array in a sample of 857 unrelated, familial PD cases. Subsequently, a meta-analysis of imputed SNPs was performed combining the familial PD data with that from a previous GWAS of 440 idiopathic PD cases. The SNPs from the meta-analysis with the lowest p-values and consistency in the direction of effect for onset age were then genotyped in a replication sample of 747 idiopathic PD cases from the Parkinson Institute Biobank of Milan, Italy.

**Results:** Meta-analysis across the three studies detected consistent association (p < 1 × 10^-5) with five SNPs, none of which reached genomewide significance. On chromosome 11, the SNP with the lowest p-value (rs10767971; p = 5.4 × 10^-7) lies between the genes *QSER1* and *PRRG4*. Near the PARK3 linkage region on chromosome 2p13, association was observed with a SNP (rs7577851; p = 8.7 × 10^-6) which lies in an intron of the *AAK1* gene. This gene is closely related to *GAK*, identified as a possible PD susceptibility gene in the GWAS of the familial PD cases.

**Conclusion:** Taken together, these results suggest an influence of genes involved in endocytosis and lysosomal sorting in PD pathogenesis.
Background
Parkinson disease (PD), the second most common neurodegenerative disorder, is characterized by debilitating symptoms of tremor, rigidity, and bradykinesia, usually occurring late in life. PD incidence increases with age from 1.7/10,000 person-years between ages 50 to 59 to 9.3/10,000 person-years between ages 70 to 79 and has a prevalence of approximately 1.8% among people over the age of 65 [1]. While the average age of onset of PD is approximately 60 years, there is wide variation, with some individuals having onset before age 20 and others not until after age 90 [2,3].

Five monogenic forms of PD have been identified and characterized by mode of transmission, onset age and protein affected by mutation. These include α-synuclein (SNCA or PARK1) [4], parkin (PARK2) [5], Pten-induced putative kinase 1 (PINK1 or PARK6) [6], DJ-1 (PARK7) [7], and leucine-rich repeat kinase 2 (LRRK2 or PARK8) [8]. Onset for PARK1 is younger than that seen for idiopathic PD [4]. PARK2 (parkin) is a recessive form with young onset, commonly before age 40. Heterozygous mutations in parkin are also associated with earlier onset of PD, typically in the early to mid sixth decade [9,10]. By contrast, PD associated with LRRK2 mutations presents an onset distribution very similar to that seen in idiopathic PD, as well as clear age-dependent penetrance [11-13].

Onset of PD has been shown to be correlated between siblings with PD [14] suggesting that genetic modifiers influence onset age. Segregation analyses in three independent studies showed evidence of a genetic effect influencing age of onset of PD [15-17]. Notably, all three of these segregation analyses showed stronger evidence for the presence of "major genes" influencing onset age or penetrance, than for genes influencing susceptibility. Furthermore, age is one of the strongest risk factors for PD, suggesting that age related penetrance is strongly associated with disease expression. By identifying genes related to onset age, it may be possible to identify pathogenic mechanisms and therapeutic targets capable of delaying onset of disease symptoms. Effectively postponing disease onset will reduce disease prevalence and ease the burden of PD in our aging population.

All prior PD genome wide association studies (GWAS) have focused exclusively on the detection of susceptibility genes and none has investigated association to genes influencing onset age [18-20]. In this study, we describe the first GWAS of onset age. This GWAS included 857 PD cases with a positive family history of PD. In addition, we performed a second GWAS with onset age as the phenotype using publicly available data from 440 randomly ascertained PD cases [19]. We conducted a meta-analysis of the two studies comprising approximately 2 million SNPs imputed using HapMap data. Finally, a replication study of the top findings from the meta-analysis was performed in an independent sample of 747 randomly ascertained PD cases from Milan, Italy.

Methods
PD Cases
One PD case (n = 935) from each family recruited from two ongoing studies of familial PD, the GenePD study and the PROGENI study, was selected for the GWAS. Both studies recruited families consisting of at least two members meeting diagnostic criteria for PD. PD cases underwent a uniform neurological evaluation that employed PD diagnostic criteria based on a modified version of the United Kingdom PD Society Brain Bank Criteria [21]. Detailed descriptions of the inclusion and exclusion criteria are described elsewhere for the PROGENI [22] and GenePD [14] studies. The analyzed sample was exclusively white, non-Hispanic.

No subject known to have a disease-producing mutation was included in this analysis. All cases were known to be negative for the LRRK2 G2019S mutation, and many, but not all, were also screened for PARK1 (SNCA) (N screened = 702), PARK2 (parkin)(N = 593), PARK7 (DJ1)(N = 328), and NR4A2 (N = 550) [9,10,12,23-27]. PD onset age was determined by interview and reflected the age of first symptom of PD, which commonly preceded age of physician diagnosis. The reliability of ascertaining age of onset through interview compared to medical records has been estimated as 0.94 [28].

In addition, control samples (n = 895) obtained from the NINDS Human Genetics Resource Center DNA and Cell Line Repository (Camden, NJ) were intermixed on the same plates as the cases when genotyped. All selected control samples were reported to be white, non-Hispanic. While not used in the analysis of onset age of PD, these samples were used for SNP and sample quality assessment and genotype imputation as described below.

Appropriate written informed consent was obtained for all samples included in this study.

Microarray Genotyping and Quality Assessment
Genotyping was performed by the Center for Inherited Disease Research (CIDR) using the Illumina HumanCNV370 version1_C BeadChips (Illumina, San Diego, CA, USA) and the Illumina Infinium II assay protocol [29]. As previously described [20], 78 cases and 28 controls were removed due to a low genotype call rate (<98% of SNPs), cryptic relatives in the sample, or population stratification. The final sample included 857 PD cases (all derived from whole blood) and 867 control DNA samples (all derived from lymphoblastoid cell
lines). Genotype calls and quality scores were determined from allele cluster definitions for each SNP as determined by the Illumina BeadStudio Genotyping Module version 3.1.14 and the combined intensity data from 96% of study samples as previously described [20]. Genotype calls with a quality score (Gencall value) of 0.25 or higher were considered acceptable. Blind duplicate reproducibility was 99.98%. Although we performed additional SNP filtering described below, the CICDR released dataset contained 344,301 SNPs and is available on dbGaP (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap; Accession number: phs000126.v1.p1).

SNPs were removed if: 1) the call rate of the SNPs was lower than 98% (n = 7,764), 2) the minor allele frequency was less than 0.01 in the combined case and control dataset (n = 7,667), 3) there were differential rates of missing genotypes in the cases and controls (n = 75) or males and females (n = 271), or 4) significant deviation from Hardy Weinberg equilibrium was observed in the control sample (n = 906). Many markers failed multiple tests. The final dataset consisted of 328,189 SNPs that passed all quality control measures (94.6% of all attempted SNPs).

Mayo-Perlegen LEAPS sample

Publicly available SNP genotype data for 443 PD cases was accessed from dbGaP for the GWAS "Mayo-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson’s Solutions) Collaboration" (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap; Accession number: phs000048.v1.p1; [19]). This data set included genotyping for the 198,345 SNPs meeting the quality control standards described by Maraganore et al. [19]. We examined these data for population stratification and for cryptic relatedness. While no population outliers were observed, three individuals were removed due to apparent second-degree relationships identified during our review of the data, leaving a sample of 440 PD cases included in these analyses. While data are also publicly available for 267 PD cases from the Fung et al. GWAS [18], that study only included cases whose age of PD onset was 55 years or greater resulting in an age distribution that was significantly different from the other two samples (p < 0.0001). Therefore, we did not include the cases from that study in our meta-analysis due to limited variability in the onset distribution.

Population stratification

Both the GenePD-PROGENI and Mayo-Perlegen LEAPS sample sets were screened for population outliers during initial QC. In addition, after the final study sample sets were determined, principle components were recalculated independently for each study using only the samples included in the final analyses. Association to onset age was then tested for the first six principle components for each study. No association (p < 0.05) between the first six principle components and onset age was seen in either study sample.

Imputation

Imputation was performed to increase the power of the meta-analysis and to facilitate the joint analysis of results generated from Mayo-Perlegen LEAPS and our GWAS, which were genotyped on different platforms. The program MACH 1.0 (compiled using Intel’s optimized compiler) was used to impute genotypes for 2,543,887 autosomal SNPs characterized in the HapMap project [30-32]. The GenePD-PROGENI sample and the Mayo-Perlegen LEAPS cohort were imputed separately using phased haplotype data downloaded from the HapMap project website [33].

In the GenePD-PROGENI sample, imputation was performed using both cases and controls. A subset of 200 cases and controls with high call rates was selected to perform the initial model parameter calculation. Next, imputation was performed on all participants using all autosomal SNPs where the strand was not ambiguous (i.e. not an A/T or G/C SNP) and that passed all other quality control measures described previously. In the Mayo-Perlegen LEAPS cohort, all PD cases were used to compute the initial model parameters. All unambiguous autosomal SNPs passing quality review, as described in the original Mayo-Perlegen LEAPS GWAS [19], were used for this imputation. In addition, because the genotyping platform used in the Mayo-Perlegen LEAPS study included a high percentage of ambiguous SNPs, those ambiguous SNPs which could be confidently matched to HapMap strands through the comparison of minor allele frequencies were also used for imputation. For each study, initial model parameters were calculated using 100 iterations. SNP quality was assessed using the Rsq metric, which estimates the squared correlation between imputed and actual genotypes with SNPs having an Rsq <0.3 excluded from further study (N = 61,271).

At the time of this study, only autosomal imputation is supported by MACH; therefore, the program IMPUTE was used to impute SNPs on the X chromosome [34,35]http://www.stats.ox.ac.uk/~marchini/software/gwas/impute.html. Haplotype, legend, and recombination rate files based upon HapMap (rel#21-NCBI build 35) were downloaded from IMPUTE’s website for use in imputing 64,621 SNPs in the non-pseudoautosomal region of the X chromosome. Imputation was run in all participants in the GenePD-PROGENI study and separately in all PD cases in the Mayo-Perlegen LEAPS GWAS, using all X-chromosome SNPs that passed quality assessment as described above and were found in the legend file provided by IMPUTE.
Statistical Analyses
In order to meta-analyze results from both studies, association to onset age was performed using exclusively imputed, not genotyped data, for 1,861,750 SNPs passing imputation QC and with minor allele frequencies greater than 10%. Imputed SNPs with minor allele frequencies less than 10% in the GenePD-PROGENI sample (N = 620,866) were excluded from the association analyses to avoid false positives that can occur in the analysis of low minor allele frequency SNPs. To evaluate SNP association under an additive mode of inheritance, the predicted allele dosage for each genotype estimated by MACH for autosomes or by IMPUTE for the X chromosome was used. To model the recessive and dominant modes of inheritance, the genotype probabilities calculated by MACH were used. While a combination of additive and recessive models can have high power to detect dominant effects, this power decreases as the minor allele frequency of the SNP nears 0.5 [36]. Therefore, we studied additive, dominant and recessive models. The total probability of having either one copy or two copies of the minor allele was used for the dominant model and the probability of having two copies of the minor allele was used for the recessive model. Recessive and dominant models were not studied for the X chromosome. Linear regression analyses were performed using SAS v9.1.

Meta-Analysis
Since the genotypic data was generated on different arrays (Illumina and Perlegen) with few SNPs in common, we employed a conservative meta-analytic approach to combine results from the two studies. Meta-analysis of the results of the linear regression of the imputed data for the GenePD-PROGENI sample and the Mayo-Perlegen LEAPS cohort was performed using METAL http://www.sph.umich.edu/csg/abecasis/metal/. As is common in GWAS meta-analyses, a fixed effects model with standard error weighting was used, as random-effects models may be too conservative in GWA studies with a small number of studies[37].

Replication Study
In order to validate the top findings from the GWAS meta-analysis, an additional sample of 896 PD cases with reported ages of PD onset was provided by the Parkinson Institute - Istituti Clinici di Perfezionamento, Milan, Italy from the “Human Genetic Bank of Patients Affected by PD and Parkinsonisms”. These cases were recruited irrespective of family history or onset age, and similar to the GenePD-PROGENI, used the UK Parkinson’s Disease Society Brain Bank criteria to confirm idiopathic PD [38] and defined onset as the age of first symptom of PD.

Twenty-four SNPs were selected based upon the following criteria (1) a p-value less than 0.00001 in the meta-analysis of the two GWAS, with (2) a consistent direction of effect in both studies. Nine SNPs meeting these criteria were identified from the additive inheritance model, thirteen from the dominant model and eleven from the recessive model. Four SNPs were identified in both the additive and dominant model, one SNP was identified in both the additive and recessive model and one SNP was identified by all three models. Three gene regions were identified under two different genetic models with different SNPs. For each multiply-nominated gene region, the SNP from the model with the smaller p-value was selected for replication. These SNPs were genotyped using TaqMan technology implemented on the ABI PRISM® 7900HT Sequence Detection system (Applied Biosystems: Foster City, CA) at Boston University School of Medicine. Individual samples (149) with genotyping call rates of less than 95% were excluded from further analysis.

Association of the 24 SNPs to onset age was evaluated in the 747 remaining Italian cases using linear regression performed with the software Plink v1.01 http://pngu.mgh.harvard.edu/purcell/plink/[39] using the corresponding genetic model (additive, recessive or dominant) by which each SNP was originally identified. A final fixed effects meta-analysis of all three studies was performed using METAL.

To distinguish whether associations observed were to age in general, as opposed to age at onset of PD, linear regression to censoring age was performed in the 867 NINDS control samples genotyped with the GenePD-PROGENI cases.

Results
Demographic characteristics of the three samples studied are shown in Table 1. All three studies have a similar percentage of male participants. The GenePD-PROGENI and Mayo-Perlegen LEAPS samples have similar mean ages of PD onset, while the Italian sample has a somewhat younger average age at onset. The GenePD-PROGENI sample has the widest range of onset ages from 19 to 90 years while the Mayo-Perlegen LEAPS has no participants under 30 years of age and the Italian sample has no participants over 81 years of age. No significant differences in onset age are seen between men and women in any of the studies.

Supplementary Tables S1 (additive model), S2 (dominant model) and S3 (recessive model) (see Additional file 1) present the top SNPs from each region with a meta-analysis p < 0.0001 for imputed SNP data in the GenePD-PROGENI and Mayo-Perlegen LEAPS studies. Twenty-four SNPs with meta-analysis p-values of less than 0.00001 and with a consistent direction of effect for both GWAS were genotyped in the Italian replication sample of 747 PD cases (Table 2). SNPs genotyped in either the GenePD-PROGENI or Mayo-Perlegen LEAPS GWAS platforms are
Table 1: Age and sex distribution of the GenePD-PROGENI, Mayo-Perlegen LEAPS, and Milan, Italian study samples.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>% Male</th>
<th>Mean Onset</th>
<th>Std Dev Onset</th>
<th>Range Onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>GenePD-PROGENI</td>
<td>857</td>
<td>59.2</td>
<td>61.9</td>
<td>10.9</td>
<td>19-90</td>
</tr>
<tr>
<td>Mayo-Perlegen LEAPS</td>
<td>440</td>
<td>61.6</td>
<td>61.0</td>
<td>11.2</td>
<td>30-94</td>
</tr>
<tr>
<td>Milan, Italian</td>
<td>747</td>
<td>59.4</td>
<td>55.2</td>
<td>10.8</td>
<td>20-81</td>
</tr>
</tbody>
</table>

Although ten of the replication SNPs showed a consistent direction of effect across all three studies, only two SNPs (both of which were genotyped on the Mayo-Perlegen LEAPS GWAS platforms) resulted in increased statistical evidence of association when combining the three studies, and both of these were in a recessive model. The most strongly associated SNP, rs10767971, located on chromosome 11 between the genes QSER1 and PRRG4, was associated with a 3.2 year older PD onset in individuals with 2 copies of the minor allele ($p = 5.4 \times 10^{-7}$ in the 3 sample meta-analysis, compared to $4.3 \times 10^{-6}$ in the 2 sample meta-analysis). Conversely, an estimated 6.9 year earlier age of onset ($p = 8.7 \times 10^{-6}$ in the 3 sample meta-analysis compared to $9.8 \times 10^{-6}$ in the 2 sample meta-analysis) was observed for the SNP rs7577851, located in the 16th intron of the gene AAK1 on chromosome 2.

The most highly associated SNP in both the additive and dominant three sample meta-analyses, rs17565841, is located approximately 3 kb from the 3’ end of the gene OCA2 on chromosome 15. This SNP was associated with an average 2.8 years younger onset age ($p = 2.6 \times 10^{-6}$) under an additive model and a 3.3 years younger onset age ($p = 2.1 \times 10^{-6}$) under a dominant model. Inclusion of the Italian cases in the meta-analysis did not strengthen the evidence of association compared to the results seen in the two sample meta-analyses ($9.1 \times 10^{-7}$ for additive and $1.9 \times 10^{-6}$ for dominant). However, in the Italian replication sample, this SNP did provide modest statistical association to onset age ($p = 0.05$ for additive and $p = 0.04$ for dominant) with the same direction of effect seen in the two other studies.

Also showing consistent directions of effect and p-values in the three-sample meta-analysis at the level of $p < 1 \times 10^{-5}$ were two SNPs located in the genes DSG3 and ATF6. The SNP rs1941184, had the second best p-value identified under the dominant model and is located in the third intron of the gene DSG3 on chromosome 18. This SNP was associated with an average 2.3 year younger age of onset of PD across the three studies ($p = 4.3 \times 10^{-6}$). The SNP rs10918270, located in the 15th intron of the gene ATF6 on chromosome 1, was identified under both additive and dominant modes of inheritance, but showed stronger association in the three sample meta-analysis under the dominant model ($p = 7.5 \times 10^{-6}$) with an average 2.3 year younger onset of PD. No association to age with direction of effect consistent with that observed for onset age of PD was seen in the control sample for any of the 24 SNPs at a significance level equal to 0.05.

**Discussion**

We present results from the first GWAS for age at onset of PD, including a meta-analysis with the publicly available Mayo-Perlegen LEAPS GWAS data (dbGaP Study Accession: phs000048.v1.p1) and a follow-up replication study in an independent PD sample recruited in Milan, Italy. Differences were observed in the age distributions in the three populations used in this study (Table 1). However, there were no imposed age restrictions in any of the studies and a wide distribution of ages was represented in all three populations, each with a range of greater than 60 years.

No SNP reached the commonly accepted criterion for genome-wide significance of $p < 5 \times 10^{-8}$ [40], which is based on recent estimates of independent genomewide sequence variation to maintain 5% genomewide type I error rate [41,42]. While this criterion provides an appropriate cutoff for determining significance for the large number of SNPs provided by imputation, this measure does not account for the testing of multiple genetic models, as was performed in this study. Despite the lack of genomewide significance, the meta-analysis in this study showed evidence of several interesting loci with consistent effects on onset age of PD across the three independent populations studied.

The SNP with the strongest evidence for association to onset age, rs10767971, is associated with a later age of onset under a recessive model. This SNP is nearly equidistant between the genes PRRG4 and QSER1 on chromosome 11, at just under 20 kb from each. The association results for nearby SNPs studied in the meta-analysis of the two GWAS as well as the LD structure and recombination rates for the region are shown in Figure 2A and 2B under...
recessive and dominant models. While the strongest evidence of association appeared with a recessive model, many nearby SNPs in strong LD showed similar association under the dominant model. The pattern suggests a single association, spanning several SNPs that may be more evident around the PRGG4 gene and may extend toward the CCDC73 and EIF3M genes.

The SNP showing the strongest association to PD onset across the three populations under both the additive and dominant models lies 3 kb from the 3’ end of the gene OCA2 (oculocutaneous albinism II). Mutations in OCA2 are responsible for the most common form of albinism, oculocutaneous albinism type 2 [43] and several studies have shown this gene to be associated with common skin, hair and eye color variation found in European popula-

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>SNP position</th>
<th>Minor Allele</th>
<th>MAF</th>
<th>Genes in Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs7556447</td>
<td>2,335,959</td>
<td>G</td>
<td>0.24</td>
<td>PEX10</td>
</tr>
<tr>
<td>1</td>
<td>rs1355637</td>
<td>148,849,062</td>
<td>T</td>
<td>0.19</td>
<td>MCL1</td>
</tr>
<tr>
<td>1</td>
<td>rs10918270</td>
<td>160,182,125</td>
<td>A</td>
<td>0.41</td>
<td>ATF6</td>
</tr>
<tr>
<td>2</td>
<td>rs1899121</td>
<td>20,231,454</td>
<td>C</td>
<td>0.50</td>
<td>SDC1</td>
</tr>
<tr>
<td>2</td>
<td>rs7577851</td>
<td>69,577,214</td>
<td>T</td>
<td>0.17</td>
<td>AAK1</td>
</tr>
<tr>
<td>2</td>
<td>rs17817190</td>
<td>134,068,118</td>
<td>G</td>
<td>0.11</td>
<td>NAP5</td>
</tr>
<tr>
<td>6</td>
<td>rs6936388</td>
<td>42,275,464</td>
<td>T</td>
<td>0.25</td>
<td>CCND3</td>
</tr>
<tr>
<td>6</td>
<td>rs1572662</td>
<td>154,853,289</td>
<td>A</td>
<td>0.36</td>
<td>CNKSR3</td>
</tr>
<tr>
<td>7</td>
<td>rs1420143</td>
<td>29,514,849</td>
<td>C</td>
<td>0.44</td>
<td>CHN2</td>
</tr>
<tr>
<td>7</td>
<td>rs17663983</td>
<td>76,243,177</td>
<td>T</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs12261736</td>
<td>52,738,610</td>
<td>T</td>
<td>0.28</td>
<td>PRKCG1</td>
</tr>
<tr>
<td>10</td>
<td>rs7076519</td>
<td>127,738,829</td>
<td>C</td>
<td>0.33</td>
<td>ADAM12</td>
</tr>
<tr>
<td>11</td>
<td>rs10767971</td>
<td>32,852,240</td>
<td>T</td>
<td>0.42</td>
<td>PRGG4</td>
</tr>
<tr>
<td>12</td>
<td>rs12829697</td>
<td>30,067,223</td>
<td>G</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs10773917</td>
<td>130,572,784</td>
<td>T</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>rs4771006</td>
<td>26,180,684</td>
<td>A</td>
<td>0.49</td>
<td>WASF3</td>
</tr>
<tr>
<td>13</td>
<td>rs676495</td>
<td>29,561,754</td>
<td>T</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>rs10147486</td>
<td>47,054,769</td>
<td>T</td>
<td>0.27</td>
<td>MDGA2</td>
</tr>
<tr>
<td>15</td>
<td>rs17565841</td>
<td>25,670,842</td>
<td>A</td>
<td>0.10</td>
<td>OCA2</td>
</tr>
<tr>
<td>17</td>
<td>rs9904572</td>
<td>12,803,084</td>
<td>A</td>
<td>0.34</td>
<td>RICH2</td>
</tr>
<tr>
<td>17</td>
<td>rs4791571</td>
<td>14,155,801</td>
<td>A</td>
<td>0.42</td>
<td>HS3ST3B1</td>
</tr>
<tr>
<td>18</td>
<td>rs1941184</td>
<td>27,292,121</td>
<td>C</td>
<td>0.30</td>
<td>DSG3</td>
</tr>
<tr>
<td>19</td>
<td>rs10420134</td>
<td>40,953,768</td>
<td>G</td>
<td>0.18</td>
<td>C19orf55</td>
</tr>
<tr>
<td>20</td>
<td>rs3887942</td>
<td>23,925,072</td>
<td>G</td>
<td>0.19</td>
<td>GGTLC1</td>
</tr>
</tbody>
</table>

\[a\] From NCBI Build 36 reference; [b] Nearest genes within 100 kb of SNP; [c] SNP was genotyped in Mayo-Perlegen LEAPS GWAS; [d] SNP was genotyped in GenePD-PROGENI GWAS

Table 2: Location and minor allele frequency (MAF) information for 24 SNPs selected for replication study in Italian sample
Variations in \textit{OCA2} have been associated with susceptibility to melanoma \cite{49}, which has also been reported to occur with increased frequency among PD cases \cite{50-52}. Because of the targeted degradation of pigmented neurons in PD brains, these associations have lead to a hypothesized link between genes involved in pigmentation, such as \textit{OCA2}, and PD, likely mediated through common elements in melanin and neuromelanin synthesis\cite{53}. The protein encoded by \textit{OCA2}, 'P' protein, is involved in the transport of tyrosine, a precursor to melanin, as well as in the regulation of melanosomal pH which may be key to the initiation of the enzyme controlling melanin synthesis in melanocytes \cite{54}. It is not clear whether synthesis of neuromelanin is also regulated.

### Table 3: Additive model study-specific and meta-analysis results

<table>
<thead>
<tr>
<th>SNP</th>
<th>PD-PROGENI</th>
<th>Mayo-Perlegen</th>
<th>LEAPS</th>
<th>GWAS Meta-analysis</th>
<th>Milan, Italian</th>
<th>Three Sample Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17565841</td>
<td>-4.66</td>
<td>2.7 x 10^{-4}</td>
<td>-3.64</td>
<td>1.7 x 10^{-1}</td>
<td>-4.53</td>
<td>9.1 x 10^{-7}</td>
</tr>
<tr>
<td>rs10918270</td>
<td>-2.12</td>
<td>1.1 x 10^{-4}</td>
<td>-1.90</td>
<td>1.6 x 10^{-2}</td>
<td>-2.05</td>
<td>4.8 x 10^{-6}</td>
</tr>
<tr>
<td>rs12261736</td>
<td>-3.61</td>
<td>1.8 x 10^{-7}</td>
<td>-1.17</td>
<td>4.0 x 10^{-1}</td>
<td>-3.13</td>
<td>3.5 x 10^{-7}</td>
</tr>
<tr>
<td>rs4771006</td>
<td>2.35</td>
<td>1.6 x 10^{-3}</td>
<td>2.59</td>
<td>1.6 x 10^{-3}</td>
<td>2.46</td>
<td>7.3 x 10^{-6}</td>
</tr>
<tr>
<td>rs1420143</td>
<td>2.08</td>
<td>1.4 x 10^{-4}</td>
<td>2.58</td>
<td>8.1 x 10^{-3}</td>
<td>2.20</td>
<td>3.6 x 10^{-6}</td>
</tr>
<tr>
<td>rs6936388</td>
<td>2.45</td>
<td>4.2 x 10^{-4}</td>
<td>2.98</td>
<td>7.2 x 10^{-3}</td>
<td>2.60</td>
<td>9.2 x 10^{-6}</td>
</tr>
<tr>
<td>rs10147486</td>
<td>2.21</td>
<td>3.1 x 10^{-4}</td>
<td>2.89</td>
<td>8.0 x 10^{-4}</td>
<td>2.44</td>
<td>9.2 x 10^{-7}</td>
</tr>
<tr>
<td>rs7076519</td>
<td>-2.71</td>
<td>9.3 x 10^{-6}</td>
<td>-2.53</td>
<td>2.5 x 10^{-2}</td>
<td>-2.67</td>
<td>6.0 x 10^{-7}</td>
</tr>
</tbody>
</table>

### Table 4: Dominant model study-specific and meta-analysis results

<table>
<thead>
<tr>
<th>SNP</th>
<th>PD-PROGENI</th>
<th>Mayo-Perlegen</th>
<th>LEAPS</th>
<th>GWAS Meta-analysis</th>
<th>Milan, Italian</th>
<th>Three Sample Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17565841</td>
<td>-5.08</td>
<td>4.7 x 10^{-4}</td>
<td>-3.68</td>
<td>2.0 x 10^{-1}</td>
<td>-4.90</td>
<td>1.9 x 10^{-6}</td>
</tr>
<tr>
<td>rs1941184</td>
<td>-2.94</td>
<td>1.3 x 10^{-4}</td>
<td>-2.94</td>
<td>1.0 x 10^{-2}</td>
<td>-2.94</td>
<td>3.7 x 10^{-6}</td>
</tr>
<tr>
<td>rs10918270</td>
<td>-3.26</td>
<td>2.9 x 10^{-5}</td>
<td>-2.74</td>
<td>1.5 x 10^{-2}</td>
<td>-3.09</td>
<td>1.2 x 10^{-6}</td>
</tr>
<tr>
<td>rs3887942</td>
<td>4.17</td>
<td>7.2 x 10^{-4}</td>
<td>7.10</td>
<td>1.5 x 10^{-3}</td>
<td>4.85</td>
<td>6.4 x 10^{-6}</td>
</tr>
<tr>
<td>rs12261736</td>
<td>-4.03</td>
<td>9.5 x 10^{-4}</td>
<td>-2.04</td>
<td>2.8 x 10^{-1}</td>
<td>-3.66</td>
<td>7.3 x 10^{-6}</td>
</tr>
<tr>
<td>rs10773917</td>
<td>3.15</td>
<td>3.8 x 10^{-5}</td>
<td>2.58</td>
<td>3.5 x 10^{-2}</td>
<td>2.99</td>
<td>3.7 x 10^{-6}</td>
</tr>
<tr>
<td>rs10420134</td>
<td>-3.21</td>
<td>5.8 x 10^{-5}</td>
<td>-3.12</td>
<td>5.1 x 10^{-2}</td>
<td>-3.19</td>
<td>7.1 x 10^{-6}</td>
</tr>
<tr>
<td>rs9904572</td>
<td>3.57</td>
<td>4.3 x 10^{-4}</td>
<td>2.72</td>
<td>6.4 x 10^{-2}</td>
<td>3.38</td>
<td>7.2 x 10^{-7}</td>
</tr>
<tr>
<td>rs17817190</td>
<td>-3.87</td>
<td>3.9 x 10^{-5}</td>
<td>-5.27</td>
<td>2.1 x 10^{-2}</td>
<td>-4.07</td>
<td>2.5 x 10^{-6}</td>
</tr>
<tr>
<td>rs1572662</td>
<td>3.57</td>
<td>4.1 x 10^{-5}</td>
<td>4.89</td>
<td>7.5 x 10^{-2}</td>
<td>3.69</td>
<td>7.8 x 10^{-6}</td>
</tr>
<tr>
<td>rs7076519</td>
<td>-3.73</td>
<td>1.2 x 10^{-5}</td>
<td>-3.81</td>
<td>3.0 x 10^{-2}</td>
<td>-3.74</td>
<td>9.3 x 10^{-7}</td>
</tr>
</tbody>
</table>
through melanosomal pH or is affected by the 'P' protein in a similar way [55]. Nevertheless, the association between OCA2 and younger age of onset are suggestive of a neuromelanin-related mechanism of effect. The association of an intronic SNP in the gene DSG3 (desmoglein 3 (pemphigus vulgaris antigen)) may also be indicative of a neuromelanin related effect on onset age of PD. The protein encoded by DSG3 is the autoantigen for the autoimmune skin disease pemphigus vulgaris and this gene is expressed primarily in skin. Interestingly, some reports have demonstrated increased expression of DSG3 in melanocytes compared to keratinocytes (the most common cell type in the epidermis) [56].

An intronic SNP in the gene ATF6 also showed strong association to earlier age of onset of PD. ATF6 (activating transcription factor 6) transcribes a transcription factor localized to the endoplasmic reticulum (ER). The ATF6 protein is a critical regulator of the unfolded protein response (UPR), a highly conserved pathway activated in response to ER stress, and is a protective cellular response to the accumulation of misfolded proteins [57-59]. The UPR has been implicated in neurotoxin based cellular models of PD [60] and has also been shown to be activated by the over-expression of α-synuclein in yeast cells [61]. More recently, postmortem studies of PD case and control brains have shown activation of the UPR in cases, but not controls and that this activation is associated with the aggregation of α-synuclein [62].

The finding of an intronic SNP in AAK1 associated with PD onset age is intriguing because of the gene's genomic location, its function, and its close relation to a gene identified for PD susceptibility. The AAK1 (AP2 associated kinase 1) gene is located on chromosome 2p14, near 2p13, which has previously been implicated as the PARK3 locus[63], and for which linkage to onset age was demonstrated in both the GenePD [64] and PROGENI [65] studies. The AAK1 gene itself has not been previously implicated by positional mapping, but a microarray study of PD brain compared to controls demonstrated differential expression of AAK1 [66]. In our past GWAS of PD susceptibility in the GenePD-PROGENI cohort [20], the region containing the gene GAK (cyclin G-associated kinase) had the strongest evidence for association (chromosome 4p). The AAK1 and GAK genes both function at multiple steps in clathrin-mediated vesicular transport and the two kinases likely have some redundant functions related to their homologous serine/threonine-kinase domain [68]. Recently, cathepsin D was implicated as the main lysosomal enzyme involved in α-synuclein degradation [69], and depletion of GAK was shown to impair the lysosomal sorting of cathepsin D [68]. Thus, the finding
that AAK1 influences onset age and GAK influences risk in familial PD suggest that pathways involving lysosomal activity influence PD risk.

Several previous genome scans have provided evidence of loci linked to onset age, including 2p13 seen in both the GenePD and PROGENI studies, as mentioned above. Evidence of linkage has also been reported on chromosomes 1p, 1q, 8q, 9q, 10q, 20 and 21[64,65,70], but there is little overlap between these linkage regions and our top meta-analysis association results. Aside from the identification of the AAK1 gene near 2p13, the strongest association result observed under one of the previously reported linkage regions occurs near the gene sortilin-related VPS10 domain containing receptor 3 (SORCS3) located near the LOD score peak at 10q, originally identified in a combined linkage scan of onset age in PD and Alzheimer’s disease [70]. This association is seen most strongly under a dominant model (p = 2.5 × 10^-5) with minor allele carriers having an estimated older onset age by 3.3 years (see Additional file 1: Table S2).

A final region of interest is 15q26.2 that includes the gene MCTP2. Several SNPs in this region showed association to earlier onset age of PD in the meta-analysis of the two GWAS under the recessive model (see Additional file 1: Table S3). These SNPs did not reach the criteria for inclusion in the replication study, as the strongest p-value seen was 2.2 × 10^-5 with rs17504636. This SNP was associated with a 9.2 year earlier average PD onset. This region overlaps with a SNP reported in the susceptibility GWAS including these cases [20]. In that study the SNP rs4476132 was associated with PD susceptibility with an odds ratio of 1.3 (p = 7.7 × 10^-5- meta-analysis with additive model). This overlap is consistent with a locus associated with a risk for younger-onset PD or with an effect modifying age dependent penetrance. The gene in this region, MCTP2 (multiple C2 domains, transmembrane 2), is expressed in the brain and has been implicated in linkage and association studies of abdominal fat [71] and major depression [72].

Important distinctions can be made between those genes that influence susceptibility for developing disease, and the genetic modifiers that influence penetrance or, as studied here, onset age. Perhaps the best examples for genetic modifiers are seen for Huntington’s disease (HD) where an expanded CAG trinucleotide repeat on chromosome 4p16.3
causes the disease, but wide variation in onset age is evident for individuals with identical repeat lengths. The identification of the genes that presumably interact with huntingtin to produce relatively younger or older onset for a given repeat size provide insight into the pathogenic mechanisms for HD, as well as therapeutic targets for intervention [73-75]. Similarly, identifying those genes and their products that are associated with older onset in PD may provide insight into the disease mechanisms and processes for delaying onset with implications for novel treatments.

**Conclusion**

The identification of the 15q26.2 region as well as the related genes AAK1 and GAK in the studies of PD onset age and susceptibility highlights the importance of the continued study of both of these traits, both separately and in combination. The direct overlap in affection and onset age association results in the 15q26.2 region shows this to be a candidate region that would benefit from further examination with consideration of important age-related effects, for example in studies correlating expression of genes in this region with onset age. The identification of association to onset age with the gene AAK1, in the same pathway as a previously identified susceptibility-associated gene GAK highlights the importance of genetic pathways in PD etiology, showing that the genes along the same pathway may have redundant effects or may modify disease pathology different ways, observed by differences in disease onset and progression. Studying PD in the context of onset age provides fundamental insight into the disease process and is essential to understanding mechanisms that modify disease penetrance and therefore may be key in identifying therapeutic targets.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

JCL participated in conception and design of the study, conducted statistical analyses, participated in the interpretation of data, and drafted the manuscript. NP, AD and JBW participated in the conception and design of the study, conducted statistical analyses, participated in the interpretation of data, and revised the article critically for important intellectual content. SG, GP, CBM, ALD, CH, JFG, WCN, RHM, and TF participated in the conception and design of the study, participated in the interpretation of data, and revised article critically for important intellectual content. All authors read and approved the final manuscript.

**Additional material**

**Additional file 1**

Supplementary Tables. The excel file includes three supplementary tables. Supplementary Table S1 includes regions showing associations at p < 0.0001 for the additive model GWAS meta-analysis and study-specific results. Supplementary Table S2 includes regions showing associations at p < 0.0001 for the dominant model GWAS meta-analysis and study-specific results. Supplementary Table S3 includes Regions showing associations at p < 0.0001 for the recessive model GWAS meta-analysis and study-specific results

Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2350-10-98-S1.XLS]
Acknowledgements
This project was supported by R01 NS37167, R01 NS036711, the Robert P. & Judith N. Goldberg Foundation, the Bumpus Foundation and the Harvard NeuroDiscovery Center. This study used samples from the NINDS Human Genetics Resource Center DNA and Cell Line Repository http://ccr.coriell.org/ninds, as well as clinical data.

DNA samples contributed by the Parkinson Institute - Istituti Clinici di Perfezionamento, Milan, Italy were from the "Human genetic bank of patients affected by PD and parkinsonisms", supported by Italian Telethon grant n. GTB07001 and by the “Fondazione Grigioni per il Morbo di Parkinson”.

A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal grant from the National Institutes of Health to The Johns Hopkins University, contract number HHSN26820072096C. We particularly thank Drs. Kimberly Doheny and Elizabeth Pugh from CIDR and Mr. Justin Paschall from NCBi for their assistance.


Funding support for the Mayo-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson’s Solutions) Collaboration Study was provided by the Michael J. Fox Foundation. Additional support was provided by the NIH grants ES10751 and NS33978. The datasets used for the analyses described in this manuscript were obtained from dbGaP found at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000048.v1.p1 through dbGaP study accession numbers phs000048. Data for the Mayo-Perlegen LEAPS Collaboration Study were provided by Demetrius M. Maraganore, MD on behalf of his collaborators at the Mayo Clinic and at Perlegen Sciences, Inc.

The following are members of the PROGENI Steering Committee. University of Tennessee Health Science Center: R. F. Pfeiffer; University of Rochester: F. Marshall, D. Oakes, A. Rudolph, A. Shniman; Columbia University Medical Center: K. Marder; Indiana University School of Medicine: P. M. Conneally, T. Foroud, C. Halter; University of Kansas Medical Center: K. Lyons; Eli Lilly & Company: E. Siemers; Medical College of Ohio: L. Emers; University of California, Irvine: N. Hermanowicz.

The following are members of the GenePD Steering Committee. University of Virginia Health System: G. F. Wooten; UMDNJ-Robert Wood Johnson Medical School: L. Golbe; Center for Human Genetic Research, Massachusettes General Hospital, Harvard Medical School: J. F. Gusella; Boston University School of Medicine: R. H. Myers.

We thank the subjects for their participation in this research study.


PROGENI Molecular Genetic Laboratory: Division of Human Genetics, Cincinnati Children’s Hospital Medical Center: Michael W. Pauculo, Diane K. Marek, Veronika E. Elsaesser

GenePD Investigators and Coordinators: University Southern California School of Medicine: M. Lew; University of Calgary: O. Suchowsky, S. Furtado; University of Lubeck, Germany: C. Klein; UMDNJ-Robert Wood Johnson Medical School: L. Golbe, M. H. Mark; Massachusetts Gen-


Pre-publication history
The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2350/10/98/pre-pub