Network-based metaanalysis identifies HNF4A and PTBP1 as longitudinally dynamic biomarkers for Parkinson's disease

Jose A. Santiago and Judith A. Potashkin[1]

Cellular and Molecular Pharmacology Department, Chicago Medical School, Rosalind Franklin University of Medicine and Science, North Chicago, IL 60064

Edited by Gregory A. Petsko, Weill Cornell Medical College, New York, NY, and approved January 7, 2015 (received for review December 11, 2014)

Environmental and genetic factors are likely to be involved in the pathogenesis of Parkinson's disease (PD), the second most prevalent neurodegenerative disease among the elderly. Network-based metaanalysis of four independent microarray studies identified the hepatocyte nuclear factor 4 alpha (*HNF4A*), a transcription factor associated with gluconeogenesis and diabetes, as a central regulatory hub gene up-regulated in blood of PD patients. In parallel, the polypyrimidine tract binding protein 1 (*PTBP1*), involved in the stabilization and mRNA translation of insulin, was identified as the most down-regulated gene. Quantitative PCR assays revealed that *HNF4A* and *PTBP1* mRNAs were up- and down-regulated, respectively, in blood of 51 PD patients and 45 controls nested in the Diagnostic and Prognostic Biomarkers for Parkinson's Disease. These results were confirmed in blood of 50 PD patients compared with 46 healthy controls nested in the Harvard Biomarker Study. Relative abundance of *HNF4A* mRNA correlated with the Hoehn and Yahr stage at baseline, suggesting its clinical utility to monitor disease severity. Using both markers, PD patients were classified with 90% sensitivity and 80% specificity. Longitudinal performance analysis demonstrated that relative abundance of *HNF4A* and *PTBP1* mRNAs significantly decreased and increased, respectively, in PD patients during the 3-y follow-up period. The inverse regulation of *HNF4A* and *PTBP1* provides a molecular rationale for the altered insulin signaling observed in PD patients. The longitudinally dynamic biomarkers identified in this study may be useful for monitoring disease-modifying therapies for PD.

Parkinson's disease | HNF4A | PTBP1 | network analysis | blood biomarkers

Substantial efforts have been devoted to the development of diagnostic strategies for Parkinson's disease (PD). In particular, changes in mRNA from cellular whole blood can facilitate the identification of dysregulated processes and diagnostic biomarkers for PD (1, 2). Several molecular signatures in blood have been identified. For example, 22 unique genes were found differentially expressed in blood of PD patients compared with healthy controls (1). Likewise, specific splice variants in blood were associated with PD in samples obtained from two independent clinical trials (2, 3). In addition, altered expression of the vitamin D receptor (VDR) in blood and reduced plasma levels of 25-hydroxy vitamin D_3 have been associated with PD (1, 4). Furthermore, plasma levels of the epidermal growth factor have been associated with cognitive decline in PD (5).

Environmental stressors and genetic factors are most likely involved in the pathogenesis of PD. Among the genetic factors associated with PD, mutations in the gene encoding leucine-rich repeat kinase 2 (*LRRK2*) are the most common cause of autosomal dominant PD (6) and a considerable risk factor in idiopathic forms of the disease (7, 8). Given the complex interaction between environmental and genetic factors in sporadic PD, we integrated four independent microarray studies from patients harboring a mutation in the LRRK2 gene (G2019S; glycine-to-serine substitution at amino acid 2019), sporadic patients, and untreated PD patients to identify a universal signature in blood associated with PD. We performed a transcriptomic and network-based metaanalysis to identify key regulators and potential diagnostic biomarkers. This is a powerful approach to integrate gene expression data and to gain insight into complex diseases (9). The utility of network biology to identify biologically relevant biomarkers for neurodegenerative diseases has been demonstrated recently. In this context, network analyses identified the amyloid precursor protein (*APP*) and superoxide dismutase 2 (*SOD2*) mRNAs as blood biomarkers of PD (10, 11) and protein tyrosine phosphatase 1 (*PTPN1*) mRNA as a diagnostic tool for progressive supranuclear palsy (12; reviewed in ref. 9).

In this study, network-based metaanalysis identified hepatocyte nuclear factor 4 alpha (*HNF4A*) and polypyrimidine tract binding protein 1 (*PTBP1*), previously implicated in diabetes, as the most significant up- and down-regulated genes in blood of PD patients. These results were confirmed in blood samples from two independent clinical trials. Relative abundance of *HNF4A* mRNA correlated with disease severity in PD patients. Moreover, longitudinal analysis of *HNF4A* and *PTBP1* revealed that their relative abundance changed over time, thus suggesting their potential use in tracking the clinical course of PD patients. Further evaluation of *HNF4A* and *PTBP1* mRNAs in patients at risk for PD is warranted.

Results

Metaanalysis of Blood Microarrays in PD. To identify a common transcriptional signature in blood of PD patients, four microarray

Significance

Development of therapeutic strategies for Parkinson's disease (PD) is hampered by the lack of reliable biomarkers to identify patients at early stages of the disease and track the therapeutic effect of potential drugs and neuroprotective agents. Readily accessible biomarkers capable of providing information about disease status are expected to accelerate this progress. We identified hepatocyte nuclear factor (*HNF4A*) and polypyrimidine tract binding protein 1 (*PTBP1*) mRNAs as promising blood biomarkers for identifying early stage PD patients with high diagnostic accuracy. Furthermore, *HNF4A* was identified as a potential biomarker to monitor disease severity. Longitudinal analysis demonstrated that *HNF4A* and *PTBP1* are longitudinally dynamic biomarkers that provide insights into the molecular mechanisms underlying the altered insulin signaling in PD patients and may enable novel therapeutic strategies.

Author contributions: J.A.S. and J.A.P. conceived and designed the project; J.A.S. conducted microarray metaanalysis, network analyses, qPCR experiments, and statistical analyses; and J.A.S. and J.A.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: judy.potashkin@rosalindfranklin.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1423573112/-/DCSupplemental.

studies (Table S1) were analyzed using Integrative Meta-Analysis of Expression Data (INMEX), a web interface for the integrative metaanalysis (13). The overall metaanalysis workflow used in this study is shown in Fig. 1*A*. Metaanalysis using a Fisher's test identified a total of 2,781 genes differentially expressed consistently across four microarray studies. Among this group, 680 genes were up-regulated and 2,101 were down-regulated in PD compared with healthy controls. The thy-1 cell-surface antigen (*THY1*) and *HNF4A* were the most significant up-regulated genes across the four microarray datasets. The complete list of differentially expressed genes is provided in Dataset S1. There were 921 gained genes uniquely identified in the metaanalysis that show relatively weak but consistent expression across the four datasets. A total of 491 genes were classified as lost genes (i.e., genes identified as differentially expressed genes in individual datasets but not in the metaanalysis). A Venn diagram of metaanalysis results is shown in Fig. 1*B*, and a heat map visualization of the top 20 genes across the different studies is displayed in Fig. 1*C*.

Biological and Functional Analysis. To identify the overrepresented biological processes dysregulated in blood of PD patients, we performed a gene pathway analysis using NetworkAnalyst (14). Pathway analysis was performed using the set of up- and down-regulated genes separately. Up-regulated genes in blood of PD were associated with the Kyoto Encyclopedia of Genes (KEGG) pathways ($P < 0.05$), including bacterial invasion of epithelial cells, mitogen-activated protein kinase-signaling pathway, fructose and mannose metabolism, T-cell receptor-signaling pathway, mammalian target of rapamycin-signaling pathway, type 2 diabetes mellitus, and colorectal cancer. The most important hub gene in terms of network topology measures of betweeness (BC) and degree of centrality (DC) was *HNF4A* (BC = 2,213; DC = 84) (Fig. 2*A*).

In parallel, down-regulated genes in blood of PD patients were associated with the KEGG pathways, including protein processing in the endoplasmic reticulum (ER), Epstein–Barr virus infection, and several types of cancer including prostate, endometrial, and lung cancer. The most prominent hub gene in terms of network topology measures was ubiquitin C (*UBC*) (BC = 495; DC = 1630), and *PTBP1* was the most down-regulated gene across the four microarray datasets (Fig. 2*B* and Dataset S1).

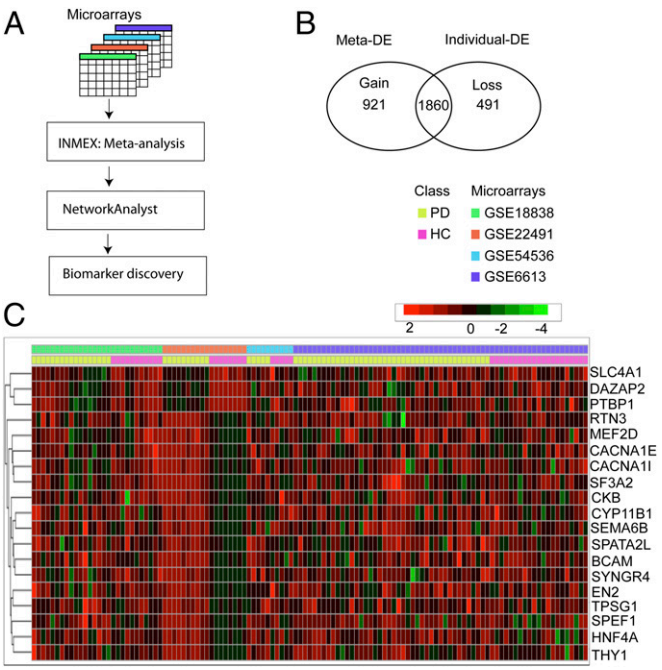

Fig. 1. Network-based and transcriptomic metaanalysis. (*A*) Four independent microarray datasets were downloaded from the GEO and preprocessed in INMEX where metaanalysis was undertaken using the Fisher's method. Datasets were subsequently uploaded into NetworkAnalyst to perform network and functional analysis and to identify key regulatory hub genes across the multiple microarray studies. Finally, the most significant genes were evaluated as biomakers for PD in blood samples obtained from two independent cohorts of patients. (*B*) Venn diagram of differentially expressed genes identified from the metaanalysis (Meta-DE) and those from each individual microarray analysis (Individual-DE). (*C*) Heat map representation of the top 20 differentially expressed genes across different microarrays identified from the metaanalysis (row-wise comparison). The heat map was rescaled to prevent domination by study-specific effects. HC, healthy controls; PD, Parkinson's disease.

Network-Based Metaanalysis. *HNF4A* was confirmed as potential key hub gene in blood of PD by network-based metaanalysis implemented in NetworkAnalyst (14). The most highly ranked node across the four datasets based on network topology measures was *HNF4A* (BC = 329; DC = 35) followed by *GATA1* (BC = 10.5; DC = 8). The resulting zero-order interaction network contained 76 nodes and 81 edges (Fig. S1). In addition, network-based metaanalysis identified the aberrant expression of several splicing factors in PD patients (Fig. S2 *A* and *B*). Among the splicing factors, *PTBP1* was the most significantly down-regulated gene in PD patients identified in the metaanalysis (Fig. 2*B*, Fig. S2*B*, and Dataset S1).

To confirm the dysregulation of HNF4A and PTBP1 at the protein level, we analyzed a protein microarray study in human serum samples of PD (GSE29654) using NetworkAnalyst (15). PTBP1 was significantly down-regulated in PD samples compared with healthy controls ($P = 0.002$). Altered expression of HNF4A was not confirmed in this protein microarray.

Evaluation of HNF4A and PTBP1 mRNAs in Blood of PD. To validate the results obtained from the network-based metaanalysis, we evaluated the most significant hub gene in the up-regulated network, *HNF4A*, and the most down-regulated gene, *PTBP1*, as potential biomarkers for PD. Relative abundance of *HNF4A* and *PTBP1* mRNAs was measured in whole blood of PD patients compared with healthy controls (HCs) from samples obtained from two independent clinical trials, the Diagnostic and Prognostic Biomarkers for Parkinson's Disease (PROBE) and the Harvard Biomarker Study (HBS). Quantitative PCR (qPCR) assays revealed that *HNF4A* and *PTBP1* mRNAs were significantly up- and down-regulated, respectively, in blood of PD patients compared with HCs in both cohorts of study participants at baseline (Fig. 3 *A*–*D*). Analysis of receiver operating characteristic (ROC) was performed to evaluate the diagnostic accuracy of both biomarkers. ROC analysis for *HNF4A* and *PTBP1* resulted in an area under the curve (AUC) of 0.72 and 0.82, respectively (Fig. S3 *A* and *B*). Combination of both biomarkers resulted in an AUC value of 0.90 (Fig. S3*C*). A step-wise linear discriminant analysis showed that PD patients can be classified with 90% sensitivity and 80% specificity according to the following canonical function: $D_{PD} = 0.10 + 0.56^{*}X_{PTBP1} - 0.20X_{HNF4A}$, where D_{PD} is the discriminant score value and X is the mRNA expression level for each biomarker multiplied by its respective canonical coefficient.

Pearson correlation analysis demonstrated that relative abundance of *HNF4A* and *PTBP1* was independent of other covariates including age (*HNF4A*: $r = -0.25$, $P = 0.9$; *PTBP1*: $r = 0.09$, $P = 0.59$) and sex (*HNF4A*: $r = -0.004$, $P = 0.97$; *PTBP1*: $r = 0.05$, $P = 0.76$) in both cohorts of patients and body mass index (BMI) (*HNF4A*: $r = -0.005$, $P = 0.96$; *PTBP1*: $r = 014$, $P = 0.37$) in the HBS cohort. Correlation analysis revealed a significant negative correlation between *HNF4A* mRNA expression and *PTBP1* mRNA ($r = -0.20$, $P = 0.008$, Fig. 4, *Upper*) and Hoehn and Yahr stage (HY) at baseline ($r = -0.32$, $P = 0.002$, Fig. 4, *Lower*).

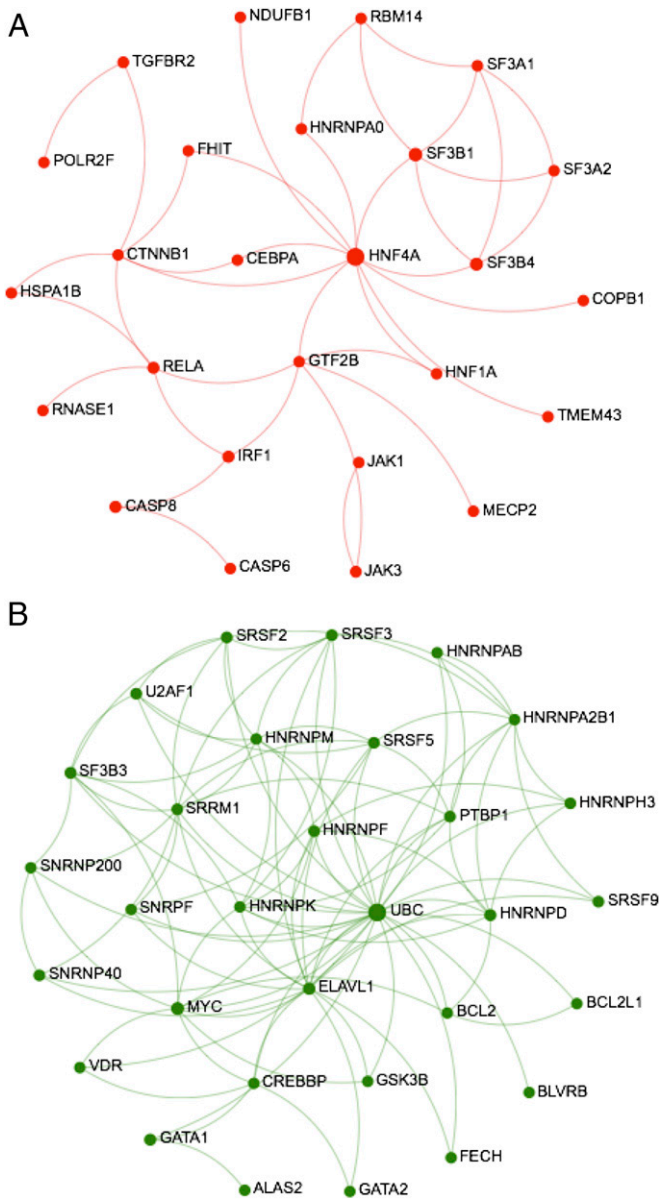

Fig. 2. Network analysis of differentially expressed genes in blood of PD patients. (*A*) Zero-order interaction network of genes up-regulated in blood of PD patients (red). (*B*) Zero-order interaction network of genes down-regulated in blood of PD patients (green).

Longitudinal Performance of HNF4A and PTBP1. To determine the longitudinal performance of *HNF4A* and *PTBP1*, we measured the relative abundance of each biomarker in HBS samples collected at two time points. The estimated rate of change for each biomarker was determined via a linear mixed-effects model using the two time points (baseline and 3-y follow-up) collected repeatedly between the same subjects adjusting for age, sex, and BMI. Relative abundance of *HNF4A* mRNA significantly decreased over time in PD patients compared with HCs ($\beta = -0.93$, $P = 0.002$) whereas *PTBP1* mRNA increased in PD patients ($\beta = 0.33$, $P = 0.004$) (Fig. 5). Relative abundance of *HNF4A* and *PTBP1* mRNAs was significantly up-regulated in PD patients compared with HCs in the follow-up period (Fig. 5 *B* and *D*). Correlation between the relative abundance of each biomarker and HY stage did not reach statistical significance in the longitudinal analysis.

Discussion

Biomarker discovery and validation is a crucial step toward the improvement of clinical management of PD. Specifically, biomarkers that are useful in tracking the clinical course of PD are essential to the development of effective therapeutics. Network analysis offers an unbiased approach to identify and prioritize biologically meaningful biomarkers for several neurodegenerative diseases (9). Here we performed a network-based metaanalysis integrating gene expression profiles of untreated, sporadic, and PD patients harboring a LRRK2 (G2019S) mutation to identify convergence among the different studies in blood of PD. Transcriptomic metaanalysis identified 2,781 genes consistently differentially expressed in blood of PD across four microarray studies.

Network-based metaanalysis identified *HNF4A* as the most significant hub gene across the four microarrray datasets. HNF4A is a master metabolic regulatory factor responsible for the activation of the hepatic gluconeogenesis (16) and has been implicated in diabetes, inflammation, and lipid metabolism (17–21). In this context, diabetes has been associated with PD in numerous epidemiological studies (22–25). Although the exact mechanism by which both diseases coexist is not clearly understood, a recent review suggests that dysregulation in common molecular processes including the insulin signaling pathway, inflammation, ER stress, and mitochondrial dysfunction may lead to both diseases (26). Numerous lines of evidence support a link between PD and diabetes, including cellular, molecular, and animal models and network-based approaches (27).

Interestingly, HNF4A was previously identified as a central regulatory node of a splicing signature in blood of PD patients (2). Furthermore, HNF4A interacts with peroxisome proliferator activator receptor gamma (PPAR-γ), a potential therapeutic target in PD (28), in the activation of genes controlling the hepatic gluconeogenesis (29). The second most prominent node across the four datasets was GATA1 (Fig. S1), a transcription factor that is known to regulate α-synuclein (SNCA) transcription in

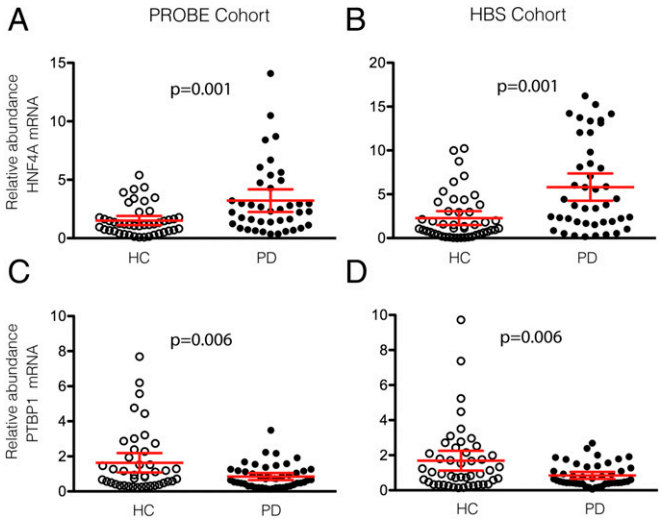

Fig. 3. Evaluation of HNF4A and PTBP1 mRNAs as biomarkers for PD at baseline. (*A*) Relative abundance of *HNF4A* mRNA in blood of PD patients (black circles) compared with HCs (white circles) in samples from the PROBE cohort. (*B*) Replication of biomarker expression in an independent set of samples from patients enrolled in the HBS study. (*C*) Relative abundance of *PTBP1* mRNA in blood of PD patients compared with HC in samples from the PROBE cohort. (*D*) Replication of *PTBP1* mRNA expression in an independent set of samples from patients enrolled in the HBS study. Relative abundance of each biomarker was calculated using *GAPDH* as a reference gene and HC as calibrator. A Student *t* test (two-tailed) was used to estimate the significance between PD cases and controls. Error bars represent 95% confidence intervals.

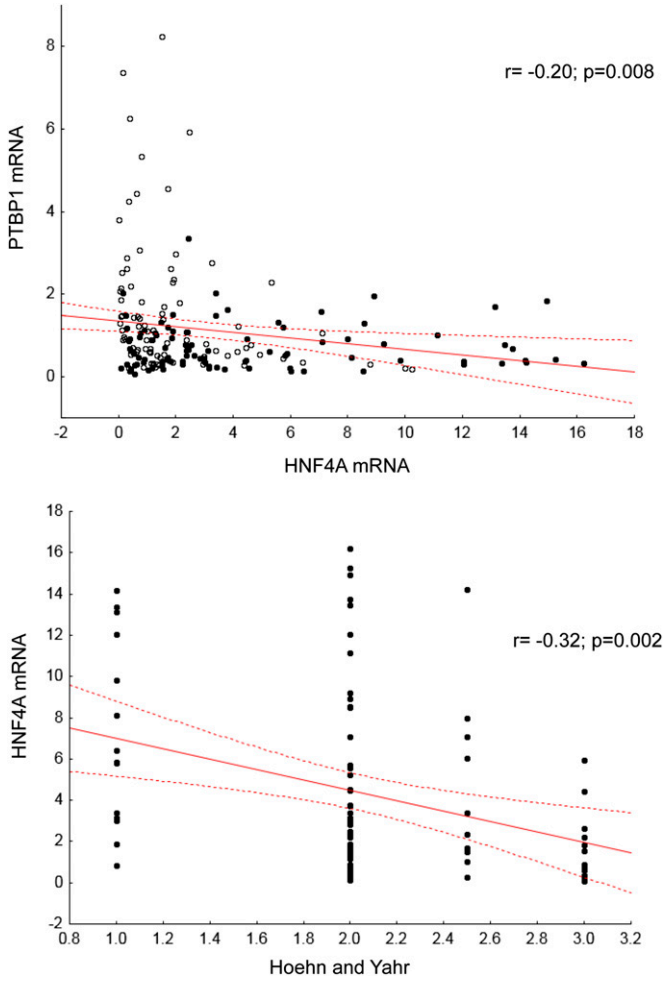

Fig. 4. Biomarker correlation analysis. (*Upper*) Pearson correlation analysis between *HNF4A* mRNA and *PTBP1* mRNA in blood of PD patients (black circles) and HC (white circles) in samples from PROBE and HBS. (*Lower*) Correlation analysis between HNF4A mRNA and Hoehn and Yahr scale in PD patients from both cohorts. Solid lines represent the linear regression of the data and dashed lines indicate the 95% confidence of intervals.

erythroid precursor cells, thereby suggesting its potential as a modifying agent in PD (30).

PTBP1 was identified as the most significant down-regulated gene across the four microarrays datasets. Interestingly, PTBP1 promotes stabilization and translation of insulin mRNA in pancreatic β-cells (31). In this context, impaired insulin signaling, decreased expression of insulin receptor mRNA, and insulin resistance have been implicated in PD in several studies (26, 32–34). Strikingly, PTBP1 regulates the transcription of the glucagon-like peptide 1 (GLP-1), a secretory granule that potentiates glucose-stimulated insulin secretion in pancreatic β-cells (31). Pharmacological stimulation of GLP-1 receptors with drugs used to treat diabetic patients, such as exenatide, elicits neuroprotective effects in animal models of PD (35), and its therapeutic potential has shown promise in clinical trials of PD (36–38).

We further evaluated *HNF4A* and *PTBP1* mRNAs as blood biomarkers for PD. Relative abundance of *HNF4A* mRNA was up-regulated whereas *PTBP1* mRNA was down-regulated in blood of PD patients compared with healthy individuals in samples obtained from two independent clinical trials (Fig. 3 *A–D*). Evaluation of biomarker performance showed that *HNF4A* and *PTBP1* can distinguish PD patients from HCs with 90% sensitivity and 80% specificity (Fig. S3*C*). The diagnostic performance of these two biomarkers is superior to the one afforded by previously identified risk markers in blood of PD patients and current clinical assessment (3, 39). The sensitivity of the two markers alone is also greater than the splice variant-specific RNA blood biosignature that included 13 risk markers (2).

HNF4A mRNA relative abundance significantly correlated with *PTBP1* mRNA. We found a significant negative correlation between *HNF4A* mRNA expression and the HY staging. Early PD patients with a low HY scale rating (HY = 1) showed a significantly higher up-regulation of *HNF4A* mRNA compared with patients with a higher HY scale (HY = 3) (Fig. 4, *Lower*). This finding suggests that *HNF4A* mRNA may be useful in identifying patients at very early stages of PD when therapeutic intervention may be most beneficial and in monitoring disease severity.

Longitudinal performance analysis showed that the relative abundance of each biomarker significantly changed over time in PD patients. For example, *HNF4A* mRNA significantly decreased whereas *PTBP1* mRNA increased in PD patients during the 3-y follow-up (Fig. 5). The correlation between the relative abundance of both biomarkers with HY stage did not reach statistical significance in follow-up samples, however. One possible explanation is that the HY stage did not change in most of the PD patients during the 3-y period, whereas the relative abundance of the transcripts did change during this time period. These results suggest that the abundance of *HNF4A* and *PTBP1* mRNAs in blood may be more sensitive than assessment of motor symptoms for monitoring disease progression. The dynamic change in expression over time of both biomarkers suggests that they may be useful biomarkers in tracking the clinical course of PD patients.

One potential caveat is that most of the PD patients were medicated in this study; therefore, a potential confounding factor introduced by PD medications cannot be ruled out. Nevertheless, this finding is interesting in light of the evidence that indicates that more than 50% of the PD patients are glucose intolerant and patients with diabetes who develop PD usually have a higher HY staging (40). Moreover, impaired glucose metabolism is suggested to be an early event in sporadic PD (41). Given that HNF4A plays a pivotal role in hepatic gluconeogenesis and

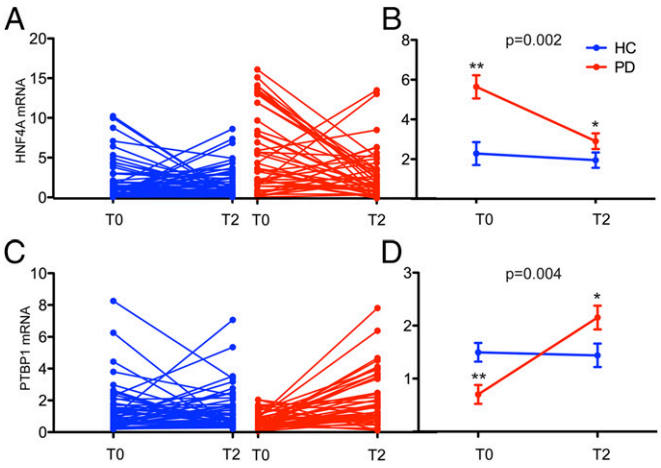

Fig. 5. Longitudinal performance of *HNF4A* and *PTBP1* mRNAs in the HBS study. (*A*) Individual trajectories for the relative abundance of *HNF4A* mRNA over time for HC (blue) and PD patients (red) in the HBS study. (*B*) Average *HNF4A* mRNA expression in PD patients compared with HCs calculated via linear mixed-effects regression analysis adjusting for sex, age, and BMI. (*C*) Individual trajectories for the relative abundance of *PTBP1* mRNA over time for HCs and PD patients. (*D*) Average *PTBP1* mRNA expression in PD patients compared with HCs calculated via linear mixed-effects regression analysis. Red and blue lines denote PD patients and HCs, respectively. T0 and T2 indicate baseline and 3-y follow-up period, respectively. Post hoc pair-wise comparisons were performed using a Tukey test (**P = 0.0001, *P = 0.001). Error bars represent SE.

that PTBP1 regulates and stabilizes mRNA translation of insulin in the pancreas, the inverse regulation of both genes provides a molecular rationale for the impairment of insulin signaling in PD patients, and these genes thus may be potential therapeutic targets.

Analysis of a previous protein microarray study of serum samples of PD patients (15) revealed that PTBP1 was significantly down-regulated in PD patients compared with controls ($P = 0.02$), but expression of HNF4A was not identified. Thus, protein levels of PTBP1 may also be a potential diagnostic biomarker for PD.

The results from this metaanalysis also highlight the dysregulation of several splicing factors in blood of PD patients. As the spliceosome assembles, protein–protein interactions are highly dynamic. One of the essential steps in the assembly of the spliceosome is the formation of new protein interactions that change the inactive B splicing complex to an active complex in which SF3B2, SF3B3, and SF3B5 form new interactions with proteins of the U5 small nuclear ribonucleic particles (snRNPs) (42). In this context, several of the core factors of the U2 snRNPs were up-regulated in PD including SF3A1, SF3A2, SF3B1, and SF3B4, whereas SF3B3 was down-regulated (Fig. 2 and Fig. S2). These results suggest that assembly of the U2 snRNP that binds to the 3′ splice site may be facilitated in PD, but the efficient formation of an active splicing complex in PD is highly unlikely. The results from the metaanalysis also revealed that many of the regulatory splicing factors, core factors of the U1, U4, U5, and U6 snRNPs, and helicases are down-regulated in PD, further supporting the idea that splicing may be both inefficient and dysregulated in PD (Fig. S2*B*). In this regard, aberrant alternative splicing in blood of PD has been highlighted in several studies (2, 3, 43). In addition, heterogeneous nuclear ribonucleoproteins, cap-binding proteins, and proteins of the exon junction complex were down-regulated in PD, suggesting that other posttranscriptional events such as cap-binding protein complex formation, localization, maturation, nonsense-mediated mRNA decay, and translation may be inefficient or dysregulated in PD (Fig. 2 and Fig. S2).

The VDR was also present in the network of down-regulated genes, thus confirming previous findings reporting lower levels of VDR in blood and plasma of PD patients (1, 4, 44). In addition, a subset of highly coexpressed genes associated with heme metabolism previously identified in blood of two independent populations (30), *ALAS2*, *FECH*, and *BLVRB* were also found to be down-regulated in the metaanalysis (Fig. 2*B*). Collectively, these results confirm the presence of a common molecular signature in blood of PD patients.

In summary, this study highlights the prominent convergence among blood microarray studies from sporadic patients, de novo patients, and the most common hereditary cause of PD, and confirms the utility of blood as a useful source of biomarkers for PD. In addition, our results strengthen the association between PD and diabetes and provide insights into the molecular mechanisms underlying the impairment of insulin signaling observed in PD patients. Furthermore, this study underscores the potential of network analysis as a powerful framework to gain insight into the mechanisms underlying PD and to identify potential therapeutic targets and biomarkers of disease severity. Evaluation of *HNF4A* and *PTBP1* mRNAs in a larger prospective study including patients at risk will be important to assess its clinical utility as a diagnostic tool for PD. Further mechanistic studies linking *HNF4A* and *PTBP1* in neurodegeneration are warranted.

Materials and Methods

Microarray Metaanalysis. Gene expression data from microarray studies was downloaded from the Gene Expression Omnibus (GEO) and GEMMA databases (45) by using the terms "Parkinson's disease" and "blood" or "transcriptional profiling" on May 31, 2014. Microarray studies using RNA prepared from human blood with 10 samples or more were included in our study. Only samples from PD patients and healthy controls were analyzed. A total of four microarray studies met the inclusion criteria and were considered for subsequent analysis. The microarray studies analyzed in this study are listed in Table S1. GSE6613 included 50 predominantly early stage sporadic PD patients with a mean Hoehn and Yahr stage of 2.3 from which 9 were untreated patients and 22 were age- and sex-matched healthy controls. GSE8838 included 18 sporadic PD patients treated with different PD medications and 12 HCs. GSE22491 included 10 PD patients carrying a LRRK2 mutation from which 1 patient was untreated and 7 HCs. GSE54536 included five untreated PD patients and five HCs. We conducted a microarray metaanalysis using INMEX (13) in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for metaanalysis (46). All gene probes were converted to a common Entrez ID using the gene/probe conversion tool in INMEX. After matching all probes to a common Entrez ID, all datasets were preprocessed using the $\log_2$ transformation and quantile normalization. Each individual dataset was visualized in box plots to ensure identical distribution among the samples. Differential expression analysis was performed with INMEX for each dataset independently using a false discovery rate of 0.05, a significance of $P < 0.05$, and moderated *t* test based on the Limma algorithm. In INMEX, the results from individual microarray dataset analyses are only for reference comparison and not required for metaanalysis in the subsequent steps (13). For metaanalysis, we used the Fisher's method with a significance level of $P < 0.05$ to combine *P* values from the multiple datasets. Fisher's method is a widely used statistical approach in metaanalysis to combine *P* values from different studies independently of the sample size (13, 47, 48). Gene ontology and functional analysis was performed using NetworkAnalyst (14).

Network-Based Metaanalysis. Network-based metaanalysis was performed using NetworkAnalyst (14). Microarray datasets were processed as described above. Briefly, microarray datasets were preprocessed by a $\log_2$ transformation followed by quantile normalization. Duplicate genes were replaced by their mean value. We used a significance value of $P < 0.05$ and a $\log_2$ fold change of 1 as a cutoff value. Network construction was restricted to contain only the original seed proteins.

Information About Study Participants. The Institutional Review Boards of Rosalind Franklin University of Medicine and Science approved the study protocol. Written informed consent was received from all participants. Clinical characteristics of the participants used in this study have been reported elsewhere (2, 3, 10). Briefly, 51 PD patients (29 men, 22 women; mean age at enrollment 63.16 ± 6.42; Hoehn and Yahr scale 2 ± 0.28) and 45 healthy age-matched controls (24 male, 21 women; mean age at enrollment 65.12 ± 8.60) enrolled in PROBE (#NCT00653783). Clinical diagnosis of PD was based on the United Kingdom Parkinson's Disease Society Brain Bank criteria. Healthy individuals had no history of neurological disease and a Mini-Mental State Examination test score higher than 27. Inclusion and exclusion criteria for patients enrolled in the PROBE study have been reported elsewhere (2). As an independent replication cohort of patients, we used 96 individuals, including 50 PD patients (31 men, 19 women; Hoehn and Yahr scale 1.97 ± 0.62; mean age at enrollment 63.12 ± 8.96; mean age at onset 58.75 ± 10.17) and 46 healthy age-matched controls (26 men, 20 women; mean age at enrollment 64.28 ± 10.42) enrolled in the HBS. Three-year follow-up samples from cases and controls enrolled in the HBS were collected and analyzed in this study. Diagnosis of cases and controls was assessed at each visit to ensure high diagnostic accuracy. Additional information about the participants enrolled in the HBS clinical trial was published previously (49) and can be found at www.neurodiscovery.harvard.edu/research/biomarkers.html.

Quantitative Real-Time Polymerase Chain Reactions. Blood was collected and prepared as described using the PAXgene Blood RNA system (Qiagen) (1, 2). Samples with RNA integrity values >7.0 and ratio of absorbances at 260/280 nm between 1.7 and 2.4 were used in the current study. A total of 1 μg of RNA was reversed-transcribed using the High Capacity RNA Transcription Kit (Life Technologies). Primers were designed using Primer Express software and ordered from Life Technologies. The primer sequences used in this study are as follows: *GAPDH*—forward, 5′-CAACGGATTTGGTCGTATTGG-3′; reverse, 5′-TGATGGCAACAATATCCACTTTACC-3′; *HNF4A*—forward, 5′-CAGAATGAGCGGGACCGGATC-3′; reverse: 5′-CAGCAGCTGCTCCTTCATGGAC-3′; and *PTBP1*—forward, 5′-GCTCAGGATCATCGTGGAGAA-3′; reverse, 5′-ATCTTCAACACTGTGCCGAACTT-3′. Quantitative PCR assays were carried out using 25-μL reactions containing Power SYBR Green master mix (Life Technologies), primer at a concentration of 5 μM and nuclease-free water. PCR reactions were amplified using a DNA engine Opticon 2 Analyzer (Bio-Rad Life Sciences). Amplification conditions and detailed description of qPCR experiments are reported elsewhere (2, 3).

Statistical Analysis. Network-based microarray metaanalysis was performed using INMEX and NetworkAnalyst. A Student *t* test (two-tailed) was used to estimate the significance between PD cases and controls for numerical variables. Pearson correlation analysis was used to determine statistical significance for HNF4A and PTBP1 adjusting for sex, age, Hoehn and Yahr scale in both cohorts, and BMI in the HBS study. An ROC curve analysis was used to evaluate the diagnostic accuracy. A step-wise linear discriminant analysis was performed to determine the sensitivity and specificity values for the linear combination of both biomarkers. Power analyses of completed experiments were performed to demonstrate that the sample size used in this study allowed the detection of a difference of 0.5 in fold change with a power of 99% and a significance of 0.05. A *P* value less than 0.05 was regarded as statistically significant. For the longitudinal analysis, we used a linear mixed-effects regression model including subjects as random effects and adjusting for sex, age, and BMI (50). Post hoc pair-wise comparisons were performed using a Tukey test. Longitudinal data were analyzed using SuperMix (Scientific Software International Inc.) and Statistica 12 (StatSoft Inc.). All other statistical analyses were undertaken using Prism4.0 (GraphPad).

ACKNOWLEDGMENTS. We thank the participants and clinicians of the Diagnostic and Prognostic Biomarkers for Parkinson's Disease (PROBE) and Harvard Biomarker Study (HBS) studies, and James P. Quinn (Q Regulating Systems) for assistance with the statistical analysis of the longitudinal data. This study was funded by the US Army Medical Research and Materiel Command under Awards W81XWH-09-0708 and W81XWH13-1-0025 (to J.A.P.).

1. Scherzer CR, et al. (2007) Molecular markers of early Parkinson's disease based on gene expression in blood. *Proc Natl Acad Sci USA* 104(3):955–960.
2. Potashkin JA, Santiago JA, Ravina BM, Watts A, Leontovich AA (2012) Biosignatures for Parkinson's disease and atypical parkinsonian disorders patients. *PLoS ONE* 7(8): e43595.
3. Santiago JA, Scherzer CR, Potashkin JA; Harvard Biomarker Study (2013) Specific splice variants are associated with Parkinson's disease. *Mov Disord* 28(12):1724–1727.
4. Ding H, et al. (2013) Unrecognized vitamin D3 deficiency is common in Parkinson disease: Harvard Biomarker Study. *Neurology* 81(17):1531–1537.
5. Chen-Plotkin AS, et al. (2011) Plasma epidermal growth factor levels predict cognitive decline in Parkinson disease. *Ann Neurol* 69(4):655–663.
6. Di Fonzo A, et al.; Italian Parkinson Genetics Network (2005) A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease. *Lancet* 365(9457): 412–415.
7. Gilks WP, et al. (2005) A common LRRK2 mutation in idiopathic Parkinson's disease. *Lancet* 365(9457):415–416.
8. Farrer M, et al. (2005) LRRK2 mutations in Parkinson disease. *Neurology* 65(5): 738–740.
9. Santiago JA, Potashkin JA (2014) A network approach to clinical intervention in neurodegenerative diseases. *Trends Mol Med* 20(12):694–703.
10. Santiago JA, Potashkin JA (2013) Integrative network analysis unveils convergent molecular pathways in Parkinson's disease and diabetes. *PLoS ONE* 8(12):e83940.
11. Santiago JA, Scherzer CR, Potashkin JA (2014) Network analysis identifies SOD2 mRNA as a potential biomarker for Parkinson's disease. *PLoS ONE* 9(10):e109042.
12. Santiago JA, Potashkin JA (2014) A network approach to diagnostic biomarkers in progressive supranuclear palsy. *Mov Disord* 29(4):550–555.
13. Xia J, et al. (2013) INMEX: A web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Res* 41(web server issue):W63–W70.
14. Xia J, Benner MJ, Hancock RE (2014) NetworkAnalyst: Integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res* 42(web server issue):W167–W174.
15. Han M, Nagele E, DeMarshall C, Acharya N, Nagele R (2012) Diagnosis of Parkinson's disease based on disease-specific autoantibody profiles in human sera. *PLoS ONE* 7(2): e32383.
16. Rhee J, et al. (2003) Regulation of hepatic fasting response by PPARgamma coactivator-1alpha (PGC-1): Requirement for hepatocyte nuclear factor 4alpha in gluconeogenesis. *Proc Natl Acad Sci USA* 100(7):4012–4017.
17. Harries LW, et al. (2008) The diabetic phenotype in HNF4A mutation carriers is moderated by the expression of HNF4A isoforms from the P1 promoter during fetal development. *Diabetes* 57(6):1745–1752.
18. Bonnefond A, Froguel P, Vaxillaire M (2010) The emerging genetics of type 2 diabetes. *Trends Mol Med* 16(9):407–416.
19. Yin L, Ma H, Ge X, Edwards PA, Zhang Y (2011) Hepatic hepatocyte nuclear factor 4α is essential for maintaining triglyceride and cholesterol homeostasis. *Arterioscler Thromb Vasc Biol* 31(2):328–336.
20. Marcil V, et al. (2010) Modification in oxidative stress, inflammation, and lipoprotein assembly in response to hepatocyte nuclear factor 4alpha knockdown in intestinal epithelial cells. *J Biol Chem* 285(52):40448–40460.
21. Konstandi M, Shah YM, Matsubara T, Gonzalez FJ (2013) Role of PPARα and HNF4α in stress-mediated alterations in lipid homeostasis. *PLoS ONE* 8(8):e70675.
22. Hu G, Jousilahti P, Bidel S, Antikainen R, Tuomilehto J (2007) Type 2 diabetes and the risk of Parkinson's disease. *Diabetes Care* 30(4):842–847.
23. Cereda E, et al. (2011) Diabetes and risk of Parkinson's disease: A systematic review and meta-analysis. *Diabetes Care* 34(12):2614–2623.
24. Cereda E, et al. (2013) Diabetes and risk of Parkinson's disease. *Mov Disord* 28(2):257.
25. Lu L, et al. (2014) Diabetes and risk of Parkinson's disease: An updated meta-analysis of case-control studies. *PLoS ONE* 9(1):e85781.
26. Santiago JA, Potashkin JA (2013) Shared dysregulated pathways lead to Parkinson's disease and diabetes. *Trends Mol Med* 19(3):176–186.
27. Santiago JA, Potashkin JA (2014) System-based approaches to decode the molecular links in Parkinson's disease and diabetes. *Neurobiol Dis* 72PA:84–91.
28. Zheng B, et al.; Global PD Gene Expression (GPEX) Consortium (2010) PGC-1α, a potential therapeutic target for early intervention in Parkinson's disease. *Sci Transl Med* 2(52):52ra73.
29. Puigserver P, et al. (2003) Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1alpha interaction. *Nature* 423(6939):550–555.
30. Scherzer CR, et al. (2008) GATA transcription factors directly regulate the Parkinson's disease-linked gene alpha-synuclein. *Proc Natl Acad Sci USA* 105(31):10907–10912.
31. Knoch KP, et al. (2006) cAMP-dependent phosphorylation of PTB1 promotes the expression of insulin secretory granule proteins in beta cells. *Cell Metab* 3(2):123–134.
32. Moroo I, et al. (1994) Loss of insulin receptor immunoreactivity from the substantia nigra pars compacta neurons in Parkinson's disease. *Acta Neuropathol* 87(4):343–348.
33. Aviles-Olmos I, Limousin P, Lees A, Foltynie T (2013) Parkinson's disease, insulin resistance and novel agents of neuroprotection. *Brain* 136(Pt 2):374–384.
34. Bassil F, Fernagut PO, Bezard E, Meissner WG (2014) Insulin, IGF-1 and GLP-1 signaling in neurodegenerative disorders: Targets for disease modification? *Prog Neurobiol* 118:1–18.
35. Li Y, et al. (2009) GLP-1 receptor stimulation preserves primary cortical and dopaminergic neurons in cellular and rodent models of stroke and Parkinsonism. *Proc Natl Acad Sci USA* 106(4):1285–1290.
36. Aviles-Olmos I, et al. (2013) Exenatide and the treatment of patients with Parkinson's disease. *J Clin Invest* 123(6):2730–2736.
37. Aviles-Olmos I, et al. (2014) Motor and cognitive advantages persist 12 months after exenatide exposure in Parkinson's disease. *J Parkinsons Dis* 4(3):337–344.
38. Foltynie T, Aviles-Olmos I (2014) Exenatide as a potential treatment for patients with Parkinson's disease: First steps into the clinic. *Alzheimers Dement* 10(1 Suppl): S38–S46.
39. Molochnikov L, et al. (2012) A molecular signature in blood identifies early Parkinson's disease. *Mol Neurodegener* 7:26.
40. Cereda E, Barichella M, Cassani E, Caccialanza R, Pezzoli G (2012) Clinical features of Parkinson disease when onset of diabetes came first: A case-control study. *Neurology* 78(19):1507–1511.
41. Dunn L, et al. (2014) Dysregulation of glucose metabolism is an early event in sporadic Parkinson's disease. *Neurobiol Aging* 35(5):1111–1115.
42. Hegele A, et al. (2012) Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell* 45(4):567–580.
43. Soreq L, Bergman H, Israel Z, Soreq H (2012) Exon arrays reveal alternative splicing aberrations in Parkinson's disease leukocytes. *Neurodegener Dis* 10(1-4):203–206.
44. Mehanna R, Scherzer CR, Ding H, Locascio JJ (2014) Unrecognized vitamin D3 deficiency is common in Parkinson disease: Harvard Biomarker study. *Neurology* 82(18): 1666, discussion 1666.
45. Zoubarev A, et al. (2012) Gemma: A resource for the reuse, sharing and meta-analysis of expression profiling data. *Bioinformatics* 28(17):2272–2273.
46. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med* 6(7):e1000097.
47. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM (2002) Meta-analysis of microarrays: Interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 62(15):4427–4433.
48. Tseng GC, Ghosh D, Feingold E (2012) Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res* 40(9):3785–3799.
49. Ding H, et al. (2011) Association of SNCA with Parkinson: Replication in the Harvard NeuroDiscovery Center Biomarker Study. *Mov Disord* 26(12):2283–2286.
50. Hedeker D, Gibbons R, du Toit M, Chen Y (2008) *SuperMix: Mixed Effects Models* (Scientific Software International, Lincolnwood, IL).