Molecular markers of early Parkinson's disease based on gene expression in blood

Clemens R. Scherzer*†‡§, Aron C. Eklund*, Lee J. Morse*, Zhixiang Liao*, Joseph J. Locascio†, Daniel Fefer*, Michael A. Schwarzschild†‡, Michael G. Schlossmacher*‡, Michael A. Hauser¶, Jeffery M. Vance¶, Lewis R. Sudarsky‡, David G. Standaert†‡, John H. Growdon†‡, Roderick V. Jensen‖, and Steven R. Gullans*

*Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, Cambridge, MA 02139; †Neurology Service, Massachusetts General Hospital, Boston, MA 02114; ‡Partners Parkinson and Movement Disorders Center, Boston, MA 02114; ¶Center for Human Genetics, Duke University Medical Center, Durham, NC 27710; and ‖Department of Physics, University of Massachusetts, Boston, MA 02125

Communicated by Gregory A. Petsko, Brandeis University, Waltham, MA, November 24, 2006 (received for review March 12, 2006)

Parkinson's disease (PD) progresses relentlessly and affects five million people worldwide. Laboratory tests for PD are critically needed for developing treatments designed to slow or prevent progression of the disease. We performed a transcriptome-wide scan in 105 individuals to interrogate the molecular processes perturbed in cellular blood of patients with early-stage PD. The molecular multigene marker here identified is associated with risk of PD in 66 samples of the training set comprising healthy and disease controls [third tertile cross-validated odds ratio of 5.7 (*P* for trend 0.005)]. It is further validated in 39 independent test samples [third tertile odds ratio of 5.1 (*P* for trend 0.04)]. Insights into disease-linked processes detectable in peripheral blood are offered by 22 unique genes differentially expressed in patients with PD versus healthy individuals. These include the cochaperone *ST13*, which stabilizes heat-shock protein 70, a modifier of α-synuclein misfolding and toxicity. *ST13* messenger RNA copies are lower in patients with PD (mean ± SE 0.59 ± 0.05) than in controls (0.96 ± 0.09) (*P* = 0.002) in two independent populations. Thus, gene expression signals measured in blood can facilitate the development of biomarkers for PD.

risk markers | biomarkers | heat shock protein 70-interacting protein (ST13) | microarray

Many fundamental decisions in medical practice are based on laboratory risk markers (1). For example, cholesterol levels, although not diagnostic, are a standard biomarker that correlate with risk for coronary heart disease and trigger interventions to prevent heart attacks (2).

Parkinson's disease (PD) is a slowly progressive disease. Tremor and slow movements develop only after ≈70% of vulnerable dopaminergic neurons in the substantia nigra have already died (3). No laboratory test that correlates with PD risk is available. This curtails our ability to test disease-modifying drugs and other neuroprotective strategies (4). Genes mutated in rare monogenic variants of PD (5) are transcribed in blood cells (6, 7). In *PARK9*, a donor splice site mutation in the predominantly neuronal gene *ATP13A2*, leads to skipping of exon 13 during RNA processing (8). The aberrant mRNA isoform can be detected in whole blood (8). Furthermore, studies in platelets (9) or lymphocytes (10–13) of patients with sporadic PD have detected subtle abnormalities in dopamine biosynthesis/signaling and mitochondrial function, two hallmarks of PD. The etiology of these changes remains unexplained and might reflect disease mechanisms, allele-specific gene expression, or reactive responses (14, 15).

Instead of testing one gene at a time, microarrays can interrogate in parallel expression levels of thousands of genes in tissues from patients with PD. This allows to rapidly scan for candidate genes and candidate biomarkers (15). In the substantia nigra of individuals affected by PD and in PD models, we (16, 17) and others (18, 19) found profound expression changes of genes involved in cellular quality control and energy metabolism. For a biomarker to be clinically useful, noninvasive detection in peripheral blood is desirable. The power of this approach for identifying disease-associated candidates in a neurodegenerative disease is precedented by an expression screen of lymphoblasts of patients with Alzheimer's disease (AD), in which we detected abnormally low signals of the neuronal sorting receptor LR11/SorLa (20). Subsequent analyses of LR11/SorLa's mechanistic role in AD demonstrated that LR11/sorLA modulates APP trafficking and its processing to amyloid-β (21, 22).

Here, first, a risk marker for PD was built based on a systematic scan of genome-wide expression changes in blood of individuals with early PD. Second, this screen identified genes differentially transcribed in PD patients versus normal controls.

Results

Identification of a Molecular Signature of PD in Blood. To identify a transcriptional profile associated with PD we probed RNA extracted from whole blood of 50 PD patients predominantly at early disease stages (mean Hoehn and Yahr stage 2.3, range 1–4) [supporting information (SI) Table 2] and 55 age-matched controls with 22,283 oligonucleotide probe sets on microarrays. The disease controls included patients with AD that may be misclassified as PD (23), as well as with progressive supranuclear palsy, multiple system atrophy, and corticobasal degeneration, which closely mimic the clinical features of PD (23) but differ in etiology, prognosis, and treatment response. We carefully assayed for shifts in differential blood counts that could bias gene expression changes and found no significant difference between PD and controls (SI Table 2).

A subset of the patient samples was randomly chosen to build the risk marker (*SI Methods*). This "training set" included ≈60% of the subjects (66/105), including 31 PD patients, 17 healthy subjects, and 18 disease controls with AD or progressive su-

Author contributions: C.R.S. designed research; C.R.S., L.J.M., Z.L., D.F., M.A.S., M.G.S., M.A.H., J.M.V., L.R.S., D.G.S., and J.H.G. performed research; C.R.S., A.C.E., Z.L., J.J.L., and R.V.J. analyzed data; and C.R.S., A.C.E., R.V.J., and S.R.G. wrote the paper.

Conflict of interest statement: C.R.S., S.R.G., and R.V.J. are listed as coinventors on a U.S. Letters Patent application for identification of dysregulated genes in patients with neurologic diseases held by Brigham and Women's Hospital. C.R.S. is a consultant to Link Medicine Corporation.

Freely available online through the PNAS open access option.

Abbreviations: PD, Parkinson's disease; AD, Alzheimer's disease; C.I., confidence interval; ROC curve, receiver-operating-characteristics curve; FDR, false discovery rate; CT, threshold cycle.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE6613).

§To whom correspondence should be addressed at: Laboratory for Functional Neurogenomics, Center for Neurologic Diseases, Harvard Medical School and Brigham and Women's Hospital, 65 Landsdowne Street, Suite 301A, Cambridge, MA 02139. E-mail: cscherzer@rics.bwh.harvard.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0610204104/DC1.

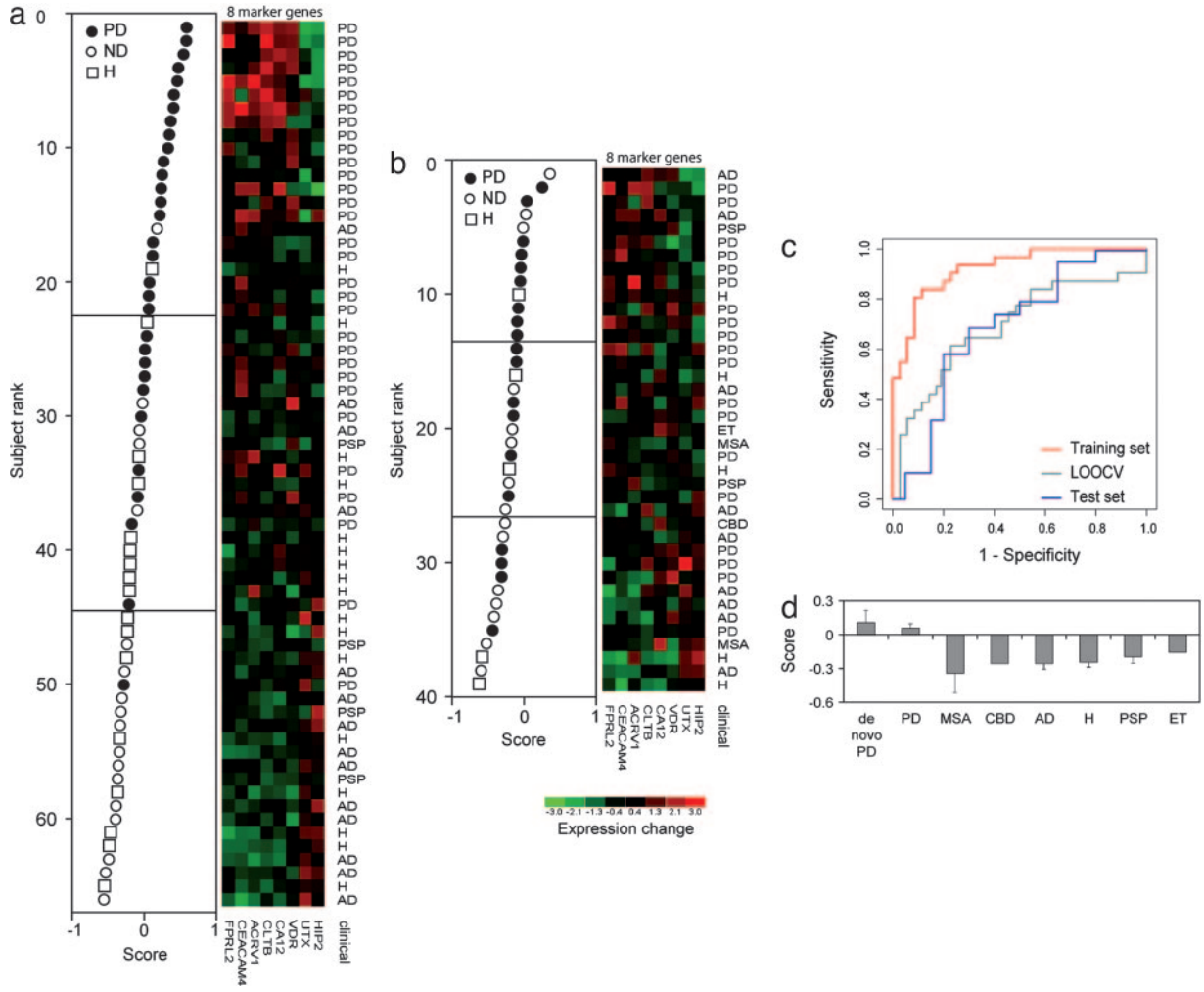

Fig. 1. Molecular marker associated with PD risk. (*a*) Expression data matrix of eight marker genes of 66 blood samples from PD and control subjects. Each row represents a blood sample, and each column represents a gene. As shown in the color bar, overexpression is displayed in red, and underexpression is displayed in green. Blood samples are ordered by their risk score (shown on the left), which is defined as the correlation with the average profile of the PD group minus the correlation with the average profile of the controls. Twenty of 22 individuals with high risk scores ranked in the top third of the list have PD (third tertile). Twenty-one of 22 individuals with low risk scores ranked in the bottom third of the list are controls (first tertile). Solid lines designate tertiles of risk score values. The clinical diagnosis for each individual is shown on the right. (*b*) Validation of the risk marker on independent test samples confirms that high scores are significantly associated with increased PD risk (P for trend = 0.04). The expression data matrix is as in *a*, except that *b* uses 39 independent samples. (*c*) The ROC curve in the test set (blue curve) is highly consistent with the ROC curve for the leave-one-out cross-validated (LOOCV) marker in the training set (gray) confirming the risk prediction observed for different cutoffs. The nominal ROC curve in the training set represents an upper limit (red). (*d*) Dopamine replacement medication does not bias the risk score. There is no difference in risk scores of PD patients on dopamine medication versus unmedicated *de novo* patients (mean ± SE 0.06 ± 0.04 and 0.11 ± 0.1, respectively; $P = 0.96$). The average risk score for the overall PD group is 0.07 ± 0.03 (data not shown). Average risk scores are low (negative) in healthy controls (−0.24 ± 0.04) and neurodegenerative disease controls [AD, −0.25 ± 0.05; progressive supranuclear palsy, −0.19 ± 0.06; multiple system atrophy (MSA), −0.34 ± 0.17; corticobasal degeneration (CBD), −0.26; essential tremor (ET), −0.15]. Individual scores range from −0.43 to 0.6 for PD patients, −0.62 to 0.12 for healthy controls, and −0.59 to 0.37 for neurodegenerative disease controls. ND, neurodegenerative disease control; H, healthy control. Error bars indicate standard errors.

pranuclear palsy. We used a powerful four-step supervised prediction method, similar to those used previously (24), to build a molecular marker for PD. The genes were rank-ordered according to the absolute value of their correlation coefficient with PD. The optimal number of genes for the marker was then determined by sequentially adding genes from the top of this rank-ordered list and evaluating its power for correct prediction accuracy based on the rank sum statistic (see *SI Methods*). The maximum rank sum was reached with an optimal number of eight marker genes (SI Fig. 3). These marker genes were significantly correlated with PD (SI Fig. 4). We then calculated the average expression value of the eight marker genes within the PD samples and the non-PD samples, forming a PD template and a non-PD template. For each sample a PD risk score was calculated, which was defined as the correlation with the PD templates minus the correlation with the non-PD template.

High Marker Scores Indicate PD. Twenty of 22 individuals with high risk scores had PD. These were ranked in the top third of the rank-ordered list (third tertile, Fig. 1*a*). Twenty-one of 22 individuals with low risk scores ranked in the bottom third of the list (first tertile) were controls. The odds for PD in individuals in the third tertile (high score) and the second tertile (intermediate scores) were compared with the odds for PD in individuals in the first tertile (low scores). The nominal odds ratio for PD of subjects in the third tertile was 210 [95% confidence interval (C.I.) 18–2,500] and in the second tertile was 18 (95% C.I. 2.0–150) with a P for trend of <0.0001. For novel risk markers

Table 1. Odds ratio for PD according to tertile of risk score (95% C.I.)

Variable	*n*	Risk score tertile 1	2	3	*P* for trend
Training set*	31 (PD) and 35 (controls)	1.0	2.2 (0.6–7.8)	5.7 (1.6–7.8)	0.005
Test set	19 (PD) and 20 (controls)	1.0	1.9 (0.4–9.6)	5.1 (1–27)	0.04

In each analysis the first tertile served as the reference group.
*Cross-validated odds ratios are shown.

the "normal value" that best discriminates high-risk from low-risk individuals is unknown. We therefore determined the receiver-operating-characteristics curve (ROC curve) of our marker and found high sensitivity and specificity at various cutoffs (Fig. 1*c*, red line). These estimates are based on the same series of patients that the marker was derived from and therefore represent an upper limit.

Confirmation by Leave-One-Out Cross-Validation. To statistically validate the predictive value of the risk score on future samples a performance leave-one-out cross-validation procedure was performed in which the left-out sample was not involved in selecting the marker genes (*SI Methods*). The cross-validated odds ratios for PD were 5.7 (95% C.I. 1.6–21) and 2.2 (95% C.I. 0.6–7.8) for persons with scores in the third and second tertile, respectively. This confirmed that high scores correlate with PD risk (P for trend = 0.005) (Table 1).

Validation in an Independent Test Set. The most rigorous test for the significance and predictive value of a risk marker is validation on independent test samples. We applied the microarray-based risk score to a test set of 39 samples, including 19 PD patients, five healthy individuals, and 15 disease controls with movement or memory disorders (Fig. 1*b*). There was a significant difference of scores in patients with PD versus healthy and disease controls (P = 0.047 by Wilcoxon test). High scores were significantly associated with increased PD risk (P for trend = 0.04). Individuals with scores in the third tertile (high score) had an odds ratio for PD of 5.1 (95% C.I. 1–27) (Table 1), and individuals with a score in the second tertile (intermediate score) had an odds ratio of 1.9 (95% C.I. 0.4–9.6). The ROC curve in the validation set (Fig. 1*c*, blue line) was highly consistent with the ROC curve for the cross-validated marker in the training set, thus confirming the risk prediction observed for different cutoffs. The average Hoehn and Yahr stage (2.1 ± 0.7) and disease duration (4.7 ± 3.6 years) in the validation set were lower (indicating an earlier disease state) than in the training set (2.4 ± 0.7 and 7.1 ± 4.9, respectively), but this did not reach statistical significance.

Overall, the risk marker predicted PD beyond the prediction afforded by the traditional risk factors of age and sex ($P < 0.0001$ by Wald χ^2 test) and was not biased by dopamine replacement therapy (Fig. 1*d* and *SI Text*).

Functionally the eight marker genes (*VDR*, *HIP2*, *CLTB*, *FPRL2*, *CA12*, *CEACAM4*, *ACRV1*, and *UTX*) do not appear to represent a single biologic pathway or process, although all are known to be expressed in the human brain (25, 26). The products of two of the eight genes, namely vitamin D receptor gene (*VDR*) and huntingtin interacting protein 2 (*HIP2*), are of particular interest to PD pathobiology (26–29). A polymorphism in the *VDR* gene is overrepresented in PD patients in Korea (29), and in rats the ligand vitamin D ameliorates 6-hydroxydopamine-induced toxicity (27). *HIP2* encodes a neuronal ubiquitin-conjugating enzyme involved in the ubiquitinylation of huntingtin, mutated in the neurodegenerative disorder Huntington disease (26). Correspondingly, the ubiquitin–proteasome system is linked to monogenic forms of PD (5). *CLTB* is implicated in dopamine transporter endocytosis (28). The remaining five genes have no known role in PD pathogenesis and include *FPRL2*, related to the G protein-coupled receptor 1 family, *CA12*, or carbonic anhydrase XII; *CEACAM4*, a cell adhesion molecule; *ACRV1*, initially described in acrosomal vesicles; and *UTX*, an X-linked gene of unknown function.

Heat-Shock Protein 70-Interacting Protein *ST13* and Biologic Insights into PD-Related Changes in Blood Cells. The molecular marker was designed to detect clinically useful gene expression changes that specifically correlate with PD. Discovery of genes dysregulated in PD compared with healthy controls, however, regardless of their expression changes in other neurodegenerative diseases, might provide detailed biologic insights into the molecular pathology underlying PD-related changes in blood cells. Significance analysis of microarrays (30) identified 22 unique genes that were most differentially expressed in 31 PD patients (including five *de novo* PD patients) compared with healthy controls at a false discovery rate (FDR) of 0.03 (Fig. 2).

One of the 22 genes most underexpressed in PD was the heat-shock protein 70-interacting protein *ST13* in the ubiquitin–proteasome pathway. *ST13* is a cofactor of heat-shock protein 70 (*HSP70*) and stabilizes its chaperone activity (31). This may be particularly relevant to PD pathobiology. *HSP70* modulates *SNCA* folding (32) and suppresses *SNCA* toxicity in cells (32), yeast (33), and flies (34). To confirm the reduction in mRNA levels for *ST13* measured by microarray (0.67- and 0.73-fold), we performed quantitative real-time PCR assays based on precise fluorogenic 5′ nuclease chemistry in a large subset of 39 PD patients and 12 age-, sex-, and blood count-matched healthy controls. We used the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase to control for input RNA and the comparative threshold cycle (CT) method (35) for analysis (Fig. 2*b*). In PD patients, *ST13* amplified at higher ΔCTs than in controls, indicating a lower abundance of *ST13* mRNA in PD (fold change = 0.78, P = 0.025 by t test).

Then we enrolled a new and independent population of 17 patients with PD and 17 age- and sex-matched healthy controls from the Partners Parkinson and Movement Disorders Center clinic population, extracted RNA from 7.5 ml of venous blood, and quantified the relative amount of *ST13* mRNA copies compared with 18S ribosomal RNA copies by a standard curve method (35). The log-transformed mean ratio of *ST13* mRNA copies to 18S ribosomal RNA copies was significantly lower in blood from patients with PD (mean ± SE 0.59 ± 0.05) than in the control group (0.96 ± 0.09) (fold change = 0.6, P = 0.002 by t test) (Fig. 2*c*). Thus, measurement of *ST13* mRNA may offer a robust, noninvasive means of identifying PD-related changes in blood.

Apoptosis-related genes such as *BCL11B* were also underexpressed in PD blood cells. The changes in mRNA levels of *BCL11B* were further confirmed by real-time PCR in a large subset of 33 individuals with PD and 12 age-, sex-, and blood count-matched healthy volunteers (P = 0.005) (Fig. 2*d*). Mitochondrial dysfunction in PD can be detected in the substantia nigra, as well as in peripheral tissues. It was remarkable to us that the nuclear encoded mitochondrial gene *LRPPRC* was among the genes underexpressed in PD. A mutation in this gene causes

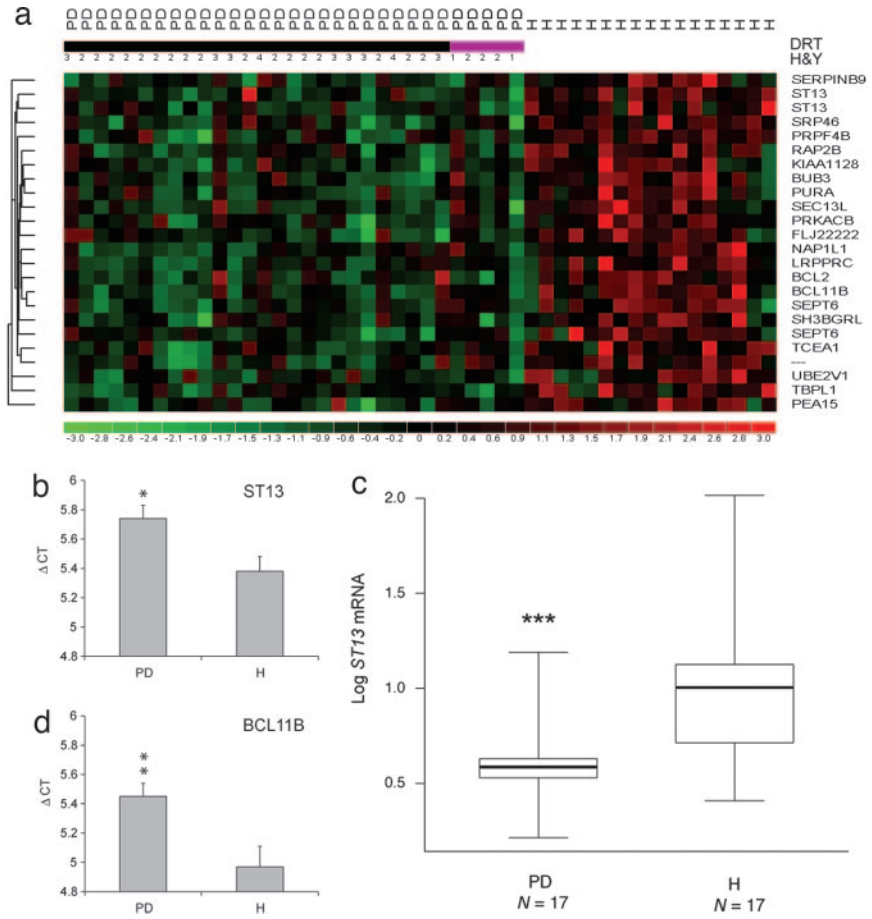

Fig. 2. Heat-shock protein 70-interacting protein *ST13* and biologic insights into PD-related changes in blood cells. Underexpression of *ST13* in PD is confirmed by three distinct assays in two independent populations (*a–c*). (*a*) Twenty-four probe sets, including two probing for *ST13*, are significantly underexpressed in cellular blood of 31 PD patients (including five *de novo* PD patients) compared with healthy controls (FDR = 0.03). Untreated *de novo* patients are identified by a pink bar, and patients treated with dopamine replacement are identified by a black bar. Expression matrices are as in Fig. 1 except that genes here are shown on the vertical axis. The majority of PD patients are at early stages of the disease process as indicated by the Hoehn and Yahr (H&Y) scale. Dendrograms illustrate genes with similar expression patterns by cluster analysis. (*b*) Underexpression of *ST13* in PD is first validated by real-time PCR in a large subset of 39 PD patients and 12 age-, sex-, and blood-count-matched healthy controls by using the comparative CT method. ΔCT values are displayed. In PD patients, *ST13* amplifies at higher ΔCTs than in controls, indicating a lower abundance of *ST13* mRNA in PD (fold change = 0.78, $P = 0.025$ by *t* test). GAPD mRNA levels are used to control for input RNA. (*c*) The log-transformed mean ratio of *ST13* mRNA copies to 18S ribosomal RNA copies is lower in patients with PD (mean ± SE 0.59 ± 0.05) than in the control group (0.96 ± 0.09) (fold change = 0.6, $P = 0.002$ by *t* test) in a second, independent population by a standard curve method. The box plot shows the median (bold line) and the 75th and 25th percentile values (top and bottom of the box) for log-transformed ratios of *ST13* mRNA copies to 18S rRNA copies in blood samples of 17 patients with PD and 17 age- and sex-matched healthy control subjects. The top and bottom of the whiskers show the maximum and minimum values. (*d*) Underexpression of the apoptosis-related gene *BCL11B* was also confirmed by the comparative CT method ($P = 0.005$).

French-Canadian-type Leigh syndrome, an early-onset progressive mitochondrial neurodegenerative disorder caused by defects in oxidative phosphorylation (36). It remains to be seen whether partial reduction in *LRPPRC* expression can contribute to mitochondrial dysfunction.

Additional biological processes represented by the 22 PD-associated genes were DNA and RNA turnover, cell cycle, carbohydrate metabolism, protein metabolism, signaling, phosphorylation, and protein transport (Fig. 2*a* and SI Fig. 5). By comparison, up to 28% of the genes most differentially transcribed in the 23 disease controls with AD (FDR = 0.04) were involved in inflammatory processes that were not perturbed in PD.

As expected in this comparison limited to healthy controls, all eight probe sets comprising the molecular multigene marker were ranked among the highly differentially expressed genes (top 305 genes by significance analysis of microarrays with FDR < 0.2; top 228 genes by *t* test with $P < 0.005$) of all 22,283 genes assayed (data not shown).

Discussion

No laboratory blood test for PD is available. Diagnosis and outcome measures rest on the physician's physical examination and clinical history. Identification of the earliest stages of PD is most unreliable (37), and detection of individuals at risk for developing PD is currently impossible. Dopamine neurons in the substantia nigra degenerate undetected for years before clinical symptoms develop (3). After 70% of vulnerable dopaminergic neurons have been lost, tremor and slow movements develop (3) and eventually PD can be diagnosed based on clinical criteria. This opens a window of opportunity for interventions designed to prevent PD or to delay the onset of the clinical disease. If a reliable risk predictor of PD can be developed, this will allow clinical trials of drugs designed to prevent PD or to delay the clinical disease onset.

Here we show that combining gene expression scans in cellular blood and linked clinical data can rapidly characterize candidate laboratory biomarkers of PD risk. Risk markers for complex

diseases such as PD are not simply present or absent (1). Rather, they have a wide range of values that overlap in persons with a disease and in those without it (1). The risk typically increases progressively with increasing levels (1). By contrast, diagnostic tests are binary classifiers designed to diagnose the presence or absence of a disease state with highest specificity and sensitivity. PD is a complex, etiologically, genetically, pathologically, and clinically heterogeneous neurodegenerative disease (5, 15, 38). Further adding complexity to the disease classification, considerable overlap is seen on autopsy between patients clinically diagnosed with PD and those clinically diagnosed with other neurodegenerative diseases, particularly AD (23). Thus, we developed a simple continuous score to identify individuals with higher or lower risk of PD that can account for the complex biology of the disease. This was derived from a supervised classification method similar to those used previously (24).

The molecular marker identified was strongly associated with PD in this cross-sectional study of 105 subjects. Individuals with the highest marker scores had a >5-fold increased risk of PD. The molecular marker was rigorously confirmed by leave-one-out cross-validation and in an independent validation set of 39 healthy as well as disease controls. Individuals with atypical parkinsonism that are frequently misdiagnosed as PD by the nonspecialist (23) had low scores. Notably, dopamine replacement therapy did not bias the risk scores.

This study was carefully designed to reduce the dual threats to molecular-marker discovery of bias and chance in accordance with published recommendations (39, 40). All PD patients met diagnostic criteria of the United Kingdom Parkinson's Disease Society Brain Bank. Controls were recruited predominantly from nonaffected spouses of the PD cases, thus adjusting for environmental biases. Blood specimens of cases and controls were collected, processed, and assayed in parallel. This contrasts with studies relying on banked convenience samples. Demographic and hematologic variables did not influence the molecular marker. Importantly, both the molecular marker assayed by eight oligonucleotide probe sets and *ST13* mRNA copy numbers assayed by quantitative real-time PCR were confirmed in independent validation studies (39, 40).

The implication of the leads here discovered for the clinical practice of medicine need to be interpreted with caution. The reliability and precision of microarray platforms in the context of personalized medicine was recently confirmed by a consortium led by the Food and Drug Administration (41). Novel biomarkers, however, must be further validated in analogy to new drugs in prospective, well controlled clinical studies of diverse patients across multiple institutions, with well established standards for all steps in the process (40). Furthermore, the list of genes identified as disease signature can vary depending on the composition of the training set (42). Distinct but equally relevant disease control populations and distinct clinical applications of the candidate molecular markers may require adaptations in the selection of discriminating probes.

In PD patients several genes differentially expressed were involved in the pathobiologically and therapeutically relevant processes of the ubiquitin–proteasome pathway, mitochondrial function, and apoptosis (5). Transcript levels of the *HSP70*-interacting gene *ST13* were low in PD and may serve as an indicator of *HSP70* function. *HSP70* suppresses *SNCA* toxicity in cellular (32), yeast (33), and fly models of PD (34). It is thought to exert its salutary effect by modulating *SNCA* folding (32). *ST13* stabilizes the ADP state of *HSP70*, which has a high affinity for substrate protein (31, 43). Through its own chaperone activity, *ST13* may contribute to the interaction of *HSP70* with various target proteins (43). In the substantia nigra of patients with PD, expression of different *HSP70* members is highly perturbed (17, 44). Furthermore, *CHIP*, a *HSP70* cofactor structurally related to *ST13*, influences *SNCA* degradation decisions between lysosomal and proteasomal pathways (45). This may suggest that gene expression profiling of blood from patients with PD captures molecular surrogates of the disease process, but more work is needed.

Materials and Methods

Clinical Study. We enrolled 50 consecutive PD patients who were diagnosed by neurology-board-certified movement disorders specialists and that met modified United Kingdom Parkinson's Disease Society Brain Bank (23) clinical diagnostic criteria and 55 age-matched healthy and neurodegenerative disease controls (SI Table 2) from the Partners Parkinson and Movement Disorders Center and the Memory Disorders Unit at Massachusetts General Hospital. To keep the number of individuals with a false positive clinical diagnosis of PD at a minimum, the United Kingdom Parkinson's Disease Society Brain Bank criteria were modified to require the presence of three (instead of only two) cardinal features (bradykinesia, and two of rigidity, rest tremor, or postural instability), none of 16 exclusion criteria, and at least three of eight supportive features (23).

Healthy controls had no personal or family history of neurodegenerative diseases. Ninety percent of the healthy controls were spouses. Because the prevalence of PD is ≈1% at age >65, we applied a previously validated PD screening questionnaire (46) to further reduce the chance of enrolling controls with undiagnosed PD. Exclusion criteria for all study subjects were age <21 years, hematologic malignancies or coagulopathies, known severe anemia (hematocrit < 30), and known pregnancy. The study protocol was approved by the Institutional Review Board of Brigham and Women's Hospital.

RNA Isolation and Quality Control. Venous blood was collected in PAXgene (Qiagen, Valencia, CA) tubes and immediately incubated at room temperature for 24 h. RNA was then extracted after the PAXgene procedure including DNase treatment. RNA quality was determined by spectrophotometry and by using the RNA 6000 NanoChip kit on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). RNA passing quality control criteria were used for further analysis.

Microarray Procedures. A modified CodeLink Bioarray protocol was used (47). Briefly, 4 μg of total RNA was used for cDNA synthesis, cleaned by a QIAquick purification kit (Qiagen), and *in vitro* transcribed by T7 MEGAscript reagents (Ambion, Austin, TX) using biotin-11-UTP (PerkinElmer/Wellesley, MA). Biotin-labeled cRNA was purified by RNeasy Mini kit (Qiagen) and hybridized at 45°C for 16–18 h to Human Genome U133A arrays (Affymetrix, Santa Clara, CA). The arrays were washed, stained, and scanned on an HP Gene Array scanner (Affymetrix). Visual inspection was performed to identify arrays with production defects.

Real-Time PCR. Briefly, for genes of interest, TaqMan Assay-on-demand and custom assay primers and probes (Applied Biosystems, Foster City, CA) were designed by using the manufacturer's "rules" including crossing exon junctions (primer/probe information is available upon request). Amplification products were analyzed for specificity by agarose gel electrophoresis. The comparative CT method and the relative standard curve method were used for analysis (35). GAPD and 18S ribosomal RNA controlled for RNA loading. Equal amplification efficiencies were confirmed for target and reference genes as appropriate. Total RNA (3 μg) was reverse-transcribed into cDNA by using TaqMan Reverse Transcription reagents and random hexamers. Real-time PCR using an ABI Prism 7000 (Applied Biosystems) and TaqMan kits was performed according to the manufacturers' protocols. Samples were loaded in duplicate or triplicate, and no-template and plate-to-plate controls were run on every

reaction plate. In the independent validation study, *ST13* mRNA copy numbers were normalized with the use of 18S rRNA copy numbers (the number of *ST13* mRNA copies in 90 ng of RNA divided by the number of 18S rRNA copies in 90 pg of RNA $\times$ 10). The ratios of *ST13* mRNA copies to 18S rRNA copies were log-transformed to reduce the skew.

Microarray Data Processing. This was previously described (15, 16). Briefly, raw CEL files were processed with the MAS5 algorithm performing global scaling with "target intensity" of 100 for all probe sets. Only high-quality arrays with GAPD 3′-to-5′ prime ratios $\leq$3 and present calls $\geq$4,000 were included in the analysis to further reduce noise due to partial RNA degradation or hybridization outliers. Because technical variation is high for genes with low average expression intensities only genes with intensities of $\geq$100 in at least one sample were considered for further analysis.

Supervised Prediction Method. The PD risk score is based on ref. 24. Genes in the training set are ranked by their absolute Pearson correlation with the binary class labels, and the top (predicting) genes are used in the marker. A template for each class is formed from the mean values of the discriminating genes. The risk score of a test case is defined as its Pearson correlation with the PD template minus its Pearson correlation with the non-PD template. The number of discriminating genes is picked by maximizing the rank sum of the PD scores in a leave-one-out cross-validation step.

Single Gene Significance Analysis. Stringent significance thresholds were set to control for false positive results due to biological and technical noise, and to correct for multiple testing. Requiring fold changes of at least 1.25 eliminated small changes in expression. To estimate the FDR, permutation analysis [significance analysis of microarrays (30)] was applied and results were visualized as described (15, 16).

Statistical Analysis. We performed multivariate analysis of covariance of relevant biological dependent variables. The third and second tertile odds ratio was the ratio of the odds in favor of PD for a subject with a risk score in the third or second tertile of score values, to the odds in favor of PD for a patient with a score in the first (lowest) tertile. The Cochran–Armitage linear trend test for proportions was applied to determine the *P* for trend.

We are grateful for our patients' support. We thank Drs. Dennis Selkoe, Peter Lansbury, Alberto Ascherio, Mel Feany, Albert Hung, and Bin Zheng for insightful comments. This study was funded by the Michael J. Fox Foundation (C.R.S. and S.R.G.). C.R.S. is funded by Beeson Grant K08AG024816 from the National Institute on Aging and the American Federation for Aging Research and a Dr. George C. Cotzias Memorial Fellowship from the American Parkinson Disease Association. The Partners Parkinson and Movement Disorders Center receives support from the Massachusetts General Hospital/Massachusetts Institute of Technology Morris K. Udall Parkinson's Disease Research Center of Excellence (Grant NS038372).

1. Manolio T (2003) *N Engl J Med* 349:1587–1589.
2. National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) (2002) *Circulation* 106:3143–3421.
3. Fearnley JM, Lees AJ (1991) *Brain* 114:2283–2301.
4. Lansbury PT, Jr (2004) *Nat Med* 10(Suppl):S51–S57.
5. Scherzer CR, Feany MB (2004) *Trends Genet* 20:273–277.
6. Miller DW, Hague SM, Clarimon J, Baptista M, Gwinn-Hardy K, Cookson MR, Singleton AB (2004) *Neurology* 62:1835–1838.
7. Sunada Y, Saito F, Matsumura K, Shimizu T (1998) *Neurosci Lett* 254:180–182.
8. Ramirez A, Heimbach A, Grundemann J, Stiller B, Hampshire D, Cid LP, Goebel I, Mubaidin AF, Wriekat AL, Roeper J, *et al*. (2006) *Nat Genet* 38:1184–1191.
9. Schulz JB, Beal MF (1994) *Curr Opin Neurol* 7:333–339.
10. Barbanti P, Fabbrini G, Ricci A, Cerbo R, Bronzetti E, Caronti B, Calderaro C, Felici L, Stocchi F, Meco G, *et al*. (1999) *Mov Disord* 14:764–771.
11. Petrozzi L, Lucetti C, Gambaccini G, Bernardini S, Del Dotto P, Migliore L, Scarpato R, Bonuccelli U (2001) *Neurol Sci* 22:83–84.
12. Caronti B, Antonini G, Calderaro C, Ruggieri S, Palladini G, Pontieri FE, Colosimo C (2001) *J Neural Transm* 108:803–807.
13. Nagai Y, Ueno S, Saeki Y, Soga F, Hirano M, Yanagihara T (1996) *Neurology* 46:791–795.
14. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, *et al*. (2005) *Nat Genet* 37:710–717.
15. Scherzer CR, Jensen RV, Gullans SR (2004) in *Principles of Molecular Neurosurgery*, eds Freese A, Simeone FA, Leone P, Janson C (Karger, Basel), Vol 18, pp 246–257.
16. Scherzer CR, Jensen RV, Gullans SR, Feany MB (2003) *Hum Mol Genet* 12:2457–2466.
17. Hauser MA, Li YJ, Xu H, Noureddine MA, Shao YS, Gullans SR, Scherzer CR, Jensen RV, McLaurin AC, Gibson JR, *et al.* (2005) *Arch Neurol* 62:917–921.
18. Duke DC, Moran LB, Kalaitzakis ME, Deprez M, Dexter DT, Pearce RK, Graeber MB (2006) *Neurogenetics* 7:139–148.
19. Mandel S, Grunblatt E, Riederer P, Amariglio N, Jacob-Hirsch J, Rechavi G, Youdim MB (2005) *Ann NY Acad Sci* 1053:356–375.
20. Scherzer CR, Offe K, Gearing M, Rees HD, Fang G, Heilman CJ, Schaller C, Bujo H, Levey AI, Lah JJ (2004) *Arch Neurol* 61:1200–1205.
21. Offe K, Dodson SE, Shoemaker JT, Fritz JJ, Gearing M, Levey AI, Lah JJ (2006) *J Neurosci* 26:1596–1603.
22. Andersen OM, Reiche J, Schmidt V, Gotthardt M, Spoelgen R, Behlke J, von Arnim CA, Breiderhoff T, Jansen P, Wu X, *et al*. (2005) *Proc Natl Acad Sci USA* 102:13461–13466.
23. Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) *J Neurol Neurosurg Psychiatry* 55:181–184.
24. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, *et al*. (2002) *Nature* 415:530–536.
25. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D (1997) *Gene Cards: Encyclopedia for Genes, Proteins and Diseases* (Weizmann Inst Sci, Rehovot, Israel), available at www.genecards.org.
26. Kalchman MA, Graham RK, Xia G, Koide HB, Hodgson JG, Graham KC, Goldberg YP, Gietz RD, Pickart CM, Hayden MR (1996) *J Biol Chem* 271:19385–19394.
27. Wang JY, Wu JN, Cherng TL, Hoffer BJ, Chen HH, Borlongan CV, Wang Y (2001) *Brain Res* 904:67–75.
28. Sorkina T, Hoover BR, Zahniser NR, Sorkin A (2005) *Traffic* 6:157–170.
29. Kim JS, Kim YI, Song C, Yoon I, Park JW, Choi YB, Kim HT, Lee KS (2005) *J Korean Med Sci* 20:495–498.
30. Tusher VG, Tibshirani R, Chu G (2001) *Proc Natl Acad Sci USA* 98:5116–5121.
31. Nollen EA, Kabakov AE, Brunsting JF, Kanon B, Hohfeld J, Kampinga HH (2001) *J Biol Chem* 276:4677–4682.
32. Klucken J, Shin Y, Masliah E, Hyman BT, McLean PJ (2004) *J Biol Chem* 279:25497–25502.
33. Flower TR, Chesnokova LS, Froelich CA, Dixon C, Witt SN (2005) *J Mol Biol* 351:1081–1100.
34. Auluck PK, Chan HY, Trojanowski JQ, Lee VM, Bonini NM (2002) *Science* 295:865–868.
35. Applied Biosystems (2001) ABI Prism 7700 Sequence Detection System (Applied Biosystems, Foster City, CA), User Bulletin No 2.
36. Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, Delmonte T, Villeneuve A, Sladek R, Xu F, *et al*. (2003) *Proc Natl Acad Sci USA* 100:605–610.
37. Hughes AJ, Ben-Shlomo Y, Daniel SE, Lees AJ (2001) *Neurology* 57:S34–S38.
38. Forman MS, Lee VM, Trojanowski JQ (2005) *Neuron* 47:479–482.
39. Ransohoff DF (2004) *Nat Rev Cancer* 4:309–314.
40. Ludwig JA, Weinstein JN (2005) *Nat Rev Cancer* 5:845–856.
41. Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, *et al*. (2006) *Nat Biotechnol* 24:1151–1161.
42. Michiels S, Koscielny S, Hill C (2005) *Lancet* 365:488–492.
43. Hohfeld J, Minami Y, Hartl FU (1995) *Cell* 83:589–598.
44. Grunblatt E, Mandel S, Maor G, Youdim MB (2001) *J Neurochem* 78:1–12.
45. Shin Y, Klucken J, Patterson C, Hyman BT, McLean PJ (2005) *J Biol Chem* 280:23727–23734.
46. Rocca WA, Maraganore DM, McDonnell SK, Schaid DJ (1998) *J Clin Epidemiol* 51:517–523.
47. Amersham Biosciences (2003) *CodeLink Expression Bioarrays Automated Target Preparation* (Amersham Biosciences, Piscataway, NJ).