HD Mouse Striatum RNA and Protein Disease Signatures

John C. Obenauer[1], Jian Chen[2], Viktoria Andreeva[1], Jeffrey S. Aaronson[2], Ramee Lee[2], Andrea Caricasole[3,4], Jim Rosinski[2]

[1]Rancho BioSciences; [2]CHDI Foundation; [3]IRBM; [4]Kedrion Biopharma

ABSTRACT

Extensive data sets from mouse models of Huntington's disease called the allelic series have been made publicly available (Alexandrov V et al., 2016), including RNA-seq data in GEO and proteomics data in HDinHD.org (Langfelder P et al., 2016). These data sets were analyzed to identify differentially expressed genes and proteins in 14 types of tissues, at up to 8 polyQ repeat lengths and at up to 3 ages. All the lists of differentially expressed genes have been added to HDinHD.org. In the striatum, many genes were observed to be recurrently significant at several Q lengths and ages. Using this data set and others, robust disease signatures were developed and tested at the RNA and protein levels. These disease signatures, Str266R (for RNA) and Str115P (for proteins), can be used to monitor disease perturbations in HD mouse experiments.

OBJECTIVE

Can reproducible disease signatures be determined in the mouse striatum for RNA-seq and proteomics experiments?

METHODS

HD and wild type striatum samples from 10 RNA-seq experiments were analyzed using DESeq2 in R with uniform criteria (fold change of at least 20% in either direction and adjusted p-value < 0.05) to get 10 lists of significant genes. These lists were grouped by Q length and age to find genes that overlapped within each group or across all groups. The genes overlapping all groups were tested on 4 validation data sets.

All the striatum proteomics data samples in PRIDE accession PXD006302 were analyzed using limma in R with uniform criteria (no fold change threshold and an adjusted p-value < 0.1). The lists of significant proteins in the Q111, Q140, and Q175 sample groups at ages 6 months or 10 months were compared. UniProt IDs that are assigned the same gene symbol were grouped together. 126 proteins were significant in at least half of the 12 comparisons and changed in consistent directions. These 126 proteins were tested using 4 validation data sets.

RESULTS

287 genes overlapped all 4 grouped RNA data sets. 21 genes were rejected as being predicted genes or poorly characterized genes, based on their names indicating gene models (like Gm10406), Riken sequences (like A830036E02Rik), or GenBank accessions (like AW495222). This left 266 genes to be validated. These genes overlap two of the striatum WGCNA modules published by Langfelder et al. (2016). 180 (68%) of these genes are in module M2, and 54 others (20%) are in module M20. In Langfelder's study, modules M2 and M20 were the top two most CAG length-dependent modules.

Experiment	Significant Genes	Group Overlap	All Overlap
Full Series Q175 10M	1,725	1,088	287
Cohort1Time1 Q175 10M	3,056		
Cohort1Time2 Q175 10M	2,379		
Full Series Q140 10M	1,593	658	
Miniseries Q140 10M	1,994		
Full Series Q140 6M	1,625	439	
Miniseries Q140 6M	984		
Cohort2Time1 Q140 6M	2,563		
HDAC R6/2 3M	6,089	4,551	
KMO R6/2 3M	5,963		

RNA Validation Expts	Significant
Cohort2Time2 Q140 12M	262
Cohort3Time1 Q175 7M	262
Cohort3Time2 Q175 7M	266
Cohort3Time3 Q175 12M	266

Protein Validation Expts	Significant
R6/2 2M (PXD013771)	108
R6/2 3M (PXD013771)	114
JNK3 R6/2 6W	91
Cohort4Time1 Q175 10M	111

Str266R

Wt1	Pcdhb12	Rgs19	Gpm6b	Car11	Ttll3	Pipox	Ddx11	Dusp14	Ptprv
Onecut1	Vwa5b2	Hook2	Clec12a	Garem1	Il17rc	Pcp4	Lmo2	Coch	Bnipl
Tnip3	Scn9a	Ltk	Ppp2r2a	Fancb	Cntn5	Zbtb18	Sfn	Atp6v1c2	Il2rb
Sfmbt2	Pcdh20	Zfp711	Gstm6	Jcad	Rasgrp2	Fsbp	Rxrg	C4a	Vwa7
Rgs13	Tmc3	Gng3	Tbc1d8	Tbc1d4	Shank3	Pde1b	Dnah1	Arpp19	Plk5
Tnfrsf13c	Pcdhb22	Pcdhb5	Rerg	Rps6ka4	Dpy19l3	Crocc	Bank1	Htr1b	Spata21
Crnde	Pcdhb3	Fam126a	Hbegf	Atf6	Rgs7bp	Itga5	Rgs4	B3gnt2	Ryr1
Ifnlr1	Asl	Hes6	Hebp1	Dusp18	Adcy5	Kdm4b	Mas1	Car12	Ffar3
Fgfr4	Rbm11	Gpr149	Slmap	Tesc	Tpm2	Ago4	Acvrl1	Epyc	Phex
Acy3	Cbx4	Sh3yl1	Bcr	Chn1	Fam184b	Hrk	Arhgef39	Cnr1	Dgat2l6
Smim24	Nagk	Syde2	Anks1b	Vrk1	Malat1	Rbp4	Cntnap3	Ankrd35	Theg
Tnnt2	Greb1l	Nsun7	Acy1	Kcnab1	Dock4	Acvr1c	Nrep	Neto2	Sec14l3
Klhl14	Pcdhb9	Ccdc177	Spock3	Gpr139	Ddn	Osbpl8	Kcnh4	Scn4b	Tmem114
Slc45a3	Lrrn3	Has1	Ppp1ca	Stk32a	Gpr83	Nrn1l	Hipk4	Tnfrsf4	Odf4
Dsp	Zfp7	Psme1	Ephx1	Ano3	Gabrd	Ppp1r1a	Ssc5d	Ccdc155	Mafa
Ccdc87	Pcdhb16	Cyp4x1	Ppp3ca	Pxdn	Ccm2	Slc39a2	Itga9	Abi3bp	Slc4a11
Vill	Smoc1	Brinp3	Cttnbp2	Hpca	Inhba	Adora2a	Ppp1r16b	Myo5c	Ddit4l
Runx2	Vps37d	Gba2	Fmnl1	Kctd1	Lzts3	Rhobtb2	Camk1g	Lrrc10b	Wnt8b
Cbx8	Galns	Sgk3	Plcxd1	Lrrk2	Wipf3	Oscar	Upb1	Penk	Tmprss6
Chdh	Pcdhb19	Ace	Sbsn	Itpr1	Slc26a10	Id4	Tcf7	Shisa2	Myo7b
Pcdhb21	Cdh18	Cap1	Dbpht2	Sh2d5	Zbtb46	Myh7	Dmkn	Cyp2a5	Gpx6
Polr2a	N4bp2	Fbln5	Ppp1r1b	D7Ertd443e	Asb2	Rspo1	Fam83d	Clspn	Ifi27l2b
Pcdhb2	Gsto1	Akt2	Atp2b1	Gsg1l	Npl	Drd2	Homer1	Krt9	Sohlh1
Rdh12	Dusp23	Baiap2	Camkk2	Camk2n1	Cd59a	Impg1	Ppp4r4	Ptpn7	
Htr2c	Trpc7	Grm4	Sec14l1	Abcc12	St8sia2	Piwil2	Pde10a	Fgf3	
Dsg2	Samd14	Zfyve28	Rnf207	Rgs9	S100a10	Gask1b	Rgs14	Abhd11os	
Insyn2b	Cep164	Wdr78	Gipc2	Ptpn5	Drd1	Gpr6	Arpp21	Plekhg4	

All 266 RNA genes were significant in at least 3 of the 4 validation data sets, so they were kept in the Str266R signature. 115 of the 126 proteins were significant in at least 2 of the 4 proteomics validation sets, so they defined the Str115P signature. Red genes increase expression in disease while blue genes decrease.

Str115P

Chdh	Pfas	Grin1	Ryr3	Camkk1	Shank3	Inf2	Ngef	Tbc1d8	Pde10a
Acy3	Pck2	Pitpnm2	Atp2b1	Cyld	Apoe	Itpka	Itpr1	Matn4	Rasd2
Ahi1	Lmnb2	Actn1	Phyhip	Rin1	Sh2d5	Rem2	Ppp1r16b	Arpp19	Tcf20
Macrod1	Psme1	Grm5	Osbpl8	Calcoco1	Gria3	Bsg	Npl	Kcnip2	Scn4b
Armcx2	Mri1	Cdkl5	Dlgap3	Anks1b	Ppp4r4	Olfm2	Rgs7bp	Ano3	Chrm4
Nagk	Rap1gap	Atp2a2	Cap1	Cbr3	Rcn1	Jcad	Kcnj4	Spata2l	Drd1
Dis3	Adcy5	Prkcb	Trim46	Fbxl16	Ptpn5	Ankrd63	Pde7b	Rgs14	Pex5l
Dus3l	Syngap1	Crocc	Inpp5j	Pde1b	Ppp1r1b	Dlgap2	Coch	Serna7a	
Fahd2	Htt	Bcr	Mast3	Cacna2d3	Sec14l1	Rasgrp2	Rps6ka4	Foxp1	
Prepl	Erc2	Baiap2	Cacnb2	Synpo	Camkk2	Hpca	Ntrk3	Arc	
Gprasp1	Dab2ip	Sorbs1	Phactr1	Pcp4	Shisa7	Them6	Mlf2	Rgs9	
Rbm3	Mink1	Kctd16	Grm1	Homer1	Camk4	Lrrtm1	Wipf3	Sh3rf2	

CONCLUSIONS

Reproducible striatum disease signatures were identified and validated at the RNA and protein levels. The gene symbols, ranked from most positive to most negative log fold change, are shown in this poster and are also provided in the supplemental file HD_Striatum_Signatures.xlsx.

REFERENCES

1. Alexandrov V et al., *Nat Biotechnol* 2016, 34:838-844.
2. Langfelder P et al., *Nat Neurosci* 2016, 19:623-633.