SUPPLEMENTARY FIGURES AND TABLE

hESC-based human glial chimeric mice reveal glial differentiation defects in Huntington disease

Mikhail Osipovitch[1*], Andrea Asenjo-Martinez[1*], John Mariani[2], Adam Cornwell[2], Simrat Dhaliwal[2], Lisa Zou[2], Devin Chandler-Militello[2], Su Wang[2], Xiaojie Li[2], Sarah-Jehanne Benraiss[2], Robert Agate[2], Andrea Lampp[1], Abdellatif Benraiss[2], Martha S. Windrem[2], Steven A. Goldman[1,2,3**]

[1]Center for Translational Neuromedicine, University of Copenhagen Faculty of Health and Medical Science, 2200 Copenhagen N, Denmark; [2]Center for Translational Neuromedicine, University of Rochester Medical Center, Rochester, NY, 10021, USA; [3]Neuroscience Center, Rigshospitalet-Copenhagen University Hospital, Copenhagen, Denmark.

SUPPLEMENTARY INFORMATION

Includes:

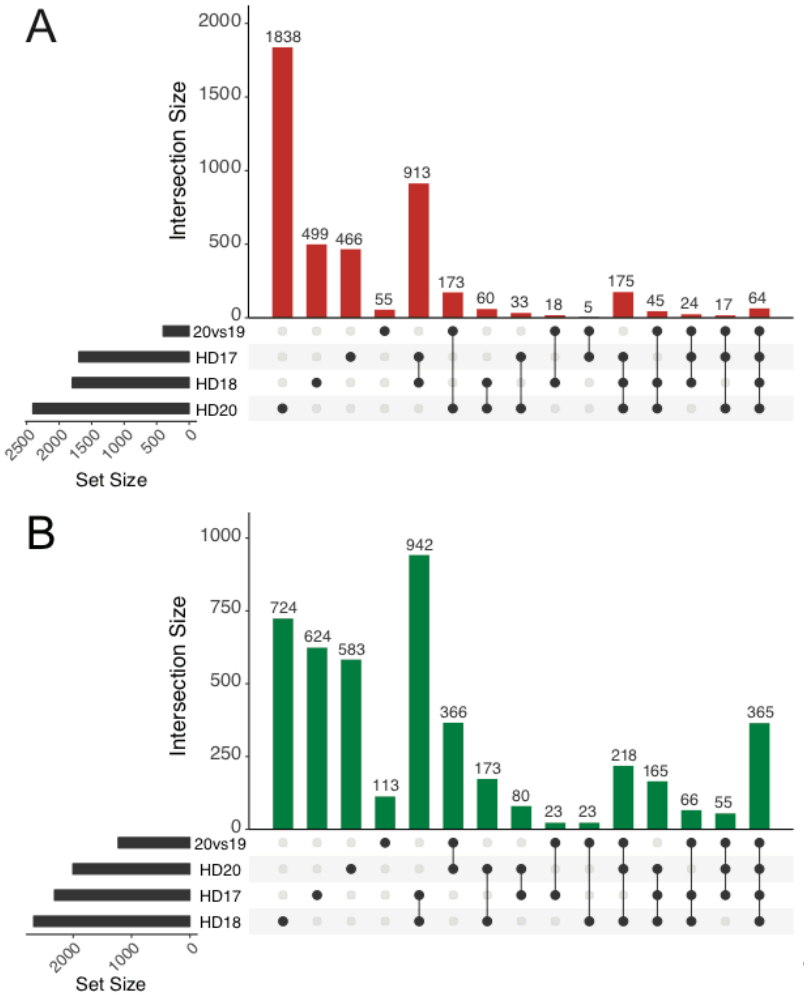
7 Supplementary Figures

2 Supplementary Tables

SUPPLEMENTARY FIGURES

Figure S1 *(Supplementary to Figure 1)*

Genes differentially-expressed between hGPCs derived from different HD hESCs vs. pooled controls

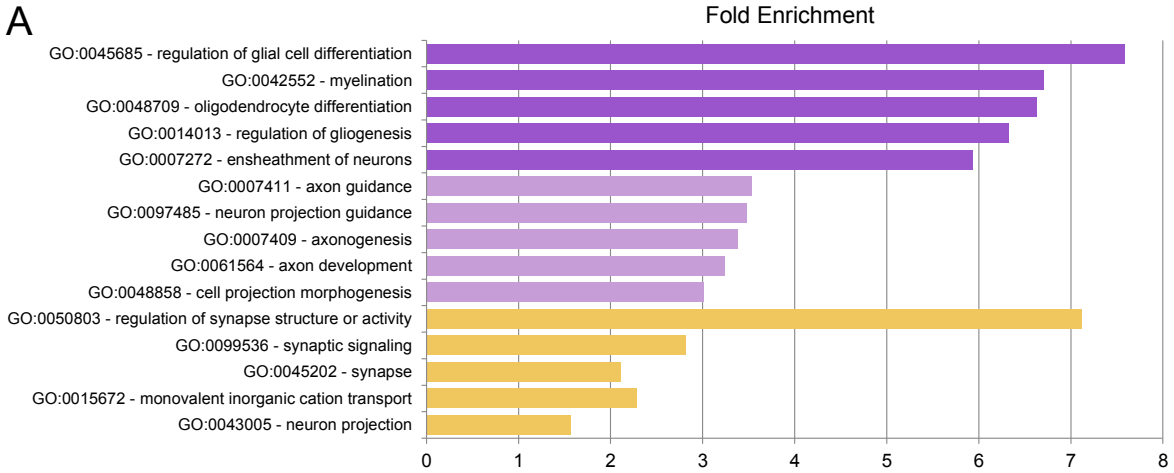


A-B, Gene set intersection plots for differentially expressed genes obtained from comparisons of each CD140a-sorted HD-derived GPC line (HD17, HD18, and HD20), compared to pooled control-derived GPCs (**A,** up-regulated genes; **B,** down-regulated genes). Differentially expressed genes in HD GPCs are significant at 1% FDR and FC > 2.00. **C-D,** CD44-sorted HD-derived APC line (HD17, HD18, and HD20) against control-derived APCs (**C,** up-regulated; **D,** down-regulated). Differentially expressed genes in HD APCs are significant at 5% FDR. In both, 20 vs 19 denotes the comparison of HD line HD20 (Genea20) against its sibling control line CTR19 (Genea19). Horizontal bars represent total sizes of gene sets, and vertical bars represent sizes of gene set intersections. Vertical bars are ordered first by the number of gene sets in the intersection, and then by the size of the intersection. The dots correspond to those gene sets comprising each intersection.

Figure S2 *(Supplementary to Figure 1)*

Functional annotation reveals HD-associated impairment in transcription of glial differentiation, myelination, and synaptic transmission-related genes

A



B

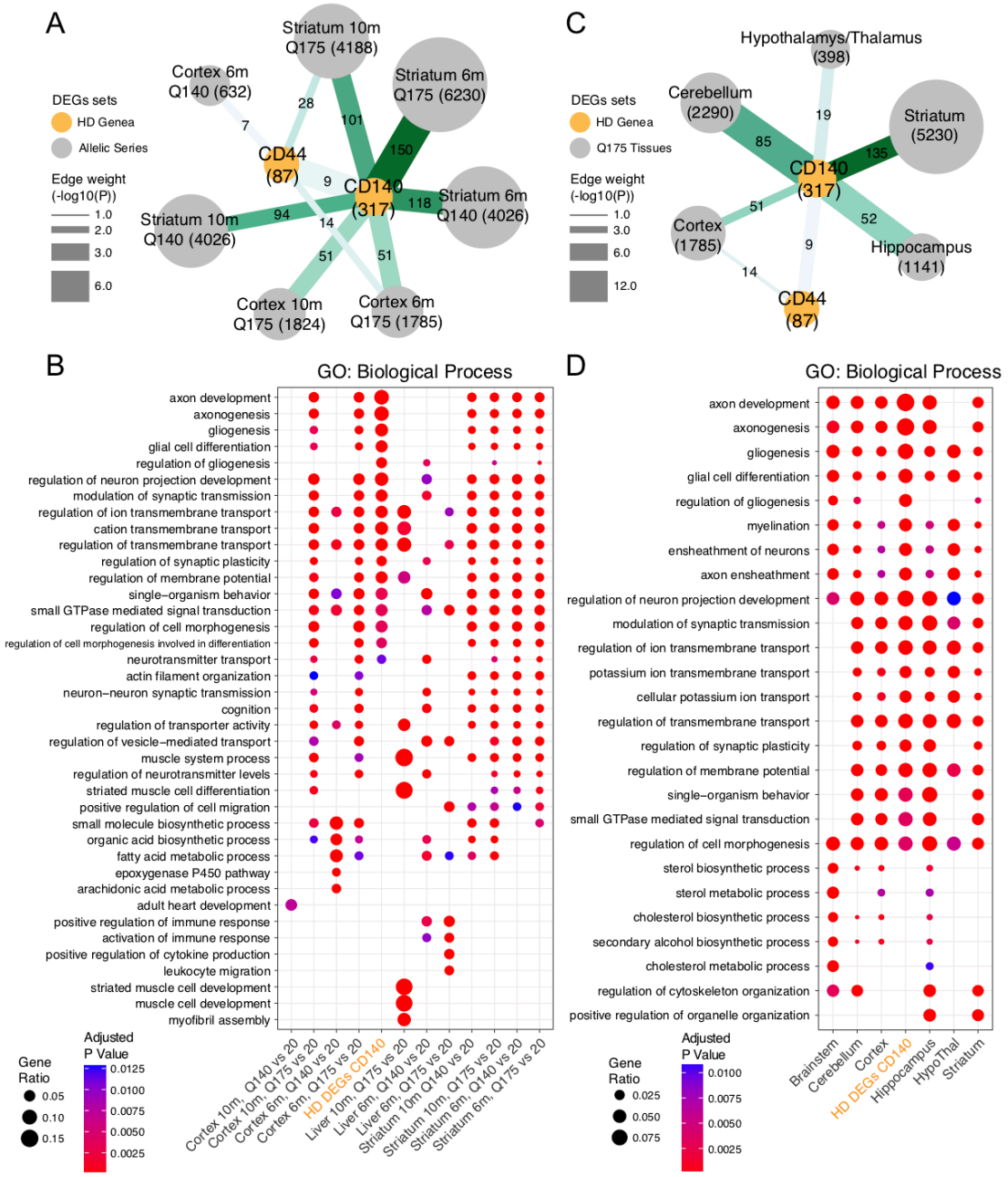
Module	Term ID	Title	P Value	Total in Term	Observed	Expected	Fold Enrichment	Associated Genes
M1	GO:0045685	regulation of glial cell differentiation	> 1.00E-10	59	10	1.32	7.58	BMP2, LINGO1, MAG, NKX2-2, NR2E1, NTRK3, OLIG2, SERPINE2, SIRT2, TCF7L2
	GO:0042552	myelination	> 1.00E-10	100	15	2.24	6.71	FA2H, GAL3ST1, MAG, MBP, MYRF, NFASC, OLIG2, OMG, PLLP, POU3F2, SIRT2, SLC8A3, TCF7L2, TF, UGT8
	GO:0048709	oligodendrocyte differentiation	> 1.00E-10	81	12	1.81	6.63	FA2H, GLI3, LINGO1, MYRF, NKX2-2, OLIG1, OLIG2, OMG, SIRT2, SLC8A3, SOX10, TCF7L2
	GO:0014013	regulation of gliogenesis	> 1.00E-10	99	14	2.21	6.32	BMP2, LINGO1, MAG, MYC, NKX2-2, NR2E1, NTRK3, OLIG2, SERPINE2, SIRT2, SOX10, TCF7L2, TF, ZCCHC24
	GO:0007272	ensheathment of neurons	> 1.00E-10	113	15	2.53	5.94	FA2H, GAL3ST1, MAG, MBP, MYRF, NFASC, OLIG2, OMG, PLLP, POU3F2, SIRT2, SLC8A3, TCF7L2, TF, UGT8
M2	GO:0007411	axon guidance	> 1.00E-10	228	18	5.10	3.53	ALCAM, BCL11B, DSCAM, FOXD1, GAS1, GLI3, HOXA1, HOXA2, MNX1, NFASC, PLXNC1, PRKCQ, PTPRO, ROBO2, SEMA6B, UNC5A, VAX1, WNT7B
	GO:0097485	neuron projection guidance	> 1.00E-10	231	18	5.16	3.49	ALCAM, BCL11B, DSCAM, FOXD1, GAS1, GLI3, HOXA1, HOXA2, MNX1, NFASC, PLXNC1, PRKCQ, PTPRO, ROBO2, SEMA6B, UNC5A, VAX1, WNT7B
	GO:0007409	axonogenesis	> 1.00E-10	423	32	9.46	3.38	ADGRB1, ALCAM, BCL11B, CACNA1A, DSCAM, FOXD1, GAS1, GLI3, HOXA1, HOXA2, LINGO1, LRRC4C, MAG, MBP, MNX1, NFASC, NR2E1, NTNG1, NTRK3, OMG, PLXNC1, POU3F2, PRKCQ, PTPRO, ROBO2, SEMA6B, SLITRK2, SLITRK3, SNAP91, UNC5A, VAX1, WNT7B
	GO:0061564	axon development	> 1.00E-10	469	34	10.49	3.24	ADGRB1, ALCAM, BCL11B, CACNA1A, DSCAM, FOXD1, GAS1, GLI3, HOXA1, HOXA2, LINGO1, LRRC4C, MAG, MBP, MNX1, NEFM, NFASC, NR2E1, NTNG1, NTRK3, OMG, PLXNC1, POU3F2, PRKCQ, PTPRO, ROBO2, RTN4RL2, SEMA6B, SLITRK2, SLITRK3, SNAP91, UNC5A, VAX1, WNT7B
	GO:0048858	cell projection morphogenesis	> 1.00E-10	608	41	13.59	3.02	ADGRB1, ALCAM, BCL11B, CACNA1A, CAMK2A, DSCAM, EHD3, FOXD1, GAS1, GLI3, HOXA1, HOXA2, KANK1, LINGO1, LRRC4C, MAG, MBP, MNX1, NEDD4L, NEURL1, NFASC, NR2E1, NTNG1, NTRK3, OMG, PCDH15, PLXNC1, POU3F2, PRKCQ, PTPRO, ROBO2, SEMA6B, SGK1, SLITRK2, SLITRK3, SNAP91, SNX10, UGT8, UNC5A, VAX1, WNT7B
M3	GO:0050803	regulation of synapse structure or activity	> 1.00E-10	132	21	2.95	7.12	ADGRB1, ADGRL3, BCAN, CALB1, CAMK2A, FGF14, LRRTM1, NCDN, NETO1, NEURL1, NR2E1, NTRK3, PPFIA3, ROBO2, SERPINE2, SHISA7, SIX4, SLC8A3, SLITRK2, SLITRK3, SYNDIG1
	GO:0099536	synaptic signaling	> 1.00E-10	604	38	13.50	2.81	BCAN, CACNA1A, CACNA1G, CALB1, CAMK2A, CHRNA4, FGF12, FGF14, GRIA2, GRIA4, GRID2, GRIK4, KCND2, LRRTM1, MBP, MPZ, NCDN, NETO1, NEURL1, NOVA1, NR2E1, P2RX7, PDE7B, PLCL1, PPFIA3, RAPGEF4, RGS8, RIT2, S1PR2, SERPINE2, SHISA7, SLC18A1, SLC1A1, SLC1A2, SLC8A3, SNAP91, SNPH, SYT6
	GO:0045202	synapse	> 1.00E-10	847	40	18.94	2.11	ADGRB1, BCAN, BCAS1, CACNA1A, CALB1, CAMK2A, CHRNA4, CTTNBP2, DSCAM, GRIA2, GRIA4, GRID1, GRID2, GRIK4, HCN2, KCND2, LGI3, LRRC4C, LRRTM1, NETO1, NEURL1, NTM, P2RX7, PCDH15, PDE4B, PPFIA3, PRIMA1, PRKCQ, PTPRO, RAPGEF4, SERPINE2, SHISA7, SLC17A8, SLC18A1, SLC1A2, SLC8A3, SNAP91, SNPH, SYNDIG1, SYT6
	GO:0015672	monovalent inorganic cation transport	1.04E-06	586	30	13.10	2.29	ABCC9, ASIC4, CACNA1A, CHRNA4, CNGB1, CNTN1, DPP10, DPP6, FGF12, FGF14, HCN2, KCND2, KCNJ9, KCNQ1, KCNS3, NALCN, NEDD4L, NKAIN4, P2RX7, PTGER3, SERPINE2, SGK1, SLC10A4, SLC17A8, SLC18A1, SLC22A3, SLC2A13, SLC5A9, SLC8A3, SLC9A7
	GO:0043005	neuron projection	4.17E-06	1251	44	27.97	1.57	ADGRL3, ALCAM, BCAN, BCL11B, CACNA1A, CACNA1G, CALB1, CAMK2A, CHRNA4, CTTNBP2, DSCAM, GRIA2, GRIA4, GRID2, GRIK4, HCN2, KCND2, LGI3, LRRTM1, MAG, MBP, MYC, NCAM2, NCDN, NEFM, NEURL1, NFASC, NTM, PDE4B, PIK3R1, PTGER3, PTPRO, RAPGEF4, RGS8, ROBO2, SGK1, SIRT2, SLC17A8, SLC1A2, SLC8A3, SNAP91, SNPH, SYNDIG1, UNC5A

Figure S2 *(cont'd)*

Gene Ontology (GO) functional annotation was performed for the 429 differentially expressed genes (DEGs) in the 3 lines of mHTT hGPCs relative to pooled control hGPCs (see **Figure 1B-C**). 50 significantly associated GO annotation terms (Biological Process and Cellular Component, Bonferroni-corrected $p<0.01$) were identified by the ToppCluster annotation tool (Kaimal et al., 2010). By network analysis, these GO terms together with their associated DEGs were grouped into three functionally related modules (M1 through M3, see **Figure 1D**). For each GO term, the expected value assumes a constant ratio, given the number of annotated DEGs and the total number of human protein-coding genes found in the term. The fold enrichment is the ratio of the number of observed DEGs found in the term, to the expected number. Within each functional module, the GO terms were ranked first by p value, then by fold-enrichment. Three GO terms, GO:0007268 (chemical synaptic transmission), GO:0098916 (anterograde trans-synaptic signaling), and GO:0099537 (trans-synaptic signaling), were respectively ranked 3 through 5 within module M3. They contained an identical set of 37 associated DEGs, which were contained within the 38 DEGs associated to GO:0099536 (synaptic signaling) ranked at number 2 in M3. To reduce redundancy, these three GO terms were thus omitted from the figure. **A**, The bar graph shows the top 5 GO terms for each functional module. **B**, The table lists the calculated values and the associated DEGs for each of the top-ranked terms. Associated DEGs are color-coded according to their direction of dysregulation in HD- vs. control-derived hGPCs (*green*, down-regulated; *red*, up-regulated).

Figure S3 *(Supplementary to Figure 1)*

Human and mouse glia exhibited overlap in genes dysregulated as a function of CAG repeat length



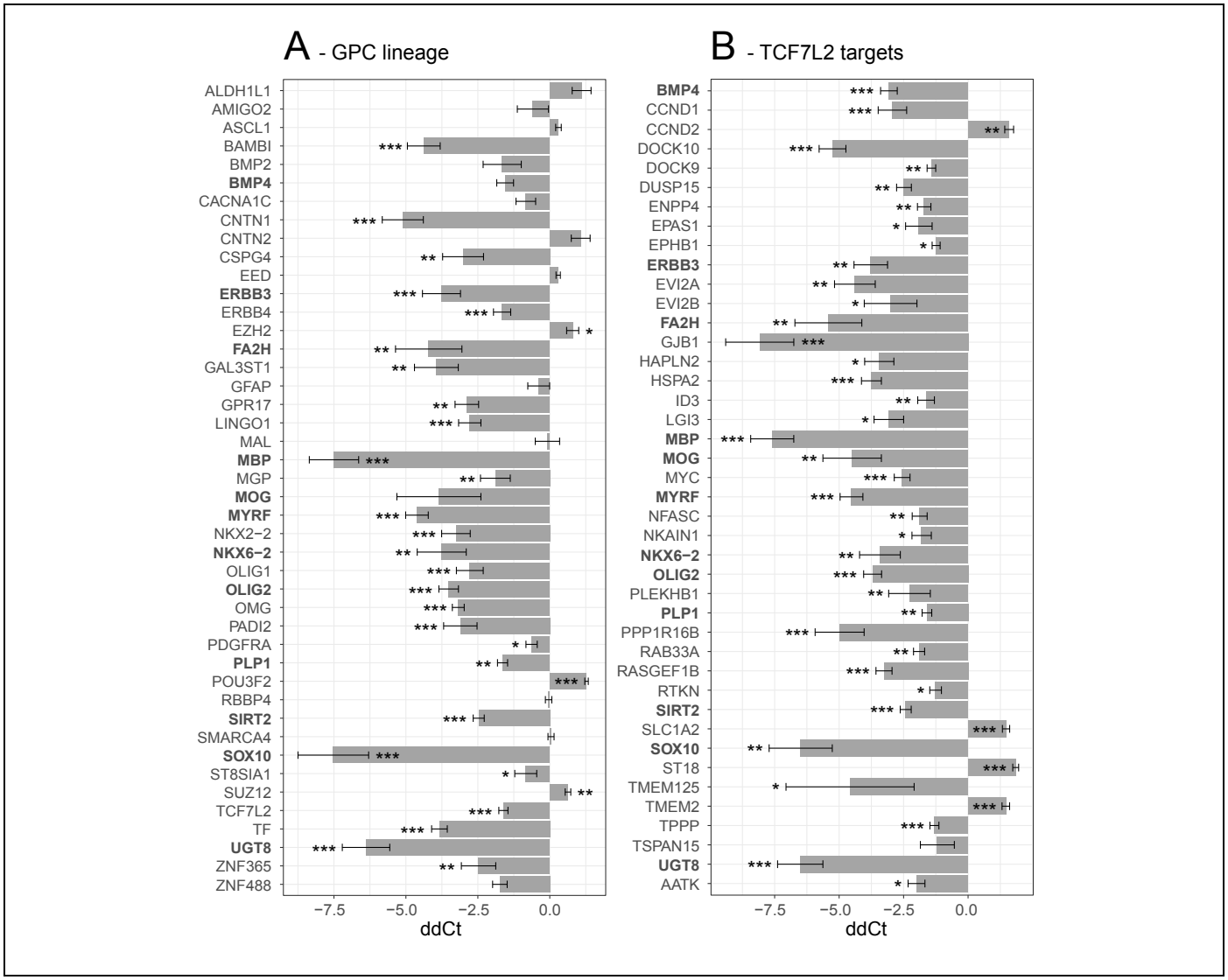
There was a high degree of overlap between those hGPC genes and ontologies found to be increasingly dysregulated with longer CAG repeat length in hGPCs, with those noted to be dysregulated with CAG repeat length in mouse brain tissue (Langfelder et al., 2016). **A,** The representative lists of differentially expressed genes (DEGs) obtained from the HD-derived CD140-sorted GPCs, and the HD-derived CD44-sorted APCs were compared against the differential expression results of the mouse *mHtt* allelic series (**A** and **B**) and the 6-month Q175 profiled tissues (**C**

and **D**) from (Langfelder et al., 2016). The network plots in **A** and **C** show the significant pairwise set intersections between the CD140 and CD44 HD Genea-derived DEGs sets (*yellow* nodes), and the DEGs sets from the Langfelder et al. (2016) analysis (*grey* nodes) (Fisher's exact test, $p<0.05$). The nodes are sized according to the total number of DEGs, indicated in parenthesis for each node. The numbers of DEGs in the HD Genea sets are post-ID conversion to mouse orthologue genes. The edge thickness indicates the significance of the gene set intersection, calculated as -log10 (Fisher's exact test p value). Edge color and label show the number of genes in the pairwise set intersection. Only the Langfelder et al. (2016) DEG sets that had a significant overlap to either of the two HD Genea sets are shown. The dot plots in **B** and **D** show the comparisons of Gene Ontology (GO): Biological Process annotation results for the DEGs sets in **A** and **C**, respectively. The dots are sized according to the gene ratio with respect to the DEGs set. The dot color represents the significance of the association to the GO term. All DEGs sets that had significant annotation (BH-corrected $p< 0.01$) are shown.

The most significant intersections were observed between the CD140 DEGs set and the DEGs in the 6-month striatum Q175 samples ($p=1.10E-06$; 150 genes) in the comparison to the allelic series DEGs and between the CD140 DEGs set and the 6-month Q175 cerebellum DEGs for the Q175 tissues ($p=9.86E-13$; 85 genes). These intersections included the glial modulators Nkx2-2, Olig1, and Olig2 as well as the genes encoding proteins involved in myelination, ion channel activity, and synaptic transmission. Overall, a number of similar significant annotations were observed for the HD Genea CD140 DEGs and the brain-derived DEGs from Langfelder et al. (2016), implicating functions that included gliogenesis, myelination, axon development, and ion channel activity.

Figure S4 *(Supplementary to Figure 1)*

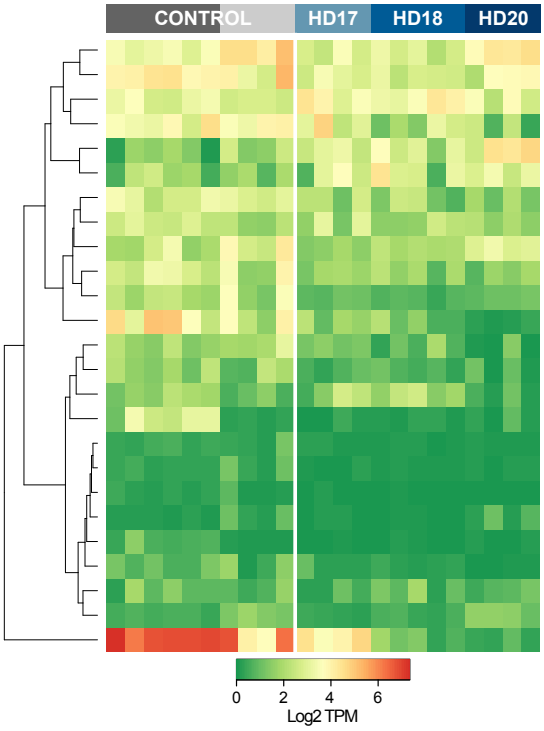
Glial differentiation-associated genes are dysregulated in mHTT-expressing GPCs



Expression of selected genes dysregulated in HD-derived GPCs, as identified by RNA-seq analysis, was assessed by TaqMan Low Density Array (TLDA) RT-qPCR and compared to that of control GPCs. Expression data were normalized to 18S and GAPDH endogenous controls. Mean ddCt values and standard error ranges calculated from 3 pooled HD GPC lines (n = 3 for lines GENEA17 and GENEA20, n = 5 for GENEA18, total n = 11) vs. 2 pooled control GPC lines (n = 6 for GENEA02 and n = 3 for GENEA19, total n = 9) are shown. The difference of expression in HD and control GPCs was assessed by paired t-tests, followed by Benjamini-Hochberg (BH) multiple testing correction (***p< 0.01, **p<0.05, *p<0.1). Genes assayed on both arrays are highlighted in bold. Analysis of TLDA data was performed in ExpressionSuite software v.1.1 (Applied Biosciences). The majority of genes identified by RNA-seq as dysregulated in HD-derived GPCs were confirmed as such by TLDA. **A,** Genes encoding key GPC lineage transcription factors and stage-regulated, myelin-related proteins. 44 genes are shown, excluding MOBP and MOG, which were noted to have a high proportion of unreliable reactions. **B,** Transcriptional targets of TCF7L2, as predicted by upstream regulator analysis in IPA (https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). A total of 42 genes are shown, excluding four genes that had a high proportion of unreliable reactions.

Figure S5 *(Supplementary to Figure 1)*

HD-derived hGPCs showed marked dysregulation of potassium channel genes

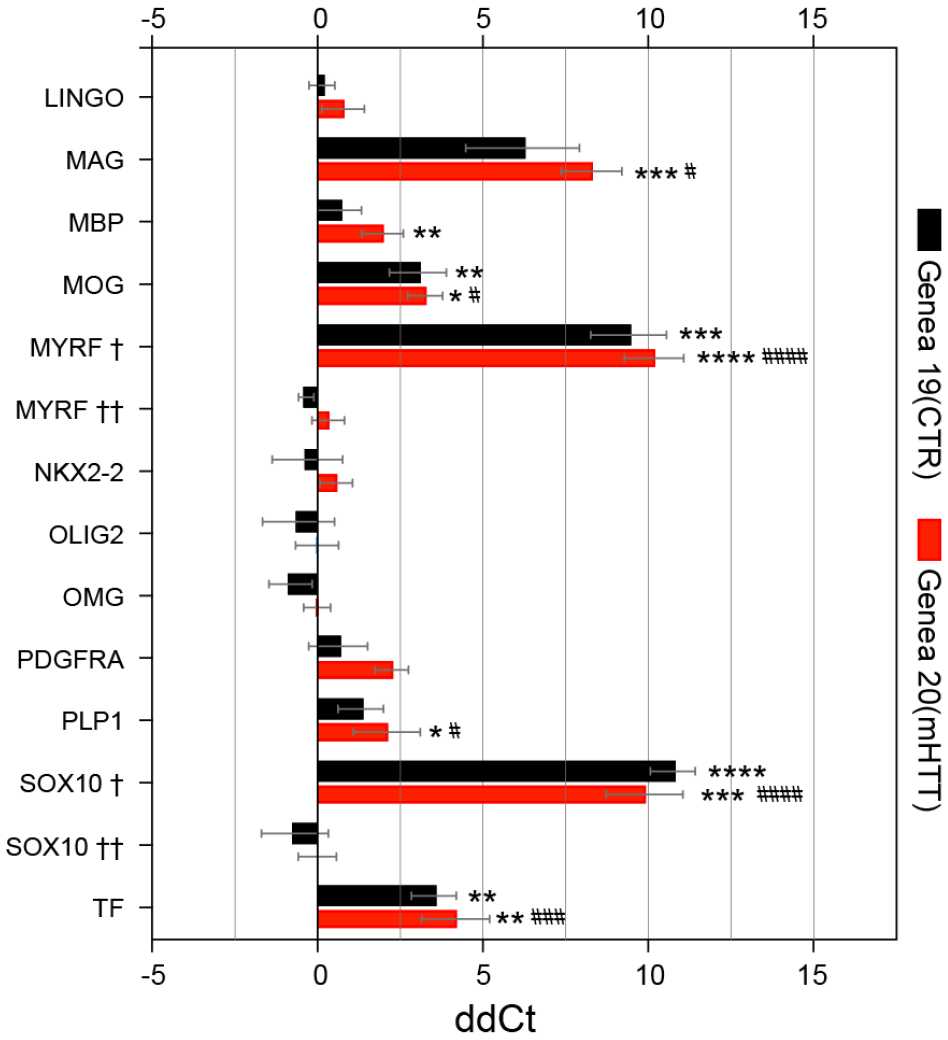


	HD17 vs. Pooled CTR		HD18 vs. Pooled CTR		HD20 vs. Pooled CTR	
	Log2 FC	P Value	Log2 FC	P Value	Log2 FC	P Value
KCNIP2	-0.852	2.16E-02	-1.179	4.07E-05	1.226	3.38E-03
KCNA2	-1.047	2.32E-02	-1.436	2.96E-04	NS	NS
KCNJ8	1.127	4.89E-04	1.081	7.77E-05	NS	NS
KCNIP1	NS	NS	-1.496	2.07E-03	-2.463	1.82E-04
KCNH7	1.735	6.12E-03	1.683	6.22E-03	2.755	1.84E-05
KCNA5	1.623	1.36E-02	1.942	2.02E-03	1.501	5.23E-03
KCNK10	NS	NS	-1.243	1.22E-02	-1.882	5.22E-06
KCNIP4	NS	NS	-4.859	5.94E-05	-0.718	2.94E-02
KCNIP3	-1.475	7.41E-03	-0.913	4.67E-02	NS	NS
KCNB1	-1.318	1.84E-02	-1.308	1.32E-02	-2.030	6.24E-04
KCNJ9	-2.000	4.50E-04	-2.715	7.15E-08	-2.283	4.61E-05
KCNS3	-2.229	1.05E-03	-2.835	3.51E-05	-6.096	2.31E-13
KCNJ16	NS	NS	-1.342	1.14E-02	-3.085	1.27E-03
KCNK1	-1.719	3.95E-03	-1.653	1.64E-03	-2.186	1.09E-02
KCNK2	NS	NS	1.208	5.12E-03	-1.682	5.81E-03
KCNV2	-3.120	4.52E-05	-2.785	9.01E-06	-1.680	3.21E-02
KCNN2	-1.389	1.44E-02	-2.140	1.94E-04	-2.631	1.37E-05
KCNQ1	-2.137	7.22E-04	-2.237	8.84E-06	-2.678	1.81E-05
KCNG4	-2.452	1.59E-03	-4.110	6.61E-07	-5.990	3.37E-07
KCNH4	-1.471	3.47E-02	-2.583	2.15E-05	NS	NS
KCNH6	-3.066	1.13E-05	-1.794	1.08E-02	-1.663	1.85E-02
KCNQ5	NS	NS	-1.971	8.12E-04	-5.332	6.58E-09
KCNMB2	NS	NS	1.117	2.32E-02	1.691	1.92E-04
KCNQ4	-1.335	1.58E-04	-0.861	4.76E-03	NS	NS
KCND2	-1.936	1.67E-04	-5.794	7.61E-18	-7.615	2.50E-24

Differential gene expression comparisons (FDR 5%, no fold change threshold) of each HD-derived hGPC line against pooled control hGPCs revealed 25 potassium channel genes that were dysregulated in at least 2 out of 3 HD-derived lines. NS = not significant.

Figure S6 *(see Table S1)*

SOX10-MYRF transduction restores myelin gene expression in mHTT GPCs



This figure shows a graphical representation of the qPCR data outlined in **Table S1**. Expression values normalized to 18S and control plasmid-transfected cells of selected oligoneogenic and myelinogenic genes in both normal (Genea19, *black bars*) and mHTT-expressing (Genea 20, *red*) hGPCs, after transfection with a bicistronic plasmid expressing SOX10 and MYRF.

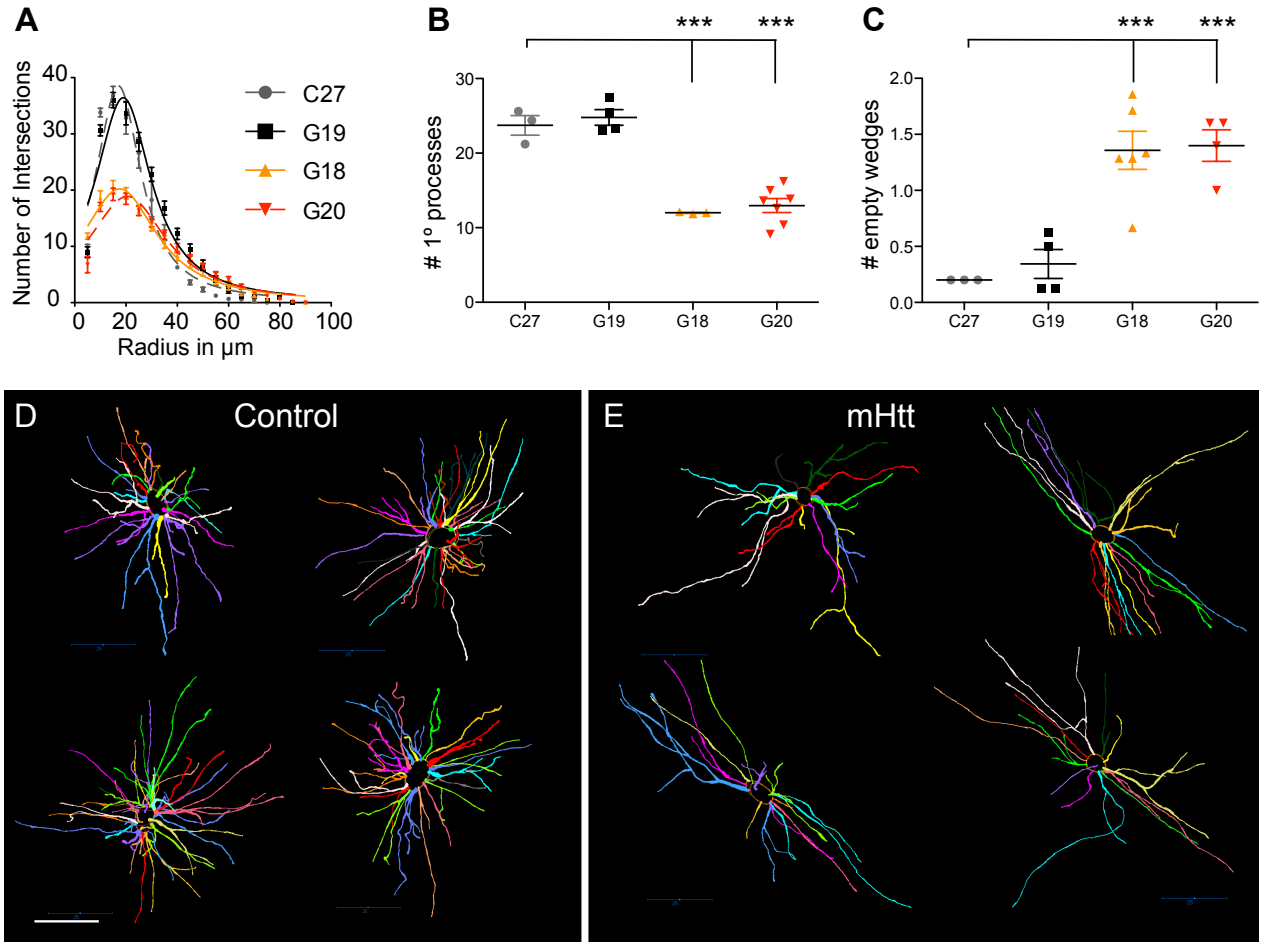
Welch's *t*-test comparisons of: 1) SOX10-MYRF- vs EGFP-transfected for each line independently, significance indicated by *asterisks*; or 2) SOX10-MYRF-transfected Genea 20, vs. EGFP control-transfected Genea19 (significance indicated by *hash marks*).

*/# $p<0.05$. **/## $p<0.01$.; ***/### $p<0.001$.; ****/#### $p<0.0001$.

†, Primers located on coding sequence; ††, primers located in 3'UTRs.

Figure S7 *(Supplementary to Figure 6)*

mHTT-expressing astrocytes exhibit diminished complexity and incomplete domain structures



A, Sholl analysis of GFAP-immunostained human cells in human glial chimeras, 18 weeks after neonatal implantation. Non-linear regression curves of radial intersections for each cell line (Lorentizan curve-fit), as a function of branch order. Comparison of control (N=7) vs mHTT mice (N=10); $p<0.0001$. **B,** Both the normal HTT control line GENEA19, and the unrelated normal HTT hiPS cell line C27 have more primary processes than the mHTT-expressing GENEA lines, GENEA18 and GENEA20. The controls GENEA19 and C27 are no different from one another, but both GENEA18 and GENEA20 are significantly different from the controls (1-way ANOVA with Dunnett's post-ttest; $p<0.0001$). **C,** The fiber distributions of astrocytes derived from the two control lines, C27 and GENEA19, are more radially symmetric than those of either mHTT line. One-way ANOVA with Dunnett's post-test, and C27 as the control, $p<0.0001$. Both GENEA18 and GENEA20 are significantly different from C27, $p<0.0001$. **A-C**, Controls: C27, *gray*; and GENEA 19, *black*. HD-derived: GENEA 18, *orange*; GENEA 20, *red*.

D, Flattened 3-dimensional coronal tracings of astrocytes from the corpus callosum of mice transplanted with C27-derived control hGPCs, compared to those of mice transplanted with GENEA 18-derived hGPCs (**E**). Scale: **D**, 25 µm.

Table S1

SOX10-MYRF transduction restores myelin gene expression in mHTT GPCs

Target gene	GENEA-20 (mHTT) ddCt ± SEM (p-value)	GENEA-19 (normal HTT) ddCt ± SEM (p-value)
LINGO1	0.78 ± 0.64 (p=0.41)	0.14 ± 0.39 (p=0.57)
MAG	8.29 ± 0.92 (p=0.0001)*	6.21 ± 1.72 (p=0.01)*
MBP	1.97 ± 0.63 (p=0.005)*	0.67 ± 0.66 (p=0.4)
MOG	3.26 ± 0.53 (p=0.02)*	3.04 ± 0.86 (p=0.009)*
MYRF-Endo†	0.33 ± 0.49 (p=0.6)	-0.34 ± 0.23 (p=0.18)
NKX2.2	0.57 ± 0.49 (p=0.6)	-0.30 ± 1.06 (p=0.85)
OLIG2	-0.01 ± 0.65 (p=0.99)	-0.57 ± 1.09 (p=0.79)
OMG	-0.01 ± 0.41 (p=0.98)	-0.81 ± 0.66 (p=0.22)
PDGFRA	2.25 ± 0.51 (p=0.05)	0.63 ± 0.89 (p=0.57)
PLP1	2.10 ± 1.01 (p=0.04)*	1.31 ± 0.69 (p=0.19)
SOX10-Endo†	0.00 ± 0.58 (p>0.99)	-0.68 ± 1.01 (p=0.59)
TF	4.18 ± 1.03 (p=0.008)*	3.52 ± 0.68 (p=0.004)*
MYRF-viral††	10.18 ± 0.90 (p<0.0001)*	9.41 ± 1.15 (p=0.0003)*
SOX10-viral††	9.89 ± 1.16 (p=0.0002)*	10.75 ± 0.68 (p<0.0001)*

These qPCR data show the ddCT values, reflecting the relative mRNA levels, of selected oligoneogenic and myelinogenic genes in normal and mHTT-expressing hGPCs, after transfection with a bicistronic plasmid expressing SOX10 and MYRF, after normalization to 18S and then control plasmid-transfected cells. Welch's *t*-test. †, Primers located on coding sequence †† Primers located in 3'UTRs. Endo: endogenous gene; Viral: viral transgene product. *p<0.05.

Supplementary Table S2

Primers used for real time PCR

Target	Forward primer	Reverse primer
LINGO1	ACCTTCGCTTTCATCTCCAAC	CGATGATGAGGGTCTTGATGTC
MAG	GGACCCTATTCTCACCATCTTC	CACACCAGTACTCTCCATCATC
MBP	CGGAGTTGTGCACGTAGTAG	ATCTTCACACAGAAAGGGACAG
MOG	CGAATCACGAGGTCAGGAGT	GCCCACCACTATGCTCAGTT
MYRF (3'UTR)	ACACTGGATGCAATGGTGTTA	CAGCAACTCCAGTGTGAAGA
MYRF (cDNA)	CATCCTGTCCTTCCGTGAAT	GAAGTGGAAGTGGTAGTCTGTG
NKX2.2	TTTATGGCCATGTAAACGTTCTG	GCAACAATCACCACCGATATT
OLIG2	GTGGGAGACTCCGGGTA	TGAGATTGGATATGACCATCAGC
OMG	GAGGGAAGAGACAACCACAAATG	GACCACAACATTGAGCAATAAGAG
PDGFRA	GAGGAGGACTTGGTTGATGTT	TGAGATGCTACTGAGGCATTG
PLP1	GTGGCTCCAACCTTCTGTCC	GCAGGGAAACCAGTGTAGC
SOX10 (3'UTR)	CCAGTTTGACTACTCTGACCA	TATAGGAGAAGGCCGAGTAGAG
SOX10 (cDNA)	AGGAATGACCCTCTATCCCA	GCATGTCAGACCCTCACTATC
TF	TGTGGTCACACGGAAAGATAAG	GTCAGTTACGTTGCTTCCAAATAG