

A Genomic Pathway Approach to a Complex Disease: Axon Guidance and Parkinson Disease

Timothy G. Lesnick¹, Spiridon Papapetropoulos², Deborah C. Mash², Jarlath Ffrench-Mullen³, Lina Shehadeh⁴, Mariza de Andrade¹, John R. Henley⁵, Walter A. Rocca^{6,7}, J. Eric Ahlskog⁷, Demetrius M. Maraganore^{7*}

1 Division of Biostatistics, Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, Minnesota, United States of America, **2** Department of Neurology, Miller School of Medicine, University of Miami, Miami, Florida, United States of America, **3** Gene Logic, Gaithersburg, Maryland, United States of America, **4** Molecular and Cellular Pharmacology, Miller School of Medicine, University of Miami, Miami, Florida, United States of America, **5** Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine, Rochester, Minnesota, United States of America, **6** Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, Minnesota, United States of America, **7** Department of Neurology, Mayo Clinic College of Medicine, Rochester, Minnesota, United States of America

While major inroads have been made in identifying the genetic causes of rare Mendelian disorders, little progress has been made in the discovery of common gene variations that predispose to complex diseases. The single gene variants that have been shown to associate reproducibly with complex diseases typically have small effect sizes or attributable risks. However, the joint actions of common gene variants within pathways may play a major role in predisposing to complex diseases (the paradigm of complex genetics). The goal of this study was to determine whether polymorphism in a candidate pathway (axon guidance) predisposed to a complex disease (Parkinson disease [PD]). We mined a whole-genome association dataset and identified single nucleotide polymorphisms (SNPs) that were within axon-guidance pathway genes. We then constructed models of axon-guidance pathway SNPs that predicted three outcomes: PD susceptibility (odds ratio = 90.8, $p = 4.64 \times 10^{-38}$), survival free of PD (hazards ratio = 19.0, $p = 5.43 \times 10^{-48}$), and PD age at onset ($R^2 = 0.68$, $p = 1.68 \times 10^{-51}$). By contrast, models constructed from thousands of random selections of genomic SNPs predicted the three PD outcomes poorly. Mining of a second whole-genome association dataset and mining of an expression profiling dataset also supported a role for many axon-guidance pathway genes in PD. These findings could have important implications regarding the pathogenesis of PD. This genomic pathway approach may also offer insights into other complex diseases such as Alzheimer disease, diabetes mellitus, nicotine and alcohol dependence, and several cancers.

Citation: Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, et al. (2007) A genomic pathway approach to a complex disease: Axon guidance and Parkinson disease. *PLoS Genet* 3(6): e98. doi:10.1371/journal.pgen.0030098

Introduction

Complex diseases occur commonly in the population and are a major source of disability and death worldwide. They are thought to arise from multiple predisposing factors, both genetic and nongenetic, and joint effects of those factors are thought to be of key importance [1,2]. Parkinson disease (PD) serves as an example of a complex disease [3,4]. Other examples include Alzheimer disease, diabetes mellitus, nicotine and alcohol dependence, and several types of cancer [5]. While major inroads have been made in identifying the genetic causes of rare Mendelian disorders, little progress has been made in the discovery of common gene variations that predispose to complex diseases [6,7]. The single gene variants that have been shown to associate reproducibly with complex diseases typically have small effect sizes or attributable risks. However, the joint actions of common gene variants within pathways may play a major role in predisposing to complex diseases (the paradigm of complex genetics), and the discovery of susceptibility genes and pathways may have sizeable public health benefits [8,9].

As early as 1997, experimental studies inferred that genetic variability in the axon guidance pathway was a possible factor contributing to the cause of PD [10]. More recently, a high-resolution, whole-genome association study of PD highlighted the semaphorin 5A gene (*SEMA5A*) as containing the single nucleotide polymorphism (SNP) most significantly associated

with PD susceptibility in that study [11]. *SEMA5A* maps to the deletion candidate interval for cri du chat syndrome, which is associated with severe abnormalities in brain development [12]. Semaphorin proteins play an important role in axon guidance and in the development of the mesencephalic dopamine neuron system during embryogenesis [13]. They interact with several other proteins from the axon guidance pathway to provide a complex and dynamic set of cues that either repel (as for the semaphorins) or attract axons toward their synaptic targets [14,15]. Indeed, several axon-guidance pathway proteins (including ephrin and netrin and slit proteins and their ephrin and deleted in colorectal carcinoma and roundabout family receptors) have been shown to also be important for dopamine axonal maintenance, regeneration, and target recognition [16,17]. Several experimental findings from the neurodevelopment literature and

Editor: Suzanne M. Leal, Baylor College of Medicine, United States of America

Received: December 29, 2006; **Accepted:** May 2, 2007; **Published:** June 15, 2007

Copyright: © 2007 Lesnick et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: CI, confidence interval; HR, hazards ratio; KEGG, Kyoto Encyclopedia of Genes and Genomes; OR, odds ratio; PD, Parkinson disease; SNP, single nucleotide polymorphism

* To whom correspondence should be addressed. E-mail: dmaraganore@mayo.edu

Author Summary

Complex diseases are common disorders that are believed to have many causes. Examples include Alzheimer disease, diabetes mellitus, nicotine and alcohol dependence, and several cancers. This study represents a paradigm shift from single gene to pathway studies of complex diseases. We present the example of Parkinson disease (PD) and a complex array of chemical signals that wires the brain during fetal development (the axon guidance pathway). We mined a dataset that studied hundreds of thousands of DNA variations (single nucleotide polymorphisms [SNPs]) in persons with and without PD and identified SNPs that were assigned to axon-guidance pathway genes. We then identified sets of SNPs that were highly predictive of PD susceptibility, survival free of PD, and age at onset of PD. The effect sizes and the statistical significance observed for the pathway were far greater than for any single gene. We validated our findings for the pathway using a second SNP dataset for PD and also a dataset for PD that studied RNA variations. There is prior evidence that the axon guidance pathway might play a role in other brain disorders (e.g., Alzheimer disease, Tourette syndrome, dyslexia, epilepsy, and schizophrenia). A genomic pathway approach may lead to important breakthroughs for many complex diseases.

our preliminary genetic findings led us to hypothesize that while the main effects of a single gene such as *SEMA5A* may be small and of limited significance, the joint effects of multiple axon-guidance pathway genes may predispose to PD.

Results

This study employed a genomic pathway approach to determine whether polymorphism in the axon guidance pathway predisposed to PD. Specifically, we employed bioinformatic methods to mine an available whole-genome association dataset for SNPs that were within brain-expressed, axon-guidance pathway genes. We then employed statistical methods to construct models of axon-guidance pathway SNPs that predicted three outcomes: PD susceptibility, survival free of PD, and age at onset of PD. The primary whole-genome association study dataset employed by this study included 443 PD cases and 443 unaffected sibling controls (Tier 1 sample). Details regarding the demographic and clinical characteristics of these participants have been reported previously [11]. The median age at onset of PD among the cases was 61 years (range 31–94). Details regarding the SNP markers genotyped, including call rates, Hardy-Weinberg equilibrium estimations, and resequencing concordance rates were previously reported [11]. Our bioinformatic methods identified 128 brain-expressed axon-guidance pathway genes, and our SNP dataset included 1,460 SNPs within 117 of those genes, as detailed in Table S1. There were no SNP data for 11 genes.

PD Susceptibility

Of the 1,460 SNPs within brain-expressed genes of the axon-guidance pathway, 183 SNPs (12.5%) were individually associated with susceptibility to PD. Table 1 contains results for the final model produced by running SNPs through the multistage process to predict PD susceptibility. This model used data from 442 matched PD patients/sibling controls (one pair was missing data on one or more SNPs). The odds ratios (ORs) (95% confidence intervals [CIs]) for the groups defined by predicted PD probability of <0.25, 0.25–0.50, 0.50–0.75, and >0.75 were as follows: 1 (reference), 4.58 (2.27–9.23), 15.42

(6.19–38.45), and 90.76 (32.60–252.67) respectively. Since we were interested in the significance of the pathway, rather than individual SNPs, the p value for the overall model was of primary importance. In this case, the model had an overall p value of 4.64×10^{-38} (95% CI $6.94 \times 10^{-28} - 5.39 \times 10^{-40}$). This model significantly predicted whether or not an individual was a case or an unaffected sibling. The predicted probabilities of PD were high (towards 1) for most of the cases, and low (towards 0) for most of the unaffected siblings (Figure 1). Indeed 35% of the cases had predicted probabilities above 0.9, and 34% of unaffected siblings had predicted probabilities below 0.1. However, as shown by Figure 1, the model did not completely distinguish the two groups; some cases had low predicted probabilities, and some controls had high predicted probabilities. The concordance for the model was about 0.70, indicating good but not complete agreement between predicted and observed case/sibling status.

Survival Free of PD

Of the 1,460 SNPs, 175 (12.0%) were individually associated with survival free of PD (hazard function) using Cox proportional hazards models, as detailed in Table S1. Table 2 contains results for the final proportional hazards model produced by running SNPs through the multistage process to predict survival free of PD. This model used data from 400 PD patients (43 patients were missing data on one or more SNPs). In this case, the model had an overall p value of 5.43×10^{-48} (95% CI $3.19 \times 10^{-37} - 2.36 \times 10^{-61}$). By contrast, the model was not significant at predicting survival (age at study) of the matched sibling controls ($p = 0.73$). This last finding suggests that the model predicts survival free of PD (hazard function), but not survival in general, and that the model is specific for PD cases.

Figure 2 shows a Kaplan-Meier plot to describe the results of the model. The groups were formed by calculating a risk score for each PD patient using the equation from the proportional hazards model, then categorizing the score at the 25th (Q1), 50th (Q2), and 75th (Q3) percentiles. The survival curves separated nicely right from the earliest ages of onset. By age 60, only 9% of PD patients in the predicted highest risk group were still free of PD, whereas 89% of PD patients in the predicted lowest risk group were free of PD. By age 70, none of the PD patients in the predicted highest risk group were still free of PD, whereas 66% of PD patients in the predicted lowest risk group were free of PD. The median ages at onset for each group, from lowest risk group to highest risk group, were 72.1, 66.5, 59.7, and 51.7, a difference in survival free of PD of more than 20 years from lowest to highest. The concordance for this model was 0.76. The hazards ratios (HRs) (95% CIs) for the four groups, from lowest to highest risk, were 1 (reference), 2.96 (2.14–4.07), 6.91 (4.87–9.81), and 19.04 (13.11–27.65).

Age at Onset of PD

Of the 1,460 SNPs, 160 (11.0%) were individually associated with age at onset of PD using linear regression models, as detailed in Table S1. Table 3 contains results for the final model produced by running SNPs through the multistage process to predict PD age at onset. This model used data from 395 PD patients (48 patients were missing data on one or more SNPs). In this case, the model had an overall p value of 1.68×10^{-51} (95% CI $3.56 \times 10^{-42} - 1.16 \times 10^{-51}$). By contrast, the set

Table 1. SNPs in Axon-Guidance Pathway Genes Predicting PD Susceptibility

Gene ^a (rs or ss Identification) Final Coding Scheme ^b	Odds Ratio (95% CI) ^c	p Value ^c	Interaction ^a	Odds Ratio (95% CI) ^c	p Value ^c
<i>CDC42</i> (rs12740705), d	0.16 (0.06–0.41)	1.3×10^{-04}	<i>DCC*PAK4</i>	0.02 (0.002–0.19)	6.0×10^{-04}
<i>CHP</i> (rs6492998), d	0.21 (0.09–0.47)	1.6×10^{-04}	<i>EPHA4*FYN</i>	0.16 (0.05–0.53)	2.8×10^{-03}
<i>DCC</i> (rs17468382), d	0.07 (0.01–0.67)	2.1×10^{-02}	<i>EPHA4*PAK7</i>	0.26 (0.11–0.62)	2.7×10^{-03}
<i>EFNA5</i> (ss46558780), d	0.80 (0.29–2.21)	6.7×10^{-01}	<i>EPHB2*EFNA5</i>	5.25 (2.10–13.13)	4.0×10^{-04}
<i>EPHA4</i> (rs13386128), a	16.29 (5.95–44.59)	5.6×10^{-08}	<i>FYN*RRAS2</i>	0.04 (0.003–0.50)	1.3×10^{-02}
<i>EPHB1</i> (rs2030737), d	2.60 (1.23–5.51)	1.2×10^{-02}	<i>FYN*SLIT3</i>	4.52 (1.32–15.47)	1.6×10^{-02}
<i>EPHB2</i> (rs10917325), a	0.11 (0.05–0.28)	1.6×10^{-06}	<i>MRAS*SLIT3</i>	0.18 (0.08–0.41)	5.9×10^{-05}
<i>FYN</i> (rs6910116), r	59.95 (5.01–717.33)	1.2×10^{-03}	<i>PAK7*CHP</i>	2.94 (1.23–7.01)	1.5×10^{-02}
<i>GNAI3</i> (rs6692804), r	0.07 (0.01–0.32)	6.3×10^{-04}	<i>SEMA5A*RAC2</i>	4.95 (1.37–17.91)	1.5×10^{-02}
<i>GSK3B</i> (rs16830689), r	0.09 (0.02–0.44)	2.9×10^{-03}	<i>UNC5C1*DCC</i>	6.89 (1.20–39.51)	3.0×10^{-02}
<i>MRAS</i> (rs4678260), a	0.97 (0.57–1.65)	9.1×10^{-01}			
<i>NTNG1</i> (rs11185076), r	0.02 (0.001–0.32)	5.7×10^{-03}			
<i>PAK4</i> (rs17641276), r	9.55 (3.73–24.43)	2.5×10^{-06}			
<i>PAK7</i> (rs2072952), a	1.70 (0.75–3.90)	2.1×10^{-01}			
<i>PLXNA2</i> (rs6656034), a	2.92 (1.75–4.87)	4.0×10^{-05}			
<i>PLXNC1</i> (rs2068435), a	0.51 (0.25–1.02)	5.5×10^{-02}			
<i>PPP3CA</i> (rs2044041), d	5.72 (2.84–11.51)	1.1×10^{-06}			
<i>RAC2</i> (rs739043), r	0.16 (0.069–0.39)	4.8×10^{-05}			
<i>RRAS2</i> (rs2970332), d	8.78 (2.45–31.44)	8.5×10^{-04}			
<i>SEMA5A</i> (ss46559487), d	0.22 (0.10–0.49)	2.0×10^{-04}			
<i>SLIT3</i> (rs9688032), d	17.46 (4.61–66.07)	2.5×10^{-05}			
<i>UNC5C1</i> (rs11097458), a	2.25 (1.41–3.61)	7.2×10^{-04}			
<i>UNC5C2</i> (rs4444836), d	0.38 (0.17–0.85)	1.8×10^{-02}			

Data from our primary whole-genome association dataset [11].

^aSubscript is used to identify individual SNPs in interactions; asterisk denotes multiplicative interaction.

^ba, log additive; d, Mendelian dominant; r, Mendelian recessive.

^cResults adjusted for age and gender.

doi:10.1371/journal.pgen.0030098.t001

of SNPs was not significant at predicting age at study of the matched sibling controls ($p = 0.34$). This last finding suggests that the model predicts age at onset of PD, not age at the time of the study, and that the model is specific for PD cases.

Figure 3 shows a plot of predicted age at onset versus reported age at onset to summarize the results of the model. The plot showed a nice elliptical pattern, reflecting the model R^2 of 0.683 (95% CI 0.62–0.69). The model explained about 68% of the variability in age at onset of PD.

Figure S1 shows the distributions of the test statistics from the models with randomly selected SNPs and the values of the test statistics from our final models for PD susceptibility (A), survival free of PD (B), and age at onset of PD (C). The test statistics from our final models were in every case much greater than those observed from the other models.

Other combinations of SNPs from the axon guidance pathway also performed quite well in predicting PD susceptibility, survival free of PD, and age at onset of PD. Although the models reported in this manuscript provided good fits to our data, our results do not preclude other combinations of axon-guidance pathway SNPs as significant predictors of PD. The SNPs in the final models that we selected showed no significant linkage disequilibrium in unaffected siblings.

Validation with a Second Whole-Genome Association Dataset

We also mined a second available whole-genome association dataset for PD, to determine whether the genes in each of the predictive genetic models in our primary whole-

genome association dataset were also predictive of the same PD outcomes in the secondary dataset (same genes and outcomes, different SNPs and samples). Details regarding the participants and genotyping procedures for that study have been recently reported [18].

The secondary whole-genome association dataset included 1,195 SNPs in the 22 genes from the model predicting PD susceptibility in our primary dataset. Of those SNPs, 127 (10.6%) were individually associated with susceptibility to PD, as detailed in Table S2. Table S3 contains results for the final model produced by running SNPs through the multistage process to predict PD susceptibility. This model used data from 528 individuals (264 PD patients and 264 unrelated controls; eight individuals were missing data on one or more SNPs). The model had an overall p value of 3.93×10^{-44} . The ORs (95% CIs) for the groups defined by predicted PD probability of <0.25 , 0.25–0.50, 0.50–0.75, and >0.75 were as follows: 1 (reference), 7.86 (3.94–15.71), 16.14 (8.13–32.05), and 121.14 (56.63–259.14), respectively. The predicted probabilities of PD were high (towards 1) for most of the cases and low (towards 0) for most of the unaffected siblings (Figure S3). Indeed 38% of the cases had predicted probabilities above 0.9, and 39% of unaffected siblings had predicted probabilities below 0.1. However, as shown by Figure S3, the model did not completely distinguish the two groups; some cases had low predicted probabilities, and some controls had high predicted probabilities. The concordance for the model was 0.90.

The secondary whole-genome association dataset included 1,411 SNPs in the 26 genes from the model predicting age at onset of PD in our primary dataset. Of those SNPs, 142

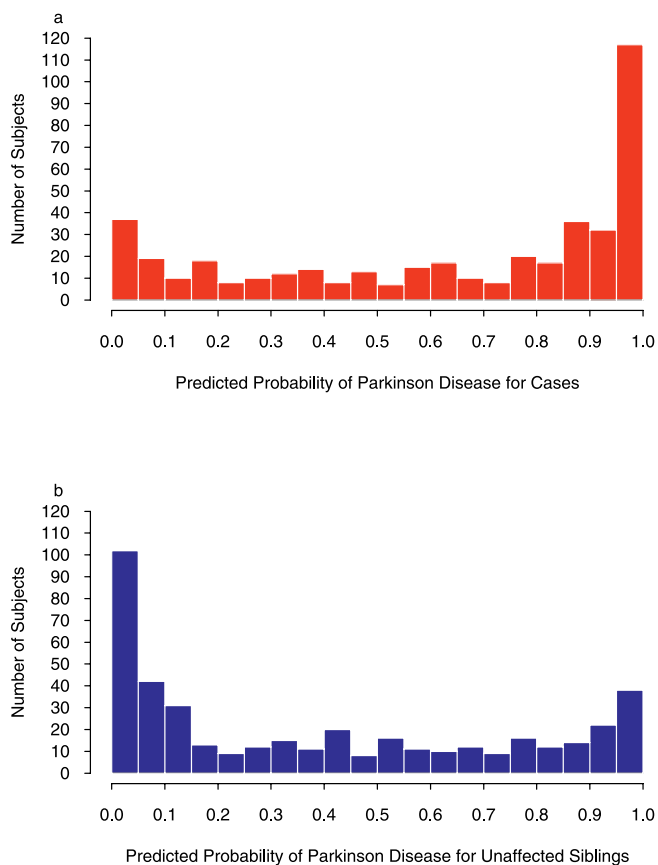


Figure 1. Goodness-of-Fit of Final Model Using Axon Guidance Genes to Predict Susceptibility to PD

Data presented are within our primary whole-genome association dataset [11]. Histogram of predicted probabilities of PD in cases (A) and histogram of predicted probabilities of PD in unaffected siblings (B) are presented. A perfect fit would have all predicted probabilities for cases equal to one and all predicted probabilities for unaffected siblings equal to zero. A model with no explanatory value would have histograms that were indistinguishable from each other. Conditional logistic regression analyses were used to generate the probabilities, and hence the data do not fit a single exponential curve (one end close to zero). doi:10.1371/journal.pgen.0030098.g001

(10.1%) were individually associated with survival free of PD (hazard function) using Cox proportional hazards models, as detailed in Table S2. Table S4 contains results for the final proportional hazards model produced by running SNPs through the multistage process to predict survival free of PD. This model used data from 263 PD patients (five patients were missing data on one or more SNPs). In this case, the model had an overall p value of 6.30×10^{-35} . However, the model was not significant at predicting survival (age at study) of the matched sibling controls ($p = 0.14$).

Figure S4 uses a Kaplan-Meier plot to describe the results of the model. The groups were formed by calculating a risk score for each PD patient using the equation from the proportional hazards model, then categorizing the score at the 25th (Q1), 50th (Q2), and 75th (Q3) percentiles. The survival curves separated nicely right from the earliest ages of onset. By age 60, only 41% of PD patients in the predicted highest risk group were still free of PD, whereas 94% of PD patients in the predicted lowest risk group were free of PD. By age 70, none of the PD patients in the predicted highest risk group were

still free of PD, whereas 68% of PD patients in the predicted lowest risk group were free of PD. The median ages at onset for each group, from lowest risk group to highest risk group, were 74, 68, 63, and 60, a difference in survival free of PD of 14 years from lowest to highest. The concordance for this model was 0.78. The HRs (95% CIs) for the four groups, from lowest to highest risk, were 1 (reference), 4.02 (2.62–6.16), 10.69 (6.72–16.98), and 21.05 (12.85–34.48).

The secondary whole-genome association dataset included 1,605 SNPs in 28 of the 29 genes from the final model predicting age at onset of PD in our primary dataset. Of those SNPs, 157 (9.8%) were individually associated with age at onset of PD using linear regression models, as detailed in Table S2. Table S5 contains results for the final model produced by running SNPs through the multistage process to predict PD age at onset-squared. This model used data from 265 PD patients (three patients were missing data on one or more SNPs). In this case, the model had an overall p value of 4.72×10^{-40} . However, the set of SNPs was not significant at predicting age at study-squared of the matched sibling controls ($p = 0.527$).

Figure S5 shows a plot of predicted age at onset versus reported age at onset to summarize the results of the model. The plot showed a nice elliptical pattern, reflecting the model R^2 of 0.714. The model explained about 71% of the variability in age at onset of PD.

Validation with a Gene-Expression Profiling Dataset

We also mined an available gene-expression profiling dataset for PD that considered 21 different brain regions. Details regarding the participants, biological samples, and microarray experiments have been recently reported [19]. For this study, we limited our data analyses to the substantia nigra and the striatum (putamen and caudate nuclei), since these are the brain regions contributing most significantly to the nigrostriatal dopamine deficiency that is characteristic of PD and since we defined our three PD outcomes according to the corresponding motor phenotype. There was a total of 45 genes represented by the SNPs listed for the three predictive genetic models (Tables 1–3), and the gene expression dataset had informative probe sets for 32 of those genes in the substantia nigra, 34 in the putamen, and 35 in the caudate. Table S6 provides a detailed listing of the multiregional expression data for the 45 genes. In each region, there were more differentially expressed genes observed than expected by chance: substantia nigra, seven observed (22%) versus 1.6 expected (5%); putamen, five observed (15%) versus 1.7 expected (5%); and caudate, five observed (14%) versus 1.8 expected (5%). Overall, 36 genes had data in at least one of the three regions, and 14 (39%) of those were differentially expressed in at least one region. Figure 4 provides a detailed summary of the multiregional expression data for the 45 axon-guidance pathway genes that were predictive of PD outcomes. We observed no significant differences in the number of probe sets for differentially expressed versus normally expressed genes, except for the putamen where two genes with eight fragments each were both differentially expressed (unpublished data). Removing those two genes (outliers) removed the significance (no differences in the median numbers of informative probe sets).

We found very similar results from our sensitivity analyses treating accurate-type and cross-reacting type fragments

Table 2. SNPs in Axon-Guidance Pathway Genes Predicting Survival Free of PD

Gene ^a (rs or ss Identification), Final Coding Scheme ^b	HR (95% CI)	p Value	Interaction ^a	HR (95% CI)	p Value
<i>ABLIM2</i> (rs10008860), r	1.68 (1.30–2.18)	7.1×10^{-05}	<i>DCC1*NTNG1</i>	2.91 (1.30–6.52)	9.3×10^{-03}
<i>DCC1</i> (rs16954702), a	0.78 (0.38–1.57)	4.8×10^{-01}	<i>EPHA8*EPHB11</i>	0.12 (0.02–0.76)	2.5×10^{-02}
<i>DCC2</i> (rs4940245), d	0.46 (0.33–0.64)	4.5×10^{-06}	<i>EPHB11*DPYSL2</i>	2.89 (1.26–6.62)	1.2×10^{-02}
<i>DPYSL2</i> (rs327234), r	0.30 (0.13–0.66)	2.7×10^{-03}	<i>PAK6*NFATC4</i>	1.46 (1.08–1.97)	1.3×10^{-02}
<i>EPHA8</i> (ss46536672), r	1.81 (0.38–8.68)	4.6×10^{-01}	<i>ROBO12*PLXNC12</i>	0.43 (0.25–0.75)	2.9×10^{-03}
<i>EPHB11</i> (rs1554675), r	1.08 (0.71–1.64)	7.1×10^{-01}	<i>ROBO2*DCC2</i>	1.82 (1.18–2.82)	7.2×10^{-03}
<i>EPHB12</i> (rs36183), r	1.65 (1.28–2.13)	1.1×10^{-04}	<i>ROBO2*SEMA3A</i>	0.40 (0.26–0.63)	4.7×10^{-05}
<i>KRAS</i> (rs11047902), d	0.62 (0.48–0.80)	3.0×10^{-04}	<i>SEMA3A*RRAS2</i>	3.39 (1.66–6.90)	7.8×10^{-04}
<i>NFATC2</i> (ss46555247), a	1.47 (1.20–1.79)	1.6×10^{-04}	<i>SLIT32*NTNG1</i>	0.47 (0.28–0.79)	3.9×10^{-03}
<i>NFATC4</i> (rs2295298), a	1.29 (1.04–1.60)	2.3×10^{-02}	<i>SLIT33*SEMA3A</i>	0.57 (0.35–0.91)	1.9×10^{-02}
<i>NTNG1</i> (rs12731259), d	2.41 (1.55–3.75)	9.1×10^{-05}			
<i>PAK3</i> (rs10521534), r	5.89 (2.26–15.31)	2.8×10^{-04}			
<i>PAK6</i> (ss46552269), a	0.46 (0.30–0.70)	2.6×10^{-04}			
<i>PLXNC11</i> (rs2068435), r	3.27 (1.61–6.65)	1.1×10^{-03}			
<i>PLXNC12</i> (rs3803071), a	1.45 (1.22–1.72)	2.9×10^{-05}			
<i>PPP3CA</i> (rs1506801), a	1.58 (1.34–1.85)	3.6×10^{-08}			
<i>ROBO11</i> (rs6769328), d	0.49 (0.35–0.69)	2.8×10^{-05}			
<i>ROBO12</i> (ss46540991), r	1.24 (0.64–2.41)	5.2×10^{-01}			
<i>ROBO2</i> (rs17015294), d	0.63 (0.42–0.93)	2.0×10^{-02}			
<i>ROCK2</i> (rs6721445), d	0.26 (0.12–0.60)	1.4×10^{-03}			
<i>RRAS2</i> (rs1431374), r	0.36 (0.23–0.59)	3.2×10^{-05}			
<i>SEMA3A</i> (ss46545340), d	1.13 (0.80–1.60)	5.0×10^{-01}			
<i>SEMA3D</i> (rs12176601), d	1.32 (1.06–1.64)	1.3×10^{-02}			
<i>SEMA3E</i> (rs215285), r	1.98 (1.45–2.71)	1.8×10^{-05}			
<i>SEMA5A</i> (rs3822787), r	5.28 (2.47–11.3)	1.8×10^{-05}			
<i>SLIT31</i> (rs1432899), d	2.21 (1.60–3.06)	1.7×10^{-06}			
<i>SLIT32</i> (rs7715777), d	1.21 (0.80–1.83)	3.6×10^{-01}			
<i>SLIT33</i> (ss46557969), r	0.54 (0.42–0.71)	5.2×10^{-06}			
<i>SRGAP1</i> (rs1840664), d	2.30 (1.69–3.12)	9.4×10^{-08}			
<i>SRGAP3</i> (rs6762693), a	0.66 (0.54–0.82)	1.2×10^{-04}			
<i>UNC5A</i> (rs539822), d	2.51 (1.23–5.09)	1.1×10^{-02}			
<i>UNC5D</i> (rs13256961), a	1.12 (0.96–1.31)	1.6×10^{-01}			

Data from our primary whole-genome association dataset [11].

^aSubscript is used to identify individual SNPs in interactions; asterisk denotes multiplicative interaction.

^ba, log additive; d, Mendelian dominant; r, Mendelian recessive.

doi:10.1371/journal.pgen.0030098.t002

equally. Again, in each region there were more differentially expressed genes than expected by chance. Restricting to the 45 genes represented in the three predictive genetic models, the results were: substantia nigra, seven observed (22%) versus 1.6 expected (5%); putamen, five observed (15%) versus 1.7 expected (5%); and caudate, seven observed (20%) versus 1.8 expected (5%). Overall, of the 36 genes with data in at least one of the three regions, 14 (39%) were still differentially expressed in at least one region.

Finally, we renormalized and reanalyzed the raw data from the original gene expression profiling study using a different standard software package, to obtain results for all probe sets in all 45 genes and in all three regions. We again observed differential expression of axon-guidance pathway genes in PD, although the findings were more modest. A total of 13 genes were differentially expressed in at least one region (eight in the substantia nigra, one in the putamen, and four in the caudate).

For all analyses of differential gene expression, we were able to code the differential expression unambiguously as increased or reduced. For genes with multiple informative and differentially expressed probe sets, the direction of effect was always the same.

Discussion

Although the genetic model findings from this study have not been replicated using identical SNPs in an independent sample, they do generate some interesting hypotheses and provide evidence for the possible role of the axon guidance pathway in PD. As such, our study represents a paradigm shift from single-gene to pathway analyses of a complex disease. Replication of genetic association findings can be challenging, particularly when the effect sizes are small [18, 20–27]. Although a second whole-genome association study of PD was recently published and has data that could be similarly mined to test the hypothesis that the axon guidance pathway is associated with PD, that dataset included only 16% of the SNPs included in the dataset that we utilized, precluding an exact replication of our models [11,18]. Nevertheless, for the 32,192 SNPs common to the two studies, we identified three SNPs that were associated with PD in all meta-analysis strategies. One of the SNPs was in the axon-guidance pathway gene *UNC5C*, which encodes a netrin receptor [28].

Furthermore, within the secondary whole-genome association dataset [18], we were able to construct predictive models for PD susceptibility, survival free of PD, and

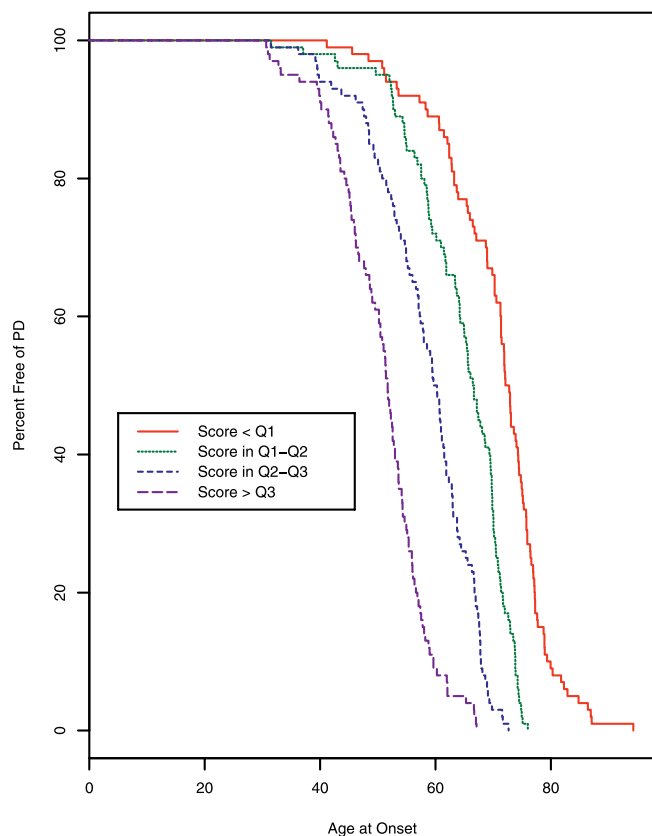


Figure 2. Kaplan-Meier Survival Plot for Age at Onset in PD Patients

Data are grouped by categorized risk score from proportional hazards model within our primary whole-genome association dataset [11]. The model clearly differentiates between early-age-at-onset cases and late-age-at-onset cases. Cases generating the long-dashed line were predicted to be at highest risk for early onset of PD, followed by the medium-dashed, short-dashed, and continuous lines, respectively. Since only cases were included, all of the lines end at 0% free of PD. doi:10.1371/journal.pgen.0030098.g002

predicted age at onset of PD. All three models in the secondary dataset were also highly significant. This is remarkable, as the two datasets employed independent samples and different study designs. Namely, the study of Fung et al. employed fewer participants, unrelated controls, and more and different SNP markers compared to the study of Maraganore et al. [11,18]. This would further attest to the robustness of our predictive models. For validation purposes, we restricted the analyses of the secondary dataset to the genes that were represented by SNPs in each of the three predictive models for the primary dataset. This biases the findings for the secondary dataset conservatively, as it is possible that additional axon-guidance pathway genes contribute to the PD outcomes in that study sample. A more conservative approach would have been to limit the analysis of the secondary dataset to either the same SNPs or to SNPs from the same linkage disequilibrium bins, as for the first study. However, because there was only limited overlap in SNPs between the two studies (16%), and because many of the SNPs employed by the two studies are not included in a common haplotype map, this would not be possible for all of the genes in the models. Nevertheless, future studies of the

two whole-genome association datasets may wish to consider additional meta-analytic strategies [28].

We note that the magnitudes of the genetic effects were sizeable. Within our primary whole-genome association dataset [11], persons with high versus low model scores differed in susceptibility by 90-fold and in age at onset by more than 20 years, and polymorphism in the axon guidance pathway accounted for nearly 70% of the age at onset variance of PD. Similarly large effects were observed for the secondary dataset [18]. The findings of a study are more likely to be true when the effect sizes are large; however, we acknowledge that limited sample size, multiple tested relationships, and flexible analyses may have reduced the positive predictive value of the large effects that we observed [29]. Although the p values observed for our predictive models were small, these were the values determined by the actual data in independent datasets, and bootstrapping resulted in CIs smaller than 1×10^{-27} for each outcome in the primary dataset.

Within our primary whole-genome association dataset [11], models constructed from thousands of random selections of genomic SNPs predicted all three PD outcomes poorly by comparison to models constructed using axon guidance SNPs. Our models of survival free of PD and age at onset of PD did not predict age at study in sibling controls (nor for unrelated controls in the secondary dataset). This would again argue against a chance association of the models with PD outcomes within the samples. We note that the models constructed using axon guidance SNPs were robust and were not crippled by the elimination of SNPs. For example, in our primary dataset, even after removing the first ten SNPs to enter the models (via the stepwise selection process), the overall model p values were smaller than 1×10^{-15} for each of the outcomes (unpublished data). This indicates that the results were not driven by a few SNPs and that the joint effects of multiple SNPs were important.

Our analyses of an available gene-expression profiling dataset revealed that axon-guidance pathway genes were differentially expressed in the brains of PD cases as compared to controls, including several genes that were predictive of PD in the models constructed from the primary whole-genome association dataset (convergence of findings) [11,19]. Not all genes included in the models constructed from the primary whole-genome association dataset were abnormally expressed in the gene-expression profiling dataset, and not all axon-guidance pathway genes that were abnormally expressed in the gene-expression profiling dataset were predictive of PD outcomes in that whole-genome association dataset (unpublished data). This may be because: (1) for some genes, the whole-genome association dataset had no data (no informative SNPs); (2) for some genes, the gene-expression profiling dataset had no data (no informative probe sets); (3) the PD brains included in the gene-expression profiling dataset were from end-stage patients, and it is possible that for some axon-guidance pathway genes differential expression occurs only earlier in life; (4) the genes included in the predictive models may have been associated with PD via a mechanism other than differential expression (i.e., qualitative rather than quantitative effects); (5) there are multiple other possible limitations intrinsic to the whole-genome association and gene expression profiling studies, which may have led to false

Table 3. SNPs in Axon-Guidance Pathway Genes Predicting Age at Onset of PD

Gene ^a (rs or ss Identification), Final Coding Scheme ^b	Regression Coefficient (Standard Error)	p Value	Interaction ^a	Regression Coefficient (Standard Error)	p Value
(Intercept)	2,177.35 (643.06)	8.0×10^{-04}	ABLIM2 ₁ *CXCR4	3,399.00 (993.00)	7.0×10^{-04}
ABL1 (rs17147263), r	-1,845.63 (610.17)	2.7×10^{-03}	ABLIM2 ₁ *NFATC4	466.16 (169.44)	6.3×10^{-03}
ABLIM2 ₁ (rs10008860), r	-919.68 (247.79)	2.4×10^{-04}	CDC42*CXCR4	-2,274.37 (976.73)	2.0×10^{-02}
ABLIM2 ₂ (rs2386113), d	2,385.58 (530.76)	9.7×10^{-06}	DPYSL2*DCC	548.09 (236.29)	2.1×10^{-02}
CDC42 (rs2056974), d	-366.15 (212.79)	8.6×10^{-02}	DPYSL2*PLXNC1	513.80 (203.78)	1.2×10^{-02}
CXCR4 (rs2734871), r	4,797.71 (1,169.33)	5.1×10^{-05}	EFNA5 ₁ *CXCR4	-3,066.66 (750.24)	5.5×10^{-05}
DCC (rs1367637), r	-514.16 (105.45)	1.7×10^{-06}	EFNA5 ₁ *NFATC4	383.76 (182.88)	3.7×10^{-02}
DPYSL2 (rs327234), r	-287.03 (188.67)	1.3×10^{-01}	EPHA4*PPP3CA	-384.13 (105.22)	3.0×10^{-04}
EFNA5 ₁ (rs152562), d	-710.18 (243.20)	3.7×10^{-03}	NFATC2 ₁ *NFATC4	-473.13 (156.98)	2.8×10^{-03}
EFNA5 ₂ (rs3797510), r	502.49 (307.99)	1.0×10^{-01}	NFATC2 ₂ *ROBO1 ₂	739.45 (274.32)	7.4×10^{-03}
EPHA4 (rs2303901), a	-28.75 (93.95)	7.6×10^{-01}	NFATC4*PAK6	-415.00 (117.66)	4.8×10^{-04}
EPHB1 (ss46540509), a	-386.45 (65.46)	8.8×10^{-09}	PAK1*NFATC2 ₁	301.08 (139.99)	3.2×10^{-02}
NFATC2 ₁ (ss46555246), r	737.17 (350.16)	3.6×10^{-02}	PPP3CA*SLIT3 ₂	645.01 (154.72)	3.9×10^{-05}
NFATC2 ₂ (ss46555247), a	90.19 (169.86)	6.0×10^{-01}	PPP3CA*SRGAP1 ₂	-560.10 (206.16)	6.9×10^{-03}
NFATC4 (rs2295298), a	-305.84 (180.30)	9.1×10^{-02}	ROBO1 ₁ *NFATC2 ₁	-790.72 (294.22)	7.6×10^{-03}
NTNG1 (rs4579786), r	1,579.13 (539.98)	3.7×10^{-03}	ROBO1 ₁ *UNC5D	1,536.78 (383.22)	7.5×10^{-05}
PAK1 (rs503406), a	168.89 (79.54)	3.4×10^{-02}	ROBO2*PLXNC1	-588.78 (160.31)	2.8×10^{-04}
PAK3 (rs10521534), r	-956.72 (410.60)	2.0×10^{-02}	ROBO2*PPP3CA	-299.35 (139.63)	3.3×10^{-02}
PAK6 (ss46552269), a	836.89 (166.71)	8.5×10^{-07}	ROBO2*SLIT3 ₂	493.18 (199.90)	1.4×10^{-02}
PAK7 (rs6086976), r	456.37 (247.31)	6.6×10^{-02}	SEMA4D*NFATC2 ₂	-421.48 (121.09)	5.7×10^{-04}
PLXNC1 (rs10507035), a	57.35 (122.66)	6.4×10^{-01}	SEMA5A ₁ *SEMA4D	865.24 (323.95)	7.9×10^{-03}
PPP3CA (rs1506801), a	189.91 (158.71)	2.3×10^{-01}	SEMA6C*EFNA5 ₂	1,497.93 (550.81)	6.9×10^{-03}
ROBO1 ₁ (rs6769328), d	319.29 (169.17)	6.0×10^{-02}	SEMA6C*ROBO2	712.21 (184.45)	1.4×10^{-04}
ROBO1 ₂ (ss46540991), r	118.28 (188.93)	5.3×10^{-01}	SEMA6C*SEMA5A ₁	1,275.38 (429.55)	3.2×10^{-03}
ROBO2 (rs17015294), d	232.75 (182.73)	2.0×10^{-01}	SLIT3 ₁ *UNC5D	-3,038.73 (977.12)	2.0×10^{-03}
SEMA3E (rs215285), r	-416.74 (128.40)	1.3×10^{-03}	SLIT3 ₂ *UNC5D	-575.60 (229.89)	1.3×10^{-02}
SEMA4D (rs9969727), a	297.51 (78.41)	1.8×10^{-01}	SRGAP3*SLIT3 ₂	-446.03 (179.54)	1.3×10^{-02}
SEMA5A ₁ (rs16882671), a	-2,288.01 (575.23)	8.6×10^{-05}	SRGAP3*SLIT3 ₃	-397.10 (165.82)	1.7×10^{-02}
SEMA5A ₂ (rs3822787), r	-1,172.76 (314.54)	2.3×10^{-04}			
SEMA6C (rs4357530), r	-41.57 (141.18)	7.7×10^{-01}			
SLIT2 (rs1355512), r	-482.36 (192.17)	1.3×10^{-02}			
SLIT3 ₁ (rs11738097), r	278.84 (289.97)	3.4×10^{-01}			
SLIT3 ₂ (rs1432899), d	-736.37 (193.67)	1.7×10^{-04}			
SLIT3 ₃ (rs4868530), d	398.81 (112.49)	4.5×10^{-04}			
SRGAP1 ₁ (rs10784373), d	271.13 (93.81)	4.1×10^{-03}			
SRGAP1 ₂ (rs1840664), d	-247.09 (174.00)	1.6×10^{-01}			
SRGAP3 (rs6762693), a	518.98 (124.81)	4.1×10^{-05}			
UNC5D (ss46546978), a	-791.59 (375.74)	3.6×10^{-02}			

Data from our primary whole-genome association dataset [11].

^aSubscript is used to identify individual SNPs in interactions; asterisk denotes multiplicative interaction.

^ba, log additive; d, Mendelian dominant; r, Mendelian recessive.

doi:10.1371/journal.pgen.0030098.t003

positive and false negative results (divergence of findings) [11,19].

Although we observed differential expression of axon-guidance pathway genes in an available gene-expression profiling dataset, we have not as yet demonstrated the biological effects of specific SNPs included in the predictive models in either human gene transcription or translation datasets or in experimental cell or animal models (construct validity). We emphasize that the SNPs that were independent predictors of susceptibility, survival free of PD, or age at onset of PD in either whole-genome association dataset may be only indirect markers of the functional variants within the genes. Fine mapping of these gene loci in diverse populations and experimental studies in transfected cell lines or in transgenic animals may help to identify functional variants within the genes and to elucidate the pathogenic mechanism of the axon-guidance pathway variability that we observed [7]. It would be of interest to genotype a large postmortem cohort

for the SNPs that were predictive of PD outcomes and to directly correlate the ORs of those SNPs to the fold changes for matching RNA fragments (probe sets).

We find it plausible that polymorphism in the axon-guidance pathway genes could lead to aberrant trajectory of the ascending dopaminergic pathway during embryonic brain development and in turn cause congenital deficiency in nigrostriatal dopamine. In other words, some persons could be “wired differently” from birth and could thus be at greater risk to develop PD and at an earlier age. Animal experiments provide support for this pathogenic mechanism. The study of embryonic brain development in knockout mice has suggested a critical role for axon guidance proteins, including semaphorin and slit proteins, in the guidance of the ascending trajectory of dopamine pathways from cell bodies in the substantia nigra to the ipsilateral striatum [13,30,31]. By the 16th day of embryogenesis, mutant mice deficient in semaphorin and slit proteins exhibit a loss of midline

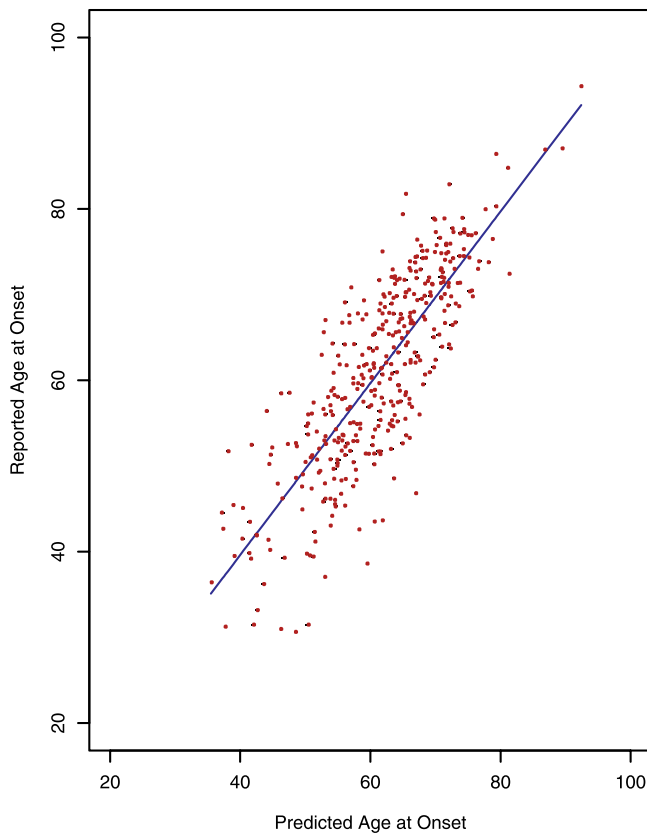


Figure 3. Predicted Versus Reported Age at Onset in PD Patients
Data are within our primary whole-genome association dataset [11]. Points indicate reported individual values, and the line represents perfect agreement. In this case, the pattern shows a fairly tight elliptical pattern, reflecting a good fit and high R^2 of 0.68. The fit appears equally good from the minimum to maximum reported age at onset.
doi:10.1371/journal.pgen.0030098.g003

repulsion of ascending dopamine pathways, so that the fibers ascend aberrantly to the contralateral striatum. This results in a visible reduction in the dopamine innervation of the striatum in semaphorin and slit deficient mice. Striatal dopamine deficiency is the neurochemical hallmark of PD [32].

Similarly, transgenic mice with disrupted ephrin signaling exhibit a 40%–50% quantitative reduction in nigrostriatal innervation and exhibit classical behavioral model characteristics of PD [33,34]. There is also some indirect support of this pathogenic “miswiring” mechanism in humans. Recently, a rare sequence variation in a single axonal-dendritic development gene was associated with Tourette syndrome, another disorder involving dopamine neurons [35]. Furthermore, developmental anatomic asymmetries in nigrostriatal innervation resulting from polymorphism in axon-guidance pathway genes could account for motor asymmetries later seen in patients with PD [36].

Another possibility is that axon-guidance pathway molecules contribute to the pathogenesis of PD via other mechanisms. For example, these proteins also play an important role in axonal repair and in apoptotic signaling. Polymorphism within the pathway may make neurons more or less vulnerable to endogenous or exogenous toxins that trigger the cell death response [37,38]. The continued

Gene	Substantia Nigra	Putamen	Caudate
ABL1			
ABLIM2			
CDC42	Green	Green	
CHP			
CXCR4	Red		Red
DCC			
DPYSL2			
EFNA5			
EPHA4			
EPHA8			
EPHB1			
EPHB2			
FYN			
GNAI3			
GSK3B			
KRAS			Red
MRAS			
NFATC2			
NFATC4			
NTNG1			
PAK1			
PAK3	Green		
PAK4			
PAK6	Green		
PAK7			
PLXNA2			Red
PLXNC1		Green	
PPP3CA		Green	
RAC2			
ROBO1			
ROBO2	Green	Red	
ROCK2			
RRAS2			
SEMA3A			
SEMA3D			
SEMA3E			
SEMA4D	Red		
SEMA5A			
SLIT2	Green		
SLIT3			
SRGAP1			Red
SRGAP3			Green
UNC5A			Green
UNC5C		Green	
UNC5D			

Figure 4. Convergence of Findings for Whole-Genome Association and Gene-Expression Profiling Datasets

Listed are genes that had one or more SNPs included in at least one of the three genetic models that predicted PD susceptibility or survival free of PD or age at onset of PD ($n = 45$ genes). Expression data are summarized for each of the three nigrostriatal regions (substantia nigra, putamen, and caudate).

Gray, no informative probe sets for the gene and region; white, no difference in expression in cases and controls; red, the gene had increased expression in cases compared to controls; green, the gene had reduced expression in cases compared to controls.

doi:10.1371/journal.pgen.0030098.g004

expression of some axon-guidance pathway genes, such as *netrin-1* and *dcc*, in nigral and striatal neurons throughout the mammalian life cycle (including embryogenesis and adulthood) would support multiple possible functions, and it has been suggested that netrin or netrin-like molecules may be of use in understanding and treating PD [10]. Prior knowledge of many biological mechanisms by which the axon guidance pathway may contribute to the pathogenesis of PD increases the probability that our genetic findings may be true (warranting validation studies) [29,39].

Significance and Innovation

In summary, we employed a genomic pathway approach to determine whether polymorphism in a candidate pathway (axon guidance) predisposed to a complex disease (PD). We

found that multiple SNPs in axon-guidance pathway genes were strong predictors of PD susceptibility, survival free of PD, and age at onset of PD in two independent whole-genome association datasets; and many axon-guidance pathway genes were differentially expressed in PD. Although an exact replication of the predictive genetic models remains to be done (same SNPs, independent samples), we consider these findings timely and potentially important. To date, predictive biomarkers for PD are lacking. Our genetic findings for the axon guidance pathway and PD might suggest a new environmental focus on exposures that occur during intra-uterine life (miswiring hypothesis).

Our findings would also provide evidence that complex diseases such as PD can be due to the joint effects of many genes that, taken singly, would show only small effects (additive effects model and epistasis) [40–43]. While the available datasets were exploratory and may have been underpowered to detect the main effects of single SNPs or probe sets, our findings are consistent with recent studies that report greater statistical power to detect the joint effects of multiple loci than the main effects of single loci [44,45]. By contrast to familial aggregation and twin studies, our findings demonstrate that a largely sporadic disease such as PD can in fact have a strong genetic component [4,46,47] and suggest that a similar genomic pathway approach might provide insights into several other complex diseases. Finally, these findings provide important insights regarding the molecular mechanisms that may control dopamine circuit formation and programmed cell death in healthy versus diseased individuals. This in turn may facilitate the development of treatments for axonal synaptic regeneration and repair and for neuroprotection in PD [17,37]. We envision that a systematic study of all defined genomic pathways will yield additional important findings for PD, even in the absence of prior evidence of strong single-gene association.

Materials and Methods

Bioinformatic methods. To formally test our hypothesis, we consulted the Kyoto Encyclopedia of Genes and Genomes (KEGG) [48–50]. The KEGG PATHWAY database is a bioinformatics resource that provides wiring diagrams of molecular interactions, reactions, and relations. There are at least 270 pathways in KEGG related to *Homo sapiens* and diseases. This includes a detailed summary of the axon guidance pathway, updated as recently as October 3, 2005 (http://www.genome.jp/dbget-bin/www_bget?path:hsa04360). We identified all of the genes that encoded proteins within the KEGG axon guidance pathway via Entrez Gene (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Gene>) and consulted the UniGene database (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>) to determine which of the genes were expressed in the human brain ($n = 128$). We then mined an available whole-genome association dataset for PD to identify those SNPs that were genotyped in brain-expressed, axon-guidance pathway genes as part of that study [11]. Specifically, in that study, (which we refer to as the primary whole-genome association dataset), we had individually genotyped 198,345 genomic SNPs that were uniformly spaced (one per 12 kb average gap distance) and informative in 443 sibling pairs discordant for PD. This included 1,460 SNPs within 117 axon-guidance pathway genes expressed in the brain.

Statistical methods. All statistical tests were two tailed and considered significant at the conventional alpha level of 0.05. All statistical analyses were performed in SAS version 9.1 (SAS Institute, <http://www.sas.com>) or S-Plus version 7 (Insightful Corporation, <http://www.insightful.com>). We considered three outcomes of interest: (1) PD susceptibility, (2) survival free of PD, and (3) age at onset of PD. We sought to identify joint action models of SNPs from the axon guidance pathway that predicted each of the three outcomes.

For the first outcome, we used conditional logistic regressions stratified on sibship to examine associations of the SNPs with PD susceptibility while adjusting for age and gender [51]. For each SNP, we calculated ORs, 95% CIs, and p values. Goodness-of-fit was assessed through measuring concordance and visually through histograms of predicted probabilities [52]. We estimated overall ORs by categorizing the predicted probability of PD from the model into four groups (<0.25 , $0.25-0.50$, $0.50-0.75$, and >0.75) and then calculating the ORs for each group relative to the <0.25 group. We used a likelihood ratio test to assess the significance of the overall model and calculated a 95% bias-corrected bootstrap CI for the associated p value using 10,000 resamples.

For the second outcome, we used Cox proportional hazards models to test for associations of the SNPs with survival free of PD [53]. For each SNP, we calculated HRs, 95% CIs, and p values. Concordance was again calculated for the proportional hazards models, and Kaplan-Meier plots of categorized scores predicting risk of PD were generated to provide visual gauges for goodness-of-fit [54]. We also calculated HRs for risk groups categorized at the quartiles, using the lowest risk group as reference. We used a likelihood ratio test to assess the significance of the overall model and calculated a 95% bias-corrected bootstrap CI for the associated p value using 10,000 resamples.

For the third outcome, we predicted the reported age at onset of PD using multiple regression models [55]. Goodness-of-fit was described through the model R^2 values and plots of the predicted versus observed ages at onset. We used an F test to assess the significance of the overall model and calculated a 95% bias-corrected bootstrap CI for the associated p value and the R^2 using 10,000 resamples. Assumptions were tested throughout. Only the linear regression models required a data transformation; age at onset-squared was used as the outcome to meet the required normality assumption. We performed tests of linkage disequilibrium in unaffected siblings for the SNPs in the final models for the three outcomes using LDSELECT version 1.0 (copyright 2004 by Deborah A. Nickerson, Mark Rieder, Chris Carlson, and Qian Yi, University of Washington, United States) with a threshold R^2 of 0.80.

Figure S2 summarizes the scheme used to develop models for each outcome. Since the modes of expression of the alleles in the SNPs of interest were not known, we first looked at each SNP using three coding schemes: log-additive, Mendelian dominant, and Mendelian recessive (Step 1). We simplified subsequent analyses by removing from further consideration those SNPs with no significant main effects in any coding scheme (Step 2). Removing those SNPs was a conservative approach, potentially biasing our tests towards the null hypotheses, since the SNPs were prevented from possibly entering the joint action models after adjustment for other variables. We generally coded the remaining SNPs using the schemes that produced the smallest p values, since these provided our best estimates of the modes of expression in our data (Step 3). For each outcome, we then created multiple sets of SNPs, where each set contained only SNPs with at least a certain number of non-missing values (Step 4). This was done to address issues due to missing values.

While most SNPs had fairly complete data, others had missing values from substantial numbers of participants: up to 18% of cases and up to 32% of case-sib pairs. We therefore chose an approach where we constructed candidate models using sets of SNPs with fairly complete data (effective sample sizes close to the maximum 443) to explain as much of the outcomes as possible, then checked to see if adding other SNPs on top of the candidate models would contribute significantly. We constructed the candidate models for each set using standard automated procedures (Step 5) and selected a final candidate model for each outcome based on significance and goodness-of-fit (Step 6). We then added other SNPs, which were significant given the candidate models (Step 7) and significant pairwise interactions (Step 8).

To compare the significance of our axon-guidance pathway SNP models to the significance of randomly selected genomic SNP models, we constructed 4,000 models for each outcome by randomly selecting the appropriate number of SNPs from the entire available dataset. We then plotted the distributions of the test statistics from those models and the values of the test statistics from our final models.

Validation with a second whole-genome association dataset. A second whole-genome association study of PD was recently published, including first stage analysis and public release of the data [18]. That study included 276 patients with PD and 276 neurologically normal and unrelated controls. The samples used for that study were derived from the National Institute of Neurological Disorders and Stroke Neurogenetics repository hosted by the Coriell Institute for Medical Research (https://lqueue.coriell.org/Q/snp_index.asp). There were

408,803 SNPs individually genotyped, and call rates and Hardy-Weinberg equilibrium p values were previously reported [23]. We downloaded the individual level data for that study from the Coriell Institute website, and we identified SNPs in that secondary dataset that were assigned to the genes represented by SNPs in each of the predictive genetic models for our primary dataset (same genes and outcomes, different SNPs and samples). We then employed the same statistical methods to construct predictive genetic models for the same outcomes in the secondary dataset, with two exceptions. First, because the secondary dataset employed unrelated controls that were not individually matched, we performed unconditional logistic regression analyses instead of conditional logistic regression analyses for the PD susceptibility outcome. Second, we restricted the age at onset of PD analyses in the secondary dataset to exclude participants with ages younger than those reported in the primary dataset, which resulted in the exclusion of one participant whose age at onset was reportedly 13 years.

Validation with a gene-expression profiling dataset. We explored an available gene-expression profiling dataset to determine if there was convergence of those functional data with the genetic association data and models [19]. That study included multi-regional gene expression data from postmortem brain specimens from 22 PD cases and 23 normal aged brain donors and represents the most comprehensive expression profiling study of PD to date (largest numbers of participants, brain regions studied, and genes assayed) [56]. Very strict RNA quality control criteria were used. We analyzed data derived from Affymetrix Human Genome U133 Plus 2.0 GeneChip arrays, which included probe set data for 126 of the 128 brain-expressed axon-guidance pathway genes that we initially identified (see Bioinformatic methods). We analyzed probe set data for the substantia nigra, putamen, and caudate regions using methods and criteria similar to the published study [19]. Supporting information regarding the study design, expression values calculations, and array normalization methods for the published study are also provided in Text S1.

For the differential expression analyses of this study, we identified probe sets from within the original dataset that were assigned to the axon-guidance pathway genes of interest (those represented by SNPs in either of the three predictive genetic models in the primary whole-genome association dataset). For each probe set we compared expression for cases and controls in each of the three nigrostriatal regions (Table S6). We considered a probe set informative in a given region if it was expressed in at least 75% of cases or 75% of controls. Although differential expression of a single probe set is most of the times enough to characterize a gene as differentially expressed (multiple polyadenylation sites are represented on Affymetrix gene chips to account for multiple gene transcripts), for this study we employed a more conservative definition of differential gene expression than for the original study [19]. We defined a gene as differentially expressed in a given region if at least one accurate-type (at) probe set assigned to the gene had a t -test p value <0.05 and absolute value of the fold expression ≥ 1.3 . Alternatively, we defined a gene as differentially expressed in a given region if at least 30% of accurate-type (at) or cross reacting-type (x_at, s_at) probe sets assigned to the gene had t -test p values <0.05 and absolute values of the fold expression ≥ 1.3 . These definitions weigh the significance of cross-reactive type probe sets lower than accurate-type probe sets because they have less specificity, and also account for the possibility that multiple probe sets for the same gene may not provide concordant gene expression measurements [57].

We also performed a sensitivity analysis where a gene was defined as differentially expressed in a given region if at least one probe set of any type assigned to the gene had a t -test p value <0.05 and absolute value of the fold expression ≥ 1.3 . This was consistent with the differential expression analyses performed for the original study [19]. Finally, we performed a second sensitivity analysis whereby we employed the same CEL file data as for the original gene expression profiling study [18], and we renormalized and reanalyzed the data using a second standard software package (GeneSpring, Agilent Technologies, <http://www.home.agilent.com>). That normalization procedure allowed us to obtain informative results for all probe sets and in all 45 genes. Additional details of this sensitivity analysis are provided as supporting information (Text S1).

For our primary differential expression analyses, we tested for possible bias due to the number of informative probe sets per gene by comparing the distributions for differentially expressed versus normally expressed genes using the Wilcoxon rank sum test. For the primary and two sensitivity analyses, we coded the differential expression of genes as increased or reduced if all probe sets assigned

to the gene had the same direction of effect or ambiguous if the probe sets had opposite directions of effect.

Supporting Information

Figure S1. Distributions of the Test Statistics from the Models with Randomly Selected SNPs and the Values of the Test Statistics

Figure S1 shows results from our final models for PD susceptibility (A), survival free of PD (B), and age at onset of PD (C), within our primary whole-genome association dataset [11].

Found at doi:10.1371/journal.pgen.0030098.sg001 (19 KB PDF).

Figure S2. Summary of the Scheme Used to Develop Models for Each Outcome

The procedure employed to build joint action models using SNPs from genes in the axon guidance pathway that predict PD susceptibility, survival free of PD, and age at onset of PD, within both whole-genome association datasets [11,18] is presented.

Found at doi:10.1371/journal.pgen.0030098.sg002 (16 KB PDF).

Figure S3. Goodness-of-Fit of Final Model Using Axon Guidance Genes to Predict Susceptibility to PD

Model is within a second whole-genome association dataset [18]. Histogram of predicted probabilities of PD in cases (A) and histogram of predicted probabilities of PD in unrelated controls (B) are presented. A perfect fit would have all predicted probabilities for cases equal to one, and all predicted probabilities for unrelated controls equal to zero. A model with no explanatory value would have histograms that were indistinguishable from each other.

Found at doi:10.1371/journal.pgen.0030098.sg003 (12 KB PDF).

Figure S4. Kaplan-Meier Survival Plot for Age at Onset in PD Patients

Data are grouped by categorized risk score from proportional hazards model, within a second whole-genome association dataset [18]. The model clearly differentiates between early-age-at-onset cases and late-age-at-onset cases. Cases generating the long-dashed line were predicted to be at highest risk for early onset of PD, followed by the medium-dashed, short-dashed, and continuous lines respectively. Since only cases were included, all of the lines end at 0% free of PD.

Found at doi:10.1371/journal.pgen.0030098.sg004 (10 KB PDF).

Figure S5. Predicted Versus Reported Age at Onset in PD Patients

Data presented are within a second whole-genome association dataset [18]. Points indicate reported individual values, and the line represents perfect agreement. In this case, the pattern shows a fairly tight elliptical pattern, reflecting a good fit and high R^2 of 0.71. The fit appears equally good from the minimum to maximum reported age at onset. By contrast to our primary dataset, which reported age at onset to the month, the second dataset reported age at onset to the year. This and smaller sample size may account for the slightly grainy appearance of the plot for this secondary study as compared to the primary study.

Found at doi:10.1371/journal.pgen.0030098.sg005 (23 KB PDF).

Table S1. Detailed SNP Identifier Information and Additional Results for Three Outcomes

For each of the 1,460 axon-guidance pathway SNPs included in our primary whole-genome association dataset [11], provided are detailed SNP identifier information and additional results for three outcomes: PD susceptibility, survival free of PD, and predicted age at onset of PD. The individual level data for the original whole-genome association study [11] can be downloaded from the NCBI website (<http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000048>, anticipated availability within two months of publication).

Found at doi:10.1371/journal.pgen.0030098.st001 (3.7 MB XLS).

Table S2. Detailed SNP Identifier Information for the 45 Axon-Guidance Pathway Genes

For the secondary whole-genome association dataset [18], provided are detailed SNP identifier information for the 45 axon-guidance pathway genes highlighted by the primary study and additional results for three outcomes: PD susceptibility, survival free of PD, and predicted age at onset of PD.

Found at doi:10.1371/journal.pgen.0030098.st002 (1.6 MB XLS).

Table S3. SNPs in Axon-Guidance Pathway Genes Predicting PD Susceptibility

Data presented are within a second whole-genome association dataset [18].

Found at doi:10.1371/journal.pgen.0030098.st003 (84 KB DOC).

Table S4. SNPs in Axon-Guidance Pathway Genes Predicting Survival Free of PD

Data presented are within a second whole-genome association dataset [18].

Found at doi:10.1371/journal.pgen.0030098.st004 (71 KB DOC).

Table S5. SNPs in Axon-Guidance Pathway Genes Predicting Age at Onset of PD

Data presented are within a second whole-genome association dataset [18].

Found at doi:10.1371/journal.pgen.0030098.st005 (77 KB DOC).

Table S6. Detailed Gene Expression Data for Each of the Three Nigrostriatal Regions Considered

For the 45 axon-guidance pathway genes represented by SNPs in the three predictive genetic models (as defined by our primary whole-genome association dataset [11]), provided are detailed gene expression data for each of the three nigrostriatal regions considered, including probe set type, *t*-test *p* value, fold expression difference in cases and controls, and percent of cases and controls with expression present for the probe set. The CEL file data for the substantia nigra from the original gene expression profiling study [19] can now be downloaded from the Gene Expression Omnibus website (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7621>).

Found at doi:10.1371/journal.pgen.0030098.st006 (47 KB XLS).

Text S1. Study Design, Expression Values Calculation, and Array Normalization

Document provides additional details regarding the methods employed for the gene expression profiling study [19], including study design, expression values calculation, and array normalization; additional details regarding the differential expression analyses for this study are presented.

Found at doi:10.1371/journal.pgen.0030098.sd001 (33 KB DOC).

References

- Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science* 265: 2037–2048.
- Kiberstis P, Roberts L (2002) It's not just the genes. *Science* 296: 685.
- Farrer MJ (2006) Genetics of Parkinson disease: Paradigm shifts and future prospects. *Nat Rev Genet* 7: 306–318.
- McDonnell SK, Schaid DJ, Elbaz A, Strain KJ, Bower JH, et al. (2006) Complex segregation analysis of Parkinson's disease: The Mayo Clinic Family Study. *Ann Neurol* 59: 788–795.
- Merikangas KR, Risch N (2003) Genomic priorities and public health. *Science* 302: 599–601.
- Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273: 1516–1517.
- Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. *Science* 298: 2345–2349.
- Collins FS, Green ED, Guttmacher AE, Guyer MSUS National Human Genome Research Institute (2003) A vision for the future of genomics research. *Nature* 422: 835–847.
- Thomas DC (2005) The need for a systematic approach to complex pathways in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 14: 557–559.
- Livesey FJ, Hunt SP (1997) Netrin and netrin receptor expression in the embryonic mammalian nervous system suggests roles in retinal, striatal, nigral, and cerebellar development. *Mol Cell Neurosci* 8: 417–429.
- Maraganore DM, de Andrade M, Lesnick T, Strain KJ, Farrer MJ, et al. (2005) High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet* 77: 685–693.
- Simmons AD, Püschel AW, McPherson JD, Overhauser J, Lovett M (1998) Molecular cloning and mapping of human semaphorin F from the cri-du-chat candidate interval. *Biochem Biophys Res Commun* 242: 685–691.
- Kawano H, Horie M, Honma S, Kawamura K, Takeuchi K, et al. (2003) Aberrant trajectory of ascending dopaminergic pathway in mice lacking Nkx2.1. *Exp Neurol* 182: 103–112.
- Chilton JK (2006) Molecular mechanisms of axon guidance. *Dev Biol* 292: 13–24.

Text S2. MIAME Guidelines Checklist

Found at doi:10.1371/journal.pgen.0030098.sd002 (76 KB DOC).

Accession Numbers

The National Center for Biotechnology Information Entrez Gene website (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>) accession numbers (GeneIDs) for the genes named in the paper include: *ABLI* (25), *ABLIM2* (84448), *CDC42* (998), *CHP* (11261), *CXCR4* (7852), *DCC* (1630), *DPYSL2* (1808), *EFNA5* (1946), *EPHA4* (2043), *EPHA8* (2046), *EPHB1* (2047), *EPHB2* (2048), *FYN* (2534), *GNAI3* (2773), *GSK3B* (2932), *KRAS* (3845), *MRAS* (22808), *NFATC2* (4773), *NFATC4* (4776), *NTNG1* (22854), *PAK1* (5058), *PAK3* (5063), *PAK4* (10298), *PAK6* (56924), *PAK7* (57144), *PLXNA2* (5362), *PLXNC1* (10154), *PPP3CA* (5530), *RAC2* (5880), *ROBO1* (6091), *ROBO2* (6092), *ROCK2* (9475), *RRAS2* (22800), *SEMA3A* (10371), *SEMA3D* (223117), *SEMA3E* (9723), *SEMA4D* (10507), *SEMA5A* (9037), *SLIT2* (9353), *SLIT3* (6586), *SRGAP1* (57522), *SRGAP3* (9901), *UNC5A* (90249), *UNC5C* (8633), and *UNC5D* (137970).

Acknowledgments

We thank Patricia J. Schultz for her assistance with preparation of the manuscript. We also thank Andrew Singleton for providing us with a list of gene assignments for the SNPs included in the study of Fung et al. [18].

Author contributions. TGL and DMM contributed to the study design; TGL, SP, DCM, JFM, LS, Mda, WAR, and DMM contributed to collection and analysis of data; TGL, SP, DCM, JFM, LS, Mda, JRH, WAR, JEA, and DMM contributed to the interpretation of data; TGL, SP, DCM, JFM, LS, Mda, JRH, WAR, JEA, and DMM contributed to the writing of the paper; and TGL, SP, DCM, JFM, Mda, JRH, WAR, JEA, and DMM contributed to the decision to submit it for publication.

Funding. This work was supported in part by grants from the National Institutes of Health (R01 ES10751 [to DMM] and R01 NS33978 [to WAR]), the Michael J. Fox Foundation (Linked Efforts to Accelerate Parkinson Solutions award to DMM), the National Parkinson Foundation (to DCM), and from Gene Logic, (to DCM).

Competing interests. TGL and DMM report a provisional application for patent under 37 CFR § 1.53 (c) entitled "Predicting Parkinson's Disease." No monies have been awarded to date. JFM is employed by Gene Logic.

- Gomez TM, Zheng JQ (2006) The molecular basis for calcium-dependent axon pathfinding. *Nat Rev Neurosci* 7: 115–125.
- Yue Y, Widmer DA, Halladay AK, Cerretti DP, Wagner GC, et al. (1999) Specification of distinct dopaminergic neural pathways: Roles of the Eph family receptor EphB1 and ligand ephrin-B2. *J Neurosci* 19: 2090–2101.
- Lin L, Rao Y, Isacson O (2005) Netrin-1 and slit-2 regulate and direct neurite growth of ventral midbrain dopaminergic neurons. *Mol Cell Neurosci* 28: 547–555.
- Fung HC, Scholz S, Matarin M, Simon-Sanchez J, Hernandez D, et al. (2006) Genome-wide genotyping in Parkinson's disease and neurologically normal controls: First stage analysis and public release of data. *Lancet Neurol* 5: 911–916.
- Papapetropoulos S, Ffrench-Mullen J, McCorquodale D, Qin Y, Pablo J, et al. (2006) Multiregional gene expression profiling identifies MRPS6 as a possible candidate gene for Parkinson's disease. *Gene Expr* 13: 205–215.
- Bialecka M, Kurzawski M, Klodowska-Duda G, Opala G, Eng-King T, et al. (2006) Polymorphism in semaphorin 5A (Sema5A) gene is not a marker of Parkinson's disease risk. *Neurosci Lett* 399: 121–123.
- Clarimon J, Scholz S, Fung HC, Hardy J, Eerola J, et al. (2006) Conflicting results regarding the semaphorin gene (SEMA5A) and the risk for Parkinson disease. *Am J Hum Genet* 78: 1082–1084.
- Elbaz A, Nelson LM, Payami H, Ioannidis JP, Fiske BK, et al. (2006) Lack of replication of thirteen single-nucleotide polymorphisms implicated in Parkinson's disease: A large-scale international study. *Lancet Neurol* 5: 917–923.
- Farrer MJ, Haugarvoll K, Ross OA, Stone JT, Milkovic NM, et al. (2006) Genomewide association, Parkinson disease, and PARK10. *Am J Hum Genet* 78: 1084–1088.
- Goris A, Williams-Gray CH, Foltynie T, Compston DA, Barker RA, et al. (2006) No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. *Am J Hum Genet* 78: 1088–1090.
- Li Y, Rowland C, Schrodi S, Laird W, Tacey K, et al. (2006) A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan. *Am J Hum Genet* 78: 1090–1092.

26. Maraganore DM, de Andrade M, Lesnick T, Krishna Pant PV, Cox DR, et al. (2006) Response from Maraganore et al. *Am J Hum Genet* 78: 1092–1094.
27. Myers RH (2006) Considerations for genomewide association studies in Parkinson disease. *Am J Hum Genet* 78: 1081–1082.
28. Evangelou E, Maraganore DM, Ioannidis JPA (2007) Meta-analysis in genome-wide association datasets: Strategies and application in Parkinson disease. *PLoS One* 2: e196. doi:10.1371/journal.pone.0000196
29. Ioannidis JP (2005) Why most published research findings are false. *PLoS Med* 2: e124. doi:10.1371/journal.pmed.0020124
30. Bagri A, Marín O, Plump AS, Mak J, Pleasure SJ, et al. (2002) Slit proteins prevent midline crossing and determine the dorsoventral position of major axonal pathways in the mammalian forebrain. *Neuron* 33: 233–248.
31. Marín O, Baker J, Puelles L, Rubenstein JLR (2002) Patterning of the basal telencephalon and hypothalamus is essential for guidance of cortical projections. *Development* 129: 761–773.
32. Hornykiewicz O (2002) Dopamine miracle: From brain homogenate to dopamine replacement. *Mov Disord* 17: 501–508.
33. Halladay AK, Tessarollo L, Zhou R, Wagner GC (2004) Neurochemical and behavioral deficits consequent to expression of a dominant negative EphA5 receptor. *Brain Res Mol Brain Res* 123: 104–111.
34. Sieber BA, Kuzmin A, Canals JM, Danielsson A, Paratcha G, et al. (2004) Disruption of EphA/ephrin-a signaling in the nigrostriatal system reduced dopaminergic innervation and dissociates behavioral responses to amphetamine and cocaine. *Mol Cell Neurosci* 26: 418–428.
35. Abelson JF, Kwan KY, O’Roak BJ, Baek DY, Stillman AA, et al. (2005) Sequence variants in SLITRK are associated with Tourette’s syndrome. *Science* 310: 317–320.
36. Djaldetti R, Ziv I, Melamed E (2006) The mystery of motor asymmetry in Parkinson’s disease. *Lancet Neurol* 5: 796–802.
37. Shirvan A, Shina R, Ziv I, Melamed E, Barzilai A (2000) Induction of neuronal apoptosis by semaphorin3A-derived peptide. *Mol Brain Res* 83: 81–93.
38. Yasahura T, Shingo T, Muraoka K, Kameda M, Agari T, et al. (2005) Toxicity of semaphorin3A for dopaminergic neurons. *Neurosci Lett* 382: 61–65.
39. Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nat Reviews Genet* 5: 251–261.
40. Bateson W (1909) Mendel’s principles of heredity. Cambridge: Cambridge University Press. 396 p.
41. Fisher R (1918) The correlation between relatives under the supposition of Mendelian inheritance. *Trans R Soc Edin* 52: 399–433.
42. Burton PR, Tobin MD, Hopper JL (2005) Genetic Epidemiology 1: Key concepts in genetic epidemiology. *Lancet* 366: 941–951.
43. Cordell HJ, Clayton DG (2005) Genetic epidemiology 3: Genetic association studies. *Lancet* 366: 1121–1131.
44. Marchini J, Donnelly P, Cardon LR (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet* 37: 413–417.
45. Evans DM, Marchini J, Morris AP, Cardon LR (2006) Two-stage two-locus models in genome-wide association. *PLoS Genet* 2: e157. doi:10.1371/journal.pgen.0020157
46. Tanner CM, Ottman R, Goldman SM, Ellenberg J, Chan P, et al. (1999) Parkinson disease in twins: An etiologic study. *JAMA* 281: 341–346.
47. Rocca WA, McDonnell SK, Strain KJ, Bower JH, Ahlskog JE, et al. (2004) Familial aggregation of Parkinson’s disease: The Mayo Clinic Family Study. *Ann Neurol* 56: 495–502.
48. Kanehisa M (1997) A database for post-genome analysis. *Trends Genet* 13: 375–376.
49. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27–30.
50. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, et al. (2006) From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res* 34: D354–D357.
51. Breslow NE, Day NE (1980) Statistical methods in cancer research. Volume I-The analysis of case-control studies. IARC Sci Publ 32: 5–338.
52. Harrell FE, Lee KL, Mark DB (1996) Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 15: 361–387.
53. Cox DR (1972) Regression models of life-tables (with discussion). *J R Stat Soc [Ser B]* 34: 187–220.
54. Kaplan EL, Meier P (1958) Non-parametric estimation from incomplete observations. *J Am Stat Assoc* 53: 457–481.
55. Draper NR, Smith H (1981) Applied regression analysis. 2nd edition. New York: Wiley. 709 p.
56. Papapetropoulos S, McCorquodale D (2007) Gene expression profiling in Parkinson’s disease: Discovery of valid biomarkers, molecular targets and biochemical pathways. *Future Neurol* 2: 29–38.
57. Elbez Y, Farkash-Amar S, Simon I (2006) An analysis of intra array repeats: The good, the bad and the non informative. *BMC Genomics* 7: 136.