Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease

F. Borovecki*†‡, L. Lovrecic*‡, J. Zhou*‡, H. Jeong*, F. Then*, H. D. Rosas*, S. M. Hersch*, P. Hogarth§, B. Bouzou¶, R. V. Jensen‖**, and D. Krainc*††

*Department of Neurology, Massachusetts General Hospital, Harvard Medical School, ¶Center for Interdisciplinary Informatics, MassGeneral Institute for Neurodegenerative Disease, Charlestown, MA 02129; †Center for Functional Genomics, Zagreb Medical School, and Clinical Hospital Center, 10000 Zagreb, Croatia; §Department of Neurology, Oregon Health and Science University, Portland, OR 97239; ‖Laboratory for Functional Genomics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02128; and **Department of Physics, University of Massachusetts, Boston, MA 02125

Communicated by Joseph E. Murray, Harvard Medical School, Boston, MA, June 14, 2005 (received for review May 18, 2005)

Huntington's disease (HD) is an autosomal dominant disorder caused by an expansion of glutamine repeats in ubiquitously distributed huntingtin protein. Recent studies have shown that mutant huntingtin interferes with the function of widely expressed transcription factors, suggesting that gene expression may be altered in a variety of tissues in HD, including peripheral blood. Affymetrix and Amersham Biosciences oligonucleotide microarrays were used to analyze global gene expression in blood samples of HD patients and matched controls. We identified 322 mRNAs that showed significantly altered expression in HD blood samples, compared with controls ($P < 0.0005$), on two different microarray platforms. A subset of up-regulated mRNAs selected from this group was able to distinguish controls, presymptomatic individuals carrying the HD mutation, and symptomatic HD patients. In addition, early presymptomatic subjects showed gene expression profiles similar to those of controls, whereas late presymptomatic subjects showed altered expression that resembled that of symptomatic HD patients. These elevated mRNAs were significantly reduced in HD patients involved in a dose-finding study of the histone deacetylase inhibitor sodium phenylbutyrate. Furthermore, expression of the marker genes was significantly up-regulated in postmortem HD caudate, suggesting that alterations in blood mRNAs may reflect disease mechanisms observed in HD brain. In conclusion, we identified changes in blood mRNAs that clearly distinguish HD patients from controls. These alterations in mRNA expression correlate with disease progression and response to experimental treatment. Such markers may provide clues to the state of HD and may be of predictive value in clinical trials.

microarrays | neurodegeneration | polyglutamine diseases

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder characterized clinically by progressive motor impairment, cognitive decline, and various psychiatric symptoms, with the typical age of onset in the third to fifth decades. The disorder is caused by the expansion of an unstable CAG triplet repeat. Changes in HD brains are widespread and include neuronal loss and gliosis, particularly in the cortex and the striatum (1). Transcriptional deregulation and loss of function of transcriptional coactivator proteins have been implicated in HD pathogenesis (2, 3). Mutant huntingtin has been shown specifically to disrupt activator-dependent transcription in the early stages of HD pathogenesis (4). Numerous microarray studies showed alterations in mRNA levels of a large number of genes in the brain of HD mice, suggesting that huntingtin may interfere with transcriptional mechanisms common to many genes (5, 6). In addition, studies in cell culture, yeast, and *Drosophila* models of polyglutamine disease have indicated that histone deacetylase (HDAC) inhibitors might provide a useful class of agents to ameliorate the transcriptional changes in HD (7–11).

Because mutant huntingtin is ubiquitously distributed and appears to affect widely expressed transcription factors, it has been postulated that transcriptional impairments in HD may exist in tissues outside of the CNS. For example, significant alterations in mRNA expression have been detected in the skeletal muscle of HD transgenic mice (6). These findings suggested that other tissues, such as peripheral blood, could be used to analyze changes in gene expression in patients with HD. Previous work indicated that lymphoblasts derived from HD patients showed increased stress-induced apoptotic cell death (12). Furthermore, both normal and mutant huntingtin have been found in peripheral blood of HD patients (13). Thus, we hypothesized that analysis of blood cells also could reveal patterns of differential gene expression that could serve as indicators of normal and abnormal biological processes in HD. Although the genetic mutation in HD serves as a definitive trait marker, differential gene expression in blood could serve as a state marker of the disease. Such biomarkers could be especially valuable in presymptomatic carriers of the HD mutation to provide an objective measure of disease state. For example, it has been hypothesized that neurons are at risk several years before motor symptoms of the disease occur, but a definitive diagnosis can be made only in the presence of unequivocal motor signs of HD. Such markers also are much needed in clinical trials of symptomatic patients, which currently rely on relatively insensitive clinical measures.

In this study, we used oligonucleotide microarrays to analyze global changes in mRNA expression in the blood samples of HD patients, compared with normal controls. In addition, we analyzed peripheral blood samples from patients who were on sodium phenylbutyrate, an HDAC inhibitor, as part of a phase I dose-finding study. Because histone acetylation and deacetylation play an important role in regulation of gene transcription, we hypothesized that the treatment of HD patients with sodium phenylbutyrate may affect gene expression in patients' blood. Our data suggest that a subset of genes altered in peripheral blood may correlate with HD progression and response to treatment.

Methods

Isolation of Blood and Brain Samples. Peripheral blood samples were collected from 62 HD subjects and 53 age- and gender-matched healthy controls. The HD group included a cohort of nine early presymptomatic carriers of the gene mutation, as determined by genetic testing, consisting of five females and four males (average age, 22.5 ± 2.6 years). An additional group of 21 late presymptomatic carriers of the HD mutation also was included in the study. This group consisted of 11 females and 10 males (average age, 39 ±

Abbreviations: HD, Huntington's disease; HDAC, histone deacetylase; QRT-PCR, quantitative RT-PCR; PCA, principal component analysis.

Data deposition: The sequences reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database (accession nos. GSE1751 and GSE1767).

‡F.B., L.L., and J.Z. contributed equally to this work.

††To whom correspondence should be addressed at: MassGeneral Institute for Neurodegenerative Disease, 114 16th Street, Room 208, Charlestown, MA 02129. E-mail: dkrainc@partners.org.

6.1 years). Further, 32 patients were symptomatic, 17 females and 15 males (average age, 49.6 ± 5.9 years). The determination of the neurological status of the HD patients used the Unified Huntington's Disease Rating Scale performed by an experienced HD neurologist. All of the symptomatic patients were in stages I or II of HD, as determined by the total functional capacity scores of 7–12. All samples were obtained in accordance with the institutional review boards, and participants gave written informed consent. Blood samples also were obtained from 12 symptomatic patients (7 males and 5 females) participating in a dose-finding study of the HDAC inhibitor, sodium phenylbutyrate (SPB11, Scandinavian Formulas, Sellersville, PA) at Oregon Health and Science University. The dosage of sodium phenylbutyrate in these subjects ranged from 12 to 18 grams per day administered in three doses. The blood samples were collected before treatment and at the end of week 4 of treatment. There were no consistent differences noted in total and differential cell blood counts between controls and HD patients or as a result of phenylbutyrate treatments. In addition, gene expression levels of known markers of lymphocytes (IFN-induced genes), neutrophils (lactotransferrin), and reticulocytes (hemoglobin δ) exhibited no systematic differences between the HD and control samples. Any and all concomitant medications were held stable throughout the sodium phenylbutyrate treatment period. For experiments involving human brain, caudate nuclei from five HD postmortem frozen brain samples (Vonsattel grades 0–2) and four age- and gender-matched controls with similar postmortem intervals were analyzed.

RNA Isolation and Gene Profiling. Total RNA from blood was extracted by using the PAXgene blood RNA kit (Qiagen, Valencia, CA), and the RNeasy lipid tissue kit (Qiagen) was used for the isolation of total RNA from striatum according to the manufacturer's protocol. All samples were treated with the RNase-free DNase set (Qiagen). The quality of total RNA was analyzed by using the RNA 6000 Nano LabChip kit on a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). Microarray analysis was performed by using U133A GeneChips (Affymetrix) and CodeLink Uniset Human I and II bioarrays (Amersham Biosciences). The U133A GeneChip contains 22,283 25-mer probe sets, and the Uniset Human I and II bioarrays contain 20,289 30-mer probes. Approximately 11,000 genes with unique LocusLink IDs are common to both platforms. The RNA for both platforms was processed by using a modified Amersham Biosciences CodeLink protocol (14). Briefly, 4 μg of high-quality total RNA was reverse-transcribed (Invitrogen), cleaned by using the QIAquick purification kit (Qiagen), and then used as a template for *in vitro* transcription by using T7 MEGA script reagents (Ambion, Austin, TX) and biotin-11-UTP (PerkinElmer/NEN). Resulting biotin-labeled cRNA was recovered and purified with the RNeasy kit (Qiagen), hybridized to the chips, and fluorescently tagged and scanned according to the manufacturer's protocol. The usual quality measures and normalization for both the Affymetrix GeneChip (3′/5′ ratios and trimmed mean normalization) and the CodeLink microarrays (detection thresholds and median normalization) were used in the experiments. All arrays were run in the same core facility.

Quantitative RT-PCR (QRT-PCR) Analysis. Reverse transcription was performed by using total RNA isolated from blood or postmortem brain and processed with the SuperScript first-strand synthesis system for RT-PCR according to the manufacturer's protocol (Invitrogen). QRT-PCR was performed with the SYBR green method by using the MyiQ Single-Color RT-PCR detection system (Bio-Rad). Primers were designed with the PRIMER3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and the sequences were as follows: ANXA1 (forward), GAGCCCCTATCCTACCTTCA; ANXA1 (reverse), GGTTGCTTCATCCACACCT; AXO (forward), ACATCAGATTCGGCTCAAGG; AXO (reverse), GGACTGCAAACCGGAATAAG; CAPZA1 (forward), TTGGAGGGCAAAGGAAGT; CAPZA1

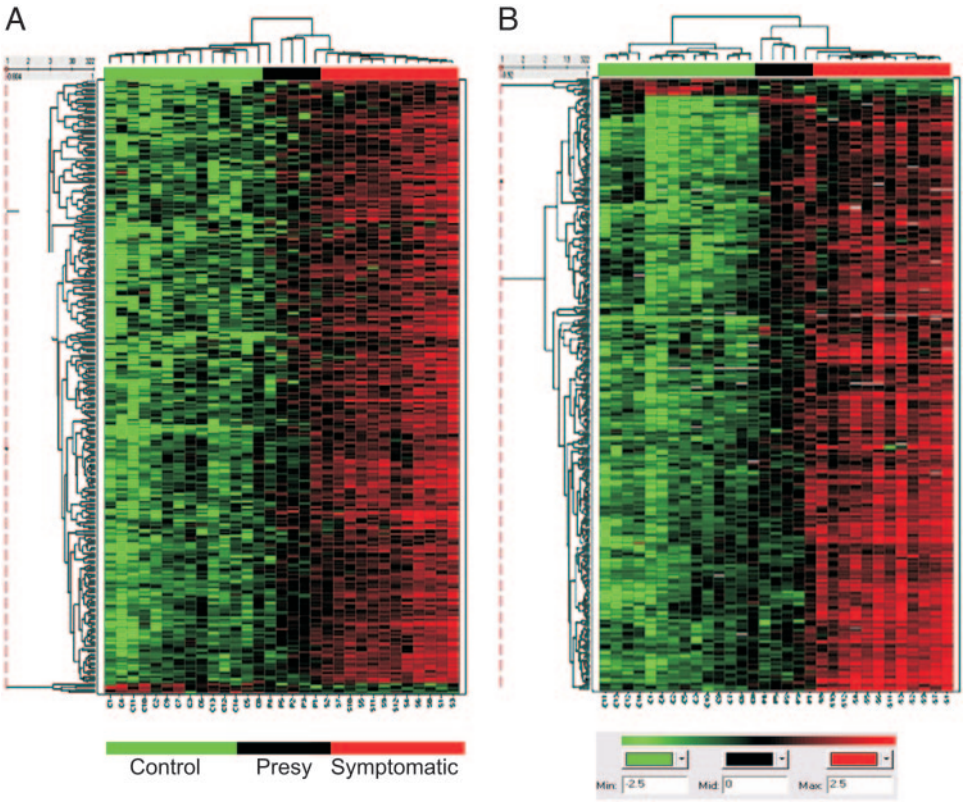
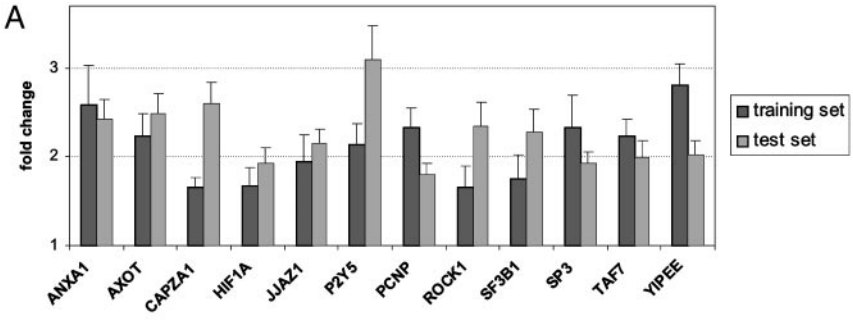


Fig. 1. Altered gene expression in the blood of HD patients. Cluster analysis of the 322 most differentially expressed genes on Affymetrix (*A*) and Amersham Biosciences (*B*) microarrays is shown. The genes were selected from 17 HD-affected subjects and 14 healthy control subjects according to *P* value ($P < 0.0005$), fold change (>1.8 or <0.6), and expression maximum >100 (Affymetrix) or >1 (Amersham Biosciences). Each column represents a sample and each row a gene. The colorgram depicts high (red) and low (green) relative levels of gene expression. The samples were normalized (median-polished) for each platform. Hierarchical clustering using cosine correlation with complete linkage was performed on the pool of all samples from both platforms to determine sample and gene clustering. The two groups were then separated in the display to compare the gene profiles between the two platforms (healthy control subjects C1–C14, late presymptomatic carriers of the HD mutation P1–P5, and symptomatic HD patients S1–S12).



B

Gene	GeneBank	LocusLink	Affymetrix Probe	Amersham Probe	Description
ANXA1	NM_000700	301	201012_at	NM_000700.1	annexin A1
AXOT	BC003404	64844	202653_s_at	NM_022826.1	Axotrophin
CAPZA1	NM_006135	829	208374_s_at	NM_006135.1	capping protein muscle Z-line, alpha 1
HIF1A	NM_001530	3091	200989_at	NM_001530.1	hypoxia-inducible factor 1, alpha subunit
JJAZ1	BF382924	23512	212287_at	NM_015355.1	joined to JAZF1
P2Y5	NM_005767	10161	218589_at	NM_005767.1	purinergic receptor (family a group 5)
PCNP	NM_020357	57092	217816_s_at	1985786CB1	PEST-containing nuclear protein
ROCK1	N22548	6093	213044_at	NM_005406.1	Rho-associated protein kinase 1
SF3B1	NM_012433	23451	201071_x_at	NM_012433.1	splicing factor 3b, subunit 1
SP3	AU145005	6670	213168_at	X68560	Sp3 transcription factor
TAF7	NM_005642	6879	201023_at	NM_005642.2	TAF7 (TAFII55)
YIPPEE	NM_016061	51646	217783_s_at	480951.23	YIPPEE protein (CGI-127)

Fig. 2. Differential expression of a subset of genes was confirmed with QRT-PCR. (*A*) The up-regulation of expression of the 12 selected genes in blood of HD subjects was confirmed on the training set of 11 HD patients and 5 controls and a test set of 30 new HD subjects and 25 controls by using QRT-PCR. Values represent average fold change $= 2^{-(\text{average } \Delta\Delta\text{Ct})}$ in mRNAs in HD relative to healthy control subjects (in every case, $P < 0.05$). Error bars are presented as (average fold change) $\times$ ($2^{\text{SEM}} - 1$). (*B*) Gene descriptions of the 12 genes with representative gene symbols, GenBank accession numbers, LocusLink ID numbers, and probe IDs for Affymetrix and Amersham Biosciences microarrays.

(reverse), GCAGGGAATGTGGTTCAAGT; HIF1A (forward), CGCAAGTCCTCAAAGCACA; HIF1A (reverse), TCAGTGGTGGCAGTGGTAGT; JJAZ1 (forward), GATGGGGAAGTAGAACAGCA; JJAZ1 (reverse), CGGAGAGGTAAGCAGGTATCA; P2Y5 (forward), TTGGACGTGCCTTTACGA; P2Y5 (reverse), TGCTGAACATGCACCCATAC; PCNP (forward), CTGTTCCAACTCTTGCTCCA; PCNP (reverse), GGCATTTCCTCTGGTTCACT; ROCK1 (forward), TGAGGTTAGGGCGAAATGGT; ROCK1 (reverse), AATCGGGTACAACTGGTGCT; SF3B1 (forward), CTTATGGGCTGTGCCATCTT; SF3B1 (reverse), GTCCGAACTTTCTGCTGCTC; SP3 (forward), CCTTACTTGCCTCTGGAACA; SP3 (reverse), CTCCCTGAACCTGGACTTGA; TAF7 (forward), CGGGAGAGTTTGTGAGTTGA; TAF7 (reverse), AGCTAGGGAACAGGAAAGCA; YIPPEE (forward), GAGAGTGAGGGCTTTGAGGA; YIPPEE (reverse), GGAGACCTGGGAAAAGATGG; ACTIN (forward), TCCCTGGAGAAGAGCTACGA; ACTIN (reverse), AGGAAGGAAGGCTGGAAGAG; and 28S (forward), AAACTCTGGTGGAGGTCCGT; 28S (reverse), CTTACCAAAAGTGGCCCACTA. Initial analysis was performed by using the ICYCLER system software (Bio-Rad). Relative gene expressions were calculated by using the $2^{-\Delta\Delta Ct}$ method, in which Ct indicates cycle threshold, the fractional cycle number where the fluorescent signal reaches detection threshold (15). β-actin was used as an internal control for blood samples and 28S rRNA as an internal control for striatum samples.

Statistical Analysis. Data analysis was performed by using the following software: MAS (version 5.0, Affymetrix), CodeLink (Amersham Biosciences), EXCEL (Microsoft), ACCESS (Microsoft), S-PLUS (version 6.1, Insightful, Seattle), and SPOTFIRE DECISIONSITE FOR FUNCTIONAL GENOMICS (version 8.0, Spotfire, Somerville, MA). Briefly, MAS (version 5.0), and CodeLink were used for the acquisition and calculation of signal values on Affymetrix and Amersham Biosciences microarrays, respectively. Calculation of maximal value of expression, two-sided *t* test, and ratio of change were performed by EXCEL, whereas ACCESS was used for filtering of the statistically significant genes on both microarray platforms. Principal component analysis (PCA) of microarray and QRT-PCR results were done with S-PLUS (version 6.1). Normalization of microarray data and hierarchical clustering were performed by using SPOTFIRE DECISIONSITE FOR FUNCTIONAL GENOMICS (version 8.0).

The genes with low expression levels have lower copy numbers of mRNA and are most susceptible to technical noise. Therefore, in the analysis of the Affymetrix microarray data, only genes with MAS (version 5.0) "signal" intensity in at least one sample above the "target intensity" of 100 were considered for further analysis. For Amersham Biosciences microarrays, a similar filter was applied by using the CodeLink median expression level of 1 as the filtering criterion. We further filtered the data by using a two-sided Student *t* test with $P < 0.0005$ and an expression ratio of average HD/average healthy control >1.8 or <0.6 as cut-off values. The lists of most significantly changed genes from both platforms were cross-referenced, and duplicate entries were excluded. Hierarchical clustering on normalized (median-polished) samples using cosine correlation with complete linkage was performed on the pool of all samples from both platforms to determine the gene clustering and to better visualize differences in expression profiles between the HD patients and healthy control subjects SPOTFIRE DECISIONSITE FOR FUNCTIONAL GENOMICS (version 8.0).

For QRT-PCR data, the average fold change $= 2^{-(\text{average } \Delta\Delta\text{Ct})}$ was computed by using the average difference in the ΔCt between the genes and internal controls. The error bars on the column plots were determined by the SEM difference of the ΔCt values so that the top of the error bar corresponds to (average fold change) $\times$ ($2^{\text{SEM}} - 1$). We also performed PCA with S-PLUS (version 6.1) by using the ΔCt values for each gene of the 12-gene marker set, compared with β-actin as control. To account for possible technical variability in different QRT-PCR experiments, all of the ΔCt values were normalized to the median of matched control samples used in each experiment. We calculated the principal components on the training set and then used the same principal components to analyze a test set. We plotted the first two principal components because they captured the most variation in the original data.

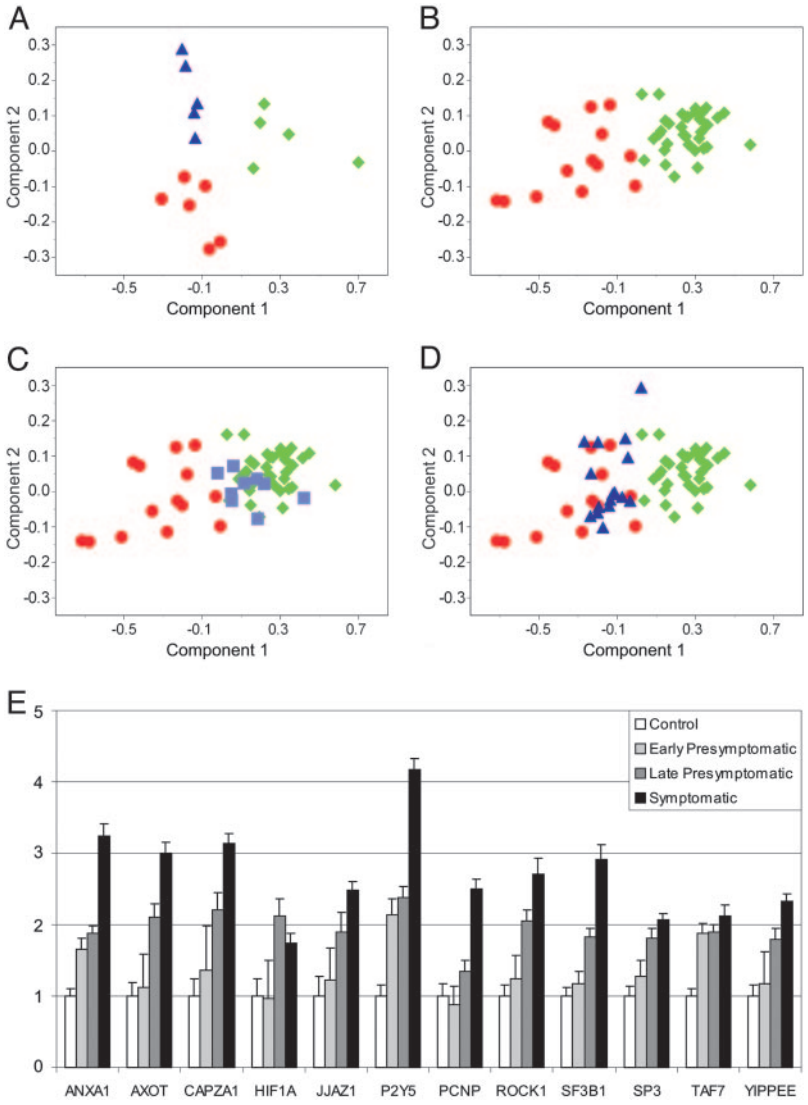


Fig. 3. The marker gene set was differentially expressed in presymptomatic and symptomatic HD subjects. PCA in the training set confirmed the separation of the presymptomatic carriers of the HD mutation and HD patients from healthy control subjects (*A*). Using the first two principal components from the QRT-PCR training set, PCA for the 12 genes in the test set confirmed the separation of symptomatic HD patients from healthy controls (*B*). Early presymptomatic carriers of the HD mutation (average age, 22.5 ± 2.6 years) cluster predominantly with control samples (*C*), whereas late presymptomatics (average age, 39 ± 6.1 years) mostly group with symptomatic HD patients (*D*). PCA was performed by using the ΔCt values. For visualization in a 2D plot, the first two principal components for the training set data were chosen representing 81% of the variance in the training set. The subgroups are colored according to their clinical classification (healthy control subjects ◆, early presymptomatics ■, late presymptomatics ▲, symptomatic HD patients ●). Analysis of the individual genes revealed up-regulation of expression with progression from early presymptomatic to symptomatic stage of HD (*E*). All of the 12 genes were significantly up-regulated in symptomatic HD patients (■) and late presymptomatic carriers of HD mutation (■), compared with controls ($P < 0.05$). In the early presymptomatic group (■), only annexin A1 (ANXA1), TAF7, and purinergic receptor P2Y (P2Y5) were significantly up-regulated ($P < 0.05$). Gene expression of the 12 selected genes in *A*–*E* was analyzed by QRT-PCR.

Results

Microarray Analysis of Global Gene Expression Changes in Blood of HD Patients. Using Affymetrix GeneChip U133A and Amersham Biosciences CodeLink arrays, we analyzed global gene expression changes in blood samples from 17 HD subjects (12 symptomatic and 5 late presymptomatic carriers of the HD mutation) and 14 healthy, age- and gender-matched control subjects. The Affymetrix platform identified 773 significantly changed genes ($P < 0.0005$, average HD/average healthy controls expression ratio >1.8 or <0.6, expression level >100). Using Amersham Biosciences CodeLink arrays for high-throughput validation of the Affymetrix data, we selected 322 genes that also were significantly changed on the Amersham Biosciences arrays ($P < 0.0005$, ratio >1.8 or <0.6, expression level >1). Cluster analysis of the Affymetrix and Amersham Biosciences expression levels of the significantly changed genes showed that the majority of the genes were up-regulated in HD, compared with control blood samples (Fig. 1). Interestingly, four of the significantly changed genes were up-regulated on Affymetrix and down-regulated on Amersham Biosciences platform. Our initial investigation suggested that probes on different platforms may hybridize to different splice variants of each gene, but further experiments will be required to confirm this observation.

Genes that had significant changes in expression belong to many different functional groups such as transcription/RNA processing, signaling, ubiquitin/proteasome, and vesicle trafficking. Those involved in transcription include transcriptional coactivators such as TAF7, splicing factors such as SF3B1, transcriptional regulators such as retinoblastoma 1, and zinc finger proteins. Other differentially expressed genes include those involved in signal transduction such as purinergic receptor P2Y5, members of the tumor necrosis factor receptor superfamily, and ubiquitin/proteasome-associated proteins such as ubiquitin-specific protease USP15. Genes involved in vesicle trafficking such as proteoglycan 1 also were significantly changed.

Selection of Biomarker Genes. To identify a subset of genes that could usefully classify HD patients and healthy controls, 322 genes were ranked according to their *P* values, highest fold change, highest expression levels, and consistency of fold change in each individual HD sample, compared with its age- and gender-matched control. Probe sequences that corresponded to duplicates of the same genes, as well as probes for unknown or hypothetical proteins, were removed. In the case of duplicate entries, probes with the most statistically significant changes were used in further analysis. Using such criteria, we selected the top 30 candidate genes, which were further tested by QRT-PCR using blood samples from 11 HD-affected subjects (6 new symptomatic patients and 5 late presymp-

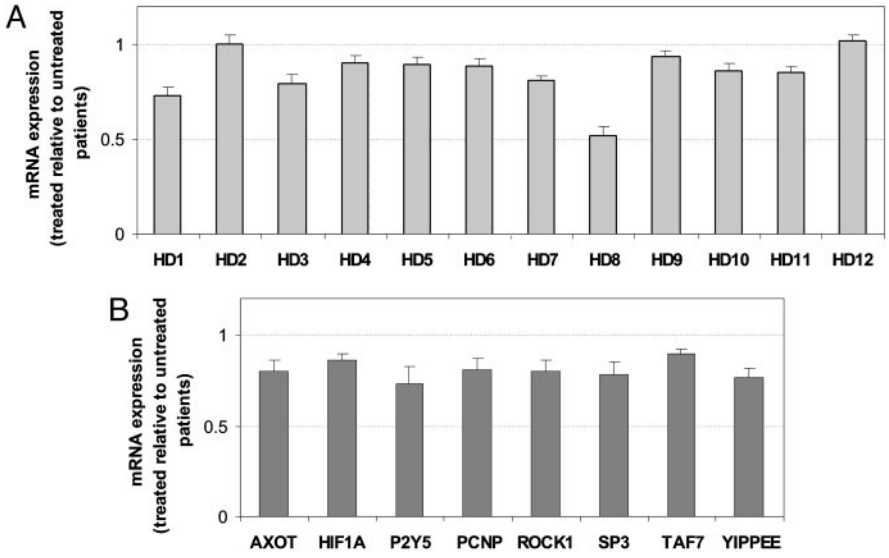


Fig. 4. Gene expression of the marker gene set was decreased in HD patients treated with phenylbutyrate. RNA was isolated from blood samples taken from 12 HD patients (HD1–HD12) before treatment and after 4 weeks of treatment with phenylbutyrate. QRT-PCR analysis of gene expression of the 12 selected genes (*A*) showed significant decrease in expression after treatment in 10 of 12 patients ($P < 0.05$). In patients HD2 and HD12, no changes were observed. The values are presented as average fold change $= 2^{-(\text{average } \Delta\Delta Ct)}$ for gene set in treated relative to untreated samples. Error bars represent (average fold change) $\times$ ($2^{SEM} - 1$). In *B*, QRT-PCR analysis of each individual gene from the marker set in patients treated with phenylbutyrate showed statistically significant decrease in expression of 8 of 12 genes ($P < 0.05$) after treatment. The values represent average fold change $= 2^{-(\text{average } \Delta\Delta Ct)}$ for the samples after treatment, compared with untreated samples. Error bars are presented as (average fold change) $\times$ ($2^{SEM} - 1$).

tomatic carriers of the HD mutation that were used for microarrays) and 5 new, age- and gender-matched controls. Using this approach, we identified 12 genes that exhibited the most statistically significant changes between the HD and control groups (Fig. 2). To confirm the ability of the selected marker set to discriminate between healthy controls and HD subjects, we used a test set consisting of a new cohort of 30 HD subjects (14 symptomatic patients and 16 late presymptomatic carriers of the HD mutation) and 25 matched controls. We found a statistically significant increase in the expression of the biomarker genes in HD subjects, compared with controls (Fig. 2*A*).

Next, PCA was performed to better visualize the differences in expression of the 12 genes in relation to the stage of disease. Such analysis of the training set confirmed that the 12 genes were able to clearly separate the HD subjects and controls in the training sets when Affymetrix, Amersham Biosciences (data not shown), and QRT-PCR data (Fig. 3*A*) were used. In addition, PCA analysis showed clear differences in mRNA expression between symptomatic and late presymptomatic HD subjects, with the presymptomatic subjects clustering between the symptomatic patients and control subjects. Although there were some differences between the Affymetrix and Amersham Biosciences analyses, namely in the classification of the presymptomatic and control samples, the concordance of the two platforms was >0.7 for each of the 12 genes as determined by Pearson's correlation, with $P < 0.00005$ (data not shown).

To further validate the biomarker genes by using the independent test set, the QRT-PCR data also were projected onto the first two principal components identified for the training set. The PCA of the HD and control samples in the test set confirmed total separation of the two groups (Fig. 3*B*). Late presymptomatic carriers of the HD mutation (average age, 39 ± 6.1 years) clustered predominantly with the symptomatic patients, although some overlap with control subjects existed (Fig. 3*C*). Next, expression of the 12 marker genes was examined in a new cohort of nine early presymptomatic HD subjects (average age, 22.5 ± 2.6 years) and nine matched controls. Interestingly, these subjects clustered between symptomatic HD and controls but overlapped considerably with the control group (Fig. 3*D*). Of the 12 genes tested, only 3 showed significant up-regulation in early presymptomatic subjects, confirming that expression profiles in early stages of HD more resemble those of controls (Fig. 3*E*). Moreover, expression of the marker genes progressively increased when HD progressed from early presymptomatic to late presymptomatic and symptomatic stage (Fig. 3*E*).

These results confirmed that the expression of the 12-gene marker set could clearly distinguish diseased blood samples from controls and also may discriminate stages of disease progression.

Biomarker Gene Changes in Response to Treatment in HD. We next examined whether the expression of the 12 marker genes could be modified by treatment with sodium phenylbutyrate, an HDAC inhibitor, conducted as part of a phase I dose-finding study. Sodium phenylbutyrate has been shown to exert significant neuroprotective effects in transgenic mouse models of HD (16). As an initial step for assessing the feasibility of phenylbutyrate administration in HD patients, a dose-finding study was performed. Blood samples were collected from 12 HD patients before treatment and at week 4 of treatment with phenylbutyrate. Using QRT-PCR, we found a small but statistically significant decrease in the expression of the 12-gene marker set in 10 of 12 patients after 4 weeks of treatment (Fig. 4*A*). When the marker genes were examined individually, 8 of 12 genes were significantly decreased in response to 4 weeks of treatment with phenylbutyrate (Fig. 4*B*).

β-Actin controls used in these experiments did not exhibit any systematic changes in expression between the untreated and treated groups of patients. There were no significant symptomatic effects of phenylbutyrate in any of the subjects as expected in a dose-finding study not powered to detect alterations in symptoms or progression.

Together, these results indicated that the selected biomarker set

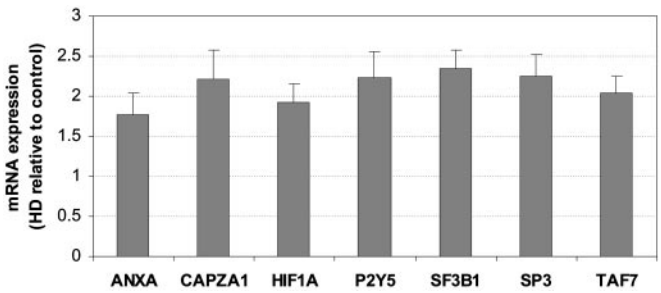


Fig. 5. Genes altered in HD blood are differentially expressed in HD brain. Expression of the 12 marker genes was analyzed in five postmortem HD caudate samples (Vonsattel grades 0–2) and four control samples by using QRT-PCR. Seven genes were significantly up-regulated in brain tissue samples from HD patients, compared with controls ($P < 0.05$). Values are presented as average fold change $= 2^{-(\text{average } \Delta\Delta Ct)}$ for HD brain samples, compared with controls. Error bars are presented as (average fold change) $\times$ ($2^{SEM} - 1$).

of 12 genes may be used to monitor early responses to treatment in patients with HD.

Expression of Biomarker Genes in Human HD Brain Tissue. To examine whether the changes in mRNAs observed in HD patient blood correlated with mRNA changes in brain, the expression of the 12 genes was analyzed by QRT-PCR in caudate nuclei isolated from five postmortem HD brains (Vonsattel grades 0–2) and four control brains. Of the 12 genes tested, 7 exhibited significant up-regulation in HD brain samples, compared with controls, whereas the other 5 genes were not significantly altered (Fig. 5). Although the mechanisms of differential gene expression between the blood and brain tissues is not fully elucidated, we show that alterations of gene expression in HD blood at least in part correlate with changes in gene expression in HD brain, indicating that mutant huntingtin may affect similar targets in these tissues. These results raise an intriguing possibility that blood markers may provide a more accessible window through which it is possible to monitor the underlying pathogenic process in HD.

Discussion

Our results demonstrate dramatic alterations in gene expression in the peripheral blood of HD patients. To analyze the observed changes in mRNAs in more detail, we focused on a subset of 12 genes that were able to clearly distinguish normal controls and patients with HD. Most importantly, the expression of these marker genes was altered in relation to disease progression from early presymptomatic to late presymptomatic and symptomatic stage of HD. Gene expression in younger carriers of the HD mutation (average age, 22 years) resembled the expression patterns observed in control subjects, whereas older presymptomatic subjects (average age, 39 years) became more similar to symptomatic HD patients (average age, 49 years). These findings indicate that a subset of mRNAs isolated from blood could be useful in monitoring the progression of HD.

Studies in cell culture, yeast, and *Drosophila* models of polyglutamine disease indicated that HDAC inhibitors might provide a useful class of agents to ameliorate the process of neurodegeneration in HD (7–11). We had the opportunity to examine the marker gene set in a phenylbutyrate dose-finding study that was recently performed in HD patients. Interestingly, we found significant down-regulation of a subset of mRNAs in treated patients, suggesting that HDAC inhibitors may affect specific targets in the blood of HD patients. Although the effects of HDAC inhibitors on the expression of the marker genes appeared rather subtle, they were statistically significant in 10 of 12 patients. At the same time, it is still not clear whether subtle or robust changes in gene expression will prove important in the pathogenesis of HD and in response to treatment. Correlating such changes in blood mRNA with the effects of the treatment on the rate of neurodegeneration and disease progression must await a much larger phase III clinical study.

To further address the question of biological relevance of the marker genes, we asked whether these mRNAs were differentially expressed in human HD brain. Of the 12 genes tested, 7 showed significantly up-regulated expression in postmortem HD brain, suggesting that mutant huntingtin may affect similar targets in blood and brain. It has been demonstrated previously that mutant huntingtin specifically interferes with the function of general transcription factors and coactivators, such as Sp1 and CBP (3, 4). Because these factors exhibit ubiquitous expression, it is not surprising that mutant huntingtin may affect gene transcription in tissues outside the CNS. At the present time, it is not clear why only neurons are predominantly affected in HD. One possible explanation may lie in the fact that neurons are postmitotic and highly specialized as opposed to dividing blood cells, which may be less susceptible to degeneration because of their short half-life.

Microarrays for gene-expression profiling are rapidly becoming important research tools for identifying potential biomarkers (17–19). In contrast to other studies related to HD that have used tissues difficult to access in live patients, we used easily accessible peripheral blood. The use of genomic research in the analysis of blood-derived mRNA may represent an important advancement for development of biomarkers in neurological and other diseases. Such a minimally invasive approach enables larger sample sizes, better matching of patients, and more standardized collection procedures. Our study identified a large number of significantly altered mRNAs from which other smaller sets of genes could be selected and validated in a variety of clinical situations. The selection and validation of marker genes could vary, depending on the purpose of the biomarker. For example, biomarkers used for prognostic purposes might differ from biomarkers linking a surrogate endpoint to a clinical endpoint and would be validated differently. Furthermore, the optimal biomarker gene set for monitoring different therapeutic targets could differ but could potentially be configured from the larger pool of genes that we have identified. One of the goals of biomarker development relevant to HD and other diseases is to enable an assessment of drug specificity, safety, and efficacy, thereby improving the efficiency and cost-effectiveness of the drug development process. Many potential therapies have now been tested in genetic models of HD, providing a rationale for a number of clinical trials that have already occurred or are planned and for which biomarkers are greatly needed.

We thank Drs. Ruth Luthi-Carter, Andy Strand, Timothy Clark, Steve Rowley, Clemens Scherzer, Nancy Wexler, Allan Tobin, Anne B. Young, and Ethan Signer for helpful discussion or comments on the manuscript. We thank Sanja Ivcevic for technical assistance. This work was supported by National Institutes of Health Grants NS-002174 (to D.K.) and NS-045242 (to D.K. and S.M.H.), the High Q Foundation (D.K. and P.H.), the New England Huntington Disease Society of America Center of Excellence (S.M.H. and H.D.R.), and U.S. Public Health Service Grant 5 M01-RR-000334 (to P.H.).

1. Vonsattel, J. P. & DiFiglia, M. (1998) *J. Neuropathol. Exp. Neurol.* **57,** 369–384.
2. Harjes, P. & Wanker, E. E. (2003) *Trends Biochem. Sci.* **28,** 425–433.
3. Sugars, K. L. & Rubinsztein, D. C. (2003) *Trends Genet.* **19,** 233–238.
4. Dunah, A. W., Jeong, H., Griffin, A., Kim, Y. M., Standaert, D. G., Hersch, S. M., Mouradian, M. M., Young, A. B., Tanese, N. & Krainc, D. (2002) *Science* **296,** 2238–2243.
5. Luthi-Carter, R., Strand, A., Peters, N. L., Solano, S.M., Hollingsworth, Z. R., Menon, A. S., Frey, A. S., Spektor, B. S., Penney, E. B., Schilling, G., *et al.* (2000) *Hum. Mol. Genet.* **9,** 1259–1271.
6. Luthi-Carter, R., Hanson, S. A., Strand, A. D., Bergstrom, D. A., Chun, W., Peters, N. L., Woods, A. M., Chan, E. Y., Kooperberg, C., Krainc, D., *et al.* (2002) *Hum. Mol. Genet.* **11,** 1911–1926.
7. Steffan, J. S., Bodai, L., Pallos, J., Poelman, M., McCampbell, A., Apostol, B. L., Kazantsev, A., Schmidt, E., Zhu, Y. Z., Greenwald, M., *et al.* (2001) *Nature* **413,** 739–743.
8. McCampbell, A., Taye, A. A., Whitty, L., Penney, E., Steffan, J. S. & Fischbeck, K. H. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 15179–15184.
9. Hughes, R. E., Lo, R. S., Davis, C., Strand, A. D., Neal, C. L., Olson, J. M. & Fields, S. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 13201–13206.
10. Hockly, E., Richon, V. M., Woodman, B., Smith, D. L., Zhou, X., Rosa, E., Sathasivam, K., Ghazi-Noori, S., Mahal, A., Lowden, P. A., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100,** 2041–2046.
11. Ferrante, R. J., Kubilus, J. K., Lee, J., Ryu, H., Beesen, A., Zucker, B., Smith, K., Kowall, N. W., Ratan, R. R., Luthi-Carter, R., *et al.* (2003) *J. Neurosci.* **23,** 9418–9427.
12. Sawa, A., Wiegand, G. W., Cooper, J., Margolis, R. L., Sharp, A. H., Lawler, J. F., Jr., Greenamyre, J. T., Snyder, S. H. & Ross, C. A. (1999) *Nat. Med.* **5,** 1194–1198.
13. Ide, K., Nukina, N., Masuda, N., Goto, J. & Kanazawa, I. (1995) *Biochem. Biophys. Res. Commun.* **209,** 1119–1125.
14. Amersham Biosciences (2003) *User Guide: CodeLink Expression Bioarrays Automated Target Preparation* (Amersham Biosciences, Piscataway, NJ).
15. Livak, K. J. & Schmittgen, T. D. (2001) *Methods* **25,** 402–408.
16. Gardian, G., Browne, S. E., Choi, D. K., Klivenyi, P., Gregorio, J., Kubilus, J. K., Ryu, H., Langley, B., Ratan, R. R., Ferrante, R. J., *et al.* (2005) *J. Biol. Chem.* **280,** 556–563.
17. Welsh, J. B., Zarrinkar, P. P., Sapinoso, L. M., Kern, S. G., Behling, C. A., Monk, B. J., Lockhart, D. J., Burger, R. A. & Hampton, G. M. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 1176–1181.
18. Staal, F. J., van der Burg, M., Wessels, L. F., Barendregt, B. H., Baert, M. R., van den Burg, C. M., van Huffel, C., Langerak, A. W., van der Velden, V. H., Reinders, M. J., *et al.* (2003) *Leukemia* **17,** 1324–1332.
19. Ilyin, S. E., Belkowski, S. M. & Plata-Salaman, C. R. (2004) *Trends Biotechnol.* **22,** 411–416.