



Published in final edited form as:

*Mov Disord.* 2015 May ; 30(6): 813–821. doi:10.1002/mds.26205.

## Low variance RNAs identify Parkinson's disease molecular signature in blood

Maria D. Chikina, PhD<sup>1</sup>, Christophe P. Gerald, PhD<sup>1</sup>, Xianting Li, PhD<sup>1</sup>, Yongchao Ge, PhD<sup>1</sup>, Hanna Pincas, PhD<sup>1</sup>, Venugopalan D. Nair, PhD<sup>1</sup>, Aaron K. Wong, PhD<sup>2</sup>, Arjun Krishnan, PhD<sup>2</sup>, Olga G. Troyanskaya, PhD<sup>2</sup>, Deborah Raymond, MS, CGC<sup>3</sup>, Rachel Saunders-Pullman, MD, MPH<sup>3</sup>, Susan B. Bressman, MD<sup>3</sup>, Zhenyu Yue, PhD<sup>1</sup>, and Stuart C. Sealfon, MD<sup>1,\*</sup>

<sup>1</sup>Department of Neurology and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York, USA

<sup>2</sup>Department of Computer Science, Princeton University, Princeton, New Jersey 08544, and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08540

<sup>3</sup>Department of Neurology, Mount Sinai Beth Israel, New York, New York, USA

### Abstract

**Background**—The diagnosis of Parkinson's disease (PD) is usually not established until advanced neurodegeneration leads to clinically detectable symptoms. Previous blood PD transcriptome studies show low concordance, possibly due to the use of microarray technology, which has high measurement variation. The Leucine-rich repeat kinase 2 (*LRRK2*) G2019S mutation predisposes to PD. Using preclinical and clinical studies, we sought to develop a novel statistically motivated transcriptomic-based approach to identify a molecular signature in the blood of Ashkenazi Jewish PD patients including *LRRK2* mutation carriers.

**Methods**—Using a digital gene expression platform to quantify 175 mRNA markers with low coefficients of variation (CV), we first compared whole blood transcript levels in mouse models 1) over-expressing wild-type (WT) *LRRK2*, 2) overexpressing G2019S *LRRK2*, 3) lacking *LRRK2* (knockout), 4) and in WT controls. We then studied an Ashkenazi Jewish cohort of 34 symptomatic PD patients (both WT *LRRK2* and G2019S *LRRK2*) and 32 asymptomatic controls.

**Results**—The expression profiles distinguished the 4 mouse groups with different genetic background. In patients, we detected significant differences in blood transcript levels both

---

\*Corresponding author: Stuart C. Sealfon, Icahn School of Medicine at Mount Sinai, Annenberg 14-94; Box 1137, One Gustave L. Levy Place, New York, NY 10029, Tel: (212) 241-7075, Fax: (212) 289-4107, stuart.sealfon@mssm.edu.

#### Authors' roles

M.D.C., Statistical Analysis: design and execution; C.P.G., Research Project: conception, organization, marker panel design and execution; X.L., Mouse Sample collection; Y.G., Statistical Analysis: design, execution, review and critique; H.P., Manuscript Preparation: writing; V.N., Research Project: execution; A.K.W., A.K. and O.G.T., Research Project: collaboration on functional gene network analysis; D.R., Clinical Sample organization; R.S.-P., Clinical Sample and Clinical Data collection; S.B.B., Clinical Sample oversight; Z.Y., Mouse Genetics oversight; S.C.S., Research Project: conception, organization and oversight. All authors reviewed and critiqued the manuscript.

**Full financial disclosures:** All other authors have nothing to report.

between individuals differing in *LRRK2* genotype and between PD patients and controls. Discriminatory PD markers included genes associated with innate and adaptive immunity and inflammatory disease. Notably, gene expression patterns in L-DOPA-treated PD patients were significantly closer to those of healthy controls in a dose-dependent manner.

**Conclusions**—We identify whole blood mRNA signatures correlating with *LRRK2* genotype and with PD disease state. This approach may provide insight into pathogenesis and a route to early disease detection.

### Keywords

Parkinson's disease; blood; low coefficient of variation markers; *LRRK2* mutation; functional genomics

---

### Introduction

Parkinson's disease (PD) shows high clinical variability, even among patients with genetic forms of the disease. Because diagnosis mainly relies on the assessment of clinical symptoms, the diagnosis is typically not established early, and misdiagnosis can occur<sup>1</sup>. Mutations in Leucine-rich repeat kinase 2 (*LRRK2*) have been associated with both familial and sporadic PD, with the G2019S substitution being the most common<sup>2</sup>. It is not possible to predict which patients carrying predisposing mutations will become symptomatic. Thus, there is interest in developing biomarker signatures for PD to improve diagnosis, management, and clinical trials of possible disease-altering treatment.

Most potential biomarkers and disease signatures reported thus far in the cerebrospinal fluid (CSF) and blood have either shown inconsistency across studies (for review, see<sup>3</sup>), or exhibited a marked expression overlap between patients and controls<sup>4</sup>. Several PD genome-wide blood expression studies reported differentially expressed genes involved in the ubiquitin-proteasome pathway, mitochondrial function and apoptosis. However, these studies found few discriminating markers in common<sup>5–10</sup>. The disappointing overall discordance of these transcriptome studies may result from clinical heterogeneity of sample cohorts, measurement variability partly due to methodological differences, reduced statistical power caused by multiple hypothesis testing (e.g. in microarray procedures) or uncontrolled confounding factors that obscure biological signals<sup>11</sup>.

Using real-time PCR assays, blood mRNA markers signatures have been reported that reproducibly distinguish psychiatric disorders - such as major depressive disorder (MDD) or borderline personality disorder - from controls with average changes of less than 2-fold<sup>12</sup>. Real-time PCR, however, has limited multiplexing capability. We have developed an approach where a limited number of blood mRNA candidates (<200) selected for their low coefficients of variation (CV) and biological disease relevance, are assayed using a digital gene expression platform<sup>13</sup>. We performed initial proof-of-principle preclinical and clinical studies in *LRRK2*-WT and *LRRK2*-G2019S mouse models<sup>14</sup> and familial (G2019S) and idiopathic PD patients.

## Methods

### Animals and subjects

Mouse studies were approved by the Institutional Animal Care and Use Committee. N=5 male mice were used in each of four groups: wild-type controls (WTC), *LRRK2* null mutant (knockout; KO), transgenic over-expressing either wild-type *LRRK2* (*LRRK2*-WT) or G2019S *LRRK2* (*LRRK2*-GS). Transgenic *LRRK2* models were previously developed using bacterial artificial chromosome (BAC)-mediated transgenesis and characterized<sup>14</sup>. *LRRK2* knockout mice were kindly provided by Dr. Huaibin Cai<sup>15</sup>.

Enrolled subjects were Ashkenazi Jews, who signed an informed consent approved by the Mount Sinai Beth Israel IRB: 34 patients had PD symptoms (17 WT and 17 G2019S *LRRK2*), and 32 healthy controls had no PD symptoms (16 WT and 16 G2019S *LRRK2*). PD patients met the United Kingdom Parkinson's Disease Society Brain clinical diagnostic criteria<sup>16</sup>, except that a positive family history was allowed; they were diagnosed by neurologists specializing in movement disorders and evaluated using the Unified Parkinson Disease Rating Scale (UPDRS). 72% of healthy controls were first-degree relatives of PD patients, to increase genetic homogeneity in the sample cohort. Clinical features are in Table 1; UPDRS motor information was available for all but one PD patient and for 26 controls (12 WT and 14 G2019S *LRRK2*).

### RNA isolation

Mouse blood samples were collected via heart puncture method. Total RNA was immediately isolated using the QIAamp RNA Blood Mini Kit (Qiagen, Valencia, CA) following the manufacturer's recommendations. Human venous blood samples were collected in PAXgene<sup>®</sup> blood RNA tubes (Beckton Dickinson, Franklin Lakes, NJ), and stored at -80°C until RNA isolation. Total RNA was extracted using the PAXgene<sup>®</sup> blood miRNA kit, following the manufacturer's instructions (Qiagen).

### Quantitative digital detection of mRNA

Using 100 ng total RNA, mRNA levels were assayed by direct digital detection following the manufacturer's instructions (NanoString Technologies, Seattle, WA)<sup>17</sup>. The marker panels were assembled using various criteria, including suspected disease relevance, suspected importance in cell regulatory functions or low variance. The mouse panel consisted of 175 markers (Supplementary Table 1), including epigenetic markers, signal transduction effectors, and markers of the immune system. The 113-marker human panel was developed based on the mouse results as well as markers selected from other published PD blood biomarker studies. More information on the selection and variance of each marker in the human panel is found in Supplementary Table 2. The assessment of variance was based on analyzing expression profiles in the healthy control subset of the public whole blood GSE17048 dataset. Low variance was defined as coefficient of variation (CV) lower than 70% in the GSE17048 dataset, with CV calculated as  $(SD * 100 / \text{mean})$ . For CV calculations, raw data (RFUs) were exported and quantile normalized. Our NanoString mouse and human expression data were deposited in Gene Expression Omnibus (GEO; GSE62471).

## Data Processing

To eliminate any possibility of batch effects, all aspects of sample preparation were done in random order, including sample collection, RNA preparation and quantification; human RNA was processed in one batch.

Using NanoString spike control data, a generalized linear model was fitted with 8 negative and 6 positive controls, where the counts for each control followed a Poisson distribution with the mean modeled as a linear function of concentration. Each sample was then normalized to the median slope and intercept of all samples. The median background value, median intercept from the fit of the generalized linear model above, was added to the data before they were log<sub>2</sub> transformed. Using the transformed data of all genes, a Bland-Altman plot<sup>18</sup> (or an MA plot<sup>19</sup> without the log<sub>2</sub> transformation) was done for each sample to compare the sample and the average of all samples. A loess curve was fitted for each Bland-Altman plot to globally normalize the transformed data<sup>20 21</sup>. A marker whose 95 percentile value over all the samples was below the 95 percentile value of the negative spike controls was deemed undetectable and removed. Six markers were removed by this criterion in the human dataset. Further details about data processing and normalization are provided in the Supplementary Methods.

Outliers were detected by reducing to 3 dimensions via PCA and calculating the probability (using a chisq distribution) of the observed sample mahalanobis distances to the center. No outliers were detected in the mouse dataset. One sample in the human dataset was a clear outlier, with the remaining p-values following a uniform distribution, and it was removed from further analysis.

## Statistical Analysis

For the mouse dataset, we used an F-test to detect which genes had altered expression in the four genetic groups. The resulting p-values were transformed to estimate false discovery rates (FDR) using the “qvalue” R function from the qvalue package<sup>22</sup> using default parameters. Human data were analyzed using a multiple linear model that included three clinical variables: PD symptoms, *LRRK2* genotype, and gender. P-values were computed from T statistics for the corresponding coefficients and were converted to q-values as above. For the PD symptomatic subjects with available L-DOPA dosage information, gene expression was fit with an additional model using the dosage as a continuous variable. Further details regarding statistical analysis are provided in the Supplementary Methods.

## Functional Network Analysis

Genes identified experimentally were studied for functional relationships using both Ingenuity Pathway Analysis and GIANT (Genome-scale Integrated Analysis of gene Networks in Tissues). Further details about GIANT are provided in the Supplementary Figure legends.

## Results

### Identification of differentially expressed genes in transgenic mice over-expressing either wild-type LRRK2 or G2019S LRRK2, and LRRK2 null mice

Previous characterization of LRRK2-GS transgenic mice revealed that they had pathological traits relevant to PD, such as decrease in striatal dopamine (DA) content, release, and uptake compared to their WT counterparts<sup>14</sup>. Transcript levels in whole blood were assayed in WTC, KO, LRRK2-WT and LRRK2-GS mice. Twelve differentially expressed markers with q-values < 0.1 were selected for PCA. Among those, DHX58, TGFB1, USP4 were up-regulated, and PLP1 was down-regulated in both LRRK2-WT and LRRK2-GS mice compared to WTC. PCA revealed a clear distinction among the four groups (Fig. 1). Another PCA based on  $p < 0.05$ , uncorrected values demonstrated that five markers best discriminated between LRRK2-GS and LRRK2-WT mice, the two groups most relevant to human studies (Supplementary Fig. S1, S2). All results were explored by principal component analysis (PCA) (Supplementary Fig. S3): the genotype effects did not correlate with major variance components. Notably, several of the differentially expressed transcripts, like PYCARD<sup>23</sup> and USP4<sup>24,25</sup>, are involved in the innate immune response. Other discriminating transcripts included the kallikrein-related peptidases KLK6 and 7, which co-localize with Lewy bodies and are SNCA inhibitors; KLK6 was previously implicated in CNS inflammation and multiple sclerosis (MS)<sup>26</sup>.

### Identification of a PD gene signature in Ashkenazi Jewish patients

Our identification of blood transcriptome signatures distinguishing the mouse *LRRK2* lines motivated us to apply this approach to PD patients. A homogenous genetic population of Ashkenazi Jews was used in this study. We assembled a 113-marker panel from: 22 most significant discriminating markers between G2019S and WT *LRRK2* in our mouse model study, 19 PD markers from the Mutez study<sup>6</sup>, 10 PD markers from the Scherzer study<sup>8</sup>, 7 PD markers from the Kynurenine review (Zinger et al., 2011)<sup>27</sup>, 21 MDD markers from the work of Antonijevic et al. (Antonijevic et al., 2010)<sup>12</sup>, 10 markers from purine/pyrimidine pathways, and other markers from PD-, MS-, and oncology-related literature.

In order to have adequate sample sizes for analysis, expression patterns were compared for clinical status (PD or asymptomatic) independently of *LRRK2* status. Fourteen markers discriminated between PD patients and asymptomatic controls (Fig. 2A) with an accuracy of 79% ( $p < 10^{-4}$ ). Similar to the mouse study results, many genes that were differentially expressed were linked to immunity, both innate (*MyD88*, *DHX58* [alias *LGP2*], *PYCARD*, *IL1B*)<sup>28-30</sup> and adaptive (*TGFB1*)<sup>31</sup>, and to inflammation (*KDM6B*)<sup>32</sup>. Hence, the human data support the implication of innate and adaptive immune response pathways in PD pathogenesis. Surprisingly, only 4 out of the 14 discriminating PD disease markers were from previous human PD studies in blood<sup>6,8</sup>, while the majority were either markers from our mouse model study, MDD markers, or other markers linked to related diseases.

We sought to determine whether drug treatment had any confounding effect on the clinical PD gene signature, as most of PD patients were treated. Namely, the symptomatic PD group differs from the asymptomatic group both in having PD symptoms and in receiving PD

medications, predominantly L-DOPA. Therefore, the putative PD signature might be due to PD, the effects of treatment on blood gene expression, or some combination of the two. The effects of treatment could not be evaluated directly because nearly all patients in the symptomatic group were taking L-DOPA. To evaluate drug effects, we fit a linear model using L-DOPA dose as the independent variable within the subset of PD patients for which dose information was available (25 patients). The regression analysis for one transcript, MyD88, is shown in Fig. 2B. A strong inverse correlation was observed between L-DOPA dosage and MyD88 mRNA levels. Overall, analysis of all genes showed a highly significant inverse correlation between L-DOPA dosage and the PD symptomatic gene signature (Fig. 2C). While the PD gene signature showed the strongest inverse correlation to L-DOPA dose, it was also inversely correlated with disease duration (data not shown). In contrast, there was no significant correlation between the expression profiles and either age or UPDRS scale ratings. Higher levels of L-DOPA tend to reverse the signature to be closer to that observed in asymptomatic individuals, supporting the conclusion that the PD symptomatic gene signature identified indeed marks the presence of the disease.

We also examined whether individuals with different *LRRK2* genotypes (WT homozygous GG vs. G2019S heterozygous GA) could be segregated based on their gene expression profiles. In this analysis, each of the two genotypes studied included both symptomatic and asymptomatic individuals. When all markers were studied by PCA, the genotype and phenotype effects did not correlate with major variance components (Supplementary Fig. S4). While the effect of genotype on blood transcript expression was smaller than that of phenotype, the overall difference between the two genotypes was highly significant ( $p < .01$ , Fig. 3). We identified eleven discriminating markers (at  $q < 0.2$ ), including genes implicated in innate immunity (*DHX58*), neuronal degeneration (*DICER1*, *GPR18*) and lymphocyte infiltration in PD models (*CD8A/B*; see Discussion). Overall, the blood transcriptional signatures discriminated the two genotypes with an accuracy of 68%.

### Functional classification of the human gene signature

We performed a Core Analysis using Ingenuity Pathway Analysis (IPA) on the human PD signature to determine which diseases and biological functions were over-represented. Six out of the 14 genes that were found to discriminate PD were linked to inflammatory disease (Fig. 4;  $p$ -values  $2.12 \times 10^{-3}$ – $4.88 \times 10^{-2}$ ). Twelve out of the 14 were associated with inflammatory response, 9 with immunological disease, and 7 with neurological disease (*IL1B*, *TGFB1*, *MyD88*, *PRKRA*, *PYCARD*, *SGK1*, and *TSPO*); Fig. 4;  $p$ -values  $2.12 \times 10^{-3}$ – $4.88 \times 10^{-2}$ ). With regards to biological functions, 9 genes were involved in cellular movement, while all 14 were associated with cell death and survival ( $p$ -values  $1.25 \times 10^{-3}$ – $4.88 \times 10^{-2}$ ; data not shown). The FDR for all of the identified diseases and biological functions was 0.255. Overall, these results support the biological significance of the PD signature.

We recently found that reanalysis of PD microarray using a new algorithm that reduces confounds caused significant concordance of the markers identified in the Scherzer et al. PD blood biomarker study [8] and several studies of midbrain PD markers [11]. We were therefore interested in exploring genes related to the blood signature in a substantia nigra



(SN) gene network. For this functional analysis we used GIANT, which provides data-driven Bayesian tissue-specific predictions of gene networks to investigate the 14 PD markers based on a SN network (Supplementary Fig. S5). Using this approach, a set of 20 novel candidate genes were detected. Candidates with the best “edge score values” in the SN network are listed in Supplementary Fig. S6. Among them, PPP2CA encodes the catalytic subunit of phosphatase 2A, whose altered activity was associated with  $\alpha$ -synuclein aggregation in neurodegeneration<sup>33</sup>; impairment of ADAM10 trafficking and its  $\alpha$ -secretase activity have been implicated in AD pathogenesis<sup>34,35</sup>; TRIM22 is involved in innate immunity<sup>36</sup>. The gene set was most significantly enriched for the following GO processes: response to cytokine, cellular response to cytokine stimulus, positive regulation of cytokine production, and positive regulation of immune system process (see Supplementary Fig. S7).

## Discussion

The present study supports the utility of whole blood low variance mRNA signatures in detecting PD, as our approach applies successfully to preclinical models, symptomatic PD, and to the presence of a PD-predisposing *LRRK2* mutation. Markers with low inter-individual biological variation are likely to be more informative than those with high variation, thus increasing the power of statistical tests and resulting in a lower frequency of clinical false negatives<sup>37,38</sup>. Although the majority of markers in our human blood-based PD signature were not found in previous PD studies, it comprises several genes linked to innate and adaptive immunity and to CNS function. The PD signature was found to be inversely correlated with L-DOPA dose and disease duration, which both correlate with each other. While the driver of this inverse correlation is uncertain, these results support the conclusion that the disease signature identified cannot be a result of L-DOPA altering blood transcription. Furthermore, the inverse correlation with disease duration suggests that this PD gene signature may be particularly useful for early disease detection.

Out of 14 markers included in both the mouse and human panels, 5 blood markers (DHX58, IL1B, KDM6B, PYCARD and TGFB1) show significant changes in both species and are novel. Interestingly, they are all involved in immune or inflammatory pathways. The mouse and human gene groups cannot be directly compared because the specific experimental groups in the mouse study and human study do not correspond: the mouse groups included knockout (KO) and *LRRK2* mutant and wild-type BAC transgenic over-expression models, whereas the human groups were single *LRRK2* wild-type and mutant alleles, and symptomatic PD versus asymptomatic individuals. Nonetheless, the identification and type of novel informative human markers discovered from analysis of mouse *LRRK2* model blood samples suggest the presence of immune-related commonalities in the regulation of blood signatures in PD-related processes in both species. Furthermore, the link between the PD markers and immune/inflammatory pathways is consistent with the association of immunity with PD pathogenesis supported by GWAS studies<sup>39,40</sup>, albeit none of the markers identified in our study were previously found in GWAS analysis. Our results do not inform the important question of whether the informative markers we identify are merely associated with disease or whether they reflect processes that contribute to disease pathogenesis.

The PD transcriptional signatures in blood may provide leads for pathogenesis or therapeutics. Genome-wide gene expression in lymphocytes correlates with gene expression in brain<sup>41</sup>, suggesting that brain and blood may be subject to similar regulatory mechanisms. These could involve immunological inflammatory responses that are disease-related in the CNS and detectable in the circulating immune system. Hence, expression of the toll-like receptor (TLR) signaling adaptor MyD88 was significantly elevated in the blood of our PD patients. Additionally, a link has been reported between *LRRK2* and the MyD88-dependent pathway, as *LRRK2* phosphorylation was induced by MyD88 agonists and prevented by MyD88 knockout<sup>42</sup>. Interestingly, in view of recent studies framing PD as an autoimmune disorder, MyD88 has been implicated in autoimmunity<sup>43,44</sup>. Neuronal MHC-1 expression and antigen display make catecholamine neurons targets for T cell-mediated cell death<sup>45</sup>. The hypothesis that PD is, in part, an autoimmune disease<sup>46</sup>, is further supported by the 33% increased risk of PD in patients with autoimmune disorders<sup>47</sup>. The PD signature also comprised *PYCARD* and *DHX58*, a major intracellular virus detector<sup>28</sup>. *PYCARD* and *DHX58* mRNAs are up-regulated in the SN of MPTP-treated mice (a toxin-based model of PD)<sup>48</sup>, and in post-mortem brains from AD patients<sup>49</sup>, respectively. Several pro-inflammatory cytokines and chemokines are increased in post-mortem brains of PD<sup>50,51</sup>. Studies in the MPTP model indicate that the TLR4 pathway and brain invasion by CD4<sup>+</sup> lymphocytes<sup>52</sup> may be involved in PD pathogenesis.<sup>53</sup>

The *LRRK2* genotype signature included *DHX58*, *DICER1*, *GPR18*, and *CD8A/B*. In the brain, specific deletion of *DICER1* results in progressive neuronal loss (for review, see<sup>54</sup>). Directed microglial cell migration in the CNS is mediated by *GPR18*, a Gi/o-coupled GPCR; dysregulated migration leads to excessive pro-inflammatory responses, which are involved in neurodegenerative diseases<sup>55</sup>. Neurodegeneration was previously associated with CD8<sup>+</sup> and CD4<sup>+</sup> lymphocytes accumulation in the brains of PD patients and MPTP-treated mice<sup>52</sup>. Hence, regulation of neuronal cell death and neuroinflammation may be altered in mutation carriers.

Our results suggest a link between PD and impairment of the microRNA machinery, which comprises *Dicer*, *PRKRA*, and transactivation response RNA-binding protein (*TRBP*)<sup>56</sup>. *PRKRA* mRNA is down-regulated in symptomatic PD patients. *DICER1* and *TARBP2* (alias *TRBP2*) are down- and up-regulated, respectively, in *LRRK2* mutation carriers. Accumulating evidence, including *Dicer* ablation in mice models and decreased expression of miR-133b and miR-34b/c in PD brain samples, suggests that microRNA deficiency contributes to neurodegeneration (for review, see<sup>57</sup>). Additionally, the *LRRK2* G2019S mutant antagonizes specific microRNAs via its increased kinase activity, presumably through increased phosphorylation of a substrate that binds to a RISC component<sup>58</sup>. Direct study of miRNA expression patterns in blood would be worthwhile in conjunction with subsequent expression profiling.

PCA demonstrated a segregation of transgenic mice over-expressing *LRRK2* (*LRRK2*-WT and *LRRK2*-GS) vs. non-over-expressing mice (WTC and *LRRK2* KO). Consistent with these observations, the G2019S mutation resulted in a gain of kinase function that alters gene expression patterns in transgenic mice compared to *LRRK2* KO mice<sup>59</sup>, and a recent study revealed that over-expression of WT *LRRK2*, and to a greater extent of the G2019S



mutant form, have an inhibitory effect on chaperone-mediated autophagy (CMA), which presumably hinders SNCA degradation by CMA; this could lead to toxic accumulation of the oligomerized protein in PD<sup>60</sup>.

In order to explore the relationship of the 14 symptomatic PD-associated mRNAs, these genes were used to seed a SN gene expression-derived Bayesian functional network. To evaluate the specificity of the SN-derived network, we compared the top ten related candidates obtained using the same genes as seeds for a muscle gene expression-derived network. There was no overlap between the two lists obtained (see Supplementary Fig. S6). Interestingly, the functional related genes identified in the SN-network are most strongly associated with immunological processes.

Although the vast majority of the 113 markers selected for the human panel had low CV, some markers having unknown or higher variance were included in the panel because of potential PD relevance (see Supplementary Table 2). As markers were selected both by variance and by brain disease relevance, we emphasize that our marker selection for analysis was not an unbiased survey of low variance mRNAs and included markers previously associated with PD. Nonetheless, the variances of the markers and of the informative markers are striking. Notably, among the 24 markers found to be informative for either LRRK2 genotype or PD phenotype, 96% showed low CV in public data and 100% showed low CV in our analysis. Indeed, the average CV of these informative markers was 31% in the public microarray dataset, supporting the value of low CV in selecting informative markers for screening.

Overall, our results have promise for the development of a PD signature that will be useful for diagnosis and clinical trials. Moreover, they strengthen the concept that blood-based transcriptomics may be used as diagnostic and patient management tools in CNS disorders. However, we emphasize the limitations of this initial study, in particular the lack of a replication sample. Further investigation is required to confirm and extend our results. In order to obtain sample sizes that were sufficient for analyses, symptomatic and asymptomatic PD patients were analyzed independently of genetic status on the one hand, and genotype was pooled independently of symptom status on the other hand. While our results support the presence of genotype-specific and PD-specific transcriptional signatures, study of groups large enough to allow independent analysis of genotype-phenotype groups is warranted. Another interesting question that could be addressed by longitudinal study is whether the PD signature precedes the development of clinical symptoms in *LRRK2* carriers.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by a grant from the Michael J. Fox Foundation (S.C.S, Z.Y., S.B.B.) and by NIH grants R01NS060809 and R01NS072359 (Z.Y.).

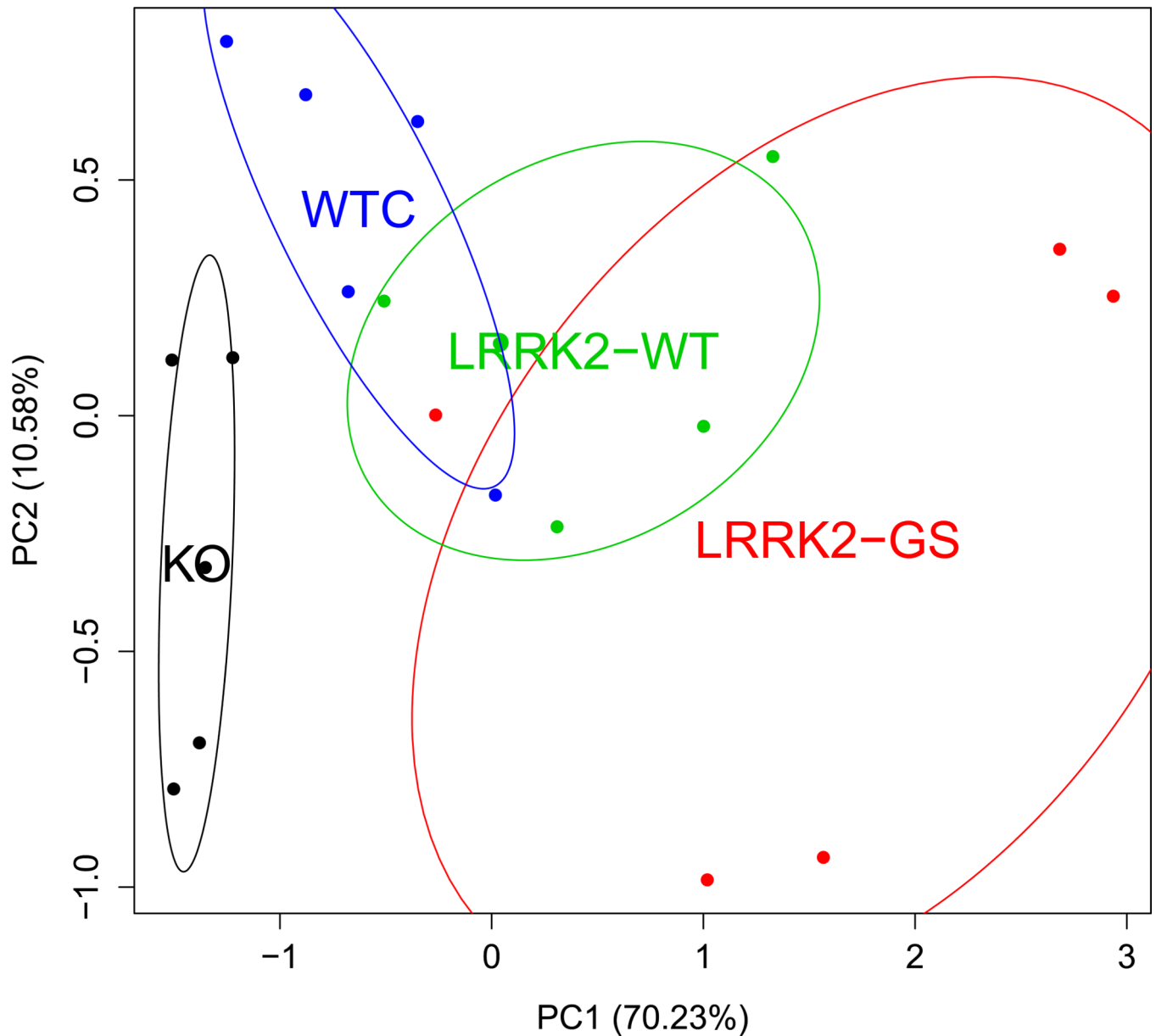
C.P.G. is the founder and owner of Transcription Diagnostics, Inc. S.B.B. has consulted for Bristol-Myers Squibb.

## References

1. Meara J, Bhowmick BK, Hobson P. Accuracy of diagnosis in patients with presumed Parkinson's disease. *Age and ageing*. 1999; 28(2):99–102. [PubMed: 10350403]
2. Corti O, Lesage S, Brice A. What genetics tells us about the causes and mechanisms of Parkinson's disease. *Physiological reviews*. 2011; 91(4):1161–1218. [PubMed: 22013209]
3. Haas BR, Stewart TH, Zhang J. Premotor biomarkers for Parkinson's disease - a promising direction of research. *Translational neurodegeneration*. 2012; 1(1):11. [PubMed: 23211054]
4. Kang JH, Irwin DJ, Chen-Plotkin AS, et al. Association of cerebrospinal fluid beta-amyloid 1–42, T-tau, P-tau181, and alpha-synuclein levels with clinical features of drug-naive patients with early Parkinson disease. *JAMA neurology*. 2013; 70(10):1277–1287. [PubMed: 23979011]
5. Molochnikov L, Rabey JM, Dobronevsky E, et al. A molecular signature in blood identifies early Parkinson's disease. *Molecular neurodegeneration*. 2012; 7:26. [PubMed: 22651796]
6. Mutez E, Larvor L, Lepretre F, et al. Transcriptional profile of Parkinson blood mononuclear cells with LRRK2 mutation. *Neurobiology of aging*. 2011; 32(10):1839–1848. [PubMed: 20096956]
7. Mutez E, Nkiliza A, Belarbi K, et al. Involvement of the immune system, endocytosis and EIF2 signaling in both genetically determined and sporadic forms of Parkinson's disease. *Neurobiology of disease*. 2014; 63:165–170. [PubMed: 24269915]
8. Scherzer CR, Eklund AC, Morse LJ, et al. Molecular markers of early Parkinson's disease based on gene expression in blood. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104(3):955–960. [PubMed: 17215369]
9. Shadrina MI, Filatova EV, Karabanov AV, et al. Expression analysis of suppression of tumorigenicity 13 gene in patients with Parkinson's disease. *Neuroscience letters*. 2010; 473(3):257–259. [PubMed: 20193743]
10. Shehadeh LA, Yu K, Wang L, et al. SRRM2, a potential blood biomarker revealing high alternative splicing in Parkinson's disease. *PloS one*. 2010; 5(2):e9104. [PubMed: 20161708]
11. Chikina MD, Sealfon SC. Increasing consistency of disease biomarker prediction across datasets. *PloS one*. 2014; 9(4):e91272. [PubMed: 24740471]
12. Antonijevic, I.; Tamm, J.; Artymyshyn, R., et al. System and methods for measuring biomarker profiles. PCT Int. Patent Application WO. 2010/025216. 2010.
13. Fortina P, Surrey S. Digital mRNA profiling. *Nature biotechnology*. 2008; 26(3):293–294.
14. Li X, Patel JC, Wang J, et al. Enhanced striatal dopamine transmission and motor performance with LRRK2 overexpression in mice is eliminated by familial Parkinson's disease mutation G2019S. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2010; 30(5):1788–1797. [PubMed: 20130188]
15. Lin X, Parisiadou L, Gu XL, et al. Leucine-rich repeat kinase 2 regulates the progression of neuropathology induced by Parkinson's-disease-related mutant alpha-synuclein. *Neuron*. 2009; 64(6):807–827. [PubMed: 20064389]
16. Hughes AJ, Daniel SE, Kilford L, et al. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of neurology, neurosurgery, and psychiatry*. 1992; 55(3):181–184.
17. Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature biotechnology*. 2008; 26(3):317–325.
18. Bland JM, Altman DG. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *Lancet*. 1986; 1(8476):307–310. [PubMed: 2868172]
19. Dudoit S, Yang YH, Callow MJ, et al. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat Sinica*. 2002; 12(1):111–139.
20. Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. *J Am Stat Assoc*. 1979; 74(368):829–836.
21. Cleveland WS. Lowess - a Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *Am Stat*. 1981; 35(1) 54-54.
22. Dabney A, Storey JD, qvalue Tutorial PRS, et al. Q-value estimation for false discovery rate control. *Medicine*. 2004; 344:539–548.

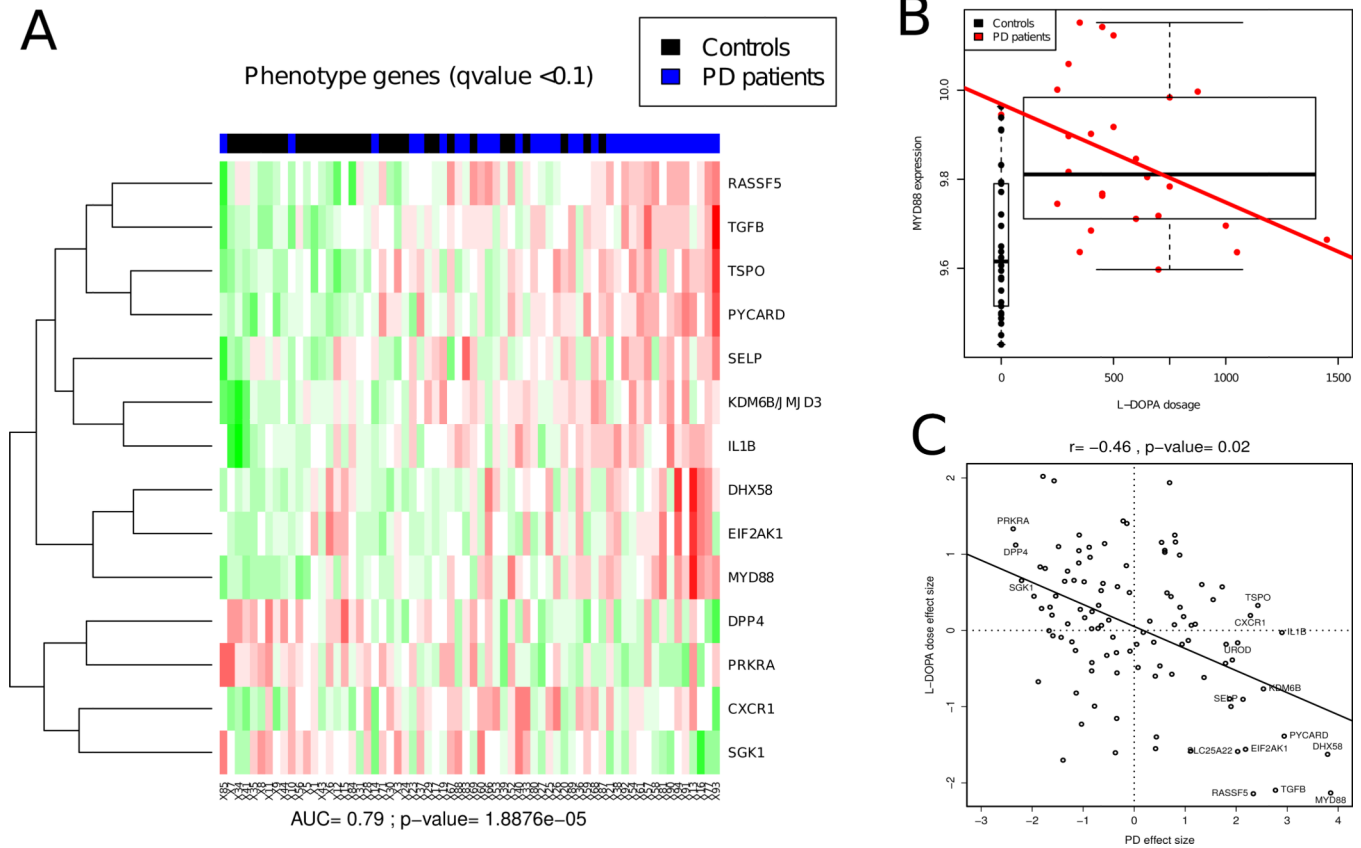
23. Taxman DJ, Holley-Guthrie EA, Huang MT, et al. The NLR adaptor ASC/PYCARD regulates DUSP10, mitogen-activated protein kinase (MAPK), and chemokine induction independent of the inflammasome. *The Journal of biological chemistry*. 2011; 286(22):19605–19616. [PubMed: 21487011]
24. Zhou F, Zhang X, van Dam H, et al. Ubiquitin-specific protease 4 mitigates Toll-like/interleukin-1 receptor signaling and regulates innate immune activation. *The Journal of biological chemistry*. 2012; 287(14):11002–11010. [PubMed: 22262844]
25. Wang L, Zhao W, Zhang M, et al. USP4 positively regulates RIG-I-mediated antiviral response through deubiquitination and stabilization of RIG-I. *Journal of virology*. 2013; 87(8):4507–4515. [PubMed: 23388719]
26. Sotiropoulou G, Pampalakis G, Diamandis EP. Functional roles of human kallikrein-related peptidases. *The Journal of biological chemistry*. 2009; 284(48):32989–35994. [PubMed: 19819870]
27. Zinger A, Barcia C, Herrero MT, et al. The involvement of neuroinflammation and kynurenine pathway in Parkinson's disease. *Parkinson's disease*. 2011; 2011:716859.
28. Li X, Ranjith-Kumar CT, Brooks MT, et al. The RIG-I-like receptor LGP2 recognizes the termini of double-stranded RNA. *The Journal of biological chemistry*. 2009; 284(20):13881–13891. [PubMed: 19278996]
29. Liu SB, Mi WL, Wang YQ. Research progress on the NLRP3 inflammasome and its role in the central nervous system. *Neuroscience bulletin*. 2013
30. Zhu J, Mohan C. Toll-like receptor signaling pathways--therapeutic opportunities. *Mediators of inflammation*. 2010; 2010:781235. [PubMed: 20981241]
31. Bommireddy R, Doetschman T. TGFbeta1 and Treg cells: alliance for tolerance. *Trends in molecular medicine*. 2007; 13(11):492–501. [PubMed: 17977791]
32. Das ND, Jung KH, Choi MR, et al. Gene networking and inflammatory pathway analysis in a JMJD3 knockdown human monocytic cell line. *Cell biochemistry and function*. 2012; 30(3):224–232. [PubMed: 22252741]
33. Wu J, Lou H, Alerte TN, et al. Lewy-like aggregation of alpha-synuclein reduces protein phosphatase 2A activity in vitro and in vivo. *Neuroscience*. 2012; 207:288–297. [PubMed: 22326202]
34. Saraceno C, Marcello E, Di Marino D, et al. SAP97-mediated ADAM10 trafficking from Golgi outposts depends on PKC phosphorylation. *Cell death & disease*. 2014; 5:e1547. [PubMed: 25429624]
35. Suh J, Choi SH, Romano DM, et al. ADAM10 missense mutations potentiate beta-amyloid accumulation by impairing prodomain chaperone function. *Neuron*. 2013; 80(2):385–401. [PubMed: 24055016]
36. Nisole S, Stoye JP, Saib A. TRIM family proteins: retroviral restriction and antiviral defence. *Nature reviews Microbiology*. 2005; 3(10):799–808.
37. Fan HP, Di Liao C, Fu BY, et al. Interindividual and interethnic variation in genomewide gene expression: insights into the biological variation of gene expression and clinical implications. *Clinical chemistry*. 2009; 55(4):774–785. [PubMed: 19233909]
38. Peters EH, Rojas-Caro S, Brigell MG, et al. Quality-controlled measurement methods for quantification of variations in transcript abundance in whole blood samples from healthy volunteers. *Clinical chemistry*. 2007; 53(6):1030–1037. [PubMed: 17434908]
39. Hamza TH, Zabetian CP, Tenesa A, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nature genetics*. 2010; 42(9):781–785. [PubMed: 20711177]
40. Nalls MA, Plagnol V, et al. International Parkinson Disease Genomics C. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*. 2011; 377(9766):641–649. [PubMed: 21292315]
41. Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 2006; 141B(3):261–268.

42. Dzamko N, Inesta-Vaquera F, Zhang J, et al. The IkappaB kinase family phosphorylates the Parkinson's disease kinase LRRK2 at Ser935 and Ser910 during Toll-like receptor signaling. *PLoS one*. 2012; 7(6):e39132. [PubMed: 22723946]
43. Liu B, Dai J, Zheng H, et al. Cell surface expression of an endoplasmic reticulum resident heat shock protein gp96 triggers MyD88-dependent systemic autoimmune diseases. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100(26):15824–15829. [PubMed: 14668429]
44. Marshak-Rothstein A. Toll-like receptors in systemic autoimmune disease. *Nature reviews Immunology*. 2006; 6(11):823–835.
45. Cebrian C, Zucca FA, Mauri P, et al. MHC-I expression renders catecholaminergic neurons susceptible to T-cell-mediated degeneration. *Nature communications*. 2014; 5:3633.
46. Monahan AJ, Warren M, Carvey PM. Neuroinflammation and peripheral immune infiltration in Parkinson's disease: an autoimmune hypothesis. *Cell transplantation*. 2008; 17(4):363–372. [PubMed: 18522239]
47. Li X, Sundquist J, Sundquist K. Subsequent risks of Parkinson disease in patients with autoimmune and related disorders: a nationwide epidemiological study from Sweden. *Neurodegenerative diseases*. 2012; 10(1–4):277–284. [PubMed: 22205172]
48. Pattarini R, Rong Y, Shepherd KR, et al. Long-lasting transcriptional refractoriness triggered by a single exposure to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyrimidine. *Neuroscience*. 2012; 214:84–105. [PubMed: 22542874]
49. Twine NA, Janitz K, Wilkins MR, et al. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS one*. 2011; 6(1):e16266. [PubMed: 21283692]
50. Mogi M, Kondo T, Mizuno Y, et al. p53 protein, interferon-gamma, and NF-kappaB levels are elevated in the parkinsonian brain. *Neuroscience letters*. 2007; 414(1):94–97. [PubMed: 17196747]
51. Shimoji M, Pagan F, Heaton EB, et al. CXCR4 and CXCL12 expression is increased in the nigrostriatal system of Parkinson's disease. *Neurotoxicity research*. 2009; 16(3):318–328. [PubMed: 19551455]
52. Brochard V, Combadiere B, Prigent A, et al. Infiltration of CD4+ lymphocytes into the brain contributes to neurodegeneration in a mouse model of Parkinson disease. *The Journal of clinical investigation*. 2009; 119(1):182–192. [PubMed: 19104149]
53. Noelker C, Morel L, Lescot T, et al. Toll like receptor 4 mediates cell death in a mouse MPTP model of Parkinson disease. *Scientific reports*. 2013; 3:1393. [PubMed: 23462811]
54. Junn E, Mouradian MM. MicroRNAs in neurodegenerative disorders. *Cell Cycle*. 2010; 9(9):1717–1721. [PubMed: 20404550]
55. McHugh D, Hu SS, Rimmerman N, et al. N-arachidonoyl glycine, an abundant endogenous lipid, potently drives directed cellular migration through GPR18, the putative abnormal cannabidiol receptor. *BMC neuroscience*. 2010; 11:44. [PubMed: 20346144]
56. Lee Y, Hur I, Park SY, et al. The role of PACT in the RNA silencing pathway. *The EMBO journal*. 2006; 25(3):522–532. [PubMed: 16424907]
57. Mouradian MM. MicroRNAs in Parkinson's disease. *Neurobiology of disease*. 2012; 46(2):279–284. [PubMed: 22245218]
58. Gehrke S, Imai Y, Sokol N, et al. Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression. *Nature*. 2010; 466(7306):637–641. [PubMed: 20671708]
59. Nikonova EV, Xiong Y, Tanis KQ, et al. Transcriptional responses to loss or gain of function of the leucine-rich repeat kinase 2 (LRRK2) gene uncover biological processes modulated by LRRK2 activity. *Human molecular genetics*. 2012; 21(1):163–174. [PubMed: 21972245]
60. Orenstein SJ, Kuo SH, Tasset I, et al. Interplay of LRRK2 with chaperone-mediated autophagy. *Nature neuroscience*. 2013; 16(4):394–406.



**Fig. 1. Principal component analyses in mice**

Blood expression profiles discriminate between mice with altered *LRRK2* genes. Twelve markers were evaluated for significant differences among the following four mouse groups using an F-test: WTC, KO, transgenic LRRK2-WT and LRRK2-GS. Markers with a q-value of 0.1 were used to construct the principal component analysis (PCA) plot. The percentage of explained variance is indicated in parenthesis on each axis. The four groups show consistent separation, with KO mice and the over-expressing transgenic lines appearing on opposite sides of WT. While the GS group is more variable than the other three, it appears to yield the most extreme expression alterations. Ellipses show the mean and 1 standard deviation for the multivariate distribution.



**Fig. 2. Human phenotype heatmap and drug effect**

**A**, Blood expression profiles discriminate between individuals with PD symptoms (PD patients, in blue) and healthy controls (Controls, in black). Marker values were used to fit a linear model using gender, genotype, and phenotype (i.e. presence or absence of PD). Markers that were associated with phenotype at a q-value of 0.1 were used to construct a heatmap. Each marker value is scaled to have 0 mean and 1 SD. Markers were clustered with hierarchical clustering using Euclidean distance. Samples are ordered according to a score that is the sum of the scaled marker values multiplied by the sign of the direction of change in the PD-affected group. This score produces significant discrimination between PD patients and healthy controls with an area under the curve (AUC) of 0.79 ( $p = 1.89 \times 10^{-5}$ ) using five-fold cross-validated support vector machine (SVM) classification using all genes.

**B**, Example of a marker associated with disease status and L-DOPA dose. Box plot and linear fit showing MyD88 gene expression by L-DOPA dosage. MyD88 gene expression is significantly higher in PD patients, yet it is lowered by L-DOPA administration. The expression profiles of PD patients who received higher doses of L-DOPA look more like those of normal controls. **C**, In order to evaluate the relative and potentially confounding effects of disease presence and L-DOPA dose on the expression of each gene, we plotted the t-statistic obtained for each gene comparing PD patients and asymptomatic individuals (X axis) and the t-statistic obtained from a regression against L-DOPA dose (Y axis, as shown for MyD88 in **B**). These paired t-values were then fit by a linear regression, showing an inverse correlation. The expression profiles of PD patients who received higher doses of L-



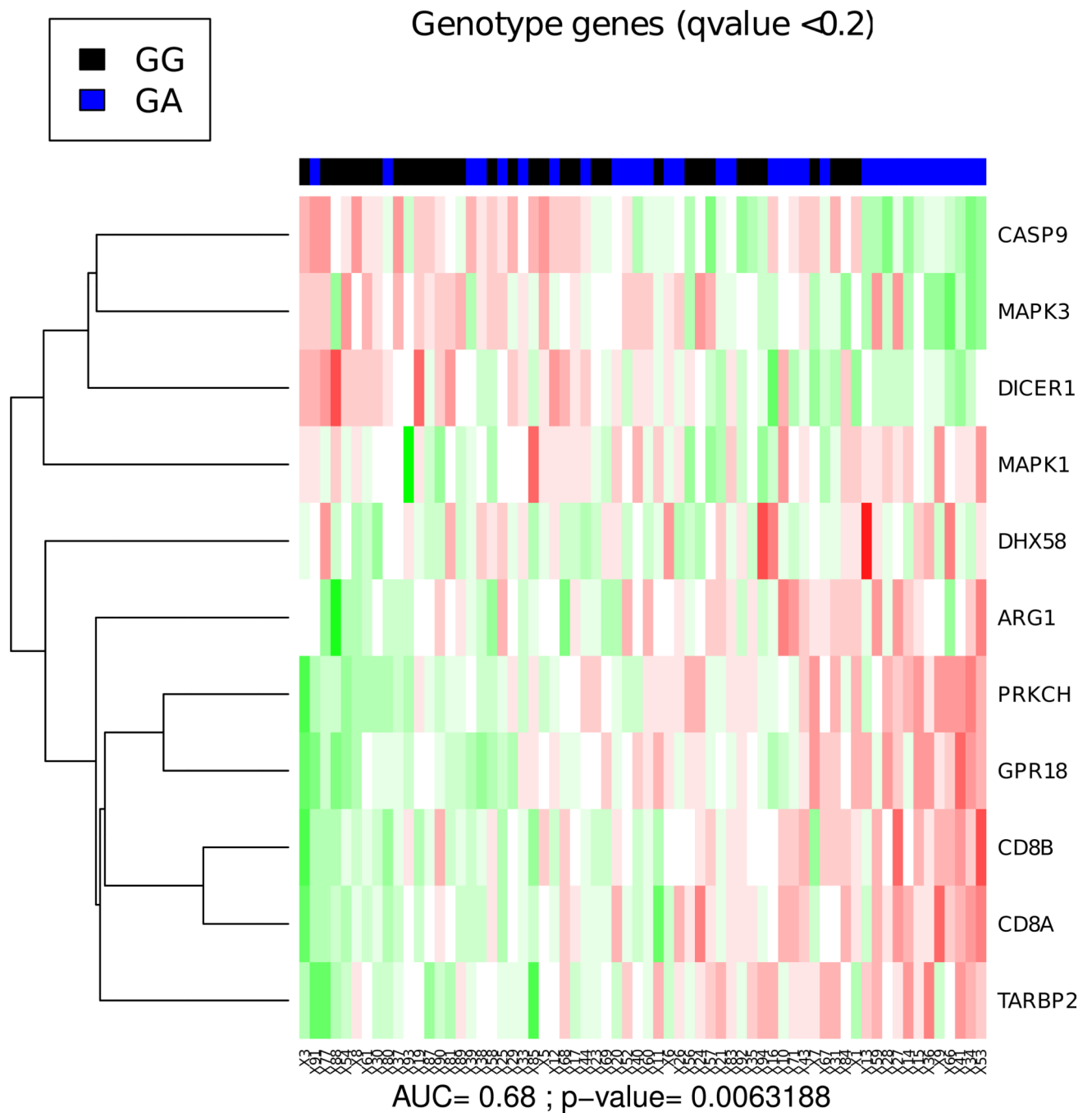
DOPA look more like those of normal controls. Since gene effect sizes are not independent, significance was evaluated by a permutation test: 500 random correlation values were generated from permuting the L-DOPA doses and deriving the corresponding effect sizes. The inverse relationship is significant ( $p= 0.02$ ) and further corroborates the argument that the observed expression profiles are due to the underlying disease biology.

Author Manuscript

Author Manuscript

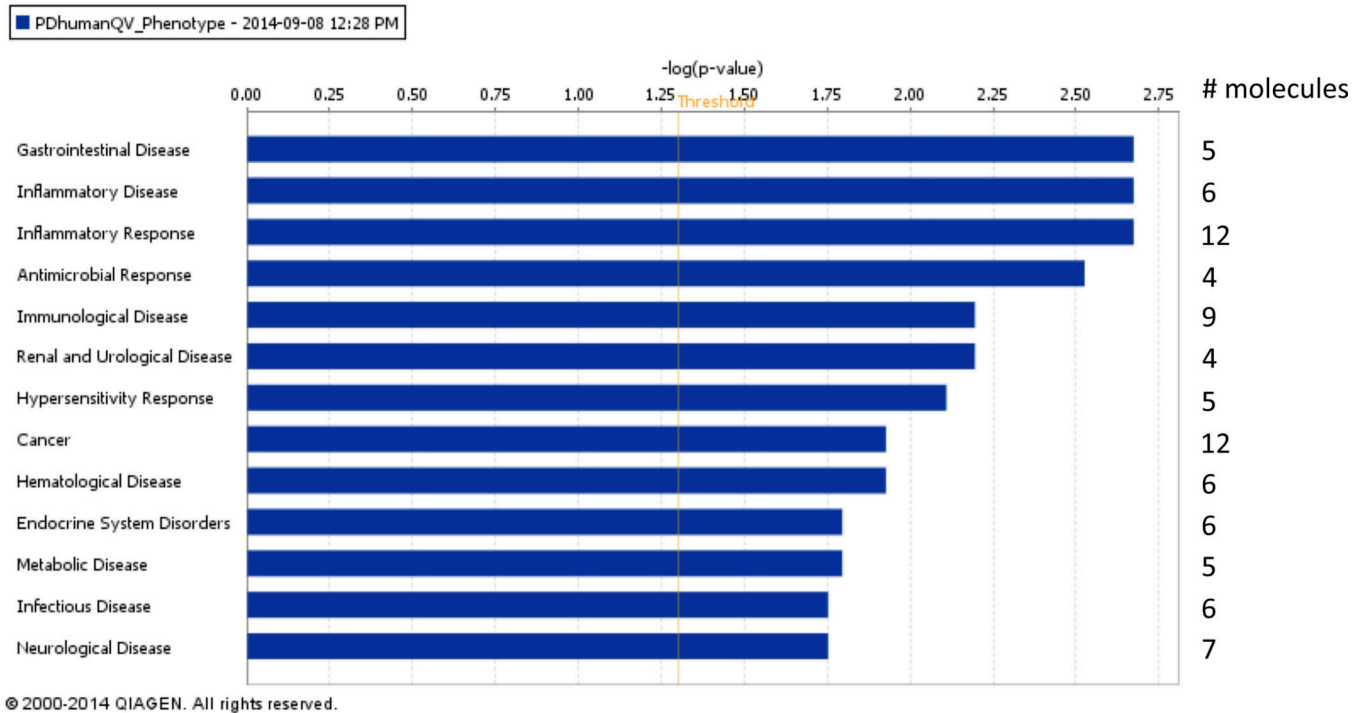
Author Manuscript

Author Manuscript



**Fig. 3. Human genotype heatmap**

Blood expression profiles discriminate between individuals with different *LRRK2* genotypes (wild-type GG vs. G2019S mutant GA) independently of the presence of PD symptoms. Markers that were associated with genotype at a q-value of 0.2 were used to construct a heatmap as in Fig. 2. The genotype effect is smaller in magnitude. Using the same score calculation as in Fig. 2, we obtain an AUC of 0.68 ( $p=6.32 \times 10^{-3}$ ) using five-fold cross-validated SVM classification using all genes.



**Fig. 4. Functional network analysis of the PD human gene signature**

Bar-plot of the diseases and disorders that are over-represented in the 14 PD-phenotype-associated markers in our human dataset. Cutoff for the expression values (i.e. q-values) was set to 0.1 in order to select the 14 signature markers. The IPA Functional Analysis shows that several genes among the 14 discriminating PD disease markers are significantly linked to inflammatory disease, inflammatory response, immunological disease, as well as neurological disease. Uncorrected p-values (Fisher's Exact Test p-value) were measured.

**Table 1**

Clinical features of the Study cohort

<b>LRRK2 genotype</b>	<b>Healthy Controls GG (WT)</b>	<b>Healthy Controls GA (G2019S)</b>	<b>PD patients GG (WT)</b>	<b>PD patients GA (G2019S)</b>
Number	12	14	17	16
Male: Female ratio	4:8	5:9	10:7	10:6
Age at blood draw, mean year +/-SD	62.3 +/- 16.1	53.9 +/- 22.7	64.7 +/- 6.4	70.6 +/- 10.8
Age at onset, mean year +/-SD	N/A	N/A	57.1 +/- 7.8	57.4 +/- 11.5
UPDRS motor, median	1 (0–12)	2 (0–17.5)	17 (5.5–53)	14.5 (1–65)
% definitely on L-DOPA *	N/A	N/A	70.6% (12/17)	81.3% (13/16)

\* L-DOPA status of some patients is not known

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript