Supporting Information

Hu et al. 10.1073/pnas.1104409108

SI Materials and Methods

Clinical Studies and Human Biospecimens. Patients with premanifest or symptomatic HD and age- and sex-matched HCs and neurological disease controls were recruited from the New England Huntington's Disease Society of America Center of Excellence (REVEAL-Huntington Disease study) into the microarray study and the two qPCR validation studies, respectively. Patients on high-dose creatine treatment were excluded. Symptomatic HD was classified into three disease stages according to a total functional capacity scale (stage I, 11–13; stage II, 7–10; stage III, 1–6) (1). Premanifest subjects were recruited on the basis of documented genetic testing for the HD mutation and the lack of clinical symptoms of HD at the time of enrollment. HCs were recruited primarily from the spouses of patients with HD and other neurodegenerative diseases seen at Partners Healthcare, who are part of the source population.

Genome-wide gene expression analysis was performed on venous cellular blood of eight patients with HD [mean age (y) ± SD: 51.2 ± 6.1, sex: 1 male (M)/7 female (F), mean RNA integrity number (RIN) ± SD: 7.1 ± 0.6] and 6 age- and sex-matched HCs without neurodegenerative disease (41.5 ± 10.6, 3 M/3 F, 7 ± 0.5) and analyzed together with genome-wide expression data of 83 neurodegenerative disease controls as well as 22 additional HCs who we had assayed using identical standard operating procedures as part of parallel PD biomarker discovery efforts (2, 3). These included individuals with corticobasal degeneration (n = 1), Alzheimer's disease (n = 23), PD (n = 50), essential tremor (n = 1), MSA (n = 2), or progressive supranuclear palsy (n = 6).

In the replication study on the qPCR platform, *H2AFY* mRNA levels were assayed in individuals with HD [n = 8, mean age (y) ± SD: 51.8 ± 6.2], dystonia (n = 6, 56.3 ± 5.5), PD (n = 14, 55.8 ± 4.8), MSA (n = 9, 59.5 ± 6.8), and HC (n = 8, 50.9 ± 15.4) enrolled in the Harvard NeuroDiscovery Center Biomarker Study or the sister REVEAL-HD study.

In the independent validation studies, the cross-sectional case–control study included subjects with premanifest HD [n = 9, mean age (y) ± SD: 48.1 ± 7.5], patients with manifest HD (n = 36, 49.4 ± 13.1), 1 subject with ataxia-1 (two visits, 51.5 ± 0.4), and HCs (n = 50, 49.5 ± 8.2) enrolled in the REVEAL-HD study. The longitudinal cohort included 25 patients with HD [mean age (y) ± SD: 53.8 ± 12] and 21 HCs (53.4 ± 9.5) followed over 2–3 y with two (9 subjects) or three (37 subjects) approximately annual follow-up visits in the REVEAL-HD study.

The Phenylbutyrate Development for Huntington's Disease (PHEND-HD) study is a multicenter, double-blind, placebo-controlled, phase II clinical trial with open-label follow-up to determine the safety and tolerability of SPB in subjects with HD. HD subjects in stage I or II were enrolled at multiple experienced Huntington Study Group investigative sites. Patients with HD under preexisting treatment of phenylbutyrate or other known transcriptionally active compounds were carefully excluded from the enrollment in this study. In a double-blind and placebo-control phase, patients with HD were randomized to two groups with a 1:1 ratio to take 15 g of SPB and matching placebo, respectively, daily for 4 wk (28 ± 3 d). Subsequently, all patients took open-label SPB for an additional 12 wk (112 ± 3 d), followed by a 4-wk washout phase (140 ± 3 d). For a subset of participants in the PHEND-HD study, blood collected and frozen in PAXgene tubes was available for analysis.

Frozen cortical tissues from controls and patients with HD were obtained from the New York Brain Bank at Columbia University. The HD group included eight grade 2/3 patients [mean age (y) ± SD: 59.1 ± 10.8, sex: 6 M/2 F] and four grade 4 patients with HD (52.8 ± 14.9; 2 M/2 F). The disease severity of patients with HD was graded with a rating scale as reported by Vonsattel et al. (4). All five controls [mean age (y) ± SD: 63.8 ± 11, sex: 2 M/3 F] were free of neurological diseases and were matched to HD cases based on age and postmortem interval (PMI). The PMI for all cases did not exceed 22 h.

All human studies were approved by the Institutional Review Boards of the Massachusetts General Hospital and Brigham and Women's Hospital.

Animals. R6/2 mice used in the study were generated by backcrossing male R6/2 mice (available from the Jackson Laboratory) with C57BL/6 × CBA F1 female mice. Mice were genotyped by PCR using tail-tip DNA (CAG repeat length of 135–140). For SPB (Scandinavian Formulas, Inc.) treatment, 8-wk-old R6/2 mice were administered SPB (150 or 300 $mg{\cdot}kg^{-1}{\cdot}d^{-1}$) or vehicle by i.p. injection for 2 wk. Mice were euthanized 2 h after the last injection; brains were removed and snap-frozen in liquid nitrogen. 140-CAG knock-in mice were originally obtained from an established colony at the University of California (Los Angeles, CA) (5) and were maintained on B6/CBA crossed background in our colony at the Massachusetts General Hospita. For Western blotting, brains were rapidly removed, snap-frozen in liquid nitrogen, and stored at −80 °C until use. All animal experiments were carried out in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory animals and were approved by local animal care committee.

RNA Extraction. Venous cellular whole blood from patients with HD and controls was collected in PAXgene tubes (Qiagen). The PAXgene tubes were shipped to the biospecimen core by courier and stored at 4 °C for 124 ± 80 h (HD) and 110 ± 36 h (HC) before RNA extraction. After 24 h of incubation at room temperature, RNA was extracted following the manufacturer's protocol, including DNase treatment for 5 min, and stored at −80 °C. RNA quality was determined by spectrophotometry and by use of the RNA 6000 NanoChip kit on an Agilent Technologies 2100 Bioanalyzer with the RIN algorithm (6). For the independent, cross-sectional and the longitudinal, case-control validation study, only high-quality RNA samples with an RIN ≥7 were included.

For a subset of participants in the PHEND-HD study, blood was collected in PAXgene tubes at three time points (i.e., week −3/visit 1, week 4/visit 4, and week 16/visit 7). The PAXgene tubes were shipped by courier to the biosample service facility of the Partners HealthCare Center for Personalized Genetic Medicine and stored at −80 °C for 3 wk before RNA extraction. Following 16 h of incubation at room temperature, RNA was isolated from whole blood using the PAXgene blood RNA kit (Qiagen) with DNase treatment at room temperature for 15 min and stored at −80 °C.

For mouse experiments, total RNA was extracted from mouse brain using Qiazol (Qiagen) and RNeasy Mini Kits (Qiagen) according to the manufacturer's instructions.

Histone Extraction and Western Blotting. Histone extraction was done essentially as previously described (7), with slight modifications. Briefly, human and murine brain tissues were homogenized in 9 vol of PBS buffer containing 4 mM sodium butyrate, 2% (vol/vol) Triton X-100, 1 mM EDTA, 3 mM DTT,

and 1× HALT protease inhibitor (Pierce Biotechnology) using a Pellet pestel (Kontes) and then centrifuged at 3,000 × g for 10 min at 4 °C. The pellet was washed twice in the buffer described above, and histones were extracted by solubilization in 0.2 M HCl. After neutralizing the pH of acid-extracted solution containing the histone pool with neutralization buffer [10 parts of 1 M Tris/HCl (pH 7.4) − 3 parts of 10 M NaOH], proteins were analyzed by SDS/PAGE and Western blotting using an antibody to histone macroH2A1 (1:3,000; Upstate Biotechnology).

Immunohistochemistry. Mice were deeply anesthetized and transcardially perfused with 2% (vol/vol) buffered paraformaldehyde. Brains were cryoprotected and serially sectioned (50 μm). Immunohistochemistry was carried out on free-floating sections at the level of crossing of the anterior commissure by means of a previously described method (8). Mouse sections were immunostained for macroH2A1 (antibody to histone macroH2A1, 1:100 dilution; Upstate Biotechonology) for 2 d, washed, and then incubated with biotinylated secondary antibody. Reactivity was developed using the Vectastain ABC kit (Vector Laboratories). Antibody complexes were visualized using 3′3′-diaminobenzidine (Sigma–Aldrich). No signal was detected in the controls in which primary antibody was omitted.

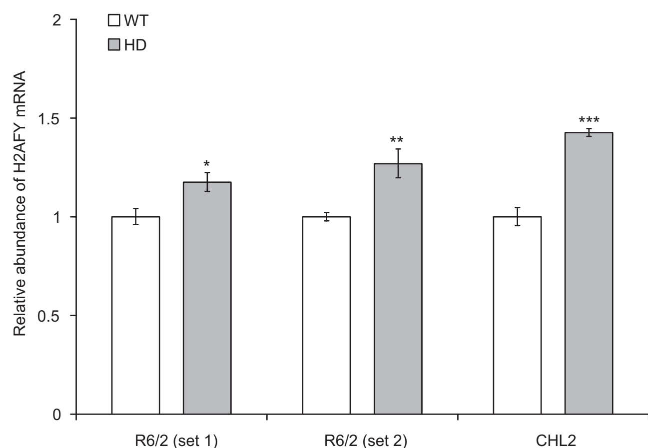
Biostatistics. The Affymetrix .CEL files for eight patients with HD and six age- and sex-matched HCs were normalized to "all probe sets" in a standardized manner and scaled to 100 by the MAS5 algorithm implemented in Bioconductor in R (9). Because technical variation is higher for genes with low average expression intensities on Affymetrix Human GeneChip U133A arrays, only genes with intensities of ≥100 in at least one sample were considered for further analysis. Significance analysis was performed using significance analysis of microarrays implemented in R on a Mac Pro computer with 8 gigabytes of memory, 2.66-GHz processor speed, and a 1-terabyte hard drive. Because of space limitations, only the 38 genes with P values $<10^{-6}$ (by a two-sample, two-sided t test) of the 99 significant genes were selected for visualization in the heat map shown in Fig. 1*A*. Unsupervised hierarchical clustering was used to group genes according to relative variation in gene expression patterns. To generate graphical representations of relative gene expression levels, all expression levels were first normalized for each gene by setting the average (mean) to 0 and the SD to 1 across all samples. Scaled levels were color-coded as a spectrum representing relative changes from the mean. The clustering was done by the "average" method implemented in hierarchical cluster analysis (hclust) method in R (expression higher than the mean was displayed as shades of red, and expression lower than the mean was displayed as shades of blue). The heat map was plotted by the "heatmap.2" method implemented in the gplots package in R. Gene expression data for the eight patients with HD and the six age- and sex-matched HCs were submitted to the Gene Expression Omnibus (GEO) database (accession no. GSE24250). Gene expression data for the 105 individuals with PD and controls were previously reported as part of our PD biomarker discovery efforts (2) and are available in the GEO database (accession no. GSE6613).

In the biomarker replication and validation studies, the general linear model procedure adjusted for age and sex with the option of Tukey's studentized range test using SAS software version 9.2 (SAS) was used to estimate statistical significance.

In the PHEND-HD study, we used a mixed random and fixed coefficient regression model, where linear and quadratic terms for weeks in the study (−3, 4, and 16) as well as for weeks on the drug were separate simultaneous fixed predictors, controlling for each other. Given the design, weeks on drug for the SPB-treated group were 0, 4, and 16, whereas the weeks on drug for the placebo group were 0, 0, and 12. Random terms were intercepts and linear and quadratic terms for weeks in the study/on drug, where the random terms were allowed to have an unstructured covariance matrix. Backward elimination (using a cutoff of $P = 0.05$ or approximately so for marginal effects) was used on both the fixed and random terms, wherein the latter were removed if their estimated variance and covariances were not significantly different from zero (nonsignificant effects were only retained in the model if a higher order predictor that included that term remained as significant).

MacroH2A1 expression changes in human brain were evaluated by two-sided t-test. For all mice experiments, the general linear model procedure for ANOVA using SAS software version 9.2 was applied to determine the P values.

1. Marder K, et al.; Huntington Study Group (2000) Rate of functional decline in Huntington's disease. *Neurology* 54:452–458.
2. Scherzer CR, et al. (2007) Molecular markers of early Parkinson's disease based on gene expression in blood. *Proc Natl Acad Sci USA* 104:955–960.
3. Scherzer CR (2009) Chipping away at diagnostics for neurodegenerative diseases. *Neurobiol Dis* 35:148–156.
4. Vonsattel JP, et al. (1985) Neuropathological classification of Huntington's disease. *J Neuropathol Exp Neurol* 44:559–577.
5. Menalled LB, Sison JD, Dragatsis I, Zeitlin S, Chesselet MF (2003) Time course of early motor and neuropathological anomalies in a knock-in mouse model of Huntington's disease with 140 CAG repeats. *J Comp Neurol* 465:11–26.
6. Imbeaud S, et al. (2005) Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res* 33:e56.
7. Ryu H, et al. (2006) ESET/SETDB1 gene expression and histone H3 (K9) trimethylation in Huntington's disease. *Proc Natl Acad Sci USA* 103:19176–19181.
8. Chopra V, et al. (2007) A small-molecule therapeutic lead for Huntington's disease: Preclinical pharmacology and efficacy of C2-8 in the R6/2 transgenic mouse. *Proc Natl Acad Sci USA* 104:16685–16689.
9. Scherzer CR, Jensen RV, Gullans SR, Feany MB (2003) Gene expression changes presage neurodegeneration in a Drosophila model of Parkinson's disease. *Hum Mol Genet* 12:2457–2466.
10. Kuhn A, et al. (2007) Mutant huntingtin's effects on striatal gene expression in mice recapitulate changes observed in human Huntington's disease brain and do not differ with mutant huntingtin length or wild-type huntingtin dosage. *Hum Mol Genet* 16: 1845–1861.



HD mouse model data sets used for *in silico* confirmation

HD model	Age	mHtt[b]	wHtt[c]	Poly Q	No. mut[d]	No. wt[e]	H2AFY abundance	*P*
R6/2 (set1)[a]	12 w	N.h	2	209	4	5	1.18	0.03
R6/2 (set2)[a]	12 w	N.h	2	209	5	4	1.27	0.01
CHL2(Hdh^{Q150})[a]	22 m	2F.m	0	155	4	4	1.43	0.0003

a. Affymetrix 430 2.0 data sets published by Kuhn et al. (2007); b. mHtt, mutant huntingtin: number of copies (one or two), protein coding length (N, N-terminal htt fragment; F, full length htt), species (h, human; m, mouse); c. wHtt, number of copies of wild-type huntingtin -encoding mouse Hdh genes; d. Number of mice with mutant huntingtin being analyzed; e. Number of wild-type mice being analyzed.

Fig. S1. *In silico* confirmation. To confirm the results from our in vivo analyses further, we analyzed *in silico* genome-wide gene expression datasets (reported in 9) from a knock-in mouse model of HD, the CHL2 mice, as well as two distinct datasets from R6/2 mice. Consistent with our in vivo studies, relative *H2AFY* abundance was 1.43-fold increased in the striatum of CHL2 knock-in mice (*P* = 0.0003). In the two R6/2 mouse datasets, relative H2AFY abundance was increased 1.18-fold and 1.27-fold, respectively (*P* = 0.03 and *P* = 0.01, respectively). WT, wild type.

Table S1. Ninety-nine genes significantly differentially expressed in blood of patients with HD

Probe ID	Gene symbol	Fold	FDR	*P*	Gene name
222375_at	*NA*	4.8	<0.00002	5.92E-11	NA
200021_at	*CFL1*	2	<0.00002	7.69E-11	Cofilin 1 (nonmuscle)
208829_at	*TAPBP*	1.9	<0.00002	4.32E-10	TAP-binding protein (tapasin)
213453_x_at	*GAPDH*	1.5	<0.00002	1.66E-09	Glyceraldehyde-3-phosphate dehydrogenase
222024_s_at	*AKAP13*	2.1	<0.00002	2.06E-09	A-kinase anchor protein 13
200001_at	*CAPNS1*	1.8	<0.00002	3.84E-09	Calpain, small subunit 1
201315_x_at	*IFITM2*	1.8	<0.00002	4.92E-09	IFN-induced transmembrane protein 2 (1–8 D)
212620_at	*ZNF609*	3.2	<0.00002	1.05E-08	Zinc finger protein 609
210916_s_at	*CD44*	1.6	<0.00002	1.05E-08	CD44 molecule (Indian blood group)
209936_at	*RBM5*	3.3	<0.00002	2.75E-08	RNA-binding motif protein 5
208518_s_at	*PER2*	3.2	<0.00002	4.49E-08	Period homolog 2 (*Drosophila*)
210706_s_at	*RNF24*	1.7	<0.00002	5.91E-08	Ring finger protein 24
203942_s_at	*MARK2*	3.5	<0.00002	8.80E-08	MAP/microtubule affinity-regulating kinase 2
222138_s_at	*WDR13*	2.6	<0.00002	1.14E-07	WD repeat domain 13
209403_at	*TBC1D3*	2.2	<0.00002	1.29E-07	TBC1 domain family, member 3
214501_s_at	*H2AFY*	1.6	<0.00002	1.30E-07	H2A histone family, member Y
90265_at	*ADAP1*	1.5	<0.00002	1.32E-07	ArfGAP with dual domains 1
211982_x_at	*XPO6*	1.5	<0.00002	1.44E-07	Exportin 6
210981_s_at	*GRK6*	2	<0.00002	1.49E-07	G protein-coupled receptor kinase 6
208998_at	*UCP2*	1.9	<0.00002	1.67E-07	Uncoupling protein 2 (mitochondrial, proton carrier)
218679_s_at	*VPS28*	1.7	<0.00002	2.08E-07	Vacuolar protein sorting 28 homolog (*Saccharomyces cerevisiae*)
213233_s_at	*KLHL9*	0.6	<0.00002	2.37E-07	Kelch-like 9 (*Drosophila*)
214035_x_at	*LOC399491*	1.7	<0.00002	2.69E-07	LOC399491 protein
213574_s_at	*KPNB1*	1.5	<0.00002	2.79E-07	Karyopherin (importin)-β1
215562_at	*TTC39A*	2.7	<0.00002	4.10E-07	Tetratricopeptide repeat domain 39A
221096_s_at	*TMCO6*	1.8	<0.00002	4.10E-07	Transmembrane and coiled-coil domains 6
32029_at	*PDPK1*	1.6	<0.00002	4.11E-07	3-Phosphoinositide dependent protein kinase-1
202156_s_at	*CUGBP2*	1.6	<0.00002	6.03E-07	CUG triplet repeat, RNA-binding protein 2
46256_at	*SPSB3*	1.5	<0.00002	6.09E-07	splA/ryanodine receptor domain and SOCS box containing 3
221513_s_at	*UTP14A*	2.3	<0.00002	6.11E-07	UTP14, U3 small nucleolar ribonucleoprotein, homolog A (yeast)
221473_x_at	*SERINC3*	1.6	<0.00002	6.12E-07	Serine incorporator 3
200819_s_at	*RPS15*	1.5	<0.00002	6.60E-07	Ribosomal protein S15
208610_s_at	*SRRM2*	1.8	<0.00002	7.26E-07	Serine/arginine repetitive matrix 2
200022_at	*RPL18*	1.7	<0.00002	7.54E-07	Ribosomal protein L18
221030_s_at	*ARHGAP24*	2	<0.00002	8.59E-07	ρ-GTPase activating protein 24
209619_at	*CD74*	1.8	<0.00002	8.70E-07	CD74 molecule, MCH complex, class II invariant chain
206150_at	*CD27*	1.6	<0.00002	9.06E-07	CD27 molecule
217926_at	*C19orf53*	1.8	<0.00002	9.94E-07	Chromosome 19 ORF 53
201901_s_at	*YY1*	1.5	<0.00002	1.06E-06	YY1 transcription factor
218080_x_at	*FAF1*	2.2	<0.00002	1.37E-06	Fas (TNFRSF6)-associated factor 1
218084_x_at	*FXYD5*	1.6	<0.00002	1.65E-06	FXYD domain containing ion transport regulator 5
218828_at	*PLSCR3*	2.7	<0.00002	1.69E-06	Phospholipid scramblase 3
218414_s_at	*NDE1*	1.5	<0.00002	1.82E-06	nudE nuclear distribution gene E homolog 1 (*Aspergillus nidulans*)
205793_x_at	*TNK1*	2.5	<0.00002	1.91E-06	Tyrosine kinase, nonreceptor, 1
208676_s_at	*PA2G4*	1.8	<0.00002	1.93E-06	Proliferation-associated 2G4, 38 kDa
201715_s_at	*ACIN1*	1.5	<0.00002	2.02E-06	Apoptotic chromatin condensation inducer 1
204908_s_at	*BCL3*	2.1	<0.00002	2.31E-06	B-cell CLL/lymphoma 3
208066_s_at	*GTF2B*	1.5	<0.00002	2.37E-06	General transcription factor IIB
218010_x_at	*C20orf149*	1.6	<0.00002	2.52E-06	Chromosome 20 ORF 149
216781_at	*KIAA1751*	2.6	<0.00002	3.22E-06	KIAA1751
216213_at	*NEK1*	1.8	<0.00002	3.39E-06	NIMA (never in mitosis gene a)-related kinase 1
218826_at	*SLC35F2*	1.6	<0.00002	3.60E-06	Solute carrier family 35, member F2
221798_x_at	*RPS2*	1.6	<0.00002	3.67E-06	Ribosomal protein S2

Table S1. Cont.

Probe ID	Gene symbol	Fold	FDR	*P*	Gene name
37028_at	*PPP1R15A*	1.6	<0.00002	3.82E-06	Protein phosphatase 1, regulatory (inhibitor) subunit 15A
206649_s_at	*TFE3*	1.6	<0.00002	4.40E-06	Transcription factor binding to IGHM enhancer 3
214196_s_at	*TPP1*	1.7	<0.00002	4.55E-06	Tripeptidyl peptidase I
221506_s_at	*TNPO2*	2.2	<0.00002	5.42E-06	Transportin 2
201500_s_at	*PPP1R11*	1.5	<0.00002	5.48E-06	Protein phosphatase 1, regulatory (inhibitor) subunit 11
218421_at	*CERK*	0.6	<0.00002	5.52E-06	Ceramide kinase
40837_at	*TLE2*	2.5	<0.00002	5.81E-06	Transducin-like enhancer of split 2 [E(sp1) homolog, *Drosophila*]
36994_at	*ATP6V0C*	1.5	<0.00002	6.48E-06	ATPase, H^+ transporting, lysosomal 16 kDa, V0 subunit c
202723_s_at	*FOXO1*	1.5	<0.00002	6.61E-06	Forkhead box O1
213650_at	*GOLGA8A*	1.7	<0.00002	7.79E-06	Golgi autoantigen, golgin subfamily a, 8A
212860_at	*ZDHHC18*	1.6	<0.00002	7.87E-06	Zinc finger, DHHC-type containing 18
218581_at	*ABHD4*	1.5	<0.00002	8.08E-06	Abhydrolase domain containing 4
202140_s_at	*CLK3*	1.6	<0.00002	8.13E-06	CDC-like kinase 3
211558_s_at	*DHPS*	1.7	<0.00002	1.00E-05	Deoxyhypusine synthase
201332_s_at	*STAT6*	2.1	<0.00002	1.17E-05	Signal transducer and activator of transcription 6, IL-4–induced
202564_x_at	*ARL2*	2.2	<0.00002	1.24E-05	ADP ribosylation factor-like 2
41047_at	*C9orf16*	1.5	<0.00002	1.33E-05	Chromosome 9 ORF 16
216242_x_at	*POLR2J2*	1.6	<0.00002	1.50E-05	Polymerase (RNA) II (DNA directed) polypeptide J2
202901_x_at	*CTSS*	0.4	<0.00002	1.51E-05	Cathepsin S
217854_s_at	*POLR2E*	1.6	<0.00002	1.91E-05	Polymerase (RNA) II (DNA directed) polypeptide E, 25 kDa
204669_s_at	*RNF24*	1.7	<0.00002	2.03E-05	Ring finger protein 24
205312_at	*SPI1*	1.9	<0.00002	2.04E-05	Spleen focus-forming virus (SFFV) proviral integration oncogene spi1
219259_at	*SEMA4A*	1.6	<0.00002	2.13E-05	Sema domain, Ig domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4A
218749_s_at	*SLC24A6*	2.4	<0.00002	2.17E-05	Solute carrier family 24 (sodium/potassium/calcium exchanger), member 6
215434_x_at	*NBPF10*	1.7	<0.00002	2.28E-05	Neuroblastoma breakpoint family, member 10
216295_s_at	*CLTA*	1.5	<0.00002	3.51E-05	Clathrin, light chain (Lca)
37793_r_at	*RAD51L3*	1.7	<0.00002	3.75E-05	RAD51-like 3 (*S. cerevisiae*)
213975_s_at	*LYZ*	0.6	<0.00002	3.89E-05	Lysozyme (renal amyloidosis)
213773_x_at	*NSUN5*	1.6	<0.00002	4.21E-05	NOL1/NOP2/Sun domain family, member 5
206491_s_at	*NAPA*	1.5	<0.00002	4.24E-05	*N*-ethylmaleimide–sensitive factor attachment protein, α
200004_at	*EIF4G2*	0.6	<0.00002	4.69E-05	Eukaryotic translation initiation factor 4 γ2
217414_x_at	*HBA1*	2	<0.00002	5.07E-05	Hemoglobin, α1
213842_x_at	*NSUN5C*	2.4	<0.00002	5.09E-05	NOL1/NOP2/Sun domain family, member 5C
212203_x_at	*IFITM3*	1.5	<0.00002	5.19E-05	IFN-induced transmembrane protein 3 (1–8 U)
218030_at	*GIT1*	1.7	<0.00002	5.26E-05	G protein-coupled receptor kinase interacting ArfGAP 1
221829_s_at	*TNPO1*	1.5	<0.00002	6.06E-05	Transportin 1
214334_x_at	*DAZAP2*	0.6	<0.00002	6.28E-05	DAZ-associated protein 2
AFFX-BioDn-5_at	*NA*	2.6	<0.00002	6.34E-05	NA
41386_i_at	*JMJD3*	1.6	<0.00002	7.22E-05	Jumonji domain containing 3, histone lysine demethylase
206337_at	*CCR7*	1.7	<0.00002	7.30E-05	Chemokine (C-C motif) receptor 7
214870_x_at	*LOC100132540*	2	<0.00002	8.86E-05	Similar to LOC339047 protein
204265_s_at	*GPSM3*	1.7	<0.00002	9.43E-05	G protein-signaling modulator 3 (AGS3-like, *Caenorhabditis elegans*)
58900_at	*LOC222070*	1.6	<0.00002	1.66E-04	Hypothetical protein LOC222070
204018_x_at	*HBA1*	1.9	<0.00002	2.08E-04	Hemoglobin, α1
209458_x_at	*HBA1*	1.8	<0.00002	4.00E-04	Hemoglobin, α1
211699_x_at	*HBA1*	1.8	<0.00002	4.16E-04	Hemoglobin, α1

ArfGAP, ADP-ribosylation factor GTPase activating protein; CDC, cell division cycle; CLL, chronic lymphocytic leukemia; DAZ, deleted in azoospermia; DHHC, Asp-His-His-Cys; IGHM, immunoglobulin heavy constant mu; NA, not annotated; SOCS, suppressor of cytokine signaling; TBC, Tre-2/Bub2p/Cdc16p.

Table S2. Overview of study subjects

No. samples	HD	Premanifest HD	Controls	
			HCs	Disease controls
Microarray	8 (8)		28 (28)	83 (83)
Initial qPCR replication	8*		8 (2[†])	29 (27[‡])
Independent, cross-sectional, case–control validation study	36 (36)	9 (9)	50 (50)	2 (1[§])
Longitudinal, case–control validation study				
Subject visit 1	25 (14[¶])		21 (10[¶])	
Subject visit 2	25[‖]		21**	
Subject visit 3	25[‖]		12**	
PHEND-HD				
Subject visit 1	28 (28)			
Subject visit 4	28[††]			
Subject visit 7	26[††]			
Subtotal	209 (86)	9 (9)	140 (90)	114 (111)
Total	472 (296)			

Numbers in parentheses represent the number of subjects unique to a particular study.
*All eight subjects were also assayed in the microarray study.
[†]Six subjects were also assayed in the microarray study.
[‡]Two subjects were also assayed in the microarray study.
[§]One patient with spinocerebellar ataxia-1 was assayed at two visits.
[¶]Eleven subjects in the HD group and 11 subjects in HC group were also included in the cross-sectional case-control validation study or in the initial qPCR replication or the microarray study.
[‖]Of the 25 subjects assayed at visit 1, all were assayed at visit 2 and visit 3.
**Of the 21 subjects assayed at visit 1, all were assayed at visit 2, whereas 12 subjects were assayed at visit 3.
[††]Of the 28 subjects assayed at visit 1, all were assayed at visit 4, whereas 26 subjects were assayed at visit 7.

Table S3. Clinical information for independent, cross-sectional, case-control validation study

	HD (n = 36)	HC (n = 50)	P
Age, y	49.4 ± 13.1	49.5 ± 8.2	0.95
Sex	14 F/22 M	32 F/18 M	0.02
RIN	7.7 ± 0.3	7.8 ± 0.5	0.11
HD stage I	n = 14		
HD stage II	n = 14		
HD stage III	n = 8		

Table S4. Clinical information for the longitudinal case–control validation study

	HD (n = 25)	HC (n = 21)	P
Age, y	53.8 ± 12	53.4 ± 9.5	0.84
Sex	13 F/12 M	10 F/11 M	0.77
RIN	7.7 ± 0.4	7.8 ± 0.5	0.18
Subject visit 1	n = 25	n = 21	
Subject visit 2	n = 25	n = 21	
Subject visit 3	n = 25	n = 12	
Average (range) days in study at visit 1	0	0	
Average (range) days in study at visit 2	349 (27–659)	429 (231–1,048)	0.14
Average (range) days in study at visit 3	705 (182–1,109)	702 (455–1,065)	0.97

Table S5. Human samples for Western blot analysis

ID	Age, y	Genotype	Sex	PMI, h
T-220	57	Control	F	3.3
T-274	65	Control	F	21.02
T-779	76	Control	F	3.36
T-206	49	Control	M	13.55
T-174	72	Control	M	4.3
T-349	61	HD grade 2	F	4.45
T-1482	76	HD grade 2	F	4.45
T-4091	46	HD grade 3	M	0.05
T-2959	47	HD grade 3	M	2.2
T-3221	58	HD grade 2	M	0.26
T-1991	53	HD grade 3	M	17.15
T-3049	72	HD grade 2	M	15.3
T-2476	60	HD grade 3	M	0
T-4110	49	HD grade 4	F	0.3
T-225	72	HD grade 4	F	17.55
T-329	54	HD grade 4	M	0.2
T-3859	36	HD grade 4	M	0.4