A pathway-based analysis provides additional support for an immune-related genetic susceptibility to Parkinson's disease

HOLMANS, Peter, et al. & International Parkinson's Disease Genomics Consortium (IPDGC)

POLLAK, Pierre (Collab.)

Abstract

Parkinson's disease (PD) is the second most common neurodegenerative disease affecting 1-2% in people >60 and 3-4% in people >80. Genome-wide association (GWA) studies have now implicated significant evidence for association in at least 18 genomic regions. We have studied a large PD-meta analysis and identified a significant excess of SNPs (P < 1 × 10(-16)) that are associated with PD but fall short of the genome-wide significance threshold. This result was independent of variants at the 18 previously implicated regions and implies the presence of additional polygenic risk alleles. To understand how these loci increase risk of PD, we applied a pathway-based analysis, testing for biological functions that were significantly enriched for genes containing variants associated with PD. Analysing two independent GWA studies, we identified that both had a significant excess in the number of functional categories enriched for PD-associated genes (minimum P = 0.014 and P = 0.006, respectively). Moreover, 58 categories were significantly enriched for associated genes in both GWA studies (P < 0.001), implicating genes involved in [...]
A pathway-based analysis provides additional support for an immune-related genetic susceptibility to Parkinson’s disease

Peter Holmans1, Valentina Moskvina1, Lesley Jones1, Manu Sharma4,10, The International Parkinson’s Disease Genomics Consortium (IPDGC)1, Alexey Vedernikov1, Finja Buchel3, Mohamad Sadd5, Jose M. Bras6, Francesco Bettella8, Nayia Nicolaou9, Javier Simón-Sánchez9, Florian Mittag3, J. Raphael Gibbs2,6, Claudia Schulte4,10, Alexandra Durr11,12, Rita Guerreiro6, Dena Hernandez2,6, Alexis Brice11,12,13,14, Hreinn Stefánsson8, Kari Majamaa15, Thomas Gasser4,10, Peter Heutink3, Nicholas W. Wood6,7, Maria Martinez5, Andrew B. Singleton2, Michael A. Nalls2, John Hardy6, Huw R. Morris1 and Nigel M. Williams1,*

1Department of Psychological Medicine and Neurology, Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre in Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff CF14 4XN, UK, 2Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA, 3Center for Bioinformatics Tuebingen (ZBIT) and 4Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tuebingen, Tuebingen, Germany, 5Paul Sabatier University, Toulouse, France, 6Department of Molecular Neuroscience, Institute of Neurology and 7UCL Genetics Institute, University College London, London, UK, 8deCODE genetics, Scientific Services, Sturlugata, 8 IS-101 Reykjavik, Iceland, 9Section of Medical Genomics, Department of Clinical Genetics, VU University Medical Centre, Amsterdam, The Netherlands, 10DZNE – Deutsches Zentrum fur Neurodegenerative Erkrankungen (German Center for Neurodegenerative Diseases), Tuebingen, Germany, 11Inserm, UMRS975, CNRS UMR 7225, CRICM, F-75013 Paris, France, 12UPMC Univ Paris 06, UMRS975, F-75013, Paris, France, 13AP-HP, Hôpital de la Salpêtrière, Département de Génétique, Paris, France, 14Institut du Cerveau et de la Moelle Epinière, F-75013, Paris, France and 15Department of Medical Biochemistry and Molecular Biology, FIN-90014 University of Oulu, Finland

Received August 10, 2012; Revised October 31, 2012; Accepted November 16, 2012

Parkinson’s disease (PD) is the second most common neurodegenerative disease affecting 1–2% in people >60 and 3–4% in people >80. Genome-wide association (GWA) studies have now implicated significant evidence for association in at least 18 genomic regions. We have studied a large PD-meta analysis and identified a significant excess of SNPs (P < 1 × 10⁻¹⁶) that are associated with PD but fall short of the genome-wide significance threshold. This result was independent of variants at the 18 previously implicated regions and implies the presence of additional polygenic risk alleles. To understand how these loci increase risk of PD, we applied a pathway-based analysis, testing for biological functions that were significantly enriched for genes containing variants associated with PD. Analysing two independent GWA studies, we identified that both had a significant excess in the number of functional categories enriched for PD-associated genes (minimum P = 0.014 and P = 0.006, respectively). Moreover, 58 categories were significantly enriched for associated genes in both GWA studies (P < 0.001), implicating genes involved in the regulation of leucocyte/lymphocyte activity and also cytokine-mediated signalling as conferring an increased susceptibility to PD. These results

*To whom correspondence should be addressed at: Department of Psychological Medicine and Neurology, Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre in Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Heath Park, Cardiff CF14 4XN, UK. Tel: +44 2920687070; Email: williamsnm@cf.ac.uk

†A full list of the IPDGC members and affiliations appears in the appendix.

© The Author 2012. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
INTRODUCTION

Parkinson’s disease (PD) is the second most common neurodegenerative disease and is characterized by bradykinesia with resting tremor, stiffness and gait disturbance. Pathologically, the disease involves the deposition of alpha-synuclein as Lewy bodies and Lewy neurites in multiple motor and non-motor brain areas. The population prevalence is estimated at 0.3%, but this increases with age—rising to 1–2% in people >60 and 3–4% in people >80 years of age (1). Although the average age at onset is 68 years old, the onset is extremely variable, ranging from adolescence to old age (2,3).

The identification of rare highly penetrant mutations in genes causing familial PD (4–8) has had a considerable impact on our understanding of the pathogenesis of this complex and common disorder. Studies of rare Mendelian forms of PD show that increased alpha-synuclein fibril formation and abnormalities of mitophagy are of central importance in at least some forms of the disease. More recently, our understanding of the idiopathic form of the illness has been greatly enhanced by a number of large genome-wide association (GWA) studies (9–19). As well as providing unequivocal evidence that SNCA and MAPT are also risk factors for idiopathic PD (9,12,17–19), these studies have collectively identified variants at over 18 loci that significantly increase risk for PD (9,10). The challenge now is to understand how these genetic loci influence the risk of PD and how their cognate biological functions are influenced by variation at these loci. It is, however, inevitable that the findings of the primary GWA analyses will be limited to a subset of the variants of the strongest genetic effect and, consequently, a large number of true polygenic risk alleles that generate weaker evidence for association may be overlooked. Thus, despite their high levels of significance, the current crop of loci that are strongly associated with PD are thought to account for only a very small amount (1–2%) of the expected heritability of PD (20). As in other complex disorders, the missing heritability in PD is expected to relate to a combination of some variants that yield only relatively weak association signals and others that are not tagged by the genotyping microarray. Given that there is a substantial increase in the estimated heritability detected in PD GWA studies (24%) when weak effect loci are also considered (20), this strongly implies that a large proportion of genetic signal must lie below the genome-wide significance thresholds set in the primary analyses.

PD risk alleles are unlikely to be randomly distributed among genes, but instead are more likely to be distributed among genes whose functions are related. Indeed, the relevance of mitochondrial dysfunction (21–24) and protein degradation pathways (25–28) to PD pathogenesis has been highlighted by converging functional studies of Mendelian genes implicated in familial PD. In applying such a model to GWA data, one would expect to see an overall excess of association signal in a series of SNPs selected from genes which code for functionally related proteins and it might be anticipated that at least some of these shared biological functions relate to the current known Mendelian and non-Mendelian loci. Such effects have been identified in other complex disorders both when the analysis was limited to genes containing robustly associated SNPs (29–32) and also when the analysis was expanded to include genes containing SNPs whose evidence for association fell short of stringent genome-wide significant thresholds (30,32,33). To date, biological pathway-based analysis in PD has been limited to relatively small GWA data sets, and has implicated the axon guidance pathway as being relevant to PD (34), a finding that was recently replicated in a small independent data set (35).

In this study, we have used the ALIGATOR algorithm (32) to analyse SNPs associated with PD in two large independent GWA studies for enrichment in pre-defined sets of functionally related genes. Importantly, we have also confirmed our results using gene set enrichment analysis and established that our findings are not dependent on the most strongly associated loci. Our study implicates specific biological categories of genes that show evidence for association with PD and is an important next step in our translation of genetic susceptibility to an understanding of disease pathogenesis.

RESULTS

Assessment of SNPs passing sub-genome-wide significance thresholds

Analyses of the large International Parkinson’s Disease Genetics Consortium (IPDGC) meta analysis consisting of 5333 PD cases and 12219 controls have previously identified 18 regions showing evidence for genome-wide significant or suggestive evidence for association (9–12). In this study, we first set out to establish whether this data set harboured further evidence for additional PD genes. Our analysis revealed that a significant excess of SNPs surpassing different thresholds of significance remained even after all 18 previously implicated regions were excluded (Table 1). At our most stringent threshold (P < 0.0001), we identified a 4.2-fold excess of independent SNPs [following correction for linkage disequilibrium (LD)] significantly associated with PD than were expected (6792 compared with 1617, P < 1 x 10^-16). This suggested the existence of genuine additional PD susceptibility loci that failed to meet the stringent significance thresholds that were set in the previous studies.
ALIGATOR analysis

As it is not possible to implicate any individual locus from the observed enrichment of sub-GWA signal, we next chose to assess whether any biological functional categories were significantly enriched for genes carrying SNPs that were nominally associated with PD. To achieve this, we a priori considered the IPDGC meta-analysis (9) as two large independent PD-GWA data sets: UK-GWA (1705 PD patients and 5200 controls) and Meta-GWA (3628 PD patients and 7019 controls) and applied ALIGATOR (32) analysis to the top 5% of associated genes (n = 1050), which captured genes containing at least one SNP surpassing P = 0.00205 and P = 0.0024 in the Meta-GWA and UK-GWA studies, respectively. In each GWA study, ALIGATOR was then used to assess the functional categories that were enriched for these 1050 genes at varying thresholds of nominal significance (P = 0.05, P = 0.01, P = 0.001). This revealed a significant excess in the number of enriched functional categories compared with the simulated data, in both the Meta-GWA (minimum P = 0.014) and UK-GWA studies (minimum P = 0.006) (Table 2). In the largest data set (Meta-GWA), our most significant observation was of 498 functional categories that were enriched at P < 0.05 for associated genes (P = 0.014).

In order to further investigate the nature of this enrichment, we next assessed the level of overlap between the two studies by running ALIGATOR in the UK-GWA study, but this time restricting the analysis to only the 498 functional categories enriched at P < 0.05 in the Meta-GWA study. This revealed a significant excess of categories enriched in both GWA study, regardless of the threshold used (Table 3). The most significant evidence implicated 58 categories enriched for associated genes at P < 0.05 in both Meta-GWA and UK-GWA (P < 0.001) (Table 3).

These 58 functional categories contained a total of 269 different genes that were associated with PD (Supplementary Material, Table S1). However, given that a single gene can be present in multiple categories, grouping together categories based on their gene membership enabled us to further condense the 58 categories into 24 functionally related groups (Supplementary Material, Table S2), such that each category within a group shared at least 50% of its genes in common with at least one other category in the group. This revealed that 26 categories (45%) were present in just two functionally related groups which, importantly, also contained 9 of the 10 categories that showed the strongest evidence for enrichment in PD in both GWA studies (P < 0.01 in both). The categories included in these two biological groups implicate genes involved in the ‘regulation of leucocyte/lymphocyte activity’ and also ‘cytokine-mediated signalling’ as conferring an increased susceptibility to PD (Table 4). In total, these two functional groups contained 132 (49%) of the 269 genes associated with PD (Supplementary Material, Table S1); however, as 38 genes are shared by both, it reflects the biological relationship between these groups.

It is plausible that our results are being artificially biased by genes whose evidence for association is merely a consequence of LD with the very strong association signal of known GWA study hits. To investigate this possibility, we repeated our analysis using identical gene-wide significance thresholds, but this time excluding all 178 genes that were present at 18 genomic regions previously reported to be strongly associated with PD, including the HLA locus. Given that each of these regions is likely to span at least one true PD susceptibility gene which would now be excluded from our ALIGATOR analysis, this approach is highly conservative. Nevertheless, this analysis again revealed a significant excess in the number of enriched functional categories compared with the simulated data, in both the Meta-GWA (minimum P = 0.016) and UK-GWA studies (minimum P = 0.022) (Table 2), as well as significant evidence in favour of a common set of functional categories that were enriched (P < 0.05) in both GWA studies (P < 0.001) (Table 3). This suggests that our findings are not dependent on either the previously identified susceptibility loci or the genes that are falsely associated with PD merely as a consequence of LD with the very strong association signals. Enrichment P-values in the absence of the 178 genes from the 18 previously associated regions are shown in Supplementary Material, Table S3 for each of the 58 categories significantly enriched in both GWA studies in the original analysis. It can be seen that, in general, the category-specific enrichment P-values are essentially unaltered by removal of the PD regions. To further confirm that the excess in significantly enriched functional categories is not a result of LD with strong signals, we repeated the analysis excluding not only the 18 genomic regions but also the entire HLA region (25–35 Mb on chromosome 6) and any gene within 1 Mb of an SNP that showed genome-wide significance for association (P < 5 × 10^{-8}) in the IPDGC meta-analysis (a total of 1058 genes). This made little difference to the significant excess of enriched functional categories in the Meta-GWA and UK-GWA samples analysed separately (Table 2) or the significance of the overlap in categories enriched in both
studies (Table 3). Enrichment $P$-values under this analysis are given in Supplementary Material, Table S3 for each of the 58 categories significantly enriched in both GWA studies. These are very similar to those obtained in the original analysis.

**Gene-set enrichment analysis**

Category-specific GSEA (gene-set enrichment analysis) $P$-values are given in Table 4 for each of the 26 categories implicating genes involved in the regulation of leucocyte/lymphocyte activity and cytokine-mediated signalling, whereas the results for all 58 categories significantly enriched for associated genes are given in Supplementary Material, Table S2. As some of the less-significant categories identified by ALIGATOR do not have significant GSEA $P$-values, it suggests that some of the enrichment may be the result of the threshold used to define significant genes. Despite this, a closer inspection of the GSEA analysis reveals that 34 (59%) and 16 (28%) of the 58 categories also have significant GSEA $P$-values in the Meta-GWA and UK-GWA samples, respectively, with 14 (24%) being significant in both GWA studies. Moreover, of the 14 categories that are significant in both GWA studies, 10 (70%) are from the two biological groups that we have termed the regulation of leucocyte/lymphocyte activity and cytokine-mediated signalling, increasing confidence that the enrichment of PD-associated genes seen in these categories is genuine.

**DISCUSSION**

In this study, we have tested whether SNPs associated with PD in two large independent GWA studies are enriched in predefined sets of functionally related genes. By restricting our analysis to the functional categories defined in five publicly available ontology databases, it is inevitable that any biological functions that are not well curated by these sources will not have been analysed by our study. Despite this limitation, our analysis of two independent GWA studies has identified significant evidence that a series of functional categories were enriched for genes associated with PD, and that a subset of 58 categories was significantly enriched in both GWA studies. This significant overlap was further condensed to just 24 biological groups, of which the two with the strongest evidence for association highlight processes involving the immune system. Based on their proposed functions, we have termed these two biological groups the regulation of leucocyte/lymphocyte activity and cytokine-mediated signalling. These findings imply that genetic variation in genes related to these biological processes has a role in increasing susceptibility to PD and is, therefore, a potential mechanism that should be the subject of further detailed genetic and functional analyses. Although in this study we were unable to replicate the previous reports of an enrichment of association signal in genes involved in axon guidance (34,35), it is worth noting that the T-cell receptor signalling pathway [Kyoto Encyclopedia of Genes and Genomes (KEGG) 04660] has previously been implicated in PD pathology by pathway analysis of GWA data (35). Clearly, the validity of our analysis is dependent on the quality with which biological categories are annotated and that this is likely to be variable. For this reason, we simultaneously analysed annotations from multiple ontology databases and as a result it is encouraging that our two strongest biological groups which contain genes involved in the regulation of leucocyte/lymphocyte activity and cytokine-mediated signalling are implicated by multiple functional categories independently curated in different ontology databases. We also obtained largely analogous results, using both ALIGATOR and GSEA analyses, implying that the biological groups implicated in this study are unlikely to be biased by our method of analysis. Moreover, as the application of ALIGATOR to GWA studies of different diseases has previously implicated biological groups that were different to those identified in this study (32,33), it suggests that the functional categories that we find enriched for genes associated with PD are not biased in favour of those that have been most comprehensively annotated.

It is important to note that this finding does not imply that all nominally significant genes in an enriched functional category are true susceptibility genes. Biological pathway-based analyses assume that variation in functionally related genes is more likely to be important in disease, so any functional category implicated in our data is more likely to be relevant to PD aetiology since it contains a significant excess of genes carrying SNPs associated with PD. With this in mind, it is worth noting that although we have demonstrated that there was a significant overlap of 58 categories enriched for genes
associated with PD from two independent GWA studies, the signal from the two studies in each functional category does not necessarily reflect the same set of genes. This observation is, however, plausible when one considers that this study focused on susceptibility alleles of small effect which are likely to be carried by affected individuals in different genes in the same functional category.

Our findings are not dependent on the genes present at loci that have been previously reported as being strongly associated with PD and their exclusion has minimal effect on the pathways identified in our analysis. Nevertheless, it might be expected that the functional pathways implicated by our analysis would be supported by a framework of genes identified through Mendelian forms of the illness or as genome-wide significant by GWA. Our data show no strong evidence for such a pattern, with only 5 of the 18 loci implicated by GWA studies harbouring genes (a total of 10 genes) that are included in the biological categories related to the regulation of leucocyte/lymphocyte activity and cytokine-mediated signalling. Moreover, SNCA is the only gene implicated in familial PD that is also included in the two biological groups implicated in our study. This observation is similar to the findings in other complex diseases (30,32,33) and demonstrates that the aggregating power of testing whole pathways can generate useful biological insights from association signals that are below the genome-wide significant by GWA. Our data show no strong evidence for such a pattern, with only 5 of the 18 loci implicated by GWA studies harbouring genes (a total of 10 genes) that are included in the biological categories related to the regulation of leucocyte/lymphocyte activity and cytokine-mediated signalling. Moreover, SNCA is the only gene implicated in familial PD that is also included in the two biological groups implicated in our study.

Table 3. ALIGATOR analysis of categories enriched in both GWA studies

<table>
<thead>
<tr>
<th>Discovery sample</th>
<th>Replication sample</th>
<th>Functional categories enriched at $P &lt; 0.05$ $N$ categories</th>
<th>$P$-value</th>
<th>Functional categories enriched at $P &lt; 0.01$ $N$ categories</th>
<th>$P$-value</th>
<th>Functional categories enriched at $P &lt; 0.001$ $N$ categories</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full data set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-GWA</td>
<td>UK-GWA*</td>
<td>58</td>
<td>$&lt;0.001$</td>
<td>18</td>
<td>0.004</td>
<td>6</td>
<td>0.004</td>
</tr>
<tr>
<td>After excluding known GWA study hits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-GWA</td>
<td>UK-GWA*</td>
<td>54</td>
<td>$&lt;0.001$</td>
<td>13</td>
<td>0.007</td>
<td>4</td>
<td>0.005</td>
</tr>
<tr>
<td>After excluding the MHC region and all genes within 1 Mb of a genome-wide significant SNP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-GWA</td>
<td>UK-GWA*</td>
<td>62</td>
<td>$&lt;0.001$</td>
<td>16</td>
<td>$&lt;0.001$</td>
<td>9</td>
<td>$&lt;0.001$</td>
</tr>
</tbody>
</table>

*UK-GWA was analysed using only the 498 functional categories enriched at $P < 0.05$ in Meta-GWA.

The strongest findings from our study show a striking enrichment of genes associated with PD being involved in biological processes related to the immune system. Specifically, these were identified as two distinct but overlapping biological groups related to the regulation of leucocyte/lymphocyte activity and cytokine-mediated signalling. Despite PD not typically being considered an immune disease, there are emerging data that are starting to demonstrate a relationship between PD progression and the immune response. First, in the degenerating PD brain, the activation of microglia and monocytes contributes to the blood–brain barrier becoming compromised, and this allows toxins and infections to reach the CNS (36). There is also evidence that altered immune components are coupled with PD progression. For example, CD4 (37) and CD3-positive (38) T-cells have been found in the substantia nigra and Lewy body lesions of PD patients, respectively, and several inflammatory cytokines such as IL1-beta, IL8, IL6, IL4, TNF-alpha, IFN-gamma are up-regulated in the CSF and sera of PD patients (39,40). LRRK2 has also been shown to regulate B2-lymphocyte function (41), potentially providing a direct link between cell-mediated immunity and PD pathology. Although these studies are typically interpreted as detecting the secondary effects of neurodegeneration in PD, they have recently been complemented by strong evidence for genetic association between SNPs at the HLA locus and PD (9,11). These genetic findings indicate that the immune system is likely to be aetiologically important in PD and not simply activated as a secondary response to neurodegeneration. It is also possible that future studies of larger GWA data sets will reveal additional biological mechanisms that are associated with PD.

Our findings, therefore, provide independent support to the strong association signal at the HLA locus which spans at least 10 immune-related genes. It is worth noting that five genes from this locus ($HLA$-$DQA1$, $HLA$-$DQB1$, $HLA$-$DRA$, $HLA$-$DRB1$, $HLA$-$DRB3$) were included in the biological groups, implicating the regulation of leucocyte/lymphocyte activity and cytokine-mediated signalling. However, the widespread LD at this locus is unlikely to have adversely influenced our results as we classed the HLA locus as a single signal in our analysis and it only counted once towards any pathway. Moreover, as previously mentioned, we obtained analogous results when previously implicated regions were excluded, demonstrating that our findings are not dependent on markers at the HLA locus. Our findings imply that the immune-related genetic susceptibility to PD is likely to be more widespread in the genome than previously appreciated. By implicating genes involved in the regulation of leucocyte/lymphocyte activity, our study indicates that components of the adaptive immune system have a role in PD pathology. However, given the large overlap in gene content with the cytokine-mediated signalling pathway which is itself related to both the innate and the adaptive immune responses as well as to inflammatory processes, the exact nature of any involvement remains to be elucidated. Future work will, therefore, be required to dissect out the exact involvement of the innate and adaptive immune system in PD. Functional and epidemiological analyses will be required to investigate whether genetic variants in these immune-related pathways influence disease progression, although the quantitative analysis of cytokine levels in peripheral blood could potentially be an avenue.
<table>
<thead>
<tr>
<th>Biological group</th>
<th>Functional category</th>
<th>Meta-GWA N significant genes observed</th>
<th>N significant genes expected</th>
<th>N significant genes ALIGATOR P-value</th>
<th>GSEA P-value</th>
<th>N significant genes observed</th>
<th>N significant genes expected</th>
<th>ALIGATOR P-value</th>
<th>GSEA P-value</th>
<th>Proposed biological function of functional category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GO:51251 190 21</td>
<td>9.16 0.0002</td>
<td>0.0138 18</td>
<td>9.29 0.0028</td>
<td>0.0538</td>
<td>Positive regulation of lymphocyte activation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:2696 208 22</td>
<td>9.89 0.0004</td>
<td>0.0032 19</td>
<td>10.03 0.003</td>
<td>0.0088</td>
<td>Positive regulation of leucocyte activation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:50867 212 22</td>
<td>10.09 0.0004</td>
<td>0.0032 19</td>
<td>10.23 0.0048</td>
<td>0.0102</td>
<td>Positive regulation of cell activation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:51249 251 23</td>
<td>11.57 0.0006</td>
<td>0.0162 23</td>
<td>11.6 0.0004</td>
<td>0.0738</td>
<td>Regulation of lymphocyte activation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:2694 288 26</td>
<td>14.36 0.0014</td>
<td>0.0004 28</td>
<td>14.26 &lt;0.0001</td>
<td>0.002</td>
<td>Regulation of leucocyte activation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:50865 300 27</td>
<td>15.02 0.0016</td>
<td>0.0004 28</td>
<td>14.89 0.0002</td>
<td>0.005</td>
<td>Regulation of cell activation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>MGI:8661 16 3</td>
<td>0.33 0.0028</td>
<td>0.0312 2</td>
<td>0.33 0.0368</td>
<td>0.0166</td>
<td>Decreased interleukin-10 secretion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>MGI:5027 153 15</td>
<td>7.37 0.0042</td>
<td>0.0128 13</td>
<td>7.49 0.0272</td>
<td>0.072</td>
<td>Increased susceptibility to parasitic infection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>MGI:8659 25 4</td>
<td>0.88 0.0009</td>
<td>0.01 3</td>
<td>0.84 0.0446</td>
<td>0.034</td>
<td>Abnormal interleukin-10 secretion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:50863 197 16</td>
<td>9.1 0.0154</td>
<td>0.0102 17</td>
<td>9.18 0.007</td>
<td>0.0336</td>
<td>Regulation of T-cell activation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:50671 84 9</td>
<td>4.08 0.0166</td>
<td>0.2736 8</td>
<td>4.17 0.043</td>
<td>0.2526</td>
<td>Positive regulation of lymphocyte proliferation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:32946 85 9</td>
<td>4.1 0.0172</td>
<td>0.2726 8</td>
<td>4.19 0.045</td>
<td>0.192</td>
<td>Positive regulation of mononuclear cell proliferation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>MGI:2406 256 17</td>
<td>10.3 0.02</td>
<td>0.4188 25</td>
<td>10.3 &lt;0.0001</td>
<td>0.3384</td>
<td>Increased susceptibility to infection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>KEGG:4672 45 1</td>
<td>1.31 0.021</td>
<td>0.006 4</td>
<td>1.31 0.0276</td>
<td>0.0002</td>
<td>Intestinal immune network for IgA production</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:50864 72 7</td>
<td>3.11 0.03</td>
<td>0.5206 8</td>
<td>3.1 0.0084</td>
<td>0.23</td>
<td>Regulation of B-cell activation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:50670 118 10</td>
<td>5.28 0.0332</td>
<td>0.1768 11</td>
<td>5.38 0.0122</td>
<td>0.2292</td>
<td>Regulation of lymphocyte proliferation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:32944 119 10</td>
<td>5.29 0.034</td>
<td>0.1774 11</td>
<td>5.4 0.0122</td>
<td>0.1922</td>
<td>Regulation of mononuclear cell proliferation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:70663 122 10</td>
<td>5.4 0.0388</td>
<td>0.1742 11</td>
<td>5.52 0.0146</td>
<td>0.219</td>
<td>Regulation of leucocyte proliferation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:30888 44 5</td>
<td>2.09 0.0444</td>
<td>0.5298 5</td>
<td>2.11 0.0454</td>
<td>0.214</td>
<td>Regulation of B-cell proliferation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:19221 236 19</td>
<td>9.69 0.0016</td>
<td>0.0012 19</td>
<td>9.69 0.0018</td>
<td>0.03</td>
<td>Cytokine-mediated signalling pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:71345 146 7</td>
<td>11.08 0.005</td>
<td>0.0056 20</td>
<td>11.08 0.004</td>
<td>0.1142</td>
<td>Cellular response to cytokine stimulus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:1637 24 2</td>
<td>0.31 0.0342</td>
<td>0.385 2</td>
<td>0.31 0.0344</td>
<td>0.2646</td>
<td>G-protein-coupled chemotactant receptor activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:4950 24 2</td>
<td>0.31 0.0342</td>
<td>0.385 2</td>
<td>0.31 0.0344</td>
<td>0.2646</td>
<td>Chemokine receptor activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GO:4896 72 7</td>
<td>3.42 0.0388</td>
<td>0.619 7</td>
<td>3.36 0.0442</td>
<td>0.3376</td>
<td>Cytokine receptor activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>KEGG:4630 146 10</td>
<td>5.46 0.039</td>
<td>0.0568 13</td>
<td>5.4 0.0022</td>
<td>0.049</td>
<td>Jak-STAT signalling pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Proposed biological functions were provided by the respective source of each ontology database. Biological groups indicate categories which share >50% of genes with at least one other category in the group.
for the investigation of biomarkers that increase risk to PD. Our findings do, however, contrast with those of a previous study which applied the same analytical strategy to GWA studies in Alzheimer’s disease and specifically highlighted the involvement of the complement and innate immune system in the disease (33). Given the importance of the immune system, it is perhaps unsurprising that it has an impact upon the presentation of both of these common late-onset diseases. However, despite implicating the involvement of immune-related genes in both Alzheimer’s disease and PD, different functional categories are highlighted by each study, suggesting that the increased risk to disease conferred by the immune system is likely to play a specific role in different neurodegenerations, and is unlikely to be a common degenerating effector end-pathway.

MATERIALS AND METHODS

Data summary

The GWA studies used in this study were performed as described (13,15,17–19,42). We have previously studied a large meta-analysis composed of these studies for allelic association (9). However, in order to provide independent discovery and replication samples, in this study we divided the meta-analysis of the IPDGC (9) into two large independent PD-GWA data sets: (i) the UK-based GWA study of Spencer et al. (17) (referred from here on as UK-GWA), which was composed of 1705 PD patients (56.7% male) and 5200 controls (49.5% male) and (ii) a trimmed meta-analysis composed of 3628 PD patients (59.5% male) and 7019 controls (52.7% male) which excluded the UK-GWA study and was, therefore, composed of the remaining four GWA studies from USA National Institute on Aging, Germany, France and the USA database of genotypes and phenotypes (13,15,18,19,42) (referred from here on as Meta-GWA).

Study-specific quality control and genotype imputation using MACH (version 1.0.16) has been previously described for the Meta-GWA study and 1941147 SNPs to 21004 genes in the whole genome excluding all spanned by genomic regions harbouring known PD GWA signals was estimated to that gene: if SNPs mapped within more than one gene, all such genes were included. Using this approach, we were able to map 1945730 SNPs to 21013 genes in the Meta-GWA study and 1941147 SNPs to 21004 genes in the UK-GWA study. The P-value of the SNP showing the strongest evidence for allelic association was then taken as the gene-wise significance measure.

Assignment of genes to functional categories. We next assigned genes to a series of functional categories defined by five independent sources: (i) Gene Ontology (GO) (44) (http://www.geneontology.org/, downloaded 8 November 2011), (ii) KEGG (45) (http://www.genome.jp/kegg/, downloaded 27 June 2011), (iii) PANTHER (http://www.pantherdb.org/pathway/, accessed on 20 August 2010), (iv) the ‘canonical pathways’ collection from the Molecular Signatures Database v3.0 (MsigDB) (http://www.broadinstitute.org/gsea/msigdb/index.jsp, accessed on 2 February 2011), (v) the Mouse Genome Informatics (MGI) database (46) (http://www.informatics.jax.org/, accessed on 22 February 2010). The MGI contains a comprehensive catalogue of behavioural, physiological and anatomical phenotypes observed in mutant mice, and as previously described (33), we extracted phenotype data for single-gene studies (excluding all transgenes) and converted mouse genes to their human orthologues, using the MGI’s mouse/human orthology assignment. This allowed us to map 6297 different phenotypic annotation terms to 5671 human genes. We restricted our analysis to a total of 15381 functional categories containing between 3 and 300 genes: 9164 in GO, 231 in KEGG, 542 in PANTHER, 687 in MsigDB and 4757 in MGI categories. Moreover, to remove the possibility of a small category being deemed significantly enriched based on just one signal, we only classed categories as being enriched if they carried at least two signals.

ALIGATOR analysis. ALIGATOR was then used to test the list of gene-wide significance measures for enrichment within functional categories as previously described in Holmans et al. (32). Briefly, ALIGATOR defines a list of significant genes as those genes containing at least one SNP surpassing a pre-defined significance criterion, and comparing the number of such genes in each category with that observed in 5000 randomly generated gene lists of the same length. This procedure gives a measure of the enrichment significance for
each category. Multiple testing correction for multiple non-independent categories is carried out by a bootstrap procedure, wherein one random gene list is selected as the ‘observed’ data and is compared with a set of 5000 gene lists selected randomly (with replacement) from the remainder, with the process being repeated 1000 times. This procedure also gives a test of whether more categories achieve a given level of enrichment significance in the real data than would be expected by chance. Further details are given in Holmans et al. (32). Unlike methods designed for gene-expression data (where there is typically only one measurement per gene), ALIGATOR corrects for variable numbers of SNPs per gene. This is done by generating the random gene lists by sampling SNPs (not genes) at random, adding the gene(s) the SNP lies in to the list. This ensures that the random gene lists account for large genes containing many SNPs being more likely to contain significant SNPs by chance. Each gene was counted once regardless of how many significant SNPs it contains, thus eliminating the influence of LD between SNPs within genes. To prevent the analysis being biased by SNPs located in multiple functionally related genes which physically overlapped, we set a restriction that limited each SNP from contributing to more than one gene in any single category. Moreover, to address the potential of multiple significant genes that are close together reflecting the same association signal due to LD, we conservatively grouped significant genes that were <1 Mb apart and located in the same functional category into one signal. Replicate gene lists of the same length as the original were generated by randomly sampling SNPs (thus correcting for variable gene size). The lists were used to obtain \( P \)-values for enrichment for each category and to correct these for testing multiple non-independent categories, and also to test whether the number of significantly enriched categories is higher than expected. To assess the potential of any bias caused by LD with strong association signals that had been previously identified in these samples, we also performed ALIGATOR analysis after excluding all genes (\( n = 178 \)) present within the list of genomic regions harbouring known PD GWA signals.

**Gene-set enrichment analysis**

As a further validation of the ALIGATOR results, and to show that the results of our analyses are not driven by the choice of \( P \)-value cut-off for defining significant genes, GSEA was performed using the method described in Wang et al. (31). Rather than defining a list of significant genes, GSEA ranks all genes in order of a gene-wide association statistic and tests whether the genes in a particular gene set have higher rank overall than would be expected by chance. Following Wang et al. (31) in order to allow for varying numbers of SNPs per gene, the gene-wide statistic used was the Simes-corrected single-SNP \( P \)-value (47).

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**REFERENCES**


**APPENDIX**

The International Parkinson’s Disease Genomics Consortium


1Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA; 2UCU Genetics Institute, Gower Place, London WC1E 6BT, UK; 3Department of Molecular Neuroscience and Rita Lila Westen Laboratories, Institute of Neurogenetics, University College London, London, UK; 4Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tuebingen and DZNE, German Center for Neurodegenerative Diseases, Tuebingen, Germany; 5UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK; 6INSERM U563, CPTP, Toulouse, France; 7Section of Medical Genomics, Department of Clinical Genetics, VU University Medical Centre, Amsterdam, The Netherlands; 8INSERM, UMR975 (formerly UMR5769), Paris, France; 9Université Pierre et Marie Curie, Paris, France; 10Department of Neurology, Addenbrooke’s Hospital, University of Cambridge, Hills Road, Cambridge CB2 2QQ, UK; 11Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 12Department of Social Medicine, Bristol University, Bristol, UK; 13Department of Neurology and Alzheimer Center, VU University Medical Center, Amsterdam, The Netherlands; 14Department of Neurology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; 15Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; 16Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA; 17Program in Medical and Population Genetics, Broad Institute, 7 Cambridge Center, Cambridge, MA 02139, USA; 18Department of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 19Department of Medical Genetics, Institute of Human Genetics, University of Tuebingen, Tuebingen, Germany; 20University of Cambridge Clinical Ageing Research Unit, Campus for Ageing and Vitality, Newcastle upon Tyne NE4 5PL, UK; 21Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, National Carolina, USA; 22Neurology M4104, The Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, UK; 23School of Clinical and Experimental Medicine, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; 24Department of Clinical Neurosciences, UCL Institute of Neurology, Royal Free Campus, Rowland Hill Street, London NW3 2PF, UK; 25INSEMER CIC-9503, Hôpital Pitié-Salpêtrière, Paris, France; 26Population Health Section, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK; 27CHU Nantes, CIC0004, Service de Neurologie, Nantes, France; 28INSEMER U897, Victor Segalen University, Bordeaux, France; 29Klinik für Neurologie, Universitätssätklinikum Schleswig-Holstein, Campus Kiel, Christian-Albrechts-Universität Kiel, Kiel, Germany; 30Parkinson’s Disease Research Group, Faculty of Medicine, Imperial College London, Fourth Floor, Burlington Danes Building, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK; 31Service de Neurologie, Hôpital Gabriel Montpied, Clermont-Ferrand, France; 32Department of Genetics and Cytogenetics, AP-HP, Pitié-Salpêtrière Hospital, Paris, France; 33Cambridge Centre for Brain Repair, Forvie Site, Robinson Way, Cambridge CB22PY, UK; 34Institute of Neurology, University College London, London, UK; 35Department of Psychiatry and Department of Neurology, Washington University School of Medicine, St Louis, MO, USA; 36deCODE genetics, Stur- lugata 8, IS-101 Reykjavik, Iceland; 37Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands; 38Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 39AARP, Washington, DC, USA; 40Queen Square Brain Bank for Neurological Disorders, Institute of Neurology, University College London, London, UK; 41Department of Clinical Neurology, West Wing, Level 3, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK; 42Institute of Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany; 43Institute of Human Genetics, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany; 44Unit of Functional Neurosurgery, Sobell Department, Institute of Neurology, Queen Square, London, UK; 45Section on Molecular Neurogenetics, Medical
Genetics Branch, NHGRI, National Institutes of Health, Bethesda, MD, USA; MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff CF14 4XN, UK; Neurogenetics Unit, UCL Institute of Neurology/National Hospital for Neurology and Neurosurgery, Queen Square, London, UK; Service de Neurologie, CHU de Grenoble, Grenoble, France; Department of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; Translational Neurology, Biogen Idec, 14 Cambridge Center, Bio 6, Cambridge, MA, USA; Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; Department of Clinical Neurosciences, University of Cambridge, Addenbrooke’s, Hospital Hills Road, Cambridge CB2 0QQ, UK; Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; Department of Neurology, University of Rochester, Rochester, New York, NY 14620, USA; Department of Pathology, Wilkie Building, Teviot Place, Edinburgh EH8 9AG, UK; Department of Neurology, Landspítali, University Hospital, Reykjavik, Iceland; Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK; Clinical Research Department, The Parkinson’s Institute and Clinical Center, Sunnyvale, CA, USA; Service de Neurologie, Hôpital Haut-Lévêque, Pessac, France; Department of Neurology, Cardiff University, Cardiff, UK; Department of Psychiatry and Medical Research Council/Wellcome Trust Behavioural and Clinical Neurosciences Institute, University of Cambridge, Cambridge, UK.