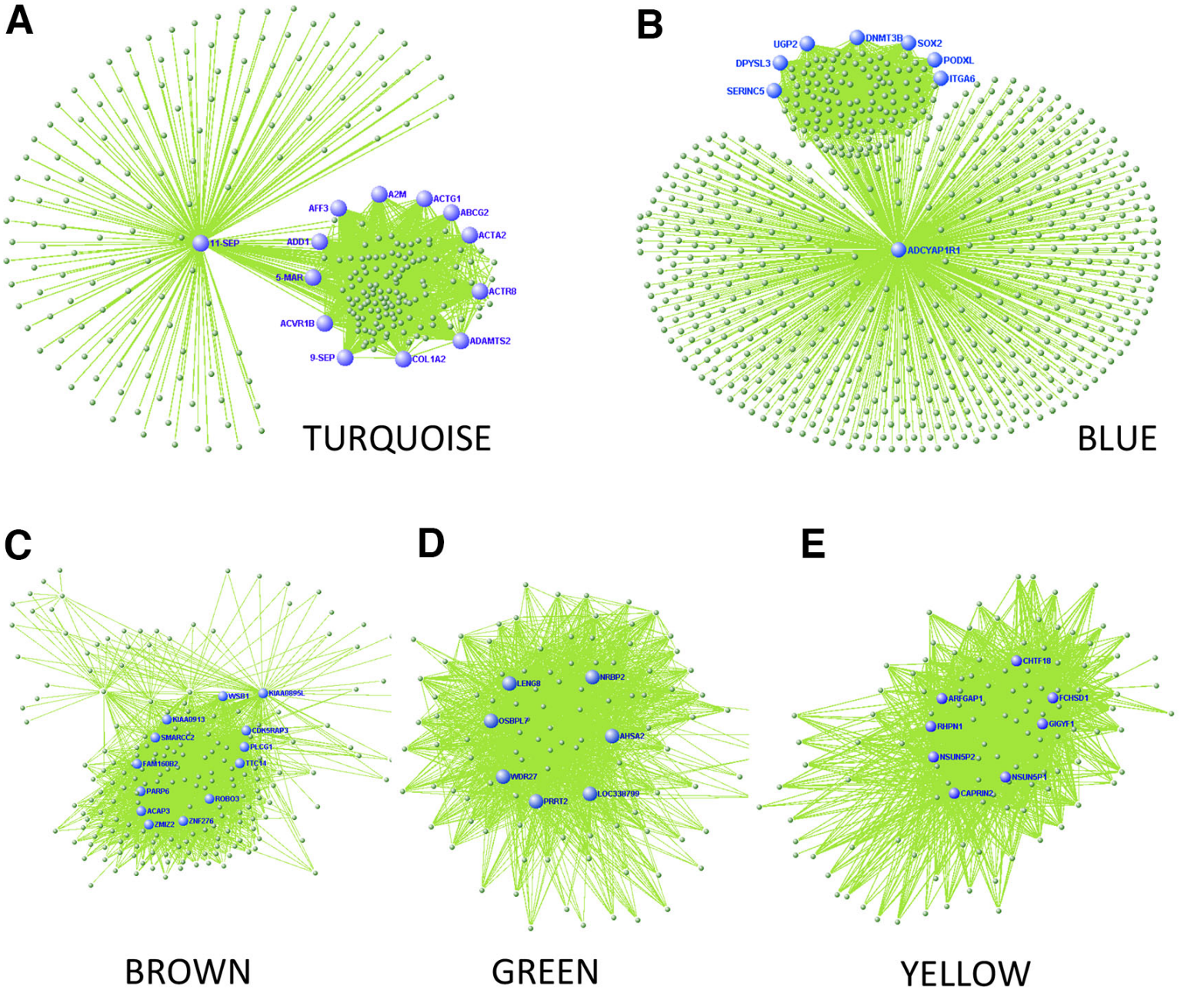
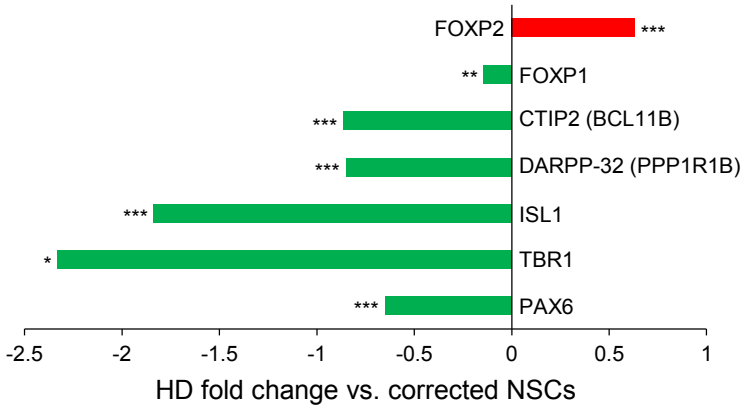
Stem Cell Reports, Volume *5*

Supplemental Information

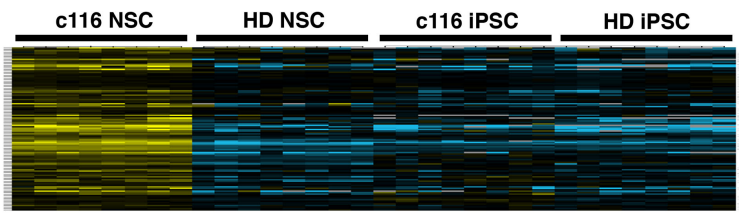
Genomic Analysis Reveals Disruption of Striatal Neuronal Development and Therapeutic Targets in Human Huntington's Disease Neural Stem Cell

Karen L. Ring, Mahru C. An, Ningzhe Zhang, Robert N. O'Brien, Eliana Marisa Ramos, Fuying Gao, Robert Atwood, Barbara J. Bailus, Simon Melov, Sean D. Mooney, Giovanni Coppola, and Lisa M. Ellerby


A
11-SEP
AFF3
A2M
ACTG1
ABCG2
ADD1
ACTA2
5-MAR
ACTR8
ACVR1B
ADAMTS2
9-SEP
COL1A2
TURQUOISE
B
UGP2
DNMT3B
SOX2
DPYSL3
PODXL
ITGA6
SERINC5
ADCYAP1R1
BLUE
C
WSB1
KIAA0895L
KIAA0913
SMARCC2
CDK5RAP3
PLCG1
FAM160B2
TTC14
PARP6
ROBO3
ACAP3
ZMIZ2
ZNF276
BROWN
D
LENG8
NRBP2
OSBPL7
AHSA2
WDR27
PRRT2
LOC338799
GREEN
E
CHTF18
ARFGAP1
FCHSD1
RHPN1
GIGYF1
NSUN5P2
NSUN5P1
CAPRIN2
YELLOW


FOXP2
***
**
FOXP1
***
CTIP2 (BCL11B)
***
DARPP-32 (PPP1R1B)
***
ISL1
*
TBR1
***
PAX6
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
HD fold change vs. corrected NSCs

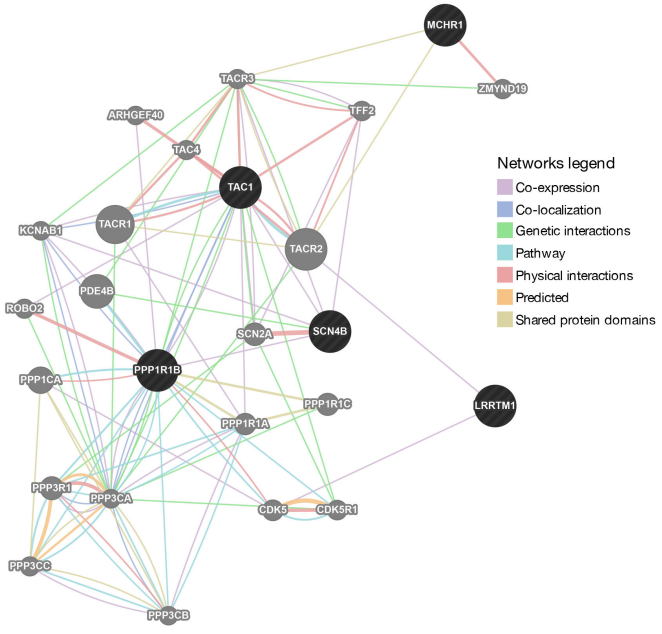
A



B

Term	P-value	Z-score	Combined Score	Genes
dorsal_striatum	0.047	-1.896	5.788	PPP1R1B;LRRTM1;TAC1;SCN4B;MCHR1
mIMCD-3	0.036	-1.697	5.647	ABCC4;HOXB9;CLCF1;CDH6
liver	0.037	-1.575	5.196	C2;LETMD1;NNMT;IGFBP7;TLCD2;HAAO;INSIG1;BCAS3;RETSAT
nih_3T3	0.086	-1.443	3.533	BDNF;TBX18;CRABP1
macrophage_bone_marrow_24h_LPS	0.191	-1.715	2.836	BCL2L2;GNGT2;PTK2B;TSPO
osteoclasts	0.077	-1.028	2.639	KCNJ2;DENND1B
mammary_gland__lact	0.070	-0.898	2.391	STAT5A;G0S2
uterus	0.116	-0.958	2.064	PADI1;COL6A3;TGM2
heart	0.206	-1.294	2.047	SLC8A1;EXT2;RBM20;MYL3
nucleus_accumbens	0.209	-1.276	1.998	PPP1R1B;MCHR1;TAC1

C



SUPPLEMENTAL FIGURE LEGENDS

Fig. S1. Hub connectivity of module genes, visualized for the five remaining modules. (**A-E**) Network plot depicting the top connections for each module. Most highly connected module members (or hub genes) are shown for each group.

Fig. S2. Expression levels of these critical striatal genes are altered in neural stem cells from HD NSCs. Colored bars indicate RNA expression level fold-change in $\log_2$ scale of human HD vs. corrected NSCs for genes prominently expressed during striatal development. All genes shown were statistically significant when adjusted for FDR, with *$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$.

Fig. S3. Hierarchical clustering approach reveals down regulated genes in HD NSCs linked to BDNF signaling. (A) Expression trait profiles across all 32 samples for module gene members. (**B**) Gene enrichment of module gene members, top hit is dorsal striatum. (**C**) Visualization of networks associated with top term gene members using the Genemania tool.

Table S4. Go Annotation of WGCNA Modules using DAVID Analysis

Module	Select Hub Genes	Top Biological Processes Associated with Module
Blue	ADCYAP1R1, SERINC5, DPYSL3, UGP2, DNMT3B, SOX2, PODXL, ITGA6	Translation, metabolic process, proteasomal protein catabolic process, proteasomal ubiquitin-dependent protein catabolic process, tRNA metabolic process, ribosome biogenesis, cellular macromolecule catabolic process
Turquoise	SEPT11, AFF3, A2M, ACTG1, ABCG2, ACTA2, ACTR8, ADAMTS2, COL1A2, SEPT9, ACVR1B, MARCH5	Regulation of transcription, transcription, regulation of RNA metabolic process, regulation of transcription, DNA-dependent, chromatin modification, chromosome organization, chromatin organization, cell cycle, response to DNA damage stimulus
Black	HSPB8, DOK5, SH3BGRL2, VGLL3, LYN, GRIP1, FAM131B, EFEMP1	None significant
Red	SEMA3D, IRS1, CDH2, APBB2	Neuron projection morphogenesis, cell projection organization, cell motion, cell projection morphogenesis, cell part morphogenesis, neuron projection development, neuron differentiation, cellular component morphogenesis, neuron development, axonogenesis
Yellow	CHTF18, FCHSD1, GIGYF1, NSUN5P1, NSUN5P2, CAPRIN2, RHPN1, ARFGAP1	Chromatin modification, DNA repair, DNA metabolic process, cellular response to stress, response to DNA damage stimulus, DNA recombination, chromatin organization, chromosome organization, striated muscle cell development, muscle cell development
Brown	WSB1, KIAA0895L, CDK5RAP3, PLCG1, TTC14, ROBO3, ZNF276, ZMIZ2, ACAP3, PARP6, FAM160B2, SMARCC2, KIAA0913	Negative regulation of gene expression, negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process, negative regulation of nitrogen compound metabolic process, negative regulation of macromolecule biosynthetic process, negative regulation of cellular biosynthetic process, negative regulation of transcription, DNA dependent, negative regulation of transcription
Green	LENG8, NRBP2, AHSA2, LOC338799, PRRT2, WDR27, OSBPL7, LENG8	Regulation of protein kinase activity, regulation of kinase activity, regulation of transferase activity

SUPPLEMENTAL PROCEDURES

Differentiation and Culture of NSCs

NSCs were generated with STEMdiff Neural Induction Medium from Stemcell Technologies according to manufacturer's manual. Briefly 70-80% confluent iPSCs were detached and dissociated to single cells by incubation in Accutase (Sigma) at 37^{o}C for 5-10 min. 4 X 10^{6} cells were seeded in one well of a AggreWell 800 plate (Stemcell Technologies) in STEMdiff Neural Induction Medium supplemented with Y27632 ROCK inhibitor (10 μM, Calbiochem). After 5 days of culture in AggreWell 800 plate, neural aggregates formed in the grids of the well were transferred and attached to poly-L-ornithine/laminin coated culture dish. After 7 more days of culture in STEMdiff Neural Induction Medium, rosettes were harvested by incubation in STEMdiff Neural Rosette Selection Reagent at 37^{o}C for 1h followed by washing in DMEM/F12 medium (Life Technologies). Rosettes were triturated with a P1000 tip for 3 times and seeded in poly-L-ornithine/laminin coated culture dish in neural proliferation medium (Neurobasal medium supplemented with 1X B27, 2mM L-Glutamine, 25 ng/ml bFGF, 10 ng/ml LIF, 100 U/ml penicillin and 100 μg/ml streptomycin). NSCs were expanded on Matrigel (Corning) coated dishes and passaged with Accutase. NSCs were over 98% nestin positive (An et al. 2012). After seven passages, NSC samples were harvested in 60 mm dishes with N = 8 for RNA-Seq.

RNA Isolation

Total RNA was purified from iPSCs, NSCs or neurons using RNeasy Mini kit (Qiagen). For RNA-Seq samples, an additional 350 μl RLT/β-mercaptoethanol solution were added to each tube, followed by vortexing and centrifugation to lyse the cells. A Qiacube instrument was then used to extract the RNA using the RNase Micro protocol. The samples were eluted in 25 μl PCR grade H_20. Sample OD was obtained by nanodrop for QC and then diluted to 4ng/40 μl.

RNA-Seq

Subsequent RNA-Seq by IIlumina library preparation on a Illumina Hiseq 2000 sequencer was performed by the core genomics facility at the University of Minnesota. 32 libraries were prepared, size selected (average insert size of ~200bp) and barcoded using TruSeq RNA Library preparation kits and sequenced using a 50bp paired-end protocol, yielding on average 20 million reads per library. Average quality scores for all libraries was above Q30. The RNAseq data set is under accession number GSE74201.

Unique reads were mapped to the genome using Bowtie or ClCBio, and differential expression analysis was performed using edgeR (Robinson et al., 2010). False discovery rate (FDR) was set at 0.05.

For WCGNA analysis, reads were aligned to the human hg19 reference genome using the STAR spliced read aligner (Dobin et al., 2013) with default parameters. Total counts of read-fragments aligned to candidate gene regions were obtained using the HTSeq program (Anders et al., 2015) and used as a basis for quantification of gene expression. Differential gene expression analysis between sample groups of interest was

performed using edgeR (Robinson et al., 2010) and the FDR set at 0.05. Normalized expression levels were used to perform WCGNA using the WGCNA package (Langfelder and Horvath, 2008) as previously described (Gu et al., 2015). Briefly, a topological overlap matrix was generated, and genes were clustered based on topological overlap, a measure of network connectivity. Using dynamic tree cutting, groups of coexpressed genes, or modules, were identified.

Enrichment and Networking Analysis

Relevant gene lists or module members were inputted into Enrichr tools (Mount Sinai School of Medicine) (Chen et al., 2013). Enrichment terms are scored by p-value, z-score, and combined score, and shown in figures sorted by combined score. Gene members of relevant terms were further investigated using Genemania (University of Toronto). In a recent paper by Onorati et al. 2014, members of the M25 WGCNA module were identified as being highly expressed in the human striatum (Onorati et al., 2014). We compared the gene symbols found in our black module with those found in the M25 module using the hypergeometric statistical test.

A permutation test was performed to assess over-representation of module genes in HD relevant gene sets curated from the literature (GeneSet lists), namely, striatum-specific mouse genes (top 100) identified from the Allen Brain Atlas (Lein et al. 2007), human caudate (Strand et al. 2007a) and developing human striatum (Onorati et al. 2014) enriched gene sets, medium spiny neuron enriched gene sets from three publications (Hodges et al. 2006, Heiman et al. 2008 and Lobo et al. 2006), and

deregulated genes in a BDNF knockout mouse model (Strand et al. 2007b). For the 2 HD-relevant modules (black and red), overlap with each GeneSet list was determined, as well as the overlap of randomly generated gene lists from the background (with the same number of genes in the module) with each GeneSet list (10,000 permutations). Z-score was calculated as the number of standard deviations our observed value was above the mean of the observed values within the empirical null distribution. The background was set to the total list of genes expressed in our data set (n=13,952 genes).

Additionally, the ENCODE ChIP-Seq significance tool (Auerbach et al. 2013) was used to identify enriched transcription factors in the 2 HD-relevant modules (black and red). Briefly, this tool determines the number of genes in each gene module list with at least one TF peak in the selected window and then calculates enrichment scores (Q-value) using one-tailed hypergeometric test followed by multiple hypothesis correction (FDR method). For our queries we used an upstream and downstream analysis window of 1,000bp from the transcription start site. The background was set to the total list of genes expressed in our data set (n=13,952 genes).

Hierarchical Clustering Analysis

In order to complement our WGCNA analysis, we performed hierarchical clustering using an independent analysis method. Genes displaying significant changes were log2 transformed, while genes with data missing in 8 or more samples (25% of total samples) were excluded. The resulting data (7290 genes, 32 arrays) were clustered by centroid linkage hierarchical clustering using Cluster 3.0 (de Hoon et al. 2004; Eisen et al. 1998).

Modules were analyzed by Genemania (Mostafavi et al., 2008).

RT-qPCR

RNA (1µg) was converted to cDNA by using the Message Sensor RT kit (Life Technologies). Real time quantitative PCR (qPCR) was performed with Universal Probe Library dye (UPL from Roche) on the LightCycler 480 system (Roche). For quantification the threshold cycle Cp of each amplification was determined by the 2nd derivative analysis provided by the LightCycler 480 software and the $2^{-\Delta\Delta Cp}$ method was used to determine the relative expression level of each gene normalized against the house-keeping gene β-actin (ActB). The primers for DARPP-32 (PPP1R1B) were forward: 5'-CACACCACCTTCGCTGAAA-3', reverse: 5'-GAAGCTCCCCCAGCTCAT-3'; netrin-1 (NTN1), forward 5'-CCCTGCATAAAGATCCCTGT-3', reverse 5'-TTGCAGTAGGAATCGCAGTC-3'; TGFβ1 (TGFB1), forward 5'-AGTGGTTGAGCCGTGGAG-3', reverse 5'-TGCAGTGTGTTATCCCTGCT-3'. The primers for β-actin (ACTB) were forward: 5'-CCAACCGCGAGAAGATG-3', reverse: 5'-CCAGAGGCGTACAGGGATAG-3'.

Western Blot Analysis

HD or corrected NSCs were cultured to 90-100% confluency. Media was removed and cells were lysed by sonication in an M-PER (Thermo-Scientific) solution containing protease inhibitors (Complete mini tablet per 10 ml buffer; Roche), phosphatase inhibitors (100 µl/10 ml buffer; Calbiochem # III and IV), and DNase (New England Biolabs; 100 units/10 ml buffer;). For TGF-β western blotting, 20 µg of each protein

sample was loaded onto a 12% Bis-Tris gel (Invitrogen) in MOPS buffer (Life Technologies) and run at 200 volts for 50-60 min. Protein bands were transferred onto a 0.45 PVDF membrane (Millipore) that was previously activated by soaking in 100% methanol. Transfer buffer was supplemented with 10% methanol, and the transfer was carried out at 4°C for 1h at 350 milliAmp. Following transfer, membranes were blocked in 5% milk for 1h and then probed over night with rabbit anti-TGF-β (Abcam, ab66043) at 1:100-1:500. The following day, blots were washed and then probed with anti-rabbit HRP secondary (GE Healthcare) for 1h at 1:4000. For pSMAD2 and SMAD2 western blotting, protein samples were run on a 4-12% Bis-Tris gel at 200 volts for 1h. Protein was transferred to nitrocellulose membranes (Whatman) and blots were blocked and probed in 5% milk. Anti-pSMAD2 and SMAD2 were used at 1:200-1:500 (Cell Signaling Technologies, #3101, #5339). Tubulin antibody was used at 1:2000 (Sigma, T6199) and rabbit β-actin at 1:500 (Cell Signaling, #4968S).

Caspase-3/7 Assay

HD and corrected NSCs were passaged and grown to confluency in neural proliferation medium (NPM) supplemented with 25 ng/ml FGF-2. 96 well plates were coated with matrigel for 1h at 37^{o}C before seeding 3x10^{4} NSCs per well in NPM with 25ng/ml FGF-2. 24h post seeding, media was removed and cells were washed once with PBS. Cells were then treated for 48h with complete medium (NPM, FGF-2), basic medium (neurobasal media only), or basic medium with TGF-β1 or TGF-β2 at 10 ng/ml (Peprotech). After 48h,

caspase-3/7 activity assays were conducted using the APO3 HTS kit and the Fusion Alpha Universal Microplate Analyzer, as previously described (Zhang et al., 2010).

Respirometry Analysis

The respirometry analysis of NSCs was carried out using an XF24 extracellular flux analyzer (Seahorse Bioscience). Both HD NSCs and corrected C116 NSCs were seeded in an XF24 cell culture microplate with standard culture conditions. 90% confluent cells were pre-treated with 100ng/ml netrin-1 or 50ng/ml TGF-β1 for 24h in normal full medium followed by 12h treatment in basic medium (Neurobasal medium only) with same doses of netrin-1 or TGF-β1. Right before respirometry analysis, cells were washed three times with "Seahorse buffer" (120mM NaCl, 3.5mM KCl, 1.3mM $CaCl_2$, 0.4mM KH_2PO_4, 20mM TES, 1.2mM Na_2SO_4, 2mM $MgCl_2$, 15mM glucose, 15mM sodium pyruvate, pH7.4) and left in "Seahorse buffer" at 37^{o}C. The plate was loaded into the XF24 analyzer for a program with 13 cycles of 1min mix, 2min wait and 3min measurement. 2μg/ml oligomycin was added into cells after the 4th cycle. 1uM FCCP was added after the 7th cycle. 2uM rotenone and 1uM antimycin-A were added after the 10th cycle. The oxygen consumption was shown as pMole O_2/min after normalization against protein concentration measured by BCA protein assay kit.

zQ175 Knockin Mice

The mouse strain was housed at the Animal Facility of the Buck Institute for Research on Aging. We are an AAALAC international accredited institution (Unit #001070) and all

procedures were approved by the Institutional Animal and Care and Use Committee (A4213-01). Wild type and zQ175 mouse brains were dissected into the indicated regions. Samples were flash frozen on dry ice and stored immediately at -80^{o}C until tissue homogenization. Homogenization buffer consisted of 10ml TPER (Thermo Scientific, cat#78510) supplemented with protease inhibitors (Roche, 1 tablet/10 ml, Complete Mini, EDTA-free, cat#11836153001), 1µl DNAse (Invitrogen, cat#18047-019, 250U/µl), 1.2 mM $MgCl_2$ (Fluka, cat# 63543), 1 µM Epoxomycin (Sigma, cat#D4321-1MG), 100 µl phosphatase inhibitor cocktail II (Calbiochem PPI II, cat#524625). HDAC activity was inhibited by a cocktail of HDAC inhibitors composed of 50 µM TSA (Sigma, cat#T8552-5MG), 30 mM Nicotinamide (Sigma, cat#N0636-100G) and 30 µM sodium butyrate (Sigma, cat#19364) added to the resuspension buffer. A 2ml dounce homogenizer was used for cell lysis on ice (2 x 60 pumps with a 30 s interval). Samples were stored at -80^{o}C prior to sonication. Tissue lysates were sonicated with continuous 40 mA pulses for 5s x5 on ice. Sonicated samples were centrifuged at 14,000 g and 4^{o}C on a 5417 Eppendorf centrifuge for 20 min. Supernatant was collected for western blot analysis as described above.

Statistical Comparisions

Statistical significance between the different groups was performed by 2-way ANOVA Graphpad Prism software (La Jolla, CA) and comparisons were significant when * $p < 0.05$, ** $p< 0.005$, *** $p< 0.001$ and **** $p< 0.0005$. Data is presented as mean and SD. Biological replicates (BR) refers to the sample grown in separate culture plates under

the similar conditions. Statistical analysis was performed on at least 3 independent replicates or animal samples.

REFERENCES

An, M.C., Zhang, N., Scott, G., Montoro, D., Wittkop, T., Mooney, S., Melov, S., and Ellerby, L.M. (2012). Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells. Cell Stem Cell *11*, 253-263.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Auerbach RK, Chen B, Butte AJ (2013). Relating genes to function: identifying enriched transcription factors using the ENCODE ChIP-Seq significance tool. Bioinformatics *29*, 1922-1924.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC bioinformatics *14*, 128.

de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453-1454.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. PNAS *95*, 14863-14868.

Gu, X., Cantle, J.P., Greiner, E.R., Lee, C.Y., Barth, A.M., Gao, F., Park, C.S., Zhang, Z., Sandoval-Miller, S., Zhang, R.L*., et al.* (2015). N17 Modifies Mutant Huntingtin Nuclear Pathogenesis and Severity of Disease in HD BAC Transgenic Mice. Neuron *85*, 726-741.

Heiman M, Schaefer A, Gong S, Peterson JD, Day M, Ramsey KE, Suárez-Fariñas M, Schwarz C, Stephan DA, Surmeier DJ et al. (2008). A translational profiling approach for the molecular characterization of CNS cell types. Cell *135*,738-748.

Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, Elliston LA, Hartog C, Goldstein DR, Thu D et al. (2006) Regional and cellular gene expression changes in human Huntington's disease brain. Hum Mol Genet *15*, 965-977.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168-176.

Lobo MK, Karsten SL, Gray M, Geschwind DH, Yang XW (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. Nat Neurosci *9*, 443-452.

Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C., and Morris, Q. (2008). GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome biology *9 Suppl 1*, S4.

Onorati, M., Castiglioni, V., Biasci, D., Cesana, E., Menon, R., Vuono, R., Talpo, F., Goya, R.L., Lyons, P.A., Bulfamante, G.P.*, et al.* (2014). Molecular and functional definition of the developing human striatum. Nat Neurosci *17*, 1804-1815.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Strand AD, Aragaki AK, Baquet ZC, Hodges A, Cunningham P, Holmans P, Jones KR, Jones L, Kooperberg C, Olson JM (2007a). Conservation of regional gene expression in mouse and human brain. PLoS Gen 3(4):e59.

Strand, A.D., Baquet, Z.C., Aragaki, A.K., Holmans, P., Yang, L., Cleren, C., Beal, M.F., Jones, L., Kooperberg, C., Olson, J.M.*, et al.* (2007b). Expression profiling of Huntington's disease models suggests that brain-derived neurotrophic factor depletion plays a major role in striatal degeneration. J Neurosci *27*, 11758-11768.

Zhang, N., An, M.C., Montoro, D., and Ellerby, L.M. (2010). Characterization of Human Huntington's Disease Cell Model from Induced Pluripotent Stem Cells. PLoS Curr *2*, RRN1193.