Cell Type Specific CAG Repeat Expansion and Toxicity of Mutant Huntingtin in Human Striatal and Cerebellar Cell Types

Kert Mätlik[1], Matthew Baffuto[1], Laura Kus[1], David A. Davis[2], Matthew R. Paul[3], Thomas S. Carroll[3], Nathaniel Heintz[1]

[1] Laboratory of Molecular Biology, The Rockefeller University, New York, United States

[2] Miller School of Medicine, University of Miami, Miami, United States

[3] Bioinformatics Resource Center, The Rockefeller University, New York, United States

SUMMARY

Tissue-specific somatic expansion of the mutant Huntingtin (*mHTT*) CAG tract and regional degeneration of the brain are key features of Huntington's disease (HD). However, the relationships between somatic CAG expansion, death of specific cell types, and molecular events associated with these processes have not been established. Here we employed fluorescence-activated nuclear sorting (FANS) and deep molecular profiling to gain insight into the properties of cell types of the human striatum and cerebellum in HD and control donors. Expansion of the *mHTT* CAG tract occurs in striatal MSNs and cholinergic interneurons, in cerebellar Purkinje cells, and at the mutant *ATXN3* locus in MSN nuclei from SCA3 donors. Somatic CAG tract instability in MSNs is associated with higher levels of MSH2 and MSH3. Our data indicate that somatic CAG tract expansion is not sufficient for cell death, and identify transcriptional changes associated with somatic CAG expansion and toxicity in the human striatum.

INTRODUCTION

Huntington's disease (HD) is a fatal neurodegenerative disease characterized by delayed age at onset, vulnerability of specific neuronal populations, and accumulation of protein aggregates, features that are shared with other, more prevalent late-onset neurodegenerative diseases. HD is caused by an abnormally long CAG tract in exon 1 of the Huntingtin gene (HTT). The mutation is autosomal dominant, variably penetrant if the CAG tract is 36 - 39 repeat units long, and fully penetrant at repeat lengths of 40 CAGs and above (Gusella, Lee et al. 2021). Abnormalities in affected patients include motor, cognitive and psychiatric problems that arise earlier or manifest at higher frequency in patients with longer CAG tracts (Genetic Modifiers of Huntington's Disease Consortium 2015, Ellis, Tee et al. 2020). HD age at onset is most often defined as the onset of motor symptoms, which are thought to arise as a consequence of early degeneration of the caudate nucleus and putamen, primarily due to loss of projection neurons of these structures, known as medium-sized spiny neurons (MSNs) (Heinsen, Strik et al. 1994). Remarkably, other neuron types within these same structures are largely spared (Ferrante, Kowall et al. 1987, Cicchetti, Prensa et al. 2000, Massouh, Wallman et al. 2008). The mutant *HTT* allele (*mHTT*) encoded CAG repeats give rise to a long polyglutamine stretch that makes the mHTT protein prone to the formation of cytoplasmic and nuclear aggregates (DiFiglia, Sapp et al. 1997). However, *HTT* is ubiquitously expressed and the reason for selective vulnerability of specific cell types in HD is largely unknown (Landwehrmeyer, McNeil et al. 1995).

The *mHTT* CAG repeat tract is unstable intergenerationally and somatically (The Huntington's Disease Collaborative Research Group 1993, Telenius, Kremer et al. 1994). Expansion of the inherited *mHTT* allele to very long CAG tracts has been observed sporadically in various brain structures, including the caudate nucleus and putamen (Kennedy, Evans et al. 2003). A causal role for somatic instability of the CAG repeat in HD pathogenesis is supported by two findings from a genome-wide association study looking for genetic modifiers of HD motor symptom onset other than CAG tract length itself (Genetic Modifiers of Huntington's Disease Consortium 2015, Genetic Modifiers of Huntington's Disease Consortium 2019). First, CAG repeat tract length predicts the age at motor symptom onset better than polyglutamine tract length. Second, several of the candidate genes identified in these studies are known to be involved in DNA repair. The latter finding is consistent with previous studies demonstrating that several of these DNA repair genes alter somatic instability of long CAG repeat tracts in brain tissue of transgenic mouse models of HD (Wheeler, Lebel et al. 2003, Dragileva, Hendricks et al. 2009, Pinto, Dragileva et al. 2013). These data have led to a two-component model of HD pathogenesis that first involves

somatic expansion of the *mHTT* CAG tract, followed by a toxic effect of the mHTT protein encoding an expanded polyglutamine tract (Hong, MacDonald et al. 2021). Although this model is consistent with most studies of HD mouse models and human neuropathology, it is not known why somatic CAG tract instability occurs in specific brain regions, whether somatic CAG tract instability occurs in specific cell types in these regions, whether somatic CAG expansion is sufficient to explain selective cellular vulnerability in HD, and if cell-specific factors in addition to mHTT carrying an expanded polyglutamine tract are required for toxicity.

To gain further insight into somatic CAG tract expansion and toxicity in HD, we have developed FANS methods for isolation of large numbers of nuclei from human striatal and cerebellar cell types, and examined the relationships between selective cellular vulnerability, somatic CAG tract expansion, and transcriptional responses in HD. We find that extensive somatic expansion of the *mHTT* CAG tract occurs in both medium spiny neuron (MSN) populations that are selectively vulnerable in HD, as well as in cholinergic interneurons that are not lost in the HD striatum. CAG tract expansion is observed also at the mutant *ATXN3* locus in MSN nuclei isolated from the post-mortem brains of spinocerebellar ataxia 3 (SCA3) donors, suggesting that MSNs are intrinsically prone to somatic expansion of CAG tracts even in the absence of mHTT protein. We demonstrate that the levels of DNA mismatch repair proteins MSH2 and MSH3 are elevated in MSN nuclei, suggesting that these proteins may contribute to preferential somatic CAG tract expansion in MSNs. The *mHTT* CAG tract also undergoes selective expansion in Purkinje cells relative to other cerebellar cell types, consistent with the well-documented loss of Purkinje cells in HD and other CAG repeat expansion disorders. Our data demonstrate that somatic CAG tract expansion and mHTT toxicity are not inextricably coupled in human striatal cell types, and they identify transcriptional programs that are associated with selective cellular vulnerability and toxicity in the human HD brain.

RESULTS

FANSseq profiling of human striatal cell types

To characterize the expression profile, somatic mutations and chromatin accessibility of human striatal cell types, we further developed the method of fluorescence-activated nuclei sorting (FANS) (Xu, Stoyanova et al. 2018) for isolating nuclei from human post-mortem caudate nucleus and putamen (Figure 1A). Nuclei were purified from samples dissected from human post-mortem caudate nucleus and putamen, stained with either antibody or RNA probes specific for the nuclei of cell types of interest, and resolved by passage through a FACS (Figure 1A) ('Suppl table 01

List of Tissue Donors.xlsx'). We collected nuclei of oligodendrocytes, astrocytes, microglia, dopamine receptor D1-expressing (DRD1+) medium spiny projection neurons of the direct pathway (dMSNs), dopamine receptor D2-expressing (DRD2+) medium spiny projection neurons of the indirect pathway (iMSNs) and 4 different types of striatal interneurons: cholinergic interneurons (CHAT+ INs), fast-spiking interneurons expressing Parvalbumin (PVALB+ INs), somatostatin-expressing interneurons (SST+ INs) and primate-specific Tachykinin Precursor 3-expressing interneurons (TAC3+ INs) (Krienen, Goldman et al. 2020) (Figure 1B and Suppl Figure 1). The specificity of labeling probes and purity of isolated populations of nuclei were verified by generating RNA-seq libraries from their nuclear transcriptomes (FANSseq) and comparing their gene expression profile to well-known cell type markers and previously published data (Doyle, Dougherty et al. 2008, Krienen, Goldman et al. 2020) (Figure 1C and 2A).

Analysis of main sources of variance between FANSseq datasets from all putamen cell types indicated that the first three principal components (P.C.) separated neuronal datasets from glial ones (P.C. 1), projection neuron datasets from those of interneurons (P.C. 2) and datasets of different glial cell types from each other (P.C. 3) (Figure 2B and C). For FANSseq datasets from putamen interneurons the major principal components separated the datasets according to interneuron subtype (Figure 2D and E). We asked how the interneuron types collected by our FANS-based methodology corresponded to the human striatal interneuron types recently defined via single-nucleus RNA-seq (snRNA-seq)(Krienen, Goldman et al. 2020). A majority of marker genes with enriched expression in a specific interneuron type in snRNA-seq data had also enriched expression in our corresponding FANS-based interneuron populations (Suppl Figure 2). The ETV1+ TAC3- population of Parvalbumin-expressing interneuron nuclei most likely captured both the major PVALB+ interneuron population and the related smaller PVALB+ TH+ interneuron population defined as a separate subtype in snRNA-seq analysis (Krienen, Goldman et al. 2020). Analysis of variance between FANSseq datasets from striatal projection neurons of both the caudate nucleus and putamen indicated that the main principal components separated datasets according to projection neuron subtype (P.C.1) and donor sex (P.C.2) (Suppl Figure 3A). Notably, none of the top principal components related to the brain region of origin (i.e. caudate nucleus or putamen) (Suppl Figure 3A).

Important features of the FANS-based molecular studies of post-mortem human tissue are the ability to isolate tens of thousands of nuclei from the same cell type, to generate deep transcriptional profiles, and the high sensitivity that can be achieved in epigenetic studies of chromatin accessibility (Xu, Stoyanova et al. 2018). Given the recent demonstration that ambient

RNA released from the cytoplasm or nucleus is a source of contamination in single nucleus RNA-seq studies of brain tissue (Caglayan, Liu et al. 2022), the difficulty in setting an arbitrary TPM (transcripts per million) threshold for gene expression in RNAseq datasets, and the likelihood that low-level ambient RNA can confound studies of differential gene expression, we used promoter accessibility in ATACseq data as an independent measure of active genes (Buenrostro, Giresi et al. 2013). To this end, we excluded genes for which none of their annotated transcriptional start sites overlapped with regions of accessible chromatin in that cell type, as defined by the presence of a consensus peak in ATACseq data generated from these cells (Figure 1D). We reasoned that this would allow us to minimize the number of possible false-positive differences in gene expression that might have resulted from contaminating ambient transcripts, FACS sorting impurities or contamination with genomic DNA. We were not able to produce ATAC-seq datasets from cholinergic interneurons due to their low abundance in striatal tissue. Excluding genes with inaccessible promoters removed a large proportion of genes with very low expression levels (< 1 TPM) while affecting fewer genes with moderate to high expression (>10 TPM) (Suppl Figure 3B). Based on the criterion of promoter accessibility, the average number of genes expressed in each striatal cell type is about 13,000 – 14,000. This estimate roughly corresponds to the average number of genes with expression level above an arbitrary cutoff value of 5 transcripts per million (TPM) in individual FANSseq datasets of neuronal cell types (Suppl Figure 3C). Therefore, the individual datasets capture most of the transcriptome complexity of these cell types at a well-measurable level. In summary, we have used coordinated FANSseq and ATACseq profiling to generate comprehensive high-quality datasets for each neural cell type present in the human caudate nucleus and putamen.

Somatic expansion of the *mHTT* CAG repeat takes place both in vulnerable MSNs and resilient cholinergic interneurons

Tissue-specific somatic repeat tract instability on the mutant allele is a central feature of HD and other triplet repeat disorders (Telenius, Kremer et al. 1994, Thornton, Johnson et al. 1994, Lokanga, Entezam et al. 2013, Long, Napierala et al. 2017, Mouro Pinto, Arning et al. 2020). Large somatic expansions of the *mHTT* exon 1 CAG tract in the striatum and cerebral cortex have been demonstrated by small pool PCR (Kennedy, Evans et al. 2003), and laser capture studies have suggested that somatic expansion can occur in both neurons and glia, albeit more frequently in neurons (Shelbourne, Keller-McGandy et al. 2007). To address the specificity of somatic expansion of *mHTT* CAG tract quantitatively, and to understand whether it is correlated with cell loss in HD, we isolated nuclei of each striatal cell type by FANS (Suppl Figure 4), verified the purity of isolated populations of nuclei by analysis of marker gene expression in nuclear

transcriptomes from the isolated nuclei (Suppl Figure 5A and B), and measured the length of *HTT* exon 1 CAG tract in the isolated genomic DNA from these populations using Illumina-sequencing of amplicons derived from *HTT* exon 1 (Ciosi, Maxwell et al. 2019). For the great majority of samples, genomic DNA from 800 or more cells was used for CAG tract length measurement.

Analysis of gDNA of different cell types from five HD donors carrying most prevalent disease-causing CAG tract lengths (from 42 to 45 uninterrupted CAGs, see Suppl Figure 5A) revealed that *mHTT* CAG tract was relatively stable in glial cell types and SST+, TAC3+ and PVALB+ interneurons (Figure 3A-C). However, only a small fraction of dMSNs and iMSNs had *mHTT* copies with the original inherited CAG tract length and approximately half of these neurons had CAG tracts that were expanded by more than 20 repeat units (Figure 3A-C). To reveal differences in the "non-expanding" striatal cell types, we calculated their ratios of somatic expansion - a metric that has been used for describing the distribution of CAG tract sizes in cells undergoing modest somatic repeat expansion (Ciosi, Maxwell et al. 2019). These data demonstrated that striatal oligodendrocytes had consistently greater CAG tract expansions compared to microglia, astrocytes and striatal interneurons (Suppl Figure 5C). There was no evidence of somatic expansion of the normal *HTT* allele CAG repeat tract in any of the cell types analyzed (Suppl Figure 5D).

Because the scarcity of tissue available prevented us from isolating striatal cholinergic interneuron nuclei from all but one of these five initially characterized HD donors, we isolated nuclei of this very rare cell type from two other HD donors. In these samples the fraction of nuclei with large repeat length increases was large, revealing that *mHTT* CAG tract undergoes large expansions also in CHAT+ interneurons (Figure 3A and D).

Pairwise comparisons of the two striatal projection neuron types from donors carrying a wider range of initial repeat lengths revealed that the extent of repeat expansion was dependent on initial repeat tract length. Thus, more modest expansion was observed for *mHTT* alleles with reduced penetrance (CAG tract lengths < 40 repeats) relative to longer, fully penetrant alleles (Figure 3E and F). Interestingly, while iMSNs have been reported to be more vulnerable of the two striatal projection neuron subtypes (Reiner, Albin et al. 1988), there was no significant difference in the frequency of highly expanded repeat copies between dMSNs and iMSNs (Figure 3E and F).

Taken together, the *mHTT* CAG tract expands dramatically in striatal projection neurons and cholinergic interneurons, and very little in other cell types analyzed. These data support the

hypothesis that extensive somatic expansion of the *mHTT* CAG tract is required for the vulnerability of MSNs in HD. However, given previous studies demonstrating that cholinergic interneurons in the striatum are not lost in HD (Ferrante, Kowall et al. 1987, Massouh, Wallman et al. 2008), it is also evident from our results that robust expansion of the *mHTT* CAG tract is not sufficient to cause neuronal loss in HD (see also, the accompanying manuscript Pressl C. et al). In addition, our data show that differences in the rate of *mHTT* CAG tract expansion are an unlikely explanation for the reportedly greater vulnerability of iMSNs than dMSNs in this disease.

Cell type-specific instability of the *mHTT* CAG tract in the HD cerebellum

The loss of cerebellar Purkinje cells (PCs) in several spinocerebellar ataxias (SCA1, SCA2, SCA6 and SCA7) where the causal elongated CAG tracts undergo germline expansion has suggested that in these disorders, somatic CAG tract expansion may occur in PCs (Paulson 2018). Although the viability of PCs in HD has been a matter of debate, recent stereological studies have demonstrated that PC loss occurs in the cerebellum in HD cases with predominant motor symptoms (Singh-Bains, Mehrabi et al. 2019). Given these data, we measured *mHTT* CAG tract instability in cerebellar cell types in several HD donors. To isolate the nuclei of granule cells, astrocytes, microglia, oligodendrocytes and their progenitors (OPCs), we followed the method of Xu et al. using antibodies to cerebellar nuclear proteins (Xu, Stoyanova et al. 2018). To isolate sufficient numbers of nuclei to perform assays for somatic expansion for the very rare Purkinje cells (<0.01% of all nuclei), we used Primeflow probes that bind to the transcripts of carbonic anhydrase 8 (*CA8*) gene expressed specifically and abundantly in Purkinje cells. Although the degree of *mHTT* CAG tract expansion in PCs was relatively modest compared to striatal MSNs, the tract had expanded more in PCs than in any other cerebellar cell type. (Figure 4A-C). The variability in somatic CAG tract expansion observed for PC samples from different donors can most likely be attributed to the rarity of this cell type, which made it extremely difficult to collect samples entirely free of contamination with nuclei of 'non-expanding' cell types. Similar to striatal oligodendrocytes, mature cerebellar oligodendrocytes had more expanded CAG tracts compared to granule neurons, other glia types or OPCs within each donor, even though statistical significance was not reached for all of these comparisons (Figure 4C).

While the degree of somatic *mHTT* CAG tract expansion differs between Purkinje cells and MSNs, our data indicate that both in striatum and cerebellum the CAG tract is somatically unstable in selected neuron types and much more stable in other neuron types and glial cells.

Striatal MSNs are prone to somatic CAG tract expansion

Preferential somatic expansion of the *mHTT* CAG tract in selected striatal neuron types could be due to cell type-specific properties of the *HTT* locus (i.e. MSN and cholinergic interneuron-specific factors acting in *cis*), putative cell type-specific factors acting in *trans* (e.g. DNA repair proteins) or a combination of the two. Since the expansion-promoting effect of putative *trans*-factors would not necessarily be limited to the *mHTT* locus, we asked whether MSNs have a propensity to expand long and pure CAG or CTG tracts at other genomic loci as well. Because transcription through the repeat tract seems to be a prerequisite for somatic instability in the central nervous system (CNS) (Adihe Lokanga, Zhao et al. 2014), we chose to analyze the CAG repeat tract in *ATXN3* gene because, as is the case for HTT, its transcription is relatively uniform across striatal cell types (Suppl Figure 6A).

To determine whether the long CAG repeat of the mutant ATXN3 allele (*mATXN3*) present in individuals with spinocerebellar ataxia 3 (SCA3) exhibits greater instability in MSNs than in other cell types, we isolated glial cell and MSN nuclei from striatal tissue of five donors and striatal interneuron nuclei from two donors carrying the SCA3 mutation (Suppl table 01 'List of Tissue Donors.xlsx'). Despite the presence of a long uninterrupted CAG repeat tract, with initial lengths ranging from 64 to 76 CAG repeats in the SCA3 donors analyzed, there were no clear signs of MSN loss even in the oldest SCA3 donors, as judged by the abundance of large NeuN+ nuclei in striatal homogenates (Suppl Figure 6B). As shown in Figure 5, the m*ATXN3* gene CAG tract was clearly more unstable in the striatal MSNs relative to glial cells and interneurons (Figure 5A and B). However, despite the increased length of the inherited *mATXN3* repeat tract, its expansion was modest compared to that of m*HTT* CAG repeat tract in MSNs in HD. For example, there were very few copies of the *mATXN3* allele with repeat expanded by more than 20 repeat units (Figure 5A). These data indicate that MSNs have a propensity to expand long CAG tracts at other genomic loci even in the absence of mHTT protein, and suggest that in addition to the CAG repeat tract length there are other locus-specific properties that are important for the dramatic expansion of the *mHTT* CAG tract. These observations support the hypothesis that *mHTT* exon 1 CAG tract instability is modulated by rate-limiting *trans*-acting factors expressed at different levels in striatal cell types.

To identify these *trans*-acting factors that may explain preferential CAG tract expansion in MSNs, we compared the FANSseq expression profiles of striatal cell types. We focused on genes coding for proteins involved in DNA mismatch repair (MMR) and base-excision repair (BER), since these processes have been shown to affect trinucleotide repeat instability in model systems (Wheeler

and Dion 2021), and because several MMR genes are represented among candidate genes identified as age of motor symptom onset modifiers in HD mutation carriers (Genetic Modifiers of Huntington's Disease Consortium 2019). We found that transcript levels of *MSH2* and *MSH3*, encoding MMR proteins that form the MutSβ complex, were more than 2-fold higher in both projection neuron subtypes compared to other striatal neurons, including cholinergic interneurons, and this difference was consistent across neuron types in both putamen and caudate nucleus (Figure 6A, and Suppl Figure 7A and B).

To determine whether the FANSseq data accurately reflected nuclear protein levels for these factors in abundant striatal cell types, we measured MSH2 and MSH3 levels by Western blotting of nuclear lysates from MSNs, microglia, astrocytes and oligodendrocytes (Figure 6B). The ratio of both MSH2 and MSH3 to chromatin, assessed by anti-H3 signal, was significantly higher in MSN nuclei compared to glial cells (Figure 6B). Since knockout of *Msh2* or *Msh3* is sufficient to eliminate the somatic CAG tract instability in mice carrying a CAG tract of over 100 repeat units in *Htt* exon 1 (Wheeler, Lebel et al. 2003, Dragileva, Hendricks et al. 2009, Kovalenko, Dragileva et al. 2012), elevated levels of the two components of MutSβ may explain the enhanced instability of long and uninterrupted CAG tracts in the genomes of human striatal projection neurons.

MSN-specific gene expression in the human striatum

Studies of gene expression in animal models of human disease and post-mortem human tissue samples, as well as proteomic studies of protein complexes accumulating during pathological progression, have emerged as valuable approaches for the formulation of etiological hypotheses. To gain further insight into the pathophysiological processes that influence HD progression, we next studied the expression pattern of the remaining genes that were identified as candidates potentially modulating HD age at onset (Genetic Modifiers of Huntington's Disease Consortium 2019). As expected, we found that candidate genes involved in MMR (marked with an asterisk on Figure 6A and Suppl Figure 7A), as well as those lacking an obvious association with DNA repair are expressed at significant levels in human MSNs (Suppl Fig 8A and B). Notable exceptions are *RRM2B* and *SYT9* which are expressed at very low levels in MSNs, but at higher levels in microglia (*RRM2B*) and oligodendrocytes (*SYT9*) in both caudate nucleus and putamen (Figure 6C, Suppl Figure 8A and B). These FANSseq results are supported by considerably lower chromatin accessibility at the promoter regions of *RRM2B* and *SYT9* in human MSNs compared to glial cell types (Figure 6D). This result could not have been predicted from published cell type-specific gene expression analysis of mouse striatum (Suppl Figure 8C). Thus, although confirmation of these data will require additional studies, our results suggest that genetic variants

near *RRM2B* and *SYT9* may act in glia to influence disease onset through mechanisms that do not impact somatic expansion of the *mHTT* CAG tract.

To complement efforts toward a mechanistic understanding of the mHTT functions in MSNs, we analyzed next cell type-specific expression of genes identified through proteomic analysis of HTT-interacting proteins (HIPs) immunoaffinity-purified from mouse striatum (Greco, Secker et al. 2022). Although the majority of genes encoding HIPs are expressed in human MSNs, our data indicate that ~10% of them (24 out of 278, 'Suppl Table 04 HIPs not expressed or with low expression in MSNs.xlsx') are either not expressed or expressed at very low levels in human MSNs. Furthermore, for many of those HIP genes that have a much lower transcript level in MSNs compared to any striatal interneuron type (Suppl Figure 9A), this expression pattern could not have been predicted from published cell type-specific gene expression analysis of mouse striatum (Suppl Figure 9B and C) (Doyle, Dougherty et al. 2008, Saunders, Macosko et al. 2018). While these data are not a significant confound for the interpretation of the pathways identified in the mHTT interaction data, they are important for evaluating the significance of individual proteins in cell-autonomous pathophysiology in HD striatum. For example, the gene of Huntingtin-interacting protein HAP1 (Li, Li et al. 1995) is transcribed in mouse MSNs (Suppl Figure 9B-D) and depletion of HAP1 leads to loss of MSNs in a mouse model of HD (Liu, Cheng et al. 2020). However, in humans, *HAP1* transcripts are virtually absent in MSNs, yet clearly present in striatal interneurons (Suppl Figure 9E, upper panel). These observations are supported by ATACseq data indicating that *HAP1* promoter region is not accessible in human MSNs (Suppl Fig. 9E, lower panel). Given the precedent for qualitative differences in species-specific gene expression in cerebellar and cortical cell types (Xu, Stoyanova et al. 2018, Bakken, Jorstad et al. 2021), the differences in striatal gene expression reported here are not surprising. Nevertheless, in light of the selective vulnerability of MSNs in the human striatum, the expression data provided here can add important information to the studies of candidate mechanisms of mHTT toxicity in HD.

Altered gene expression during Huntington's disease progression

To gain further insight into the molecular events that may play a role in somatic expansion or contribute to mHTT toxicity, we sequenced the nuclear transcriptomes of striatal MSNs and TAC3+, SST+ and PVALB+ interneurons from the putamen or caudate nucleus of 6-7 HD donors (Suppl table 01 'List of Tissue Donors'). We again limited the comparative analysis of HD and control donors' (n=8) FANSseq data to genes that are more likely to be truly expressed in the cell type of interest by excluding transcripts for which there was no promoter accessibility as indicated by ATACseq (Suppl Figure 10A). More than 98 % of genes with accessible promoters in HD MSNs

had accessible promoters also in MSNs from control donors. Due to the low abundance of striatal interneurons and limited amounts of striatal tissue available from HD donors, we were not able to produce ATAC-seq datasets from the interneurons of HD donors. Therefore, striatal interneuron FANSseq datasets were filtered to include only those genes which had accessible promoters in control donors.

Principal component analysis of FANSseq data from MSNs indicated that disease status (P.C.1) and MSN subtype (P.C.2) were the main sources of variance in these datasets (Suppl Figure 10B). The overall number of genes with statistically significant disease-associated differences in transcript levels in each cell type is shown in Figure 7A. Although transcriptional changes occurred in both interneurons and MSNs, the number of both up- and downregulated genes was higher in MSNs. As expected, disease-associated changes were well correlated between dMSNs and iMSNs, but the correlation was poor in comparisons involving other neuron types (Suppl Figure 10C).

Gene ontology analysis. To identify cellular processes are that affected by disease-associated gene expression changes that take place only in MSNs, thereby correlating with the presence of more toxic mHTT species in these neurons, we analyzed which Gene Ontology Cellular Component (GOCC) terms were enriched for genes that were up- or down-regulated in MSNs, but did not display these changes in expression in any of the interneuron populations studied. The results indicated that many genes downregulated specifically in MSNs are involved in ribosomal biogenesis (GOCC terms ‘cytosolic ribosome’ and ‘nucleolus’), pre-mRNA maturation (GOCC terms ‘nuclear body’ and ‘spliceosomal complex’), and other nuclear functions (Figure 7B). Although transcripts of mitochondrial oxidative phosphorylation pathway genes have been reported to be downregulated in HD MSNs (Lee, Fenster et al. 2020), we noticed that this disease-associated change is much more evident in the nuclear transcriptome of PVALB+ striatal interneurons (Suppl Figure 11A and Suppl Figure 12A).

The GOCC terms enriched for genes that increase expression in MSNs in the HD donor data include many terms that indicate alterations in membrane protein function. ‘Neuron projection’, ‘Synapse’ and ‘Lysosome’ were among GOCC terms enriched for genes upregulated specifically in both striatal projection neuron subtypes (Figure 7B). In contrast, for one or more of the striatal interneuron types these gene ontology categories were instead enriched for downregulated genes (Suppl Figure 11A and Suppl Figure 12B). We observed also that genes central to the regulation of lysosomal biogenesis and autophagy were among the top upregulated genes in MSNs. For example, the transcription factor TFEB, that has been shown to be essential for regulation of

many genes in these pathways (Sardiello, Palmieri et al. 2009), is strongly increased in HD MSNs. Several genes encoding proteins essential for autophagy are also induced in HD, including *ATG9B*, *ATG9A* and the HTT-interacting protein ULK1 involved in autophagosome formation (Rui, Xu et al. 2015), *MAPK8IP1* and *MAPK8IP3* which encode proteins involved in retrograde transport of autophagosomes (Cason, Carman et al. 2021) and *SQSTM1*, the gene encoding HTT-interacting autophagy cargo receptor p62 (Rui, Xu et al. 2015).

Taken together, the GO analysis results indicate that HD-associated gene expression changes in human striatal MSNs are distinct from those of other striatal neurons in HD, and point to the different cellular processes that are either induced or repressed at the level of transcription in MSNs. The large magnitude of HD-associated down-regulation seen for many genes (median log2FoldChange -0.67 and -0.73 for dMSNs and iMSN, respectively) establishes that these changes must be occurring in the majority of nuclei being assayed, and it seems likely that these may be a direct effect of the presence of mHTT in the nucleus. In contrast, the fact that many of the genes upregulated in HD MSNs belong to GO categories involved in membrane function, and that there is strong induction of genes involved in autophagy, may reflect a response to extra-nuclear mHTT aggregation.

DNA repair pathways. Given the instability of *mHTT* CAG tract in MSNs, we investigated whether there are disease-associated changes in the transcript levels of genes encoding MMR and BER proteins. These data revealed clear distinctions in the regulation of these pathways in MSNs compared to interneurons. Notably, we found that in HD, *MSH2* and *MSH3* expression levels in MSNs are significantly reduced (Figure 7C). The expression of several DNA glycosylases (*NEIL1*, *UNG* and *TDG* genes) involved in the detection of DNA lesions is also changed in MSNs to a greater degree than in other neuronal cell types. Although further validation will be required to define and understand these changes, transcriptional regulation of DNA repair genes in response to genotoxic stress has been demonstrated in both rodent and human cells (Christmann and Kaina 2013). It seems likely, therefore, that further studies of these pathways will shed light on additional features of DNA repair that occur in concert with or as a result of *mHTT* CAG tract expansion.

Genes required for MSN viability. A second important class of genes that we chose to analyze in more detail were those previously shown to preserve MSN viability in mice expressing *mHTT* with extra-long CAG tract (Wertz, Mitchem et al. 2020). Since decreased expression of these genes in HD MSNs could contribute to their loss, we asked whether any of these genes were strongly downregulated in HD MSNs. Figure 7D and Suppl Figure 11C depict genes that were found to be

essential for MSN viability in the zQ175 and R6/2 mouse models, respectively, and had undergone the largest expression level decreases in human MSNs in HD donors. We consider these genes as candidates for further investigation as they identify molecular processes that may contribute to loss of MSN viability in HD. For example, HD-associated downregulation of TATA-binding protein-associated factor 1 gene (*TAF1*) in MSNs is an example of a change that could be detrimental to cell survival, as reduced expression and aberrant splicing of *TAF1* is thought to be the cause of MSN loss seen in X-linked dystonia parkinsonism (Makino, Kaji et al. 2007, Ito, Hendriks et al. 2016, Campion, Mejia Maza et al. 2022).

Distinctions between mouse and human transcriptional responses in HD MSNs. While extensive transcriptional dysregulation observed in human MSNs in HD donor samples is generally consistent with the "transcriptionopathy" that has been reported in HD mouse models (Veldman and Yang 2018), most of the HD-associated nuclear transcript levels changes seen by FANSseq are not recapitulated in the HD mouse models for which MSN-specific RNAseq data is available (Suppl. Fig. 13 A-C). Conversely, many individual genes and pathways reported as altered in mouse models of HD are not exhibiting an HD-associated change in human FANSseq data. In particular, human FANSseq data does not support a consistent decrease in MSN marker gene expression (Supp. Fig. 14A) or the induction of genes involved in innate immune signaling (Supp. Fig. 14B) (Lee, Fenster et al. 2020). Also, most of the genes with altered expression in the striatum of BAC-CAG mouse model are not undergoing a similar transcript level change in human MSNs (Suppl. Fig. 14 D) (Gu, Richman et al. 2022), and the derepression of PRC2-complex target genes that accompanies neurodegeneration in the striatum of MSN-specific Ezh1/Ezh2 double knockout mice is not observed for the majority of these genes expressed in human MSNs (Suppl. Fig. 14D) (von Schimmelmann, Feinberg et al. 2016). While we have not focused on the transcriptional events that are altered in cell types that do not undergo somatic expansion in the human striatum, it seems highly probable that these data will also differ significantly between HD and mouse models, given that these cell types carry pre-expanded mHTT alleles in these models. Although, it is difficult to directly compare the human FANSseq data with transcriptional changes occurring in bulk RNAseq data from the striatum or MSN translational profiling (TRAP) data from HD mouse models, the apparent discrepancies between molecular events occurring in mouse model systems and human HD donor samples place a strong emphasis on the importance of human data when initiating focused studies of candidate mechanisms of HD pathogenesis.

While we have highlighted some of the cell type-specific transcript level changes that correlate with the occurrence of somatic CAG expansion, future studies are needed to clarify the degree to

which these changes perturb the implicated biological processes *in vivo*, and whether these changes are detrimental or compensatory. We believe the comprehensive datasets we have collected will also be an important source for continued comparative analyses of the large body of transcriptional profiling data that has been collected from mouse models (Veldman and Yang 2018), and from additional studies of molecular events associated with HD in other structures of the human brain (see the accompanying manuscript by Pressl et al).

DISCUSSION

Here we have used FANS (Xu, Stoyanova et al. 2018) to isolate thousands of nuclei of each neural cell type of human caudate nucleus and putamen to generate deep, high resolution, cell type-specific transcriptional and *HTT* CAG repeat tract length-measurement data from control and HD donors. Our data are consistent with the two-component model of HD pathogenesis (Hong, MacDonald et al. 2021) and reveal several cell type-specific molecular features of the disease.

Somatic expansion of *mHTT* CAG tract is not sufficient to cause cell loss in the human striatum.

The most vulnerable cell types in the HD striatum are MSNs (Heinsen, Strik et al. 1994). While both dMSNs and iMSNs are progressively lost during the progression of the disease, iMSNs that express dopamine receptor D2 and enkephalin are most vulnerable in early stages (Reiner, Albin et al. 1988). Striatal interneurons are relatively spared early in the disease (Ferrante, Kowall et al. 1987, Cicchetti, Prensa et al. 2000). In particular, although cholinergic interneurons are clearly affected, as indicated by reduced choline acetyltransferase (CHAT) activity in histological sections, the persistent expression of acetylcholinesterase (ACHE) in these cells indicates that they do not die during the disease (Suzuki, Desmond et al. 2001, Massouh, Wallman et al. 2008).

To understand the relationships between somatic expansion and cell loss, we have analyzed the length of *HTT* CAG tract in gDNA extracted from nuclei of striatal cell types. Importantly, the extraction of RNA and gDNA from the same population of isolated nuclei has allowed us to use transcriptome analysis for the verification of gDNA sample purity (i.e. cell type specificity). While the stability of the *mHTT* CAG tract in SST+, TAC3+ and, PVALB+ interneurons can explain their relative resilience in HD, our data also demonstrate that large expansions of the *mHTT* CAG tract are not sufficient for loss of CHAT+ INs in HD. Furthermore, data collected from dMSN and iMSN nuclei isolated from HD donors (N=13 total) establish that the extent of somatic expansion is equivalent in these two projection neuron subtypes despite the increased vulnerability of iMSNs relative to dMSNs (Reiner, Albin et al. 1988). This is especially evident in data from carriers of

reduced penetrance *mHTT* alleles where the loss of MSNs is minor. While our data support the hypothesis that somatic expansion is a first step in disease progression that is necessary for the loss of neurons in HD, they indicate also that additional factors must be required. This conclusion is strongly supported by studies of the human cerebral cortex in HD (Pressl C. et al, 2023) demonstrating that extensive expansion of the *mHTT* CAG tract occurs in many deep layer pyramidal cell types despite very selective loss of a specific class of corticostriatal projection neurons. Finally, the observation that in cerebellum the *mHTT* CAG tract is unstable selectively in Purkinje cells is consistent with somatic expansion of the CAG tract being a prerequisite of mHTT toxicity since these neurons have been shown to be vulnerable in HD (Singh-Bains, Mehrabi et al. 2019).

Striatal MSNs are prone to somatic CAG tract expansion.

The preferential expansion of the *mATXN3* CAG tract we detect in striatal MSN nuclei isolated from 3 (SCA3) donor samples indicates that these neurons have a general propensity to expand long CAG tracts, which is not limited to the *mHTT* locus and does not require the presence of mHTT protein. This independence from the presence of mHTT protein is consistent with the fact that the homozygosity for *mHTT* allele does not result in earlier onset of HD (Squitieri, Gellera et al. 2003). Our results are thus suggesting that there are *trans*-acting factors that are rate-limiting in the process of repeat expansion or stabilization, and that their level of expression or activity differs between MSNs and cell types where *mHTT* CAG tract is more stable. Our finding that levels of both *MSH2* and *MSH3* are higher in MSNs compared to other striatal cell types suggests that the actions of the MutSb complex at the *mHTT* locus could be a rate-limiting step that contributes significantly to the selective vulnerability of MSNs. This idea is further supported by previous reports showing that both *MSH2* and *MSH3* are required for *mHTT* CAG tract expansion in a mouse model of HD (Wheeler, Lebel et al. 2003, Dragileva, Hendricks et al. 2009, Kovalenko, Dragileva et al. 2012). These observations however do not rule out the presence of other factors that favor the expansion of *mHTT* CAG repeat tract in MSNs and cholinergic interneurons.

Two additional features of somatic expansion are also evident from our data. First, the degree of somatic expansion of the *mHTT* CAG tract in striatal MSNs exceeds the expansion evident in PCs. Second, in MSNs, the expansion of the *mHTT* CAG tract is more extensive than the expansion of the *mATXN3* CAG tract. Notably, in HD donors carrying the fully penetrant *mHTT* alleles, about half of the remaining MSNs have *mHTT* CAG tract length increases of at least 21 CAG repeats, whereas very few tracts that gain that many repeats are evident at *mHTT* allele in Purkinje cells or in *mATXN3* allele in MSNs. These data demonstrate that both the locus at which

the expansion occurs, as well as cell-intrinsic factors determine the rate of expansion. The characteristics of *mHTT* exon 1 that make its CAG tract especially prone to expansion in MSNs remain to be determined.

Insights into mHTT toxicity.

The first conclusion we can draw is that the presence of a very long CAG repeat tract in RNA, or polyQ domain in any protein does not necessarily lead to cell loss in the human brain. In addition to the resilience of cholinergic interneurons in the face of large expansions of the *mHTT* CAG tract (discussed above), this conclusion is supported by the observation that despite expansion of *mATXN3* CAG tract to over 80 CAG repeats, we did not see a clear decrease in the number of MSN nuclei isolated from SCA3 donor samples. The latter finding is not inconsistent with neuroanatomical studies that have reported MSN loss in SCA3 to be variable (Koeppen 2018). These results are reminiscent of the detailed studies of SCA1 mouse models demonstrating that the presence of Capicua sensitizes cerebellar Purkinje cells to mutant *Atxn1* (Rousseaux, Tschumperlin et al. 2018). While these considerations place additional emphasis on proteins that interact with mHTT *in vivo* (Greco, Secker et al. 2022), delineation of the cell types and subcellular domains in which these interactions occur will be crucial for understanding their role in HD pathogenesis.

A second important conclusion that can be drawn from these data is that extensive transcriptional disturbances occur in the majority of MSNs prior to their demise in the HD striatum. While these data are not by themselves sufficient to identify mechanisms of toxicity that cause neuronal dysfunction and death in the second step of HD pathogenesis, they raise an important issue that remains to be resolved: do these transcriptional changes indicate that striatal MSN functions are significantly compromised long before they are lost in the HD brain, or are these transcriptional events homeostatic responses that allow MSNs to operate relatively well for an extended time? The strong induction TFEB, of a central regulator of lysosomal function and autophagy, and of a variety of its downstream targets argues that human MSNs mount an important defense against the mHTT misfolding and aggregation in HD. Transcriptional induction of genes involved in somatodendritic and synaptic function may also be indicative of homeostatic mechanisms that respond to problems in these important neuronal functions. However, these positive responses to mHTT must be weighed against the nuclear functions that are strongly decreased in HD MSNs, the decreased expression of several genes that have been shown to be required for MSN viability in R6/2 and Q175 mouse models of HD (Wertz, Mitchem et al. 2020), and the wide variety of additional transcriptional changes that occur in these very vulnerable neurons.

A third important point that arises from our observation that both the FANSseq and ATACseq data we have collected suggest that the *RRM2B* and *SYT9* genes, which can modify HD age at onset (Genetic Modifiers of Huntington's Disease Consortium 2019), are expressed more highly in glia. Although these data will need to be verified by further experimentation, the elevated expression of these genes in microglia (*RRM2B*) and oligodendrocytes (*SYT9*) suggest that glia may contribute to HD pathogenesis. It is noteworthy that a major role for glial dysfunction has been proposed in a variety of other late-onset neurodegenerative disorders (Hickman, Izzy et al. 2018, Brandebura, Paumier et al. 2023). However, even if the onset of motor symptoms in HD is modulated by the effect genetic variants in RRM2B and SYT9 have on glial cells, our results suggest that the role of glia in HD progression is likely to be secondary given that the *mHTT* CAG tract instability is minimal in these cell types.

Concluding remarks. The data we have presented here shed new light on several important features of HD pathogenesis. We hope that further analyses of the comprehensive datasets we have provided will stimulate others to interrogate them in the context of detailed mechanistic studies of HD. However, it is important to recognize that somatic expansion of *mHTT* CAG tract occurs only in select CNS cell types in HD, and that there are species-specific differences in cell type-specific gene expression in the striatum. Careful consideration of these findings will be important when studying HD molecular pathogenesis in models based on other species.

ACKNOWLEDGEMENTS

This study was supported by funding from the CHDI Foundation, and we are thankful to Thomas F. Vogt and Jian Chen for helpful discussions throughout the project. We thank Nicholas Didkovsky, Cuidong Wang and Dr. Hasnahana Chetia for technical assistance, and Dr. Kärt Mätlik for advice on data visualization and for critical comments on the manuscript. We are also thankful to the Rockefeller University Genomics Resource Center for advice and support. We are grateful to tissue donors and their families, and also thank the University of Miami Brain Endowment Bank, Jean Paul G. Vonsattel and Andrew F. Teich from Columbia University Alzheimer’s Disease Research Center (funded by NIH, grant P30AG066462), Dirk Keene from University of Washington BioRepository and Integrated Neuropathology Laboratory (supported by the Alzheimer's Disease Research Center, grants AG066567 and AG066509), The University of Michigan Brain Bank, Netherlands Brain Bank, Harvard Brain Tissue Resource Center and UCLA Human Brain & Spinal Fluid Resource Center (supported by National Institutes of Health and the US Department of Veterans Affairs) for assisting and providing

postmortem brain tissues. K.M. was supported by a fellowship from the Sigrid Jusélius Foundation.

Author Contributions

Conceptualization: K.M., N.H., Formal Analysis: K.M., M.R.P., T.C., Investigation: K.M., M.B., Resources: L.K., D.A.D., N.H., Visualization: K.M., Writing - original draft: K.M., N.H., Writing - review and editing: M.R.P., M.B., L.K., Supervision: K.M., N.H., Funding acquisition: N.H.

SUPPLEMENTAL MATERIALS

List supplemental files:

1 - Suppl table 01 List of Tissue Donors.xlsx

2 - Suppl Table 04 HIPs not expressed or with low expression in MSNs.xlsx

REFERENCES

The Huntington's Disease Collaborative Research Group (1993). "A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group." Cell **72**(6): 971-983.

Adihe Lokanga, R., X. N. Zhao, A. Entezam and K. Usdin (2014). "X inactivation plays a major role in the gender bias in somatic expansion in a mouse model of the fragile X-related disorders: implications for the mechanism of repeat expansion." Hum Mol Genet **23**(18): 4985-4994.

Bakken, T. E., N. L. Jorstad, Q. Hu, B. B. Lake, W. Tian, B. E. Kalmbach, M. Crow, R. D. Hodge, F. M. Krienen, S. A. Sorensen, J. Eggermont, Z. Yao, B. D. Aevermann, A. I. Aldridge, A. Bartlett, D. Bertagnolli, T. Casper, R. G. Castanon, K. Crichton, T. L. Daigle, R. Dalley, N. Dee, N. Dembrow, D. Diep, S. L. Ding, W. Dong, R. Fang, S. Fischer, M. Goldman, J. Goldy, L. T. Graybuck, B. R. Herb, X. Hou, J. Kancherla, M. Kroll, K. Lathia, B. van Lew, Y. E. Li, C. S. Liu, H. Liu, J. D. Lucero, A. Mahurkar, D. McMillen, J. A. Miller, M. Moussa, J. R. Nery, P. R. Nicovich, S. Y. Niu, J. Orvis, J. K. Osteen, S. Owen, C. R. Palmer, T. Pham, N. Plongthongkum, O. Poirion, N. M. Reed, C. Rimorin, A. Rivkin, W. J. Romanow, A. E. Sedeno-Cortes, K. Siletti, S. Somasundaram, J. Sulc, M. Tieu, A. Torkelson, H. Tung, X. Wang, F. Xie, A. M. Yanny, R. Zhang, S. A. Ament, M. M. Behrens, H. C. Bravo, J. Chun, A. Dobin, J. Gillis, R. Hertzano, P. R. Hof, T. Hollt, G. D. Horwitz, C. D. Keene, P. V. Kharchenko, A. L. Ko, B. P. Lelieveldt, C. Luo, E. A. Mukamel, A. Pinto-Duarte, S. Preissl, A. Regev, B. Ren, R. H. Scheuermann, K. Smith, W. J. Spain, O. R. White, C. Koch, M. Hawrylycz, B. Tasic, E. Z. Macosko, S. A. McCarroll, J. T. Ting, H. Zeng, K. Zhang, G. Feng, J. R. Ecker, S. Linnarsson and E. S. Lein (2021). "Comparative cellular analysis of motor cortex in human, marmoset and mouse." Nature **598**(7879): 111-119.

Brandebura, A. N., A. Paumier, T. S. Onur and N. J. Allen (2023). "Astrocyte contribution to dysfunction, risk and progression in neurodegenerative disorders." Nat Rev Neurosci **24**(1): 23-39.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang and W. J. Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nat Methods **10**(12): 1213-1218.

Caglayan, E., Y. Liu and G. Konopka (2022). "Neuronal ambient RNA contamination causes misinterpreted and masked cell types in brain single-nuclei datasets." Neuron **110**(24): 4043-4056 e4045.

Campion, L. N., A. Mejia Maza, R. Yadav, E. B. Penney, M. G. Murcar, K. Correia, T. Gillis, C. Fernandez-Cerado, M. S. Velasco-Andrada, G. P. Legarda, N. G. Ganza-Bautista, J. B. B. Lagarde, P. J. Acuna, T. Multhaupt-Buell, G. Aldykiewicz, M. L. Supnet, J. K. De Guzman, C. Go, N. Sharma, E. L. Munoz, M. C. Ang, C. C. E. Diesta, D. C. Bragg, L. J. Ozelius and V. C. Wheeler (2022). "Tissue-specific and repeat length-dependent somatic instability of the X-linked dystonia parkinsonism-associated CCCTCT repeat." Acta Neuropathol Commun **10**(1): 49.

Cason, S. E., P. J. Carman, C. Van Duyne, J. Goldsmith, R. Dominguez and E. L. F. Holzbaur (2021). "Sequential dynein effectors regulate axonal autophagosome motility in a maturation-dependent pathway." J Cell Biol **220**(7).

Christmann, M. and B. Kaina (2013). "Transcriptional regulation of human DNA repair genes following genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation." Nucleic Acids Res **41**(18): 8403-8420.

Cicchetti, F., L. Prensa, Y. Wu and A. Parent (2000). "Chemical anatomy of striatal interneurons in normal individuals and in patients with Huntington's disease." Brain Res Brain Res Rev **34**(1-2): 80-101.

Ciosi, M., A. Maxwell, S. A. Cumming, D. J. Hensman Moss, A. M. Alshammari, M. D. Flower, A. Durr, B. R. Leavitt, R. A. C. Roos, T.-H. team, H. D. t. Enroll, P. Holmans, L. Jones, D. R. Langbehn, S. Kwak, S. J. Tabrizi and D. G. Monckton (2019). "A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes." EBioMedicine **48**: 568-580.

DiFiglia, M., E. Sapp, K. O. Chase, S. W. Davies, G. P. Bates, J. P. Vonsattel and N. Aronin (1997). "Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain." Science **277**(5334): 1990-1993.

Doyle, J. P., J. D. Dougherty, M. Heiman, E. F. Schmidt, T. R. Stevens, G. Ma, S. Bupp, P. Shrestha, R. D. Shah, M. L. Doughty, S. Gong, P. Greengard and N. Heintz (2008). "Application of a translational profiling approach for the comparative analysis of CNS cell types." Cell **135**(4): 749-762.

Dragileva, E., A. Hendricks, A. Teed, T. Gillis, E. T. Lopez, E. C. Friedberg, R. Kucherlapati, W. Edelmann, K. L. Lunetta, M. E. MacDonald and V. C. Wheeler (2009). "Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes." Neurobiol Dis **33**(1): 37-47.

Ellis, N., A. Tee, B. McAllister, T. Massey, D. McLauchlan, T. Stone, K. Correia, J. Loupe, K. H. Kim, D. Barker, E. P. Hong, M. J. Chao, J. D. Long, D. Lucente, J. P. G. Vonsattel, R. M. Pinto, K. A. Elneel, E. M. Ramos, J. S. Mysore, T. Gillis, V. C. Wheeler, C. Medway, L. Hall, S. Kwak, C. Sampaio, M. Ciosi, A. Maxwell, A. Chatzi, D. G. Monckton, M. Orth, G. B. Landwehrmeyer, J. S. Paulsen, I. Shoulson, R. H. Myers, E. van Duijn, H. Rickards, M. E. MacDonald, J. M. Lee, J. F. Gusella, L. Jones and P. Holmans (2020). "Genetic Risk Underlying Psychiatric and Cognitive Symptoms in Huntington's Disease." Biol Psychiatry **87**(9): 857-865.

Ferrante, R. J., N. W. Kowall, M. F. Beal, J. B. Martin, E. D. Bird and E. P. Richardson, Jr. (1987). "Morphologic and histochemical characteristics of a spared subset of striatal neurons in Huntington's disease." J Neuropathol Exp Neurol **46**(1): 12-27.

Genetic Modifiers of Huntington's Disease Consortium. (2015). "Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease." Cell **162**(3): 516-526.

Genetic Modifiers of Huntington's Disease Consortium. (2019). "CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset." Cell **178**(4): 887-900 e814.

Greco, T. M., C. Secker, E. S. Ramos, J. D. Federspiel, J. P. Liu, A. M. Perez, I. Al-Ramahi, J. P. Cantle, J. B. Carroll, J. Botas, S. O. Zeitlin, E. E. Wanker and I. M. Cristea (2022). "Dynamics of huntingtin protein interactions in the striatum identifies candidate modifiers of Huntington disease." Cell Syst **13**(4): 304-320 e305.

Gu, X., J. Richman, P. Langfelder, N. Wang, S. Zhang, M. Banez-Coronel, H. B. Wang, L. Yang, L. Ramanathan, L. Deng, C. S. Park, C. R. Choi, J. P. Cantle, F. Gao, M. Gray, G. Coppola, G. P. Bates, L. P. W. Ranum, S. Horvath, C. S. Colwell and X. W. Yang (2022). "Uninterrupted CAG repeat drives striatum-selective transcriptionopathy and nuclear pathogenesis in human Huntingtin BAC mice." Neuron **110**(7): 1173-1192 e1177.

Gusella, J. F., J. M. Lee and M. E. MacDonald (2021). "Huntington's disease: nearly four decades of human molecular genetics." Hum Mol Genet **30**(R2): R254-R263.

Heinsen, H., M. Strik, M. Bauer, K. Luther, G. Ulmar, D. Gangnus, G. Jungkunz, W. Eisenmenger and M. Gotz (1994). "Cortical and striatal neurone number in Huntington's disease." Acta Neuropathol **88**(4): 320-333.

Hickman, S., S. Izzy, P. Sen, L. Morsett and J. El Khoury (2018). "Microglia in neurodegeneration." Nat Neurosci **21**(10): 1359-1369.

Hong, E. P., M. E. MacDonald, V. C. Wheeler, L. Jones, P. Holmans, M. Orth, D. G. Monckton, J. D. Long, S. Kwak, J. F. Gusella and J. M. Lee (2021). "Huntington's Disease Pathogenesis: Two Sequential Components." J Huntingtons Dis **10**(1): 35-51.

Ito, N., W. T. Hendriks, J. Dhakal, C. A. Vaine, C. Liu, D. Shin, K. Shin, N. Wakabayashi-Ito, M. Dy, T. Multhaupt-Buell, N. Sharma, X. O. Breakefield and D. C. Bragg (2016). "Decreased N-TAF1 expression in X-linked dystonia-parkinsonism patient-specific neural stem cells." Dis Model Mech **9**(4): 451-462.

Kennedy, L., E. Evans, C. M. Chen, L. Craven, P. J. Detloff, M. Ennis and P. F. Shelbourne (2003). "Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis." Hum Mol Genet **12**(24): 3359-3367.

Koeppen, A. H. (2018). "The Neuropathology of Spinocerebellar Ataxia Type 3/Machado-Joseph Disease." Adv Exp Med Biol **1049**: 233-241.

Kovalenko, M., E. Dragileva, J. St Claire, T. Gillis, J. R. Guide, J. New, H. Dong, R. Kucherlapati, M. H. Kucherlapati, M. E. Ehrlich, J. M. Lee and V. C. Wheeler (2012). "Msh2 acts in medium-spiny striatal neurons as an enhancer of CAG instability and mutant huntingtin phenotypes in Huntington's disease knock-in mice." PLoS One **7**(9): e44273.

Krienen, F. M., M. Goldman, Q. Zhang, C. H. D. R. R, M. Florio, R. Machold, A. Saunders, K. Levandowski, H. Zaniewski, B. Schuman, C. Wu, A. Lutservitz, C. D. Mullally, N. Reed, E. Bien, L. Bortolin, M. Fernandez-Otero, J. D. Lin, A. Wysoker, J. Nemesh, D. Kulp, M. Burns, V. Tkachev, R. Smith, C. A. Walsh, J. Dimidschstein, B. Rudy, S. K. L, S. Berretta, G. Fishell, G. Feng and S. A. McCarroll (2020). "Innovations present in the primate interneuron repertoire." Nature **586**(7828): 262-269.

Landwehrmeyer, G. B., S. M. McNeil, L. S. t. Dure, P. Ge, H. Aizawa, Q. Huang, C. M. Ambrose, M. P. Duyao, E. D. Bird, E. Bonilla and et al. (1995). "Huntington's disease gene: regional and cellular expression in brain of normal and affected individuals." Ann Neurol **37**(2): 218-230.

Lee, H., R. J. Fenster, S. S. Pineda, W. S. Gibbs, S. Mohammadi, J. Davila-Velderrain, F. J. Garcia, M. Therrien, H. S. Novis, F. Gao, H. Wilkinson, T. Vogt, M. Kellis, M. J. LaVoie and M. Heiman (2020). "Cell Type-Specific Transcriptomics Reveals that Mutant Huntingtin Leads

to Mitochondrial RNA Release and Neuronal Innate Immune Activation." Neuron **107**(5): 891-908 e898.

Li, X. J., S. H. Li, A. H. Sharp, F. C. Nucifora, Jr., G. Schilling, A. Lanahan, P. Worley, S. H. Snyder and C. A. Ross (1995). "A huntingtin-associated protein enriched in brain with implications for pathology." Nature **378**(6555): 398-402.

Liu, Q., S. Cheng, H. Yang, L. Zhu, Y. Pan, L. Jing, B. Tang, S. Li and X. J. Li (2020). "Loss of Hap1 selectively promotes striatal degeneration in Huntington disease mice." Proc Natl Acad Sci U S A **117**(33): 20265-20273.

Lokanga, R. A., A. Entezam, D. Kumari, D. Yudkin, M. Qin, C. B. Smith and K. Usdin (2013). "Somatic expansion in mouse and human carriers of fragile X premutation alleles." Hum Mutat **34**(1): 157-166.

Long, A., J. S. Napierala, U. Polak, L. Hauser, A. H. Koeppen, D. R. Lynch and M. Napierala (2017). "Somatic instability of the expanded GAA repeats in Friedreich's ataxia." PLoS One **12**(12): e0189990.

Makino, S., R. Kaji, S. Ando, M. Tomizawa, K. Yasuno, S. Goto, S. Matsumoto, M. D. Tabuena, E. Maranon, M. Dantes, L. V. Lee, K. Ogasawara, I. Tooyama, H. Akatsu, M. Nishimura and G. Tamiya (2007). "Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism." Am J Hum Genet **80**(3): 393-406.

Massouh, M., M. J. Wallman, E. Pourcher and A. Parent (2008). "The fate of the large striatal interneurons expressing calretinin in Huntington's disease." Neurosci Res **62**(4): 216-224.

Mouro Pinto, R., L. Arning, J. V. Giordano, P. Razghandi, M. A. Andrew, T. Gillis, K. Correia, J. S. Mysore, D. M. Grote Urtubey, C. R. Parwez, S. M. von Hein, H. B. Clark, H. P. Nguyen, E. Forster, A. Beller, S. Jayadaev, C. D. Keene, T. D. Bird, D. Lucente, J. P. Vonsattel, H. Orr, C. Saft, E. Petrasch-Parwez and V. C. Wheeler (2020). "Patterns of CAG repeat instability in the central nervous system and periphery in Huntington's disease and in spinocerebellar ataxia type 1." Hum Mol Genet **29**(15): 2551-2567.

Paulson, H. (2018). "Repeat expansion diseases." Handb Clin Neurol **147**: 105-123.

Pinto, R. M., E. Dragileva, A. Kirby, A. Lloret, E. Lopez, J. St Claire, G. B. Panigrahi, C. Hou, K. Holloway, T. Gillis, J. R. Guide, P. E. Cohen, G. M. Li, C. E. Pearson, M. J. Daly and V. C. Wheeler (2013). "Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches." PLoS Genet **9**(10): e1003930.

Reiner, A., R. L. Albin, K. D. Anderson, C. J. D'Amato, J. B. Penney and A. B. Young (1988). "Differential loss of striatal projection neurons in Huntington disease." Proc Natl Acad Sci U S A **85**(15): 5733-5737.

Rousseaux, M. W. C., T. Tschumperlin, H. C. Lu, E. P. Lackey, V. V. Bondar, Y. W. Wan, Q. Tan, C. J. Adamski, J. Friedrich, K. Twaroski, W. Chen, J. Tolar, C. Henzler, A. Sharma, A. Bajic, T. Lin, L. Duvick, Z. Liu, R. V. Sillitoe, H. Y. Zoghbi and H. T. Orr (2018). "ATXN1-CIC Complex Is the Primary Driver of Cerebellar Pathology in Spinocerebellar Ataxia Type 1 through a Gain-of-Function Mechanism." Neuron **97**(6): 1235-1243 e1235.

Rui, Y. N., Z. Xu, B. Patel, Z. Chen, D. Chen, A. Tito, G. David, Y. Sun, E. F. Stimming, H. J. Bellen, A. M. Cuervo and S. Zhang (2015). "Huntingtin functions as a scaffold for selective macroautophagy." Nat Cell Biol **17**(3): 262-275.

Sardiello, M., M. Palmieri, A. di Ronza, D. L. Medina, M. Valenza, V. A. Gennarino, C. Di Malta, F. Donaudy, V. Embrione, R. S. Polishchuk, S. Banfi, G. Parenti, E. Cattaneo and A. Ballabio (2009). "A gene network regulating lysosomal biogenesis and function." Science **325**(5939): 473-477.

Saunders, A., E. Z. Macosko, A. Wysoker, M. Goldman, F. M. Krienen, H. de Rivera, E. Bien, M. Baum, L. Bortolin, S. Wang, A. Goeva, J. Nemesh, N. Kamitaki, S. Brumbaugh, D. Kulp and S. A. McCarroll (2018). "Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain." Cell **174**(4): 1015-1030 e1016.

Shelbourne, P. F., C. Keller-McGandy, W. L. Bi, S. R. Yoon, L. Dubeau, N. J. Veitch, J. P. Vonsattel, N. S. Wexler, U. S.-V. C. R. Group, N. Arnheim and S. J. Augood (2007). "Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain." Hum Mol Genet **16**(10): 1133-1142.

Singh-Bains, M. K., N. F. Mehrabi, T. Sehji, M. D. R. Austria, A. Y. S. Tan, L. J. Tippett, M. Dragunow, H. J. Waldvogel and R. L. M. Faull (2019). "Cerebellar degeneration correlates with motor symptoms in Huntington disease." Ann Neurol **85**(3): 396-405.

Squitieri, F., C. Gellera, M. Cannella, C. Mariotti, G. Cislaghi, D. C. Rubinsztein, E. W. Almqvist, D. Turner, A. C. Bachoud-Levi, S. A. Simpson, M. Delatycki, V. Maglione, M. R. Hayden and S. D. Donato (2003). "Homozygosity for CAG mutation in Huntington disease is associated with a more severe clinical course." Brain **126**(Pt 4): 946-955.

Suzuki, M., T. J. Desmond, R. L. Albin and K. A. Frey (2001). "Vesicular neurotransmitter transporters in Huntington's disease: initial observations and comparison with traditional synaptic markers." Synapse **41**(4): 329-336.

Telenius, H., B. Kremer, Y. P. Goldberg, J. Theilmann, S. E. Andrew, J. Zeisler, S. Adam, C. Greenberg, E. J. Ives, L. A. Clarke and et al. (1994). "Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm." Nat Genet **6**(4): 409-414.

Thornton, C. A., K. Johnson and R. T. Moxley, 3rd (1994). "Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes." Ann Neurol **35**(1): 104-107.

Veldman, M. B. and X. W. Yang (2018). "Molecular insights into cortico-striatal miscommunications in Huntington's disease." Curr Opin Neurobiol **48**: 79-89.

von Schimmelmann, M., P. A. Feinberg, J. M. Sullivan, S. M. Ku, A. Badimon, M. K. Duff, Z. Wang, A. Lachmann, S. Dewell, A. Ma'ayan, M. H. Han, A. Tarakhovsky and A. Schaefer (2016). "Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration." Nat Neurosci **19**(10): 1321-1330.

Wertz, M. H., M. R. Mitchem, S. S. Pineda, L. J. Hachigian, H. Lee, V. Lau, A. Powers, R. Kulicke, G. K. Madan, M. Colic, M. Therrien, A. Vernon, V. F. Beja-Glasser, M. Hegde, F. Gao, M. Kellis, T. Hart, J. G. Doench and M. Heiman (2020). "Genome-wide In Vivo CNS Screening Identifies Genes that Modify CNS Neuronal Survival and mHTT Toxicity." Neuron **106**(1): 76-89 e78.

Wheeler, V. C. and V. Dion (2021). "Modifiers of CAG/CTG Repeat Instability: Insights from Mammalian Models." J Huntingtons Dis **10**(1): 123-148.

Wheeler, V. C., L. A. Lebel, V. Vrbanac, A. Teed, H. te Riele and M. E. MacDonald (2003). "Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum." Hum Mol Genet **12**(3): 273-281.

Xu, X., E. I. Stoyanova, A. E. Lemiesz, J. Xing, D. C. Mash and N. Heintz (2018). "Species and cell-type properties of classically defined human and rodent neurons and glia." Elife **7**.

METHODS

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Nathaniel Heintz (heintz@rockefeller.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

All sequencing datasets generated as part of this study will be made publicly available in NCBI GEO under accession ID GSE227729.

Experimental Model and Subject Details

Human Samples

For this work, fresh frozen brain samples were obtained from Miami's Brain Endowment Bank™, University of Washington BioRepository and Integrated Neuropathology Laboratory, Columbia University Alzheimer's Disease Research Center, The University of Michigan Brain Bank and Netherlands Brain Bank, or through the NIH NeuroBioBank and sourced from either the Harvard Brain Tissue Resource Center (HBTRC) or the NIH Brain & Tissue Repository-California, Human Brain & Spinal Fluid Resource Center, VA West LA Medical Center, Los Angeles, California. Drug addiction and schizophrenia as well as clinical evidence of brain cancers were reasons for sample exclusion, while samples from donors with a history of other non-brain cancers and diabetes were accepted. Caudate nucleus, putamen and cerebellar vermis were used for isolation of nuclei. The brain regions used from each donor and their age, race, sex and post-mortem interval, are noted in Suppl. Table 1, 'List of tissue donors'. The table includes information about the Vonsattel grade, calculated CAP100 score (Warner, Long et al. 2022), the number of uninterrupted CAG repeats in their HTT alleles as well as the sequence of the CAG tract and the adjacent CCG tract, as determined from CAG tract length measurement data.

Method Details

Isolation, labeling and sorting of glial cell nuclei

Nuclei were isolated as described previously (Xu, Stoyanova et al. 2018). For the labeling of glial cell nuclei and cerebellar granule cells, the isolated nuclei were washed once with Homogenization buffer (0.25 M sucrose, 150 mM KCl, 5 mM MgCl2, 20 mM Tricine pH 7.8, 0.15 mM spermine, 0.5 mM spermidine, EDTA-free protease inhibitor cocktail, 1 mM DTT, 20 U/mL SUPERase-In RNase Inhibitor (Thermo Fisher, #AM2696), 40 U/mL RNasin ribonuclease inhibitor (Promega #N2515). Each washing step constituted of resuspension of nuclei pellet followed by centrifugation (1000 g, 4 min, +4ºC). Resuspended nuclei were fixed in Homogenization buffer with 1% formaldehyde for 8 min at room temperature followed by quenching with 0.125 M glycine for 5 min. Following centrifugation, the nuclei were washed once with Wash buffer (PBS, 0.05% TritonX-100, 0.5% BSA, 20 U/mL Superase-In RNase Inhibitor, 40 U/mL RNasin ribonuclease inhibitor) and incubated at room temperature on a shaker in Wash buffer for permeabilization and blocking of unspecific binding. Nuclei were washed twice in Wash buffer without TritonX-100 and resuspended in 100 µl of 40% ethanol containing TrueBlack Lipofuscin Autofluorescence Quencher (Biotium, #23007) for 40-50 seconds. Nuclei were washed twice with Wash buffer (w/o TritonX-100) and incubated overnight at +4°C with the following antibodies: Rb x NeuN-Alexa647 (1:400, Abcam #ab190565), Rb x NeuN-Alexa594 (1:400, Abcam #ab207279), Mm x EAAT1 (1:2000, Santa Cruz Biotechnology, #sc-515839), Mm x IRF8-PE (1:65, Thermo Fisher, #12-9852-82) and Goat x Olig2 (1:300, R&D Systems, #AF2418). After three washes with Wash buffer (w/o TritonX-100) the nuclei were incubated for 30-45 min at room temperature with Donkey x Mm-Alexa488 (1:1000, Thermo Fisher, # A-21202) and Donkey x Goat-Alexa647 (1:300, Thermo Fisher, # A-21447). After three washes with Wash buffer (w/o TritonX-100) the nuclei were resuspended in Sorting buffer (PBS, 0.2% BSA, 40 U/mL RNasin ribonuclease inhibitor, 0.5 µg/mL DAPI) (Sigma-Aldrich, #D9542). Aggregates of nuclei were excluded based on higher DAPI signal and the following gating strategies were used: neuronal nuclei (647+, 594+, 488-, large), oligodendrocyte nuclei (647+, 594-, 488-, small), microglia nuclei (647-, 594+, 488-, small) and astrocyte nuclei (647-, 594-, 488+, small). A separate sorting experiment was performed for collecting cerebellar granule cell nuclei. For this purpose nuclei were labeled with Rb x NeuN-Alexa594 (1:400, Abcam #ab207279) and Mm x ITPR1-Alexa488 (Santa Cruz Biotechnology # sc-271197 AF488), and granule cell nuclei were collected (488-, 594+). For labeling neuronal nuclei, PrimeFlow RNA Assay Kit (Thermo Fisher, #88-18005-210) was used and fixation and permeabilization were carried out according to manufacturer's instructions but with 200 U/mL Superase-In RNase

Inhibitor and 400 U/mL RNasin ribonuclease inhibitor present at every incubation step. For sorting the nuclei were resuspended in sorting buffer (PBS, 0.2% BSA, 40 U/mL RNasin ribonuclease inhibitor, 0.5 µg/mL DAPI). Probes specific to DRD1 (Alexa-647, # VA1-3002351-PF), DRD2 (Alexa-488, # VA4-3083767-PF) and PPP1R1B (Alexa-568, # VA10-3266354-PF) were used to label dMSN (647+, 568+, 488-, large) and iMSN nuclei (647-, 568+, 488+, large). In a separate set of experiments, probes specific to TAC3 (Alexa-647, # VA1-16603-PF), ETV1 (Alexa-488, # VA4-3083818-PF), SST (Alexa-568, # VA10-3252595-PF) and PPP1R1B (Alexa-568, # VA10-3266354-PF) were used to label the nuclei of TAC3+ interneurons (647+, 568-, 488+), PVALB+ interneurons (647-, 568-, 488+), SST+ interneurons (647-, 568+++, 488-) and MSNs (647-, 568+, 488-, large). Probes specific to TRPC3 (Alexa-647, # VA1-3004835-PF), COL6A6 (Alexa-647, # VA1-3014134-PF) and PPP1R1B (Alexa-568, # VA10-3266354-PF) were used in another set of experiments to label cholinergic interneuron nuclei (647+, 568, large) and MSN nuclei (647-, 568+, large). CA8 probe (Alexa-647, # VA1-3001892-PF) was used for sorting Purkinje neuron nuclei (647+, large). Aggregates of nuclei were always excluded based on higher intensity of DAPI staining. All PrimeFlow target probes were used at a dilution of 1:40.

ATACseq library preparation

For generating ATACseq data the nuclei were treated with Tagment DNA TDE1 Enzyme (Illumina, #15027865) before fixation and labeling. The exact number of nuclei processed depended on the abundance of the population labeled and collected. Briefly, 800,000 nuclei were pelleted by centrifugation (5 min at 950xg) and resuspended in 10 mM Tris-HCl pH 7.6, 10 mM NaCl, 3 mM MgCl2, 0.01% NP-40 followed by centrifugation (500xg for 10 min at +4°C). The pellet was resuspended in 200uL of Transposition Mix (1x TD buffer containing 20 U/mL Superase-In RNase Inhibitor, 40 U/mL RNasin ribonuclease inhibitor and 1.25 µl of Illumina Tagment DNA TDE1 Enzyme per every 100 000 nuclei) and incubated at +37°C for 30 minutes. The reaction was stopped and nuclei fixed by adding 1mL of Homogenization buffer with 1mM EDTA and 1% formaldehyde. After 8 min of incubation on a shaker the fixative was quenched by adding glycine (0.125 M) for 5 min. After washing the nuclei once in Homogenization buffer and once in Wash buffer (w/o TritonX-100) the sample was processed like described above, proceeding with the steps that follow permeabilization of nuclei. After sorting, the collected nuclei were centrifuged at 1600xg for 10 min +4°C and resuspended in 200 uL of RC solution (50 mM Tris-HCl pH 7.6, 200 mM NaCl, 1 mM EDTA, 1% SDS, 5 µg/mL Proteinase K) and incubated overnight at +55°C. Genomic DNA was isolated with using MinElute Reaction

Cleanup Kit (Qiagen, #28206) and used for PCR amplification (72°C – 5 min, 98°C – 30 sec, 12-14x[98°C – 10 sec, 63°C – 30 sec, 72°C – 1 min]) with NEBNext® High-Fidelity 2X PCR Master Mix (New England Biolabs, #M0541S) and barcoded Nextera primers (1.25 µM each) (Buenrostro, Giresi et al. 2013). Following double-sided size selection by bead-purification the libraries were quantified with Qubit dsDNA HS assay kit (ThermoFisher #Q32851) and pooled for sequencing on NovaSeq 6000 (SP 2 x 100bp).

FANSseq library preparation and sequencing

RNA extraction was carried out with AllPrep DNA/RNA FFPE Kit (Qiagen, #80234) with modifications described previously (Xiao 2018). RNA-seq libraries were prepared with Trio RNA-Seq™ library preparation kit (Tecan, #0506-A01), quantified with Qubit dsDNA HS assay kit (ThermoFisher #Q32851) and pooled for sequencing on NovaSeq6000 (SP 2 x 150bp).

RNAseq Data Processing

Sequence and transcript coordinates for human hg38 UCSC genome and gene models were retrieved from the BSgenome.Hsapiens.UCSC.hg38 Bioconductor package (version 1.4.1) and TxDb.Hsapiens.UCSC.hg38.knownGene (version 3.4.0) Bioconductor libraries respectively. FANSseq reads were aligned to the genome using Rsubread's subjunc method (version 1.30.6) (Liao, Smyth et al. 2013) and exported as bigWigs normalized to reads per million using the rtracklayer package (version 1.40.6). Reads in genes were counted using the featurecounts function within the Rsubread package against the full gene bodies (Genebody.Counts) and gene exons (Gene.Counts).

ATACseq Data Processing

The ATACseq reads were aligned with the hg38 genome from the BSgenome.Hsapiens.UCSC.hg38 Bioconductor package (version 1.4.1) with Rsubread's align method in paired-end mode. Fragments between 1 and 5,000 base pairs long were considered correctly paired. Normalized, fragment signal bigWigs were created with the rtracklayer package. Peak calls were made with MACS2 software in BAMPE mode (Zhang, Liu et al. 2008, Feng, Liu et al. 2012). For each striatal interneuron type except cholinergic interneurons, the ATAC-seq consensus peaks were called from 4 ATAC-seq datasets generated from four different control donors. For MSNs, ATAC-seq consensus peaks were called from 8 dMSNs datasets from 7 different HD donors, from 9 iMSN datasets from 8 different HD donors, and 31 dMSNs datasets and 32 iMSN datasets from 8 control donors (up to four datasets from each

donor). High confidence consensus peaks were derived by creating a nonredundant peak set for each cell type and disease state, and then filtering down to peaks that were present in the majority of samples. These were then annotated to TSS based on proximity using the ChIPseeker package (version 1.28.3) (Wang, Li et al. 2022). NCBI RefSeq hg38 gene annotation was used (version 109.20211119).

Differential gene expression analysis and principal components analysis

For comparison of transcript abundance data between different cell types from control donors, the comparisons of data from caudate nucleus and putamen were done independently. For control donors from which there was data available from both posterior and anterior parts of the same structure, a single table of average raw read counts per gene was generated for each cell type. For comparison of control donor data to HD donor data, up to four separate datasets for a given cell type (anterior putamen, posterior putamen, anterior caudate nucleus, posterior caudate nucleus) were combined by calculating the average raw read counts per gene (rounding up to integer), so that each donor was represented by a single FANSseq dataset for each cell type. Principal component analysis plots for 500 most variant genes were generated with pcaExplorer (Marini and Binder 2019) using average raw read 'Genebody.Counts' tables as input data (one table for each donor per cell type). Average raw read 'Gene.Counts', one for each donor per cell type, were converted into normalized counts by DESeq2, thereby accounting for sequencing depth differences, and used for differential gene expression analysis by DESeq2 (Love, Hogenesch et al. 2016) (Love, Soneson et al. 2018) (version 1.36.0). For comparing expression across cell types, 'expression in cell type A' was calculated as the mean of DESeq2-normalized counts from each donor. 'Expression in a cell type A' was then turned into relative expression ('relative expression in cell type A' = 'expression in a cell type A' divided by 'mean of expression in all cell types compared') and the resulting values were log2-transformed for visualization by Pheatmap R package (version 1.0.12). Relative expression was calculated in the same manner when comparing expression across individual samples instead of cell types. Differential gene expression analysis results were filtered to exclude genes for which none of their annotated TSS positions in NCBI RefSeq hg38 (version 109.20211119) overlapped with ATAC-seq consensus peaks defined separately for the cell types compared. These lists were augmented with a small number of genes (<110) for which manual inspection of mapped FANSseq and ATACseq reads in Integrative Genomics Viewer (Robinson, Thorvaldsdottir et al. 2011) suggested that these genes were in fact expressed.

Pathway enrichment analysis

The filtered lists of differentially expressed genes (DEGs) (p adj < 0.05) with accessible TSS regions were used for analysis of over-representation of Gene Ontology Cellular Compartment (GOCC) terms with clusterProfiler package (Yu, Wang et al. 2012) (version 4.4.4, GOSOURCEDATE: 2022-03-10). The augmented list of all genes with accessible TSS regions was used as the 'background' list for comparison ('universe'), and the following parameters were used: qvalueCutoff = 0.05, minGSSize = 5, maxGSSize = 2000. For identifying GOCC pathways over-represented among genes that showed disease-associated up- or downregulation only in dMSNs, the list of DEGs from 'HD_dMSN vs ctrl_dMSN' comparison (P adj. < 0.05) was filtered to exclude genes that had changed expression (P adj. < 0.05) in the same direction in any of the three 'HD_interneuron vs ctrl_interneuron' comparisons (TAC3+ INs, PVALB+ INs or SST+ INs).

***HTT* and *ATXN3* CAG tract sizing**

Genomic DNA (gDNA) was purified using AllPrep DNA/RNA FFPE Kit (Qiagen, #80234) and concentrated in a vacuum concentrator if required. *HTT* CAG tract sizing was done by next generation sequencing of PCR amplicons of *HTT* exon1 using a modified version of a previously published protocol (Ciosi, Maxwell et al. 2019). Up to 10 ng of gDNA was amplified in a 20 µl volume using NEBNext® High-Fidelity 2X PCR Master Mix (New England Biolabs, #M0541S) supplemented with 5% dimethyl sulfoxide and barcoded primers specific to *HTT* Exon 1 (0.5 µM each) (Ciosi, Maxwell et al. 2019) or *ATXN3* exon 10: 1 cycle 96°C - 5 min, 30x [96°C – 45 sec, 61°C – 45 sec, 72°C – 3 min, 72°C – 10 min. The number of amplification cycles was raised to 32 cycles if the amount of gDNA input was below 4 ng. After PCR the samples were combined into small pool of 2-6 samples and size selection was carried out by adding 0.55x volume of AMPure XP beads (Beckman Coulter, #A68831). The concentrations of purified library pools were quantified with Collibri™ Library Quantification Kit (ThermoFisher, #A38524100), combined into a sequencing library and sequenced on MiSeq sequencer using a 500 cycle MiSeq Reagent Nano Kit v2 (Illumina, # MS-103-1003) with both index reads, but with 400 nt long read 1 and no read 2. Demultiplexed sequencing read data was aligned using Burrows-Wheeler Aligner (BWA MEM default settings except: -O 6,6 -E 4,4) to a set of *HTT* exon 1 or *ATXN3* exon 10 reference sequences that differed by the number of CAG repeat units in the repeat tract. The number of reads uniquely mapped to each of the reference sequences in the set was considered to reflect the distribution of CAG tract lengths in the two *HTT* or *ATXN3* alleles in the cell population analyzed. *HTT* read mapping data from each donor was inspected

manually for determining the nucleotide sequence of the adjacent polyproline tract and the presence/absence of interruptions in CAG tract. If *mHTT* exon 1 structure of the donor was atypical, then sequencing reads were realigned to a set of reference sequences matching that *mHTT* exon 1 structure. The length of CAG repeat tracts reliably mapped was limited to 113 repeat units. Uninterrupted CAG tract lengths of progenitor/unexpanded *mHTT* allele (M repeat units) and normal *HTT* allele (N repeat units) were defined from the two modes of mapped read length-distribution in CAG-sizing data from non-expanding cell types (usually striatal microglia and astrocytes, or, if available, cerebellar granule cells). R is the number of reads mapped to a reference sequence with the specified CAG tract length. The ratio of somatic expansions (RoSE) (Ciosi, Maxwell et al. 2019) and frequency of *mHTT* copies expanded by more than 20 repeat units (RU) were calculated as follows:

$$normal\ allele\ RoSE = \frac{R_{N+1} + R_{N+2} + R_{N+3}}{R_N}$$

$$mutant\ allele\ RoSE = \frac{\sum_{i=M+1}^{113} Ri}{R_M}$$

$$\%\ of mutant\ allele\ copies\ expanded\ by > 20\ RU = \frac{\sum_{i=M+21}^{113} Ri}{\sum_{i=M}^{113} Ri} \times 100\%$$

Quantification of CAG tract length changes for *mATXN3* was done in the same way. Statistical analysis of differences between cell types was carried out by comparing their ratio of somatic expansions or frequency of *mHTT* copies expanded by > 20 RU with one-way ANOVA, followed by Holm-Sidak's multiple comparisons test.

Western blotting

When isolating nuclei for western blotting the tissue homogenization and ultracentrifugation steps were carried out as described by Xiao et al. (Xu, Stoyanova et al. 2018). After washing the nuclei once in Homogenization buffer the nuclei were resuspended in 1 mL of 1x PBS, 0.05% Triton X-100, 2% BSA and incubated at room temperature on a shaker for ~ 15-20 minutes. The nuclei were labeled by adding the following antibodies: Rb x NeuN-Alexa647 (1:300, Abcam #ab190565), Rb x NeuN-Alexa594 (1:300, Abcam #ab207279), Mm x EAAT1-Alexa488 (1:200, Santa Cruz Biotechnology, #sc-515839 AF488), Mm x IRF8-PE (1:65, ThermoFisher, #12-9852-82) and Goat x Olig2 (1:200, R&Dsystems, #AF2418). After two washes with WB Wash buffer (1x PBS, 0.05% Triton X-100, 0.2% BSA) the nuclei were incubated for 30 min at room temperature with Donkey x Goat-Alexa647 (1:400, ThermoFisher, # A-21447). Nuclei were

washed twice with WB Wash buffer and resuspended in Sorting buffer (w/o RNase inhibitors). After sorting, the collected nuclei were centrifuged at 1600xg for 10 min +4°C and the residual volume was kept to a minimum. Nuclei were treated with DNase I (in the presence of 0.5 mM $MgCl_2$) at +37°C for 10 min, mixed with NuPAGE Sample Reducing Agent (ThermoFisher #NP0004) and β-mercaptoethanol (final conc. 4%, Sigma, #M3148), and heat-denaturated at +96°C for 3 minutes. Material from 25,000 to 50,000 nuclei were loaded on NuPAGE™ 4 to 12% Bis-Tris Mini gels (ThermoFisher, #NP0322BOX), aiming for equal loading in each well. After blotting the samples onto nitrocellulose membrane and blocking unspecific binding by incubating the membrane in a 5% solution of non-fat dry milk, the membranes were probed with Rb x Histone H3 antibody (1:5000, Abcam, #ab1791) and Mm x Human MSH2 (1:300 BD Biosciences, #556349) or Mm x Human MSH3 (1:300 BD Biosciences, #611390) by incubating overnight at +4°C. After three washes with TBS-T (1x TBS, 0.1% Tween-20) the membranes were probed with IRDye® 680LT Donkey anti-Rabbit IgG Secondary Antibody (1:10 000, LICOR, #926-68023) and IRDye® 800CW Goat anti-Mouse IgG Secondary Antibody (1:10 000, LICOR, #926-32210) by 1 hour at room temperature. After three washes with TBS-T the membranes were imaged with Odyssey® DLx Imaging System.

SUPPPLEMENTAL REFERENCES

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang and W. J. Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nat Methods **10**(12): 1213-1218.

Ciosi, M., A. Maxwell, S. A. Cumming, D. J. Hensman Moss, A. M. Alshammari, M. D. Flower, A. Durr, B. R. Leavitt, R. A. C. Roos, T.-H. team, H. D. t. Enroll, P. Holmans, L. Jones, D. R. Langbehn, S. Kwak, S. J. Tabrizi and D. G. Monckton (2019). "A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes." EBioMedicine **48**: 568-580.

Feng, J., T. Liu, B. Qin, Y. Zhang and X. S. Liu (2012). "Identifying ChIP-seq enrichment using MACS." Nat Protoc **7**(9): 1728-1740.

Liao, Y., G. K. Smyth and W. Shi (2013). "The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote." Nucleic Acids Res **41**(10): e108.

Love, M. I., J. B. Hogenesch and R. A. Irizarry (2016). "Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation." Nat Biotechnol **34**(12): 1287-1291.

Love, M. I., C. Soneson and R. Patro (2018). "Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification." F1000Res **7**: 952.

Marini, F. and H. Binder (2019). "pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components." BMC Bioinformatics **20**(1): 331.

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov (2011). "Integrative genomics viewer." Nat Biotechnol **29**(1): 24-26.

Wang, Q., M. Li, T. Wu, L. Zhan, L. Li, M. Chen, W. Xie, Z. Xie, E. Hu, S. Xu and G. Yu (2022). "Exploring Epigenomic Datasets by ChIPseeker." Curr Protoc **2**(10): e585.

Warner, J. H., J. D. Long, J. A. Mills, D. R. Langbehn, J. Ware, A. Mohan and C. Sampaio (2022). "Standardizing the CAP Score in Huntington's Disease by Predicting Age-at-Onset." J Huntingtons Dis **11**(2): 153-171.

Xu, X., E. I. Stoyanova, A. E. Lemiesz, J. Xing, D. C. Mash and N. Heintz (2018). "Species and cell-type properties of classically defined human and rodent neurons and glia." Elife **7**.

Yu, G., L. G. Wang, Y. Han and Q. Y. He (2012). "clusterProfiler: an R package for comparing biological themes among gene clusters." OMICS **16**(5): 284-287.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li and X. S. Liu (2008). "Model-based analysis of ChIP-Seq (MACS)." Genome Biol **9**(9): R137.

Figure 1

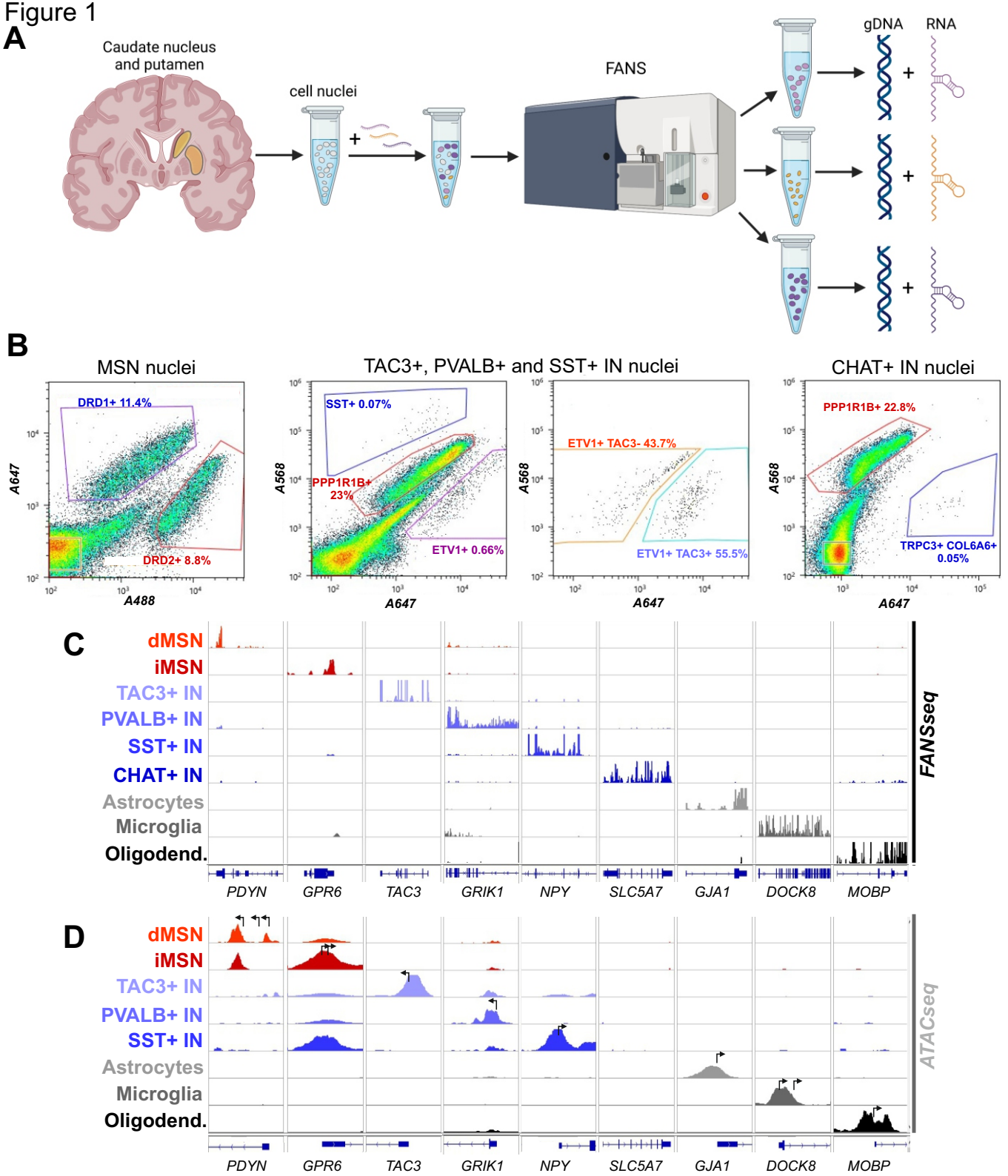

A
Caudate nucleus and putamen
cell nuclei
FANS
gDNA
RNA
B
MSN nuclei
TAC3+, PVALB+ and SST+ IN nuclei
CHAT+ IN nuclei
DRD1+ 11.4%
DRD2+ 8.8%
A647
A488
SST+ 0.07%
PPP1R1B+ 23%
ETV1+ 0.66%
A568
ETV1+ TAC3- 43.7%
ETV1+ TAC3+ 55.5%
PPP1R1B+ 22.8%
TRPC3+ COL6A6+ 0.05%
C
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
CHAT+ IN
Astrocytes
Microglia
Oligodend.
FANSseq
PDYN
GPR6
TAC3
GRIK1
NPY
SLC5A7
GJA1
DOCK8
MOBP
D
ATACseq

Figure legends

Figure 1. FANS-based isolation of nuclei of striatal cell types from human post-mortem caudate nucleus and putamen. (**A**) Schematic representation of the procedure used to extract cell type-specific genomic DNA and nuclear RNA from cell nuclei labeled with cell type-specific probes. Created with BioRender.com. (**B**) Representative FANS plots showing the labeling of striatal cell nuclei with Primeflow probes specific for dMSNs (DRD1+ nuclei), iMSNs (DRD2+ nuclei), SST+ interneurons (SST+ nuclei), PVALB+ interneurons (ETV1+ TAC3- nuclei), TAC3+ interneurons (ETV1+ TAC3+ nuclei) and CHAT+ interneurons (TRPC3+ COL6A6+ nuclei). The probe against *PPP1R1B* labels all MSN nuclei. The detailed strategy used for sorting is described in Materials and Methods. (**C-D**) Representative distribution of human FANSseq (**C**) and ATACseq (**D**) reads mapped to genes expressed specifically in each of the striatal cell types studied. In **D**, arrows mark the position of annotated transcriptional start sites. The data are from a 41-year-old male donor.

Figure 2

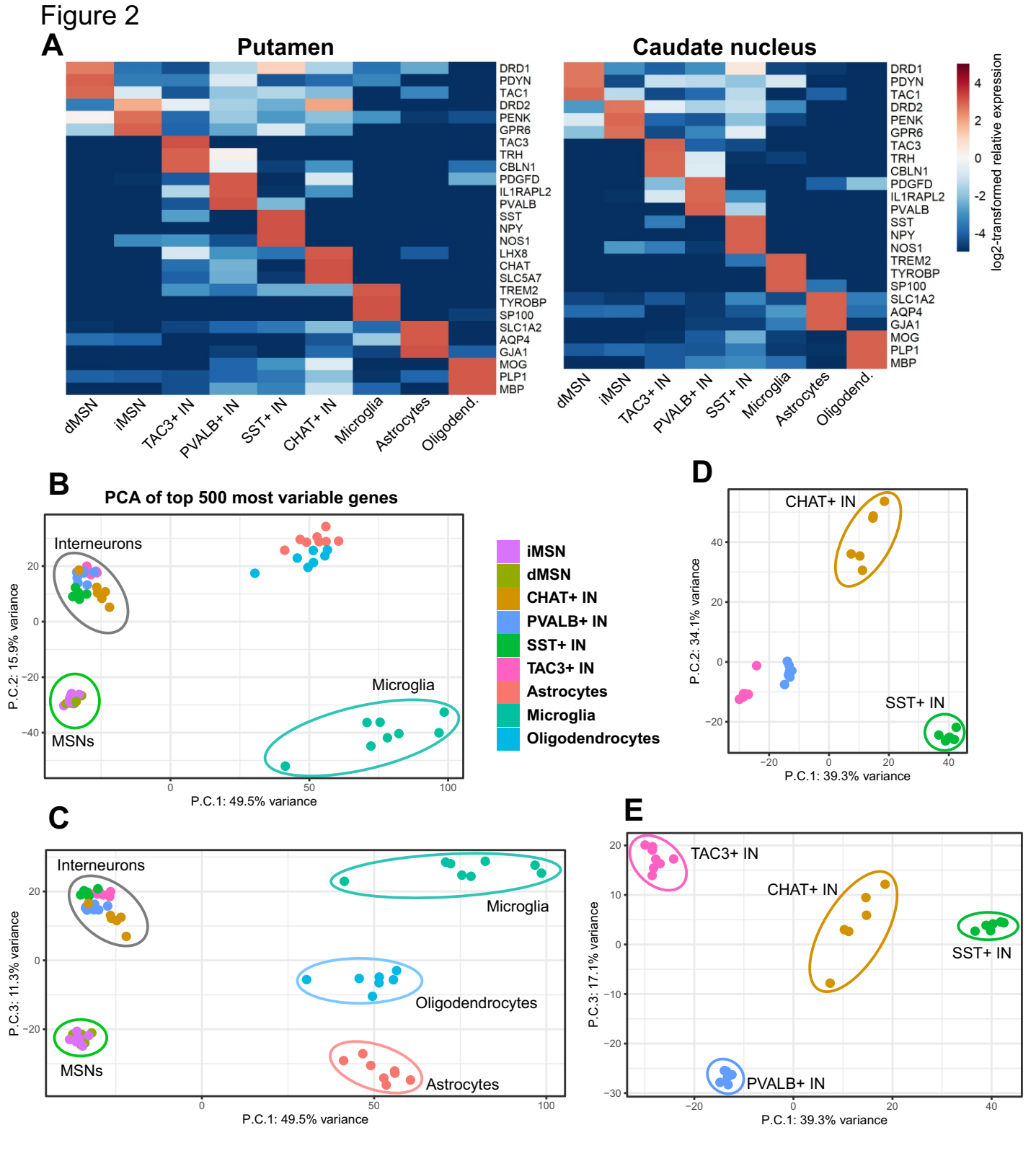

A
Putamen
Caudate nucleus
DRD1
PDYN
TAC1
DRD2
PENK
GPR6
TAC3
TRH
CBLN1
PDGFD
IL1RAPL2
PVALB
SST
NPY
NOS1
LHX8
CHAT
SLC5A7
TREM2
TYROBP
SP100
SLC1A2
AQP4
GJA1
MOG
PLP1
MBP
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
CHAT+ IN
Microglia
Astrocytes
Oligodend.
log2-transformed relative expression
B
PCA of top 500 most variable genes
Interneurons
MSNs
Microglia
P.C.2: 15.9% variance
P.C.1: 49.5% variance
iMSN
dMSN
CHAT+ IN
PVALB+ IN
SST+ IN
TAC3+ IN
Astrocytes
Microglia
Oligodendrocytes
D
CHAT+ IN
SST+ IN
P.C.2: 34.1% variance
P.C.1: 39.3% variance
C
Interneurons
Microglia
Oligodendrocytes
MSNs
Astrocytes
P.C.3: 11.3% variance
P.C.1: 49.5% variance
E
TAC3+ IN
CHAT+ IN
SST+ IN
PVALB+ IN
P.C.3: 17.1% variance
P.C.1: 39.3% variance

Figure 2. Purity and reproducibility of the isolation of striatal cell type nuclei across the two striatal brain regions studied. (**A**) Heatmaps depict log2-transformed relative expression level of cell type-specific marker genes in each cell type, calculated based on the mean of DESeq2-normalized counts from 6-8 control donors. (**B-E**) Principal component analysis of FANSseq data from all putamen cell types (**B and C**) or interneuron types (**D and E**) from 6-8 control donors.

Figure 3

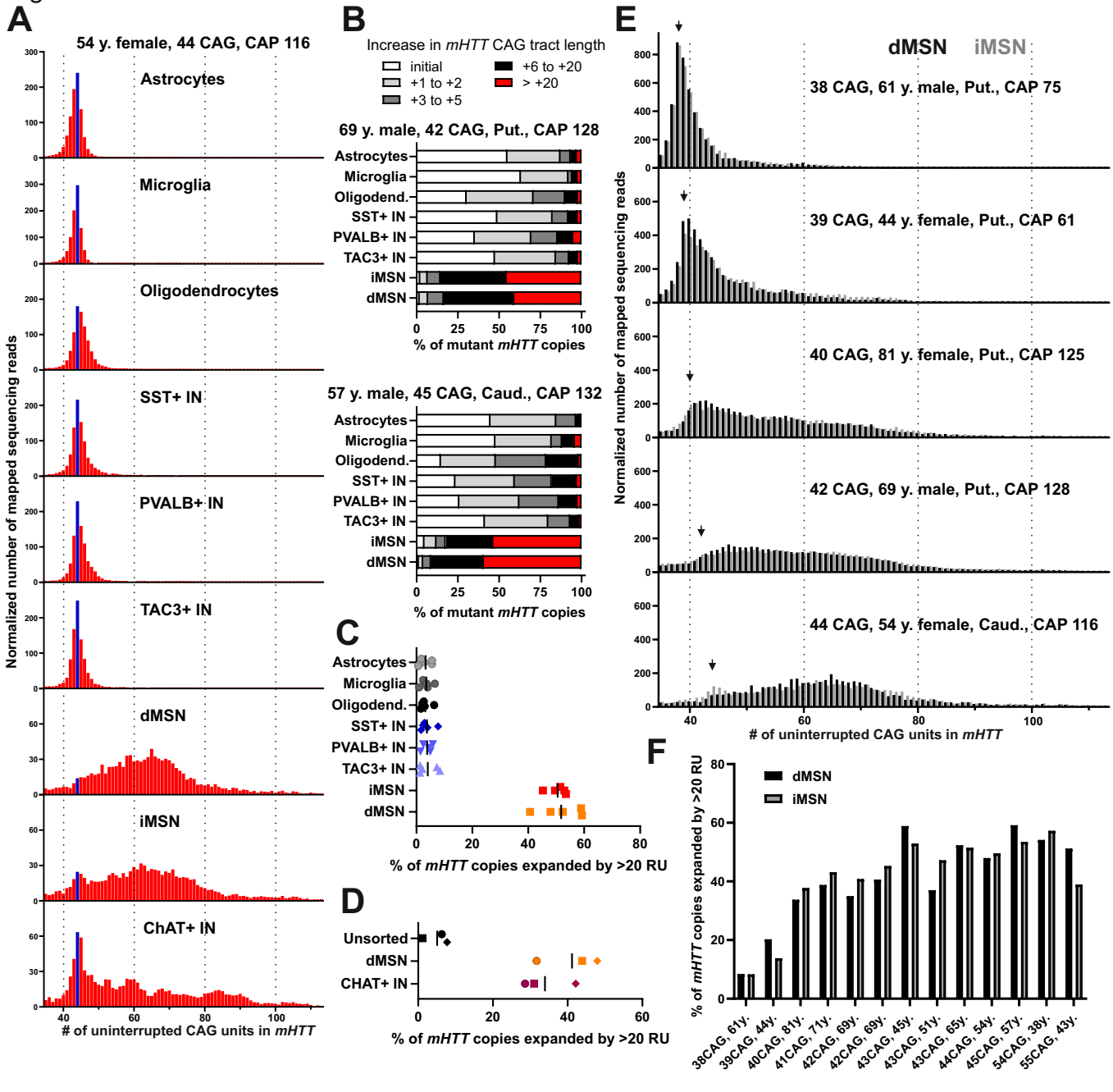
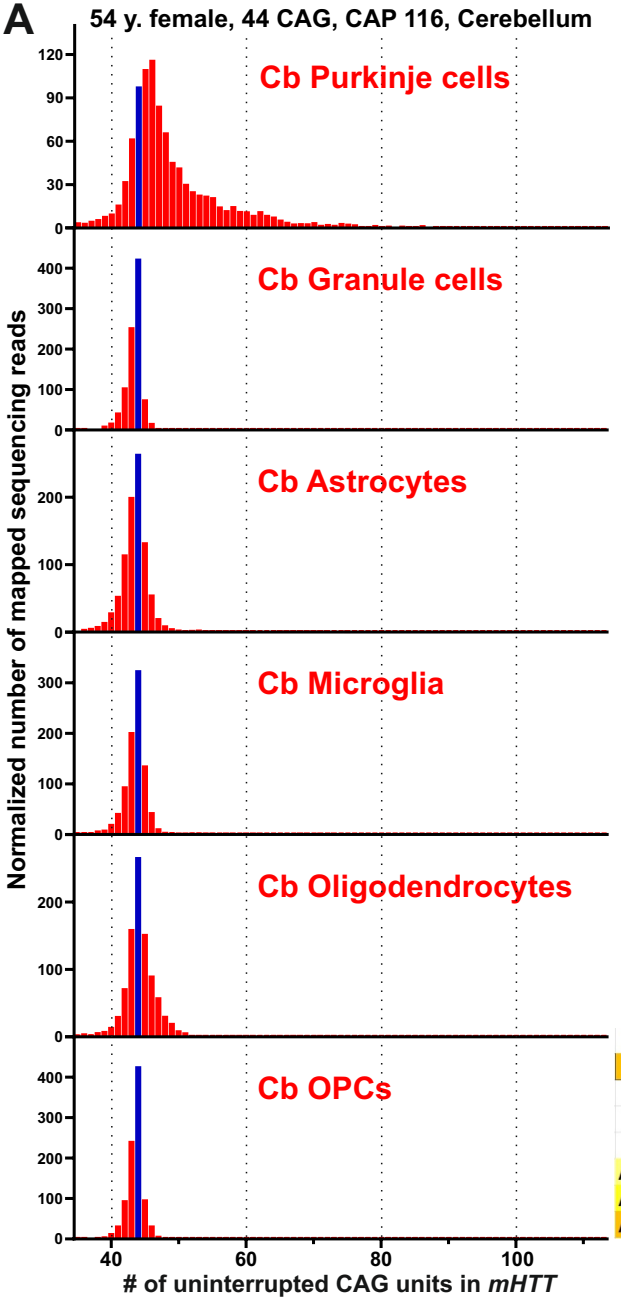
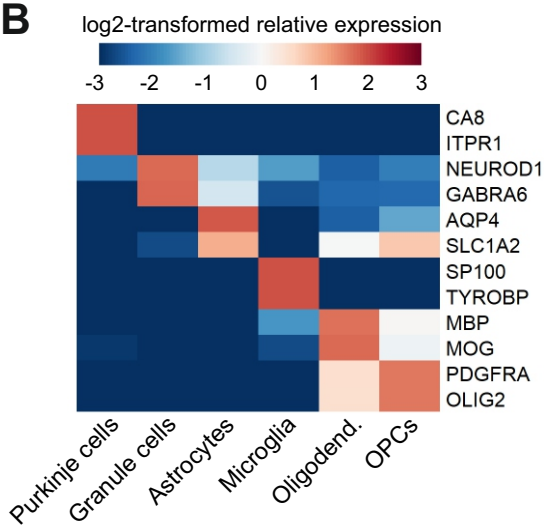
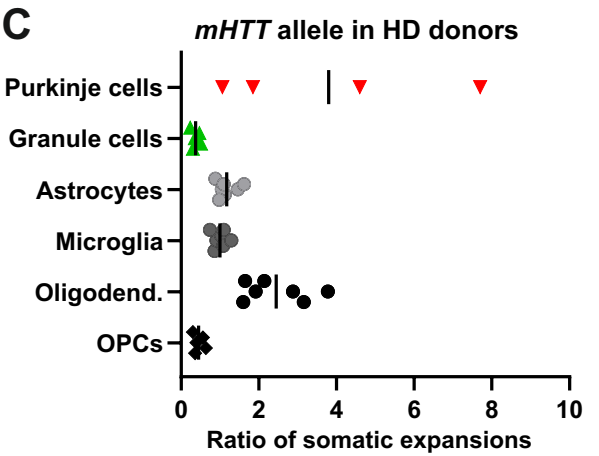

A
54 y. female, 44 CAG, CAP 116
Astrocytes
Microglia
Oligodendrocytes
SST+ IN
PVALB+ IN
TAC3+ IN
dMSN
iMSN
ChAT+ IN
Normalized number of mapped sequencing reads
# of uninterrupted CAG units in mHTT
B
Increase in mHTT CAG tract length
initial
+1 to +2
+3 to +5
+6 to +20
> +20
69 y. male, 42 CAG, Put., CAP 128
57 y. male, 45 CAG, Caud., CAP 132
Astrocytes
Microglia
Oligodend.
SST+ IN
PVALB+ IN
TAC3+ IN
iMSN
dMSN
% of mutant mHTT copies
C
% of mHTT copies expanded by >20 RU
D
Unsorted
dMSN
CHAT+ IN
% of mHTT copies expanded by >20 RU
E
dMSN
iMSN
38 CAG, 61 y. male, Put., CAP 75
39 CAG, 44 y. female, Put., CAP 61
40 CAG, 81 y. female, Put., CAP 125
42 CAG, 69 y. male, Put., CAP 128
44 CAG, 54 y. female, Caud., CAP 116
Normalized number of mapped sequencing reads
# of uninterrupted CAG units in mHTT
F
dMSN
iMSN
% of mHTT copies expanded by >20 RU
38CAG, 61y.
39CAG, 44y.
40CAG, 81y.
41CAG, 71y.
42CAG, 69y.
42CAG, 69y.
43CAG, 45y.
43CAG, 51y.
43CAG, 65y.
44CAG, 54y.
45CAG, 57y.
54CAG, 38y.
55CAG, 43y.

Figure 3. *mHTT* CAG tract undergoes somatic expansion in selected striatal neuron types. (**A**) Length distribution of *mHTT* CAG tract in studied cell types of caudate nucleus and putamen of a 44-year-old female donor that carried a tract of 44 uninterrupted CAG units. Blue bar marks sequencing reads derived from the initial unexpanded CAG tract. Y-axes denote normalized number of sequencing reads mapped to reference sequences with different CAG tract lengths (normalized by scaling to 1000 reads). Reads derived from the normal HTT allele are not shown. (**B**) Frequency distribution of *mHTT* copies with different CAG tract length increases. Data is shown for striatal cell types of two donors that carried the most common HD-causing CAG tract lengths. (**C**) Comparison of the fraction of *mHTT* copies with CAG tracts expanded by more than 20 repeat units (RU). While the frequency of most highly expanded mHTT CAG tract copies was not different between dMSN and iMSN, comparison of each of these to any other striatal cell type showed a statistically significant difference ($P < 0.0001$ by one-way ANOVA, adjusted P value <0.0001 in Holm-Sidak's multiple comparisons test). (**D**) Comparison of the fraction of *mHTT* copies with CAG tracts expanded by more than 20 repeat units (RU). While the frequency of most highly expanded mHTT CAG tract copies was not different between dMSNs and CHAT+ interneurons, comparison of each of these to unsorted nuclei showed a statistically significant difference ($P = 0.0013$ by one-way ANOVA, adjusted P value <0.004 in Holm-Sidak's multiple comparisons test). Different symbols are used for each of the three donors, all carrying *mHTT* allele with an initial length of 44 uninterrupted CAG repeats. (**E**) Length distribution of *mHTT* CAG tract in striatal projection neurons of caudate nucleus and putamen in donors carrying *mHTT* alleles of reduced and full penetrance. Arrowhead indicates the initial unexpanded size of the CAG tract. Y axes denote normalized number of sequencing reads mapped to reference sequences with different CAG tract lengths. (**F**) Comparison of the fraction of *mHTT* copies with CAG tracts expanded by more than 20 repeat units (RU). $P = 0.896$ between cell types, in ratio paired t-test.

Figure 4





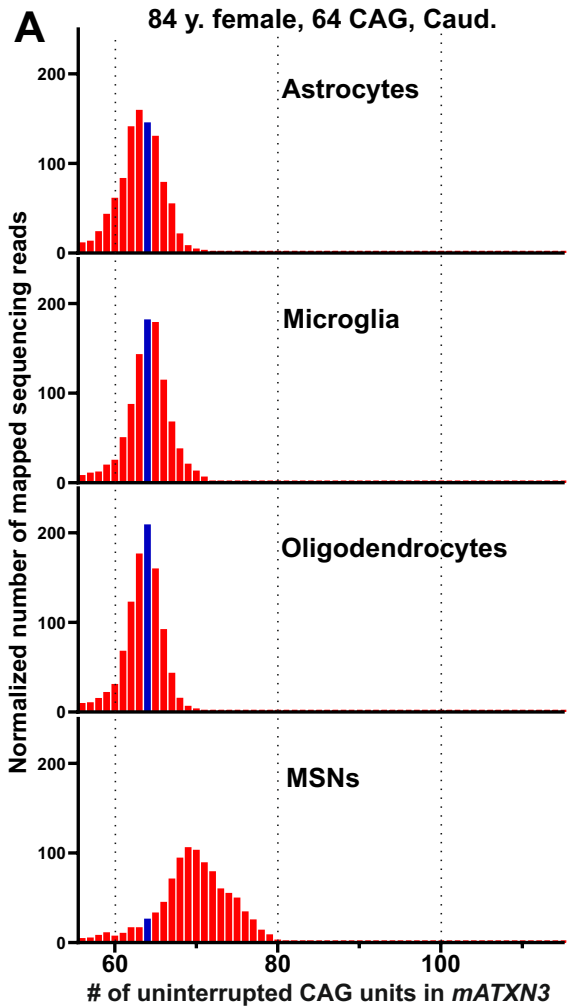


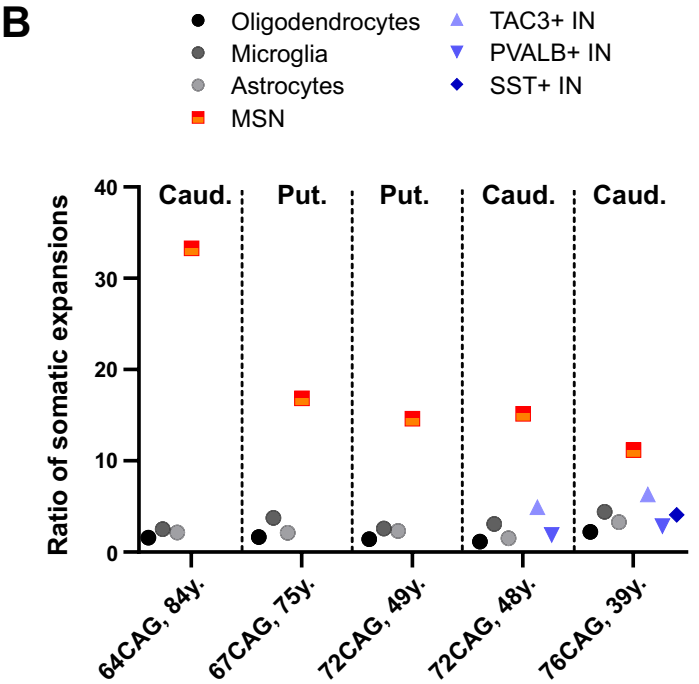
	Granule c.	Astrocytes	Microglia	Oligodend.	OPCs
Purkinje c.	0.0006	0.005	0.003	0.30	0.0007
Granule c.		0.75	0.78	0.023	0.94
Astrocytes			0.94	0.22	0.76
Microglia				0.13	0.78
Oligodend.					0.029

Adj. P < 0.05
Adj. P < 0.01
Adj. P < 0.001

Figure 4. Expansion of *mHTT* CAG tract in cerebellar Purkinje cells. (**A-B**) Length distribution of *mHTT* CAG tract (**A**) and cell type marker gene expression (**B**) in cerebellar cell nuclei isolated from a 44-year-old female donor that carried a tract of 44 uninterrupted CAG units. Blue bar marks sequencing reads derived from the initial, unexpanded CAG tract. Y axes denote normalized numbers of sequencing reads mapped to reference sequences with different CAG tract lengths. Reads derived from the shorter normal HTT allele are not shown. Heatmap depicts log2-transformed relative expression in each sample (calculated based on DESeq2-normalized counts). (**C**) Comparison of the calculated ratio of somatic expansions of *mHTT* CAG tract in cerebellar cell types. The table presents adjusted P values as calculated by Holm-Sidak's multiple comparisons test post one-way ANOVA ($P < 0.0001$).

Figure 5





Microglia	Astrocytes	MSN	TAC3+ IN	PVALB+ IN	
0.993	0.9984	<0.0001	0.9521	0.9984	Oligodend.
	0.9979	0.0003	0.993	0.9984	Microglia
		0.0001	0.9787	0.9984	Astrocytes
Adj. P < 0.001			0.02	0.0028	MSN
Adj. P < 0.05				0.9909	TAC3+ IN

Figure 5: Long *mATXN3* CAG tract expands more in striatal projection neurons than in other striatal cell types. (**A**) Length distribution of *mATXN3* CAG tract in cell types of caudate nucleus of an 84-year-old female donor that carried a tract of 64 uninterrupted CAG units. Sequencing reads derived from the initial unexpanded CAG tract are marked with a blue bar. Y axes denote normalized numbers of sequencing reads mapped to reference sequences with different CAG tract lengths (normalized by scaling to 1000 reads). Reads derived from the normal *ATXN3* allele are not shown. (**B**) Comparison of the calculated ratio of somatic expansions of *mATXN3* CAG tract in striatal cell types. The table presents adjusted P values as calculated by Holm-Sidak's multiple comparisons test following one-way ANOVA ($P < 0.0001$).

Figure 6

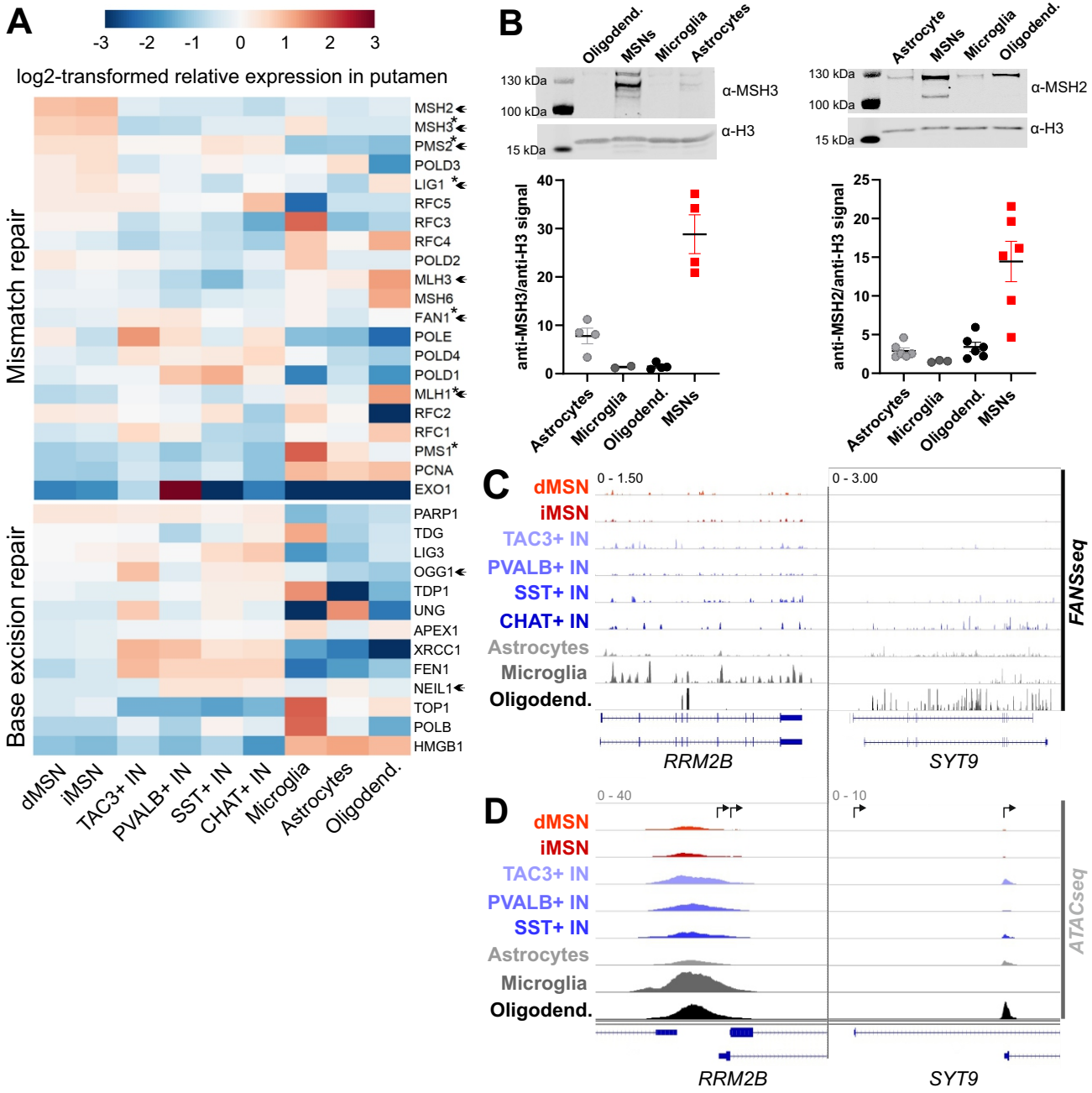

A
-3 -2 -1 0 1 2 3
log2-transformed relative expression in putamen
Mismatch repair
Base excision repair
MSH2
MSH3*
PMS2*
POLD3
LIG1 *
RFC5
RFC3
RFC4
POLD2
MLH3
MSH6
FAN1*
POLE
POLD4
POLD1
MLH1*
RFC2
RFC1
PMS1*
PCNA
EXO1
PARP1
TDG
LIG3
OGG1
TDP1
UNG
APEX1
XRCC1
FEN1
NEIL1
TOP1
POLB
HMGB1
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
CHAT+ IN
Microglia
Astrocytes
Oligodend.
B
Oligodend.
MSNs
Microglia
Astrocytes
130 kDa
100 kDa
15 kDa
α-MSH3
α-H3
Astrocyte
MSNs
Microglia
Oligodend.
α-MSH2
α-H3
anti-MSH3/anti-H3 signal
anti-MSH2/anti-H3 signal
Astrocytes
Microglia
Oligodend.
MSNs
C
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
CHAT+ IN
Astrocytes
Microglia
Oligodend.
0 - 1.50
0 - 3.00
FANSseq
RRM2B
SYT9
D
0 - 40
0 - 10
ATACseq
RRM2B
SYT9

Figure 6: Elevated expression of MSH2 and MSH3 proteins in MSNs. (**A**) Relative expression level of MMR and BER genes in cell types of putamen. Heatmaps depict log2-transformed relative expression in each cell type, calculated based on the mean of DESeq2-normalized counts from 6-8 control donors. Genes identified as HD age at onset-modifying candidates are marked with an asterisk. Genes known to influence CAG tract instability in HD mouse models are marked with an arrowhead. (**B**) Representative immunoblots and quantification of anti-MSH3/anti-H3 (left) and anti-MSH2/anti-H3 ratio (right) in unfixed nuclei isolated from the putamen of 4 or 6 control donors, respectively. The anti-MSH3/anti-H3 and anti-MSH2/anti-H3 signal ratios were higher in MSNs compared to any other analyzed cell types ($P < 0.0001$ by one-way ANOVA, adjusted P value ≤ 0.0005 in Tukey's multiple comparisons test). Data are presented as mean +/- SEM. (**C**) Representative distribution of human FANSseq and (**D**) ATACseq reads mapped to *RRM2B* and *SYT9* genes. Arrows mark the position of annotated transcriptional start sites. The data are from a 41-year-old male donor.

Figure 7

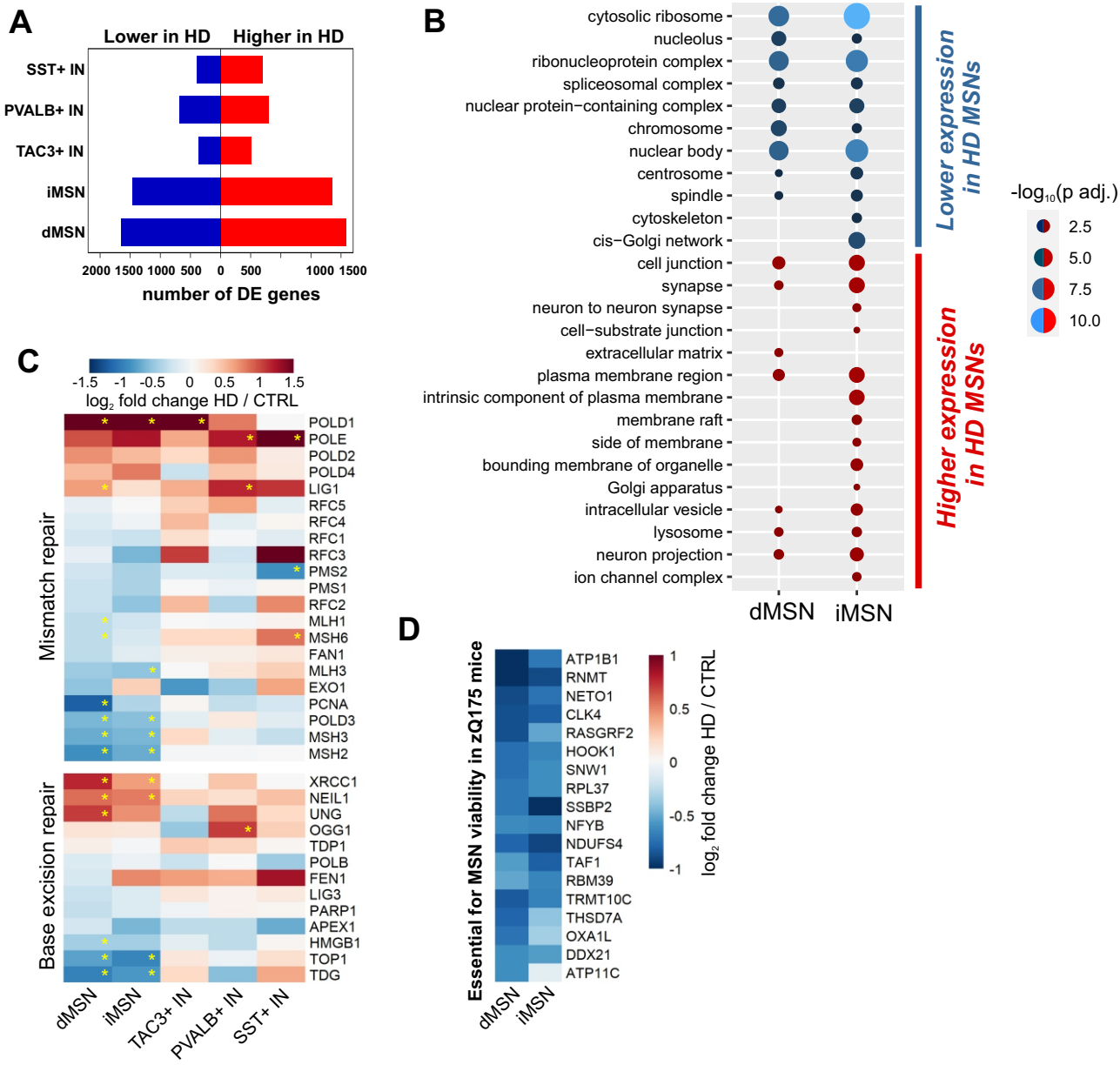
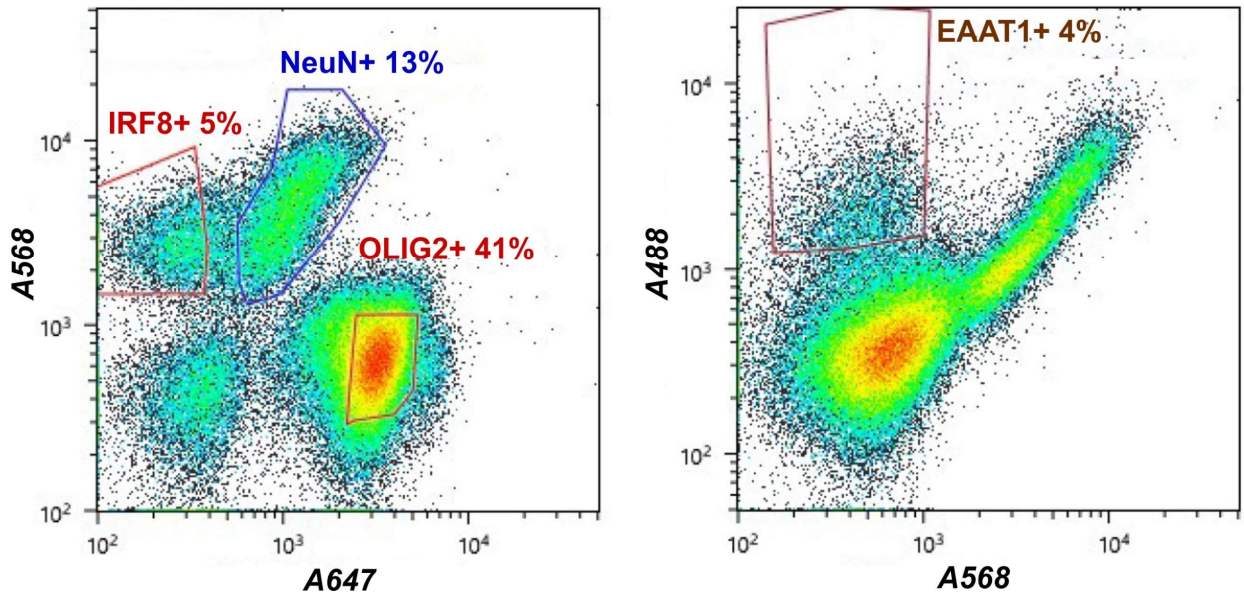

A
Lower in HD
Higher in HD
SST+ IN
PVALB+ IN
TAC3+ IN
iMSN
dMSN
2000 1500 1000 500 0 500 1000 1500
number of DE genes
B
cytosolic ribosome
nucleolus
ribonucleoprotein complex
spliceosomal complex
nuclear protein−containing complex
chromosome
nuclear body
centrosome
spindle
cytoskeleton
cis−Golgi network
cell junction
synapse
neuron to neuron synapse
cell−substrate junction
extracellular matrix
plasma membrane region
intrinsic component of plasma membrane
membrane raft
side of membrane
bounding membrane of organelle
Golgi apparatus
intracellular vesicle
lysosome
neuron projection
ion channel complex
dMSN
iMSN
Lower expression in HD MSNs
Higher expression in HD MSNs
-log10(p adj.)
2.5
5.0
7.5
10.0
C
-1.5 -1 -0.5 0 0.5 1 1.5
log2 fold change HD / CTRL
Mismatch repair
Base excision repair
POLD1
POLE
POLD2
POLD4
LIG1
RFC5
RFC4
RFC1
RFC3
PMS2
PMS1
RFC2
MLH1
MSH6
FAN1
MLH3
EXO1
PCNA
POLD3
MSH3
MSH2
XRCC1
NEIL1
UNG
OGG1
TDP1
POLB
FEN1
LIG3
PARP1
APEX1
HMGB1
TOP1
TDG
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
D
Essential for MSN viability in zQ175 mice
ATP1B1
RNMT
NETO1
CLK4
RASGRF2
HOOK1
SNW1
RPL37
SSBP2
NFYB
NDUFS4
TAF1
RBM39
TRMT10C
THSD7A
OXA1L
DDX21
ATP11C
1
0.5
0
-0.5
-1
log2 fold change HD / CTRL
dMSN
iMSN

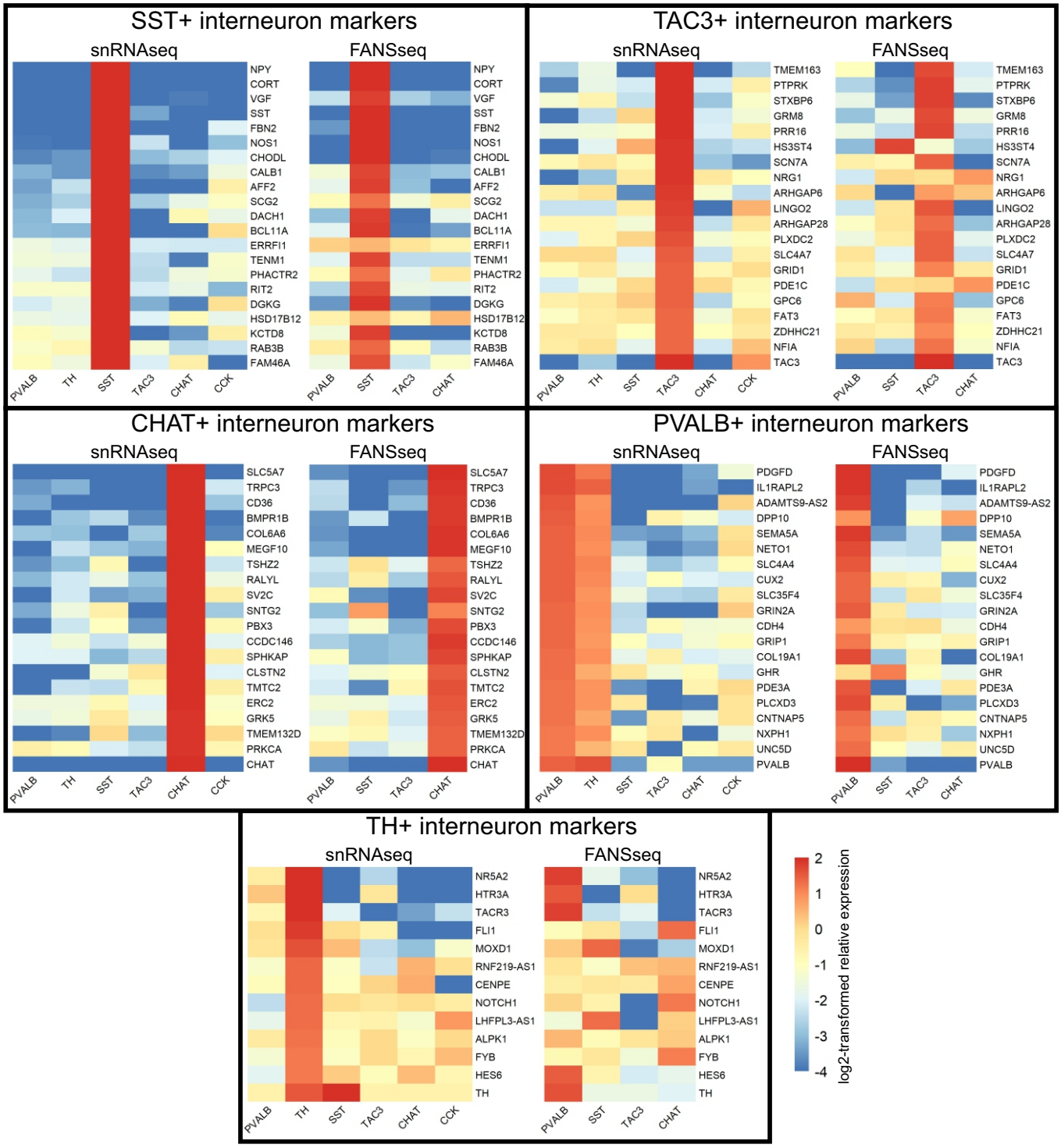
Figure 7. Disease-associated gene expression changes in striatal neuron types. (**A**) Number of differentially expressed (DE) genes (p adj. < 0.05 by DESeq2) in the comparison of HD and control donor FANSseq datasets from putamen or caudate nucleus. (**B**) Selected non-redundant Gene Ontology Cellular Component terms from enrichment analysis of genes with disease-associated expression changes in iMSN or dMSN (p adj < 0.05), but not in any interneuron type (p adj > 0.05 in all HD vs CTRL interneuron comparisons). (**C**) Heatmaps depicting disease-associated changes in transcript levels of MMR and BER genes. Statistically significant differences are marked with an asterisk (p adj. < 0.05 by DESeq2). (**D**) Genes essential for MSN viability in zQ175 mouse model of HD that are also downregulated in HD by more than a third in either iMSNs or dMSNs (log2FC<-0.6, p<0.01).

Suppl Figure 1



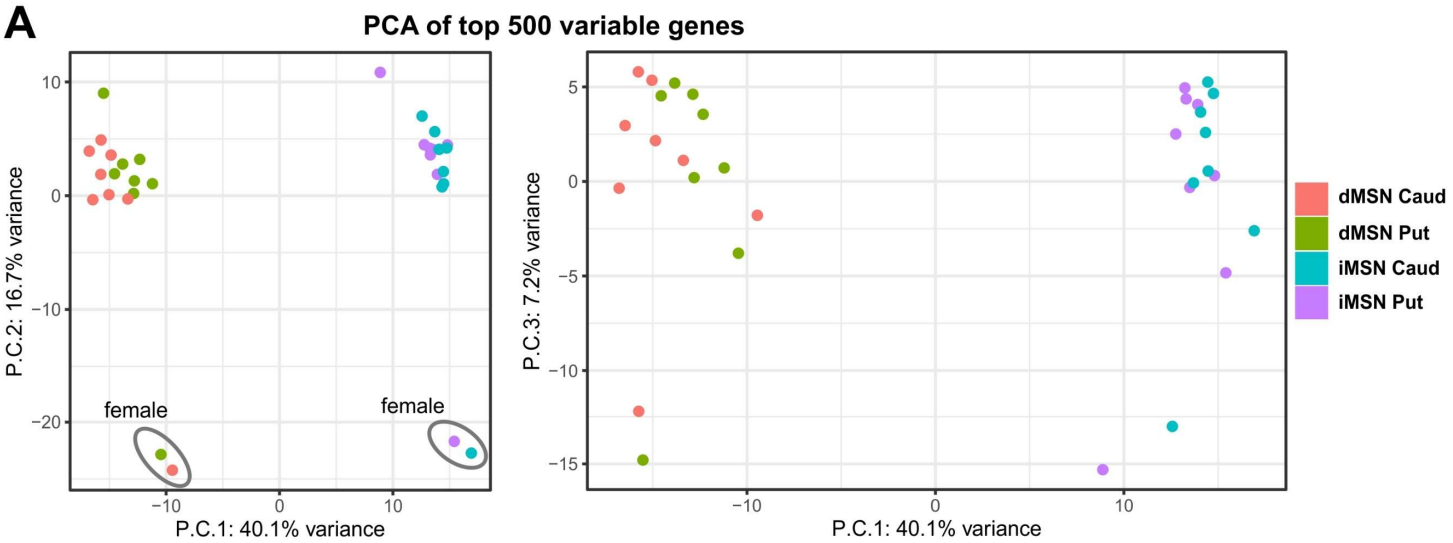
Suppl Figure 1. Representative FANS plots showing antibody-based labeling of glial cell nuclei isolated from human post-mortem caudate nucleus and putamen. The detailed strategy used for sorting is described in Materials and Methods.

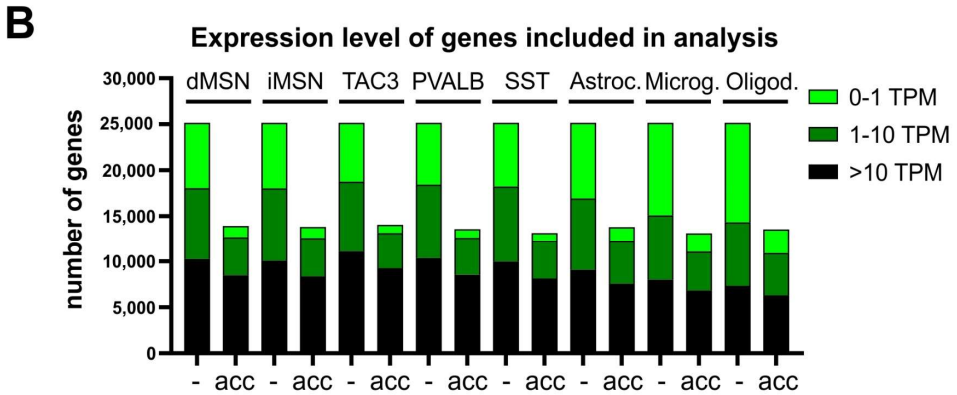
Suppl Figure 2

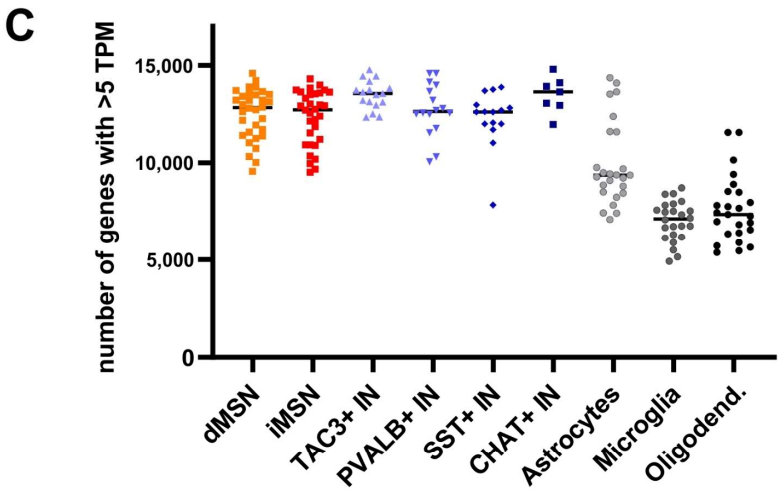


Suppl Figure 2. Comparison of interneuron populations collected using FANS to cell types defined from single-nucleus RNA sequencing of human striatum (Krienen, Goldman et al. 2020). Relative expression in each cell type was calculated based on DESeq2-normalized counts from 6-8 control donors, and was log2-transformed for visualization (see Materials and Methods). The marker genes specific for each interneuron subtype were selected based on single-nucleus RNA sequencing data.

Suppl Figure 3

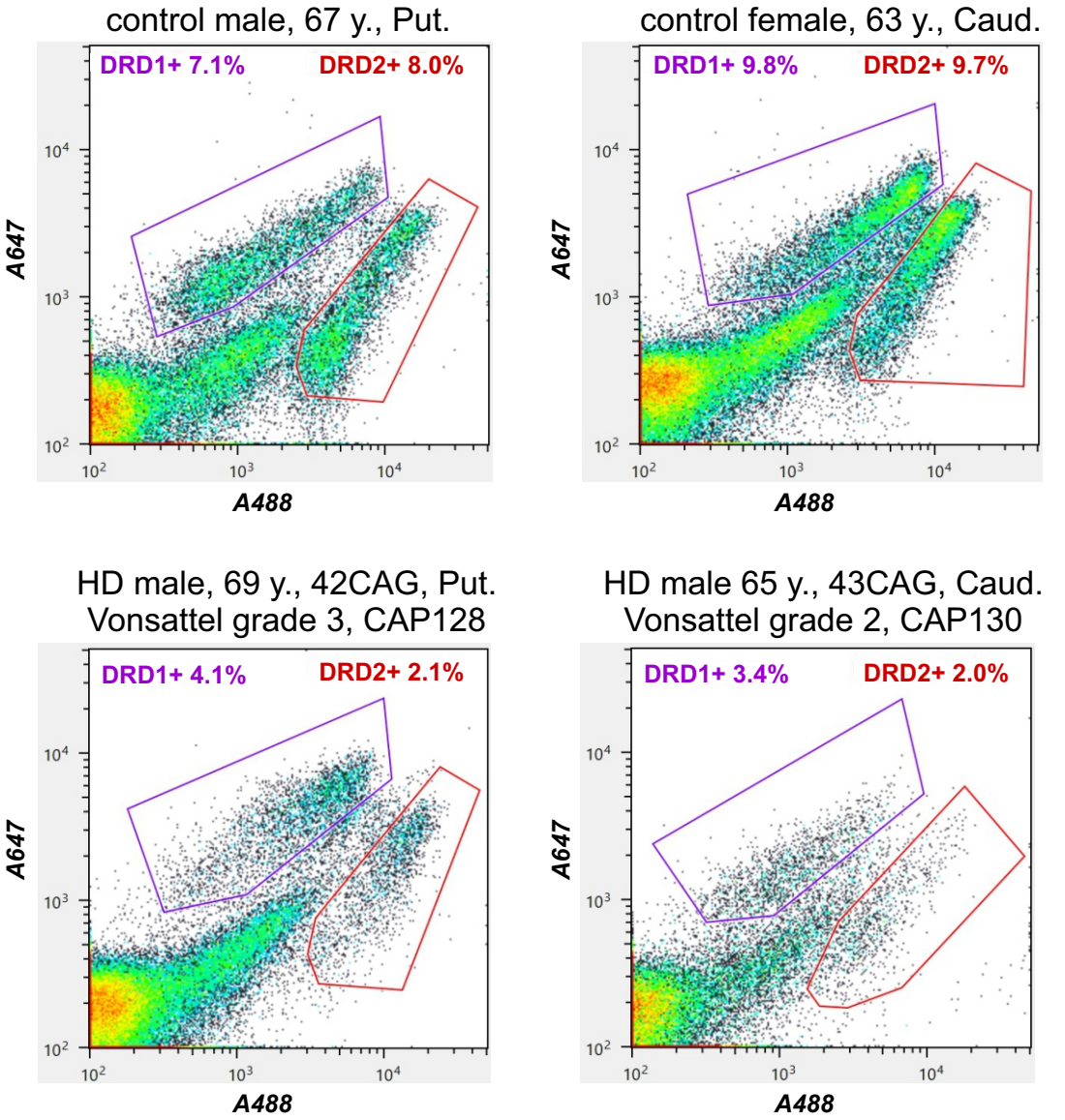






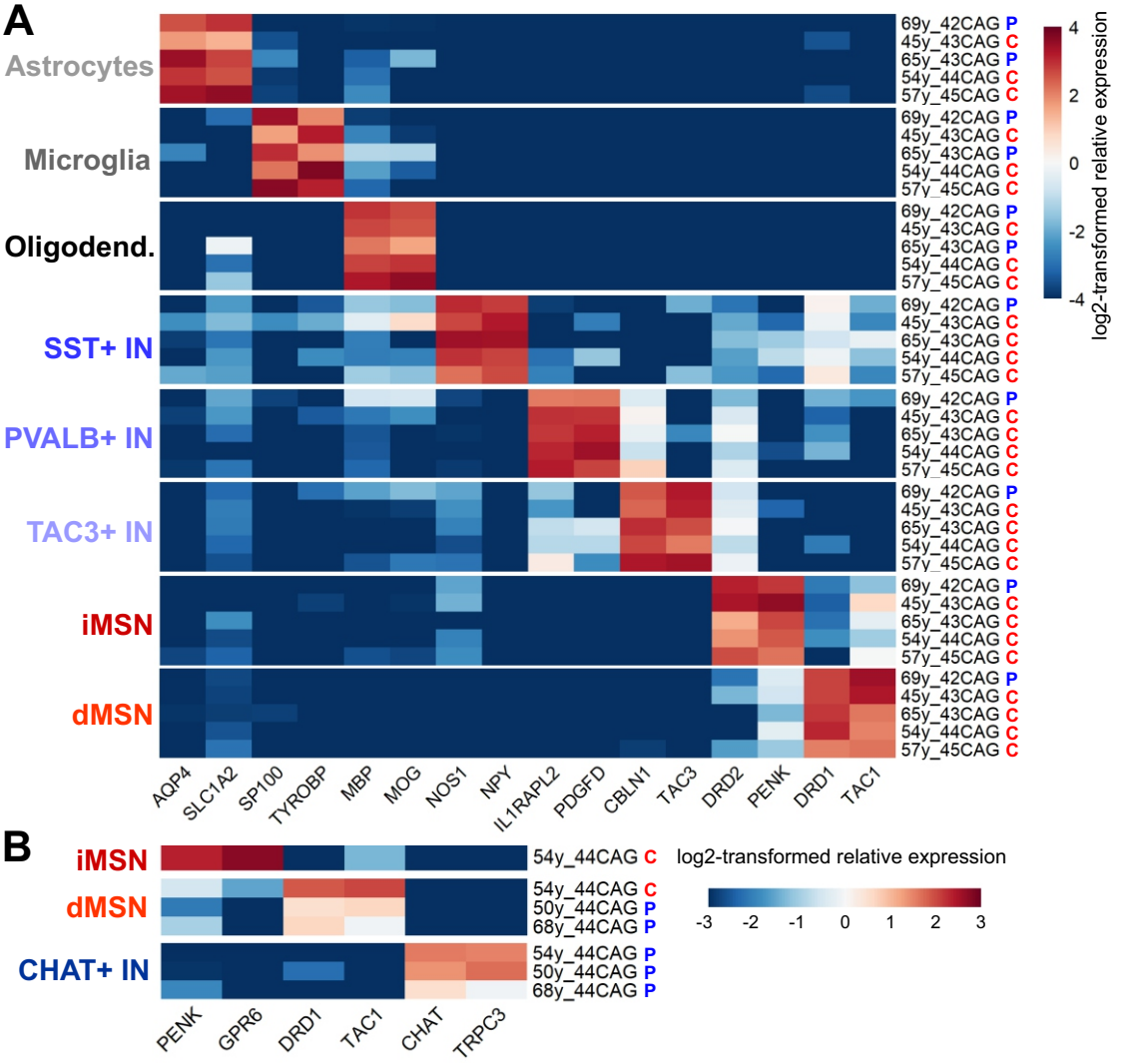
Suppl Figure 3. (**A**) Principal component analysis of FANSseq data from striatal projection neurons of caudate nucleus and putamen from 8 control donors. (**B**) The number of genes with accessible transcriptional start sites (acc.), defined as having a consensus ATACseq peak at the TSS, in striatal cell types from control donors. Genes have been grouped according to average expression level in FANSseq datasets from the specified cell type (TPM - transcripts per million). (**C**) Number of genes with expression level above 5 TPM in individual FANSseq datasets from control donors.

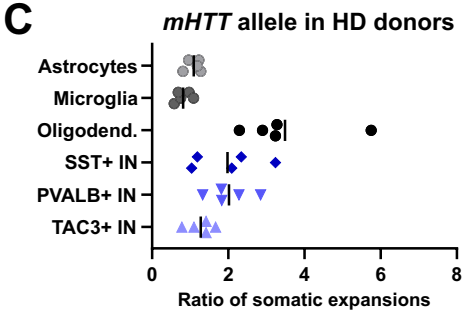
Suppl Figure 4

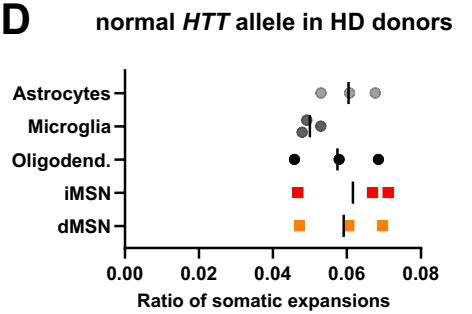


Suppl Figure 4. Representative FANS plots showing the labeling of dMSN and iMSN nuclei with Primeflow probes specific for *DRD1* and *DRD2* transcripts, respectively. Note that while there is a clear reduction in the abundance of MSN nuclei in striatal tissue from HD donors, these probes can still be used for the separation and isolation of MSN subtypes.

Suppl Fig 5







Microglia	Oligod.	SST+ IN	PVALB+ IN	TAC3+ IN	
0.91	0.0003	0.36	0.35	0.91	Astrocytes
	<0.0001	0.14	0.13	0.79	Microglia
		0.033	0.038	0.0007	Oligod.
P Adj < 0.05			0.93	0.52	SST+ IN
P Adj < 0.001				0.52	PVALB+ IN

Suppl Figure 5. (**A** and **B**) Relative expression level of marker genes of striatal cell types in the nuclear transcriptome of nuclei collected for *mHTT* CAG repeat tract analysis. Heatmap depicts log2-transformed relative expression in each sample (calculated based on DESeq2-normalized counts). The striatal region of origin is indicated by colored letters (blue P – putamen, red C – Caudate nucleus). (**C**) Comparison of the calculated ratio of somatic expansions of *mHTT* CAG tract in striatal cell types other than projection neurons. The table presents adjusted P-values as calculated by Holm-Sidak's multiple comparisons test post one-way ANOVA ($P < 0.0001$). (**D**) Ratio of somatic expansions of normal HTT allele CAG tract in striatal cell types isolated from donors with 21-25 uninterrupted repeats in normal HTT allele. $P = 0.599$ according to one-way ANOVA.

Suppl Fig 6

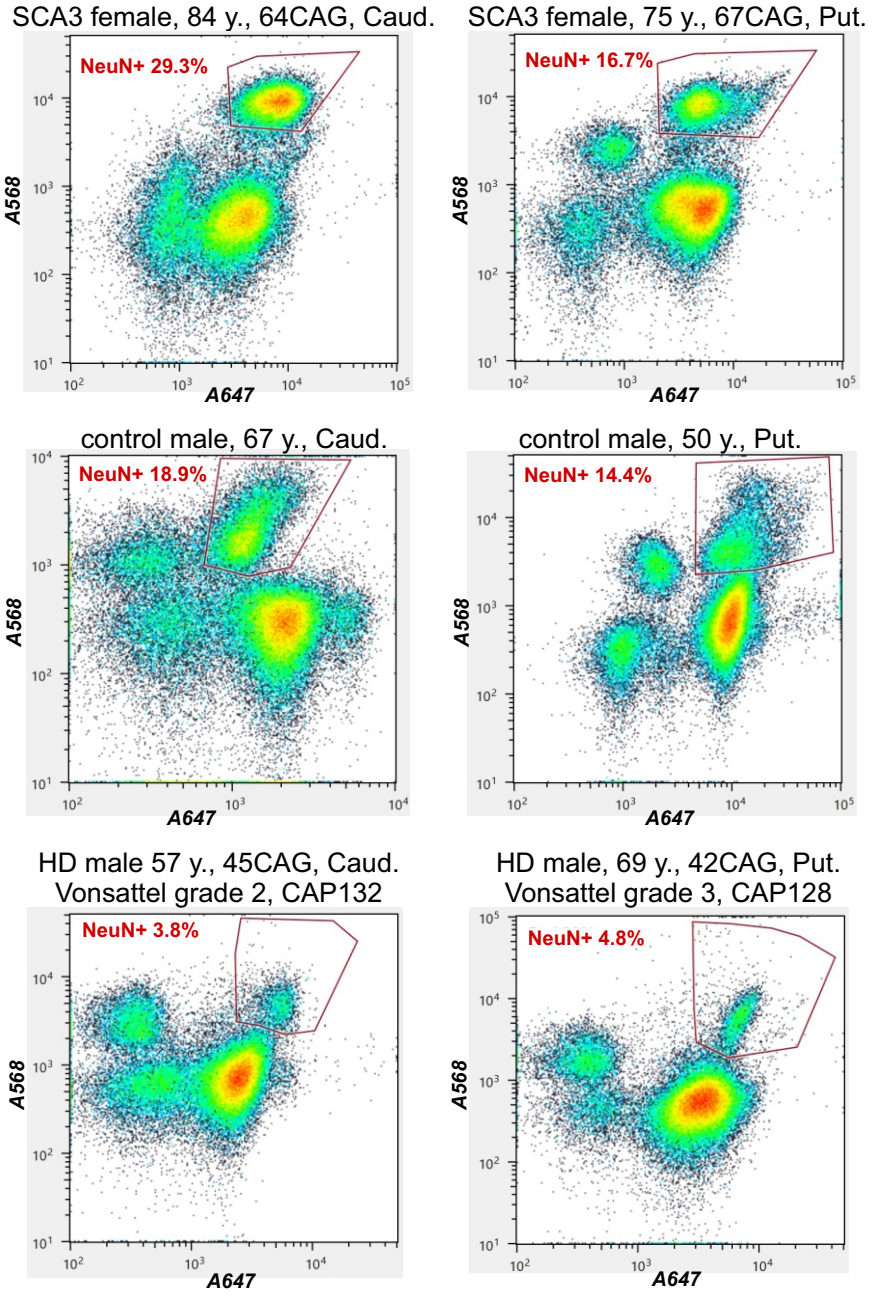
A

Average normalized counts in putamen

	dMSN	iMSN	TAC3+ IN	PVALB+ IN	SST+ IN	CHAT+ IN	Microglia	Astrocytes	Oligodend.
HTT	1813	1841	1899	2131	2034	2176	1728	934	476
ATXN3	836	846	834	638	808	866	1412	1662	892

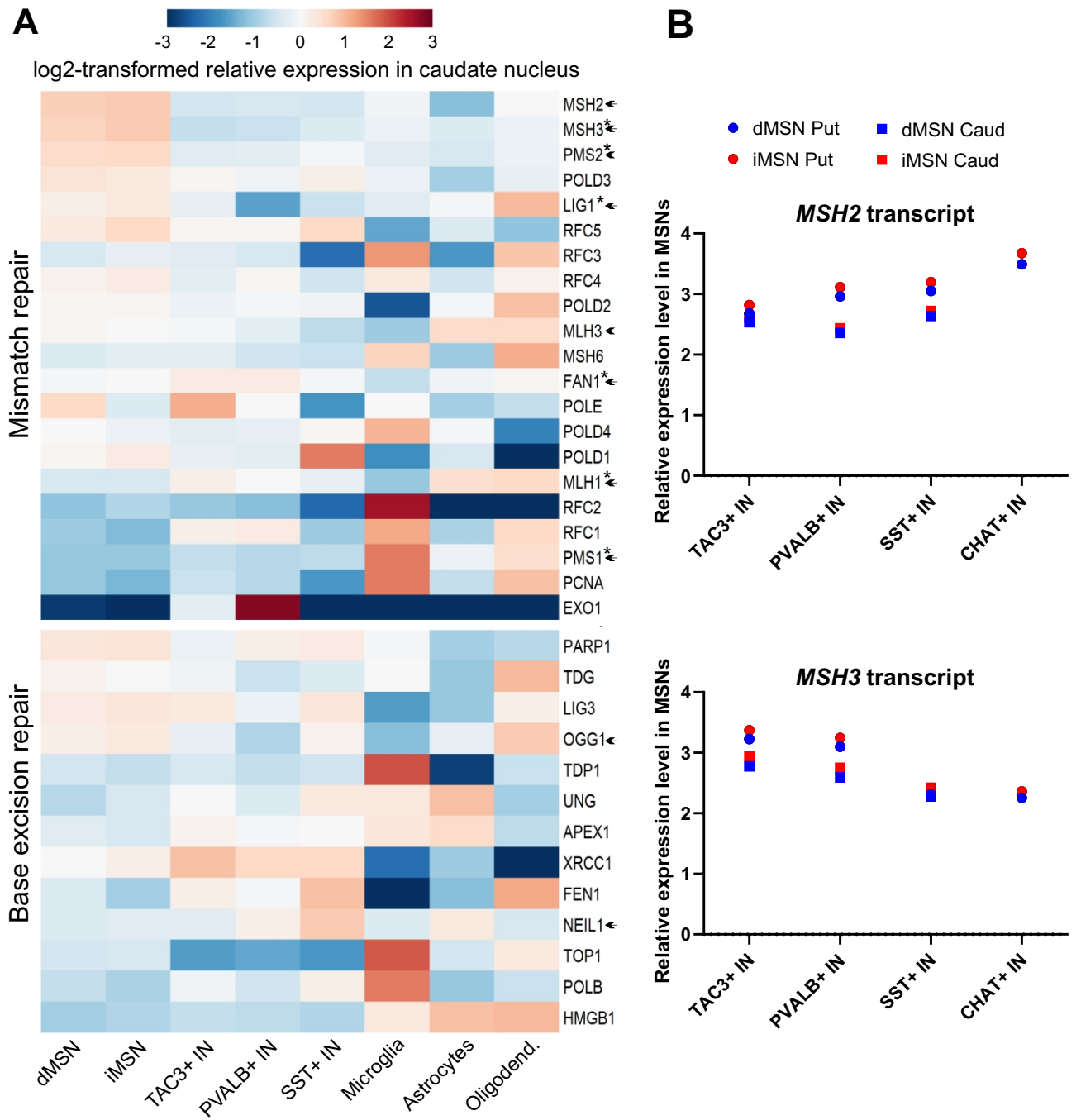
$\log_2$-transformed relative expression: 3, 2, 1, 0, -1, -2, -3

B



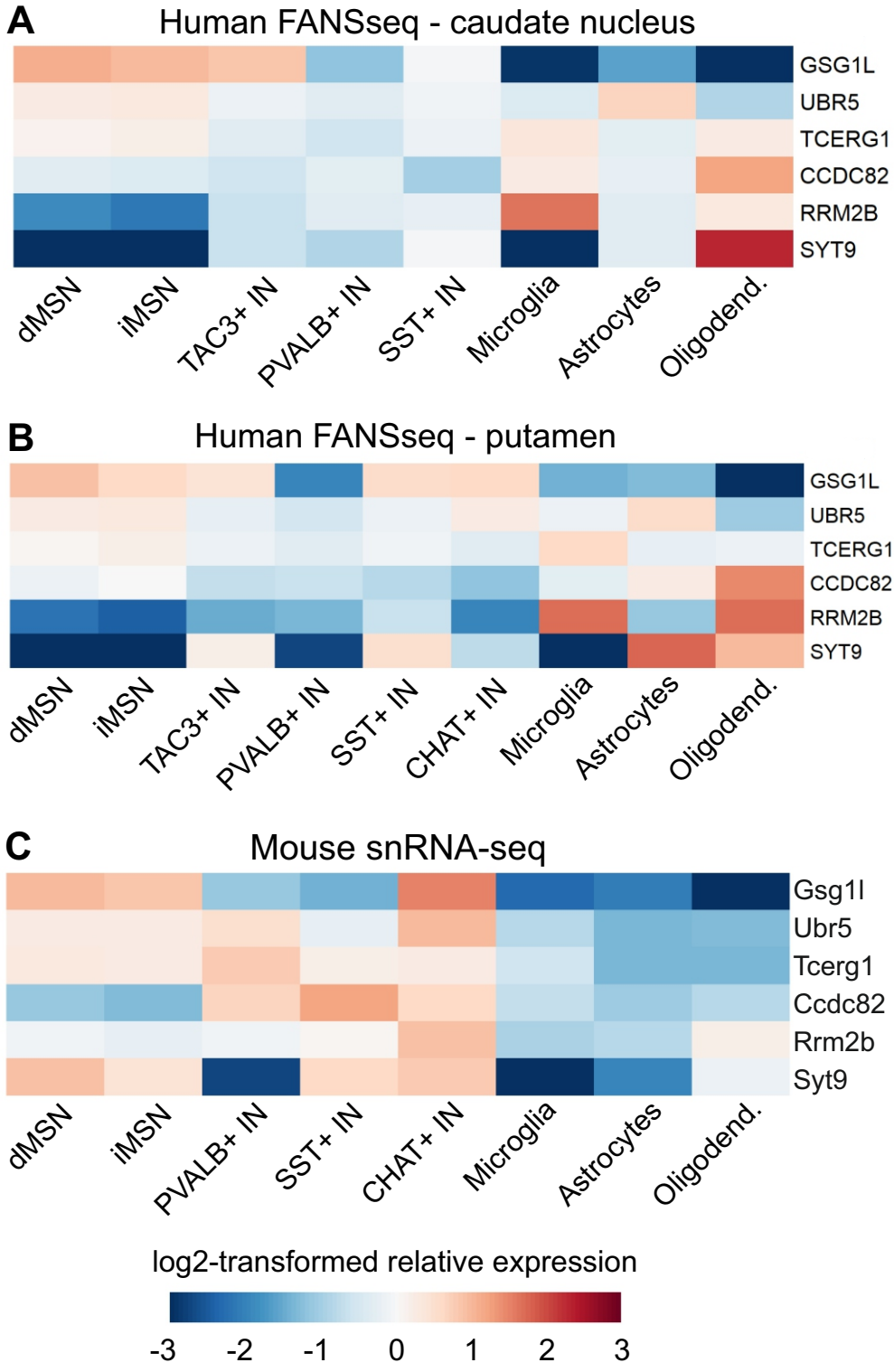
Suppl Figure 6: (**A**) Relative expression level of *HTT* and *ATXN3* in cell types of the putamen. Heatmaps depict log2-transformed relative expression in each cell type, calculated based on the mean of DESeq2-normalized counts from 6-8 control donors. (**B**) FANS plots showing the percentage of nuclei stained with anti-NeuN antibody in two oldest SCA3 donors, and in two representative HD donors and control donors.

Suppl Fig 7


A
-3 -2 -1 0 1 2 3
log2-transformed relative expression in caudate nucleus
Mismatch repair
MSH2
MSH3*
PMS2*
POLD3
LIG1*
RFC5
RFC3
RFC4
POLD2
MLH3
MSH6
FAN1*
POLE
POLD4
POLD1
MLH1*
RFC2
RFC1
PMS1*
PCNA
EXO1
Base excision repair
PARP1
TDG
LIG3
OGG1
TDP1
UNG
APEX1
XRCC1
FEN1
NEIL1
TOP1
POLB
HMGB1
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
Microglia
Astrocytes
Oligodend.
B
dMSN Put
dMSN Caud
iMSN Put
iMSN Caud
MSH2 transcript
Relative expression level in MSNs
0
1
2
3
4
TAC3+ IN
PVALB+ IN
SST+ IN
CHAT+ IN
MSH3 transcript
Relative expression level in MSNs
0
1
2
3
4
TAC3+ IN
PVALB+ IN
SST+ IN
CHAT+ IN

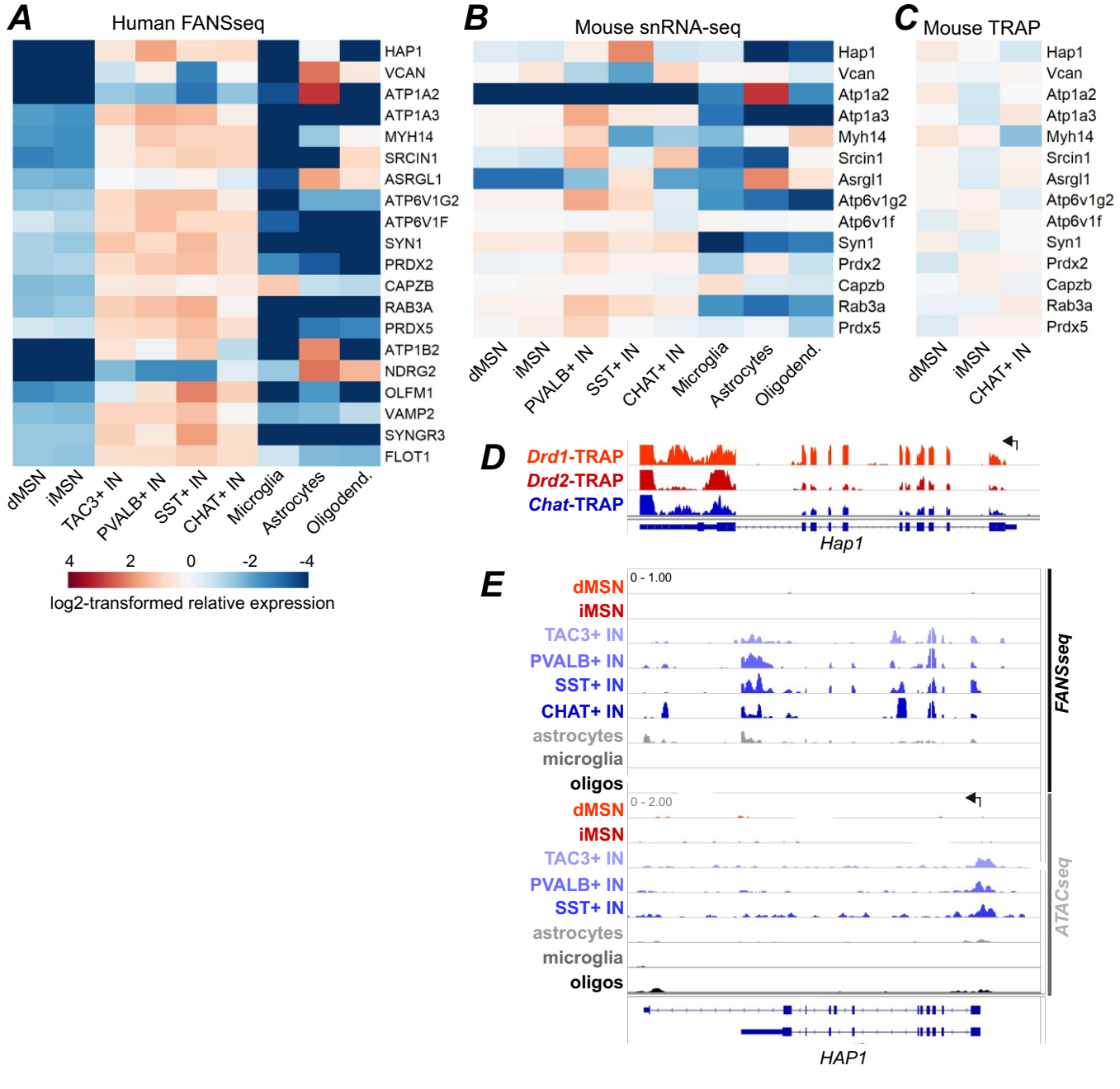
Suppl Figure 7: (**A**) Relative expression level of MMR and BER genes in cell types of caudate nucleus. Heatmaps depict log2-transformed relative expression in each cell type, calculated based on the mean of DESeq2-normalized counts from 7-8 control donors. Genes identified as HD age at onset-modifying candidates are marked with an asterisk. Genes known to influence CAG tract instability in HD mouse models are marked with an arrowhead. (**B**) Relative levels of *MSH2* and *MSH3* transcripts in striatal projection neurons compared to interneurons, calculated with DESeq2 using cell type-specific FANSseq data from control donors. P adj. < 8.2E-08 for all MSN vs interneuron comparisons (DESeq2).

Suppl Figure 8



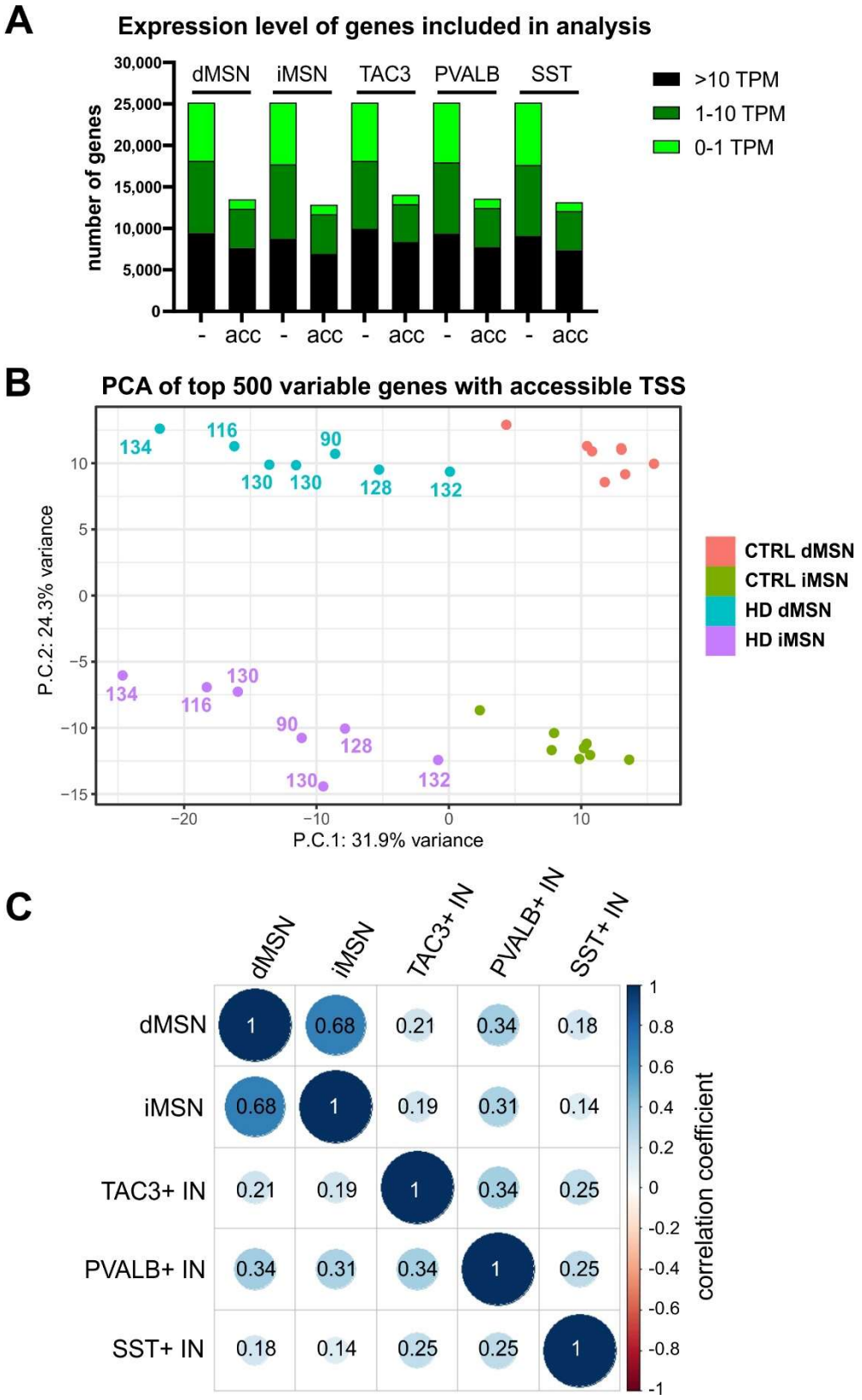
Suppl Figure 8. (**A and B**) Relative expression level of HD age-of-onset modifying candidate genes unrelated to DNA repair. Heatmaps depict log2-transformed relative expression in each cell type of (**A**) caudate nucleus and (**B**) putamen, calculated based on the mean of DESeq2-normalized counts from 6-8 control donors. (**C**) Cell type-specific expression according to snRNA-seq data from mouse striatum (Saunders, Macosko et al. 2018), calculated from 'transcripts per 100,000' values.

Suppl Fig 9


A
Human FANSseq
HAP1
VCAN
ATP1A2
ATP1A3
MYH14
SRCIN1
ASRGL1
ATP6V1G2
ATP6V1F
SYN1
PRDX2
CAPZB
RAB3A
PRDX5
ATP1B2
NDRG2
OLFM1
VAMP2
SYNGR3
FLOT1
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
CHAT+ IN
Microglia
Astrocytes
Oligodend.
4
2
0
-2
-4
log2-transformed relative expression
B
Mouse snRNA-seq
Hap1
Vcan
Atp1a2
Atp1a3
Myh14
Srcin1
Asrgl1
Atp6v1g2
Atp6v1f
Syn1
Prdx2
Capzb
Rab3a
Prdx5
dMSN
iMSN
PVALB+ IN
SST+ IN
CHAT+ IN
Microglia
Astrocytes
Oligodend.
C
Mouse TRAP
Hap1
Vcan
Atp1a2
Atp1a3
Myh14
Srcin1
Asrgl1
Atp6v1g2
Atp6v1f
Syn1
Prdx2
Capzb
Rab3a
Prdx5
dMSN
iMSN
CHAT+ IN
D
Drd1-TRAP
Drd2-TRAP
Chat-TRAP
Hap1
E
0 - 1.00
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
CHAT+ IN
astrocytes
microglia
oligos
FANSseq
0 - 2.00
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
astrocytes
microglia
oligos
ATACseq
HAP1

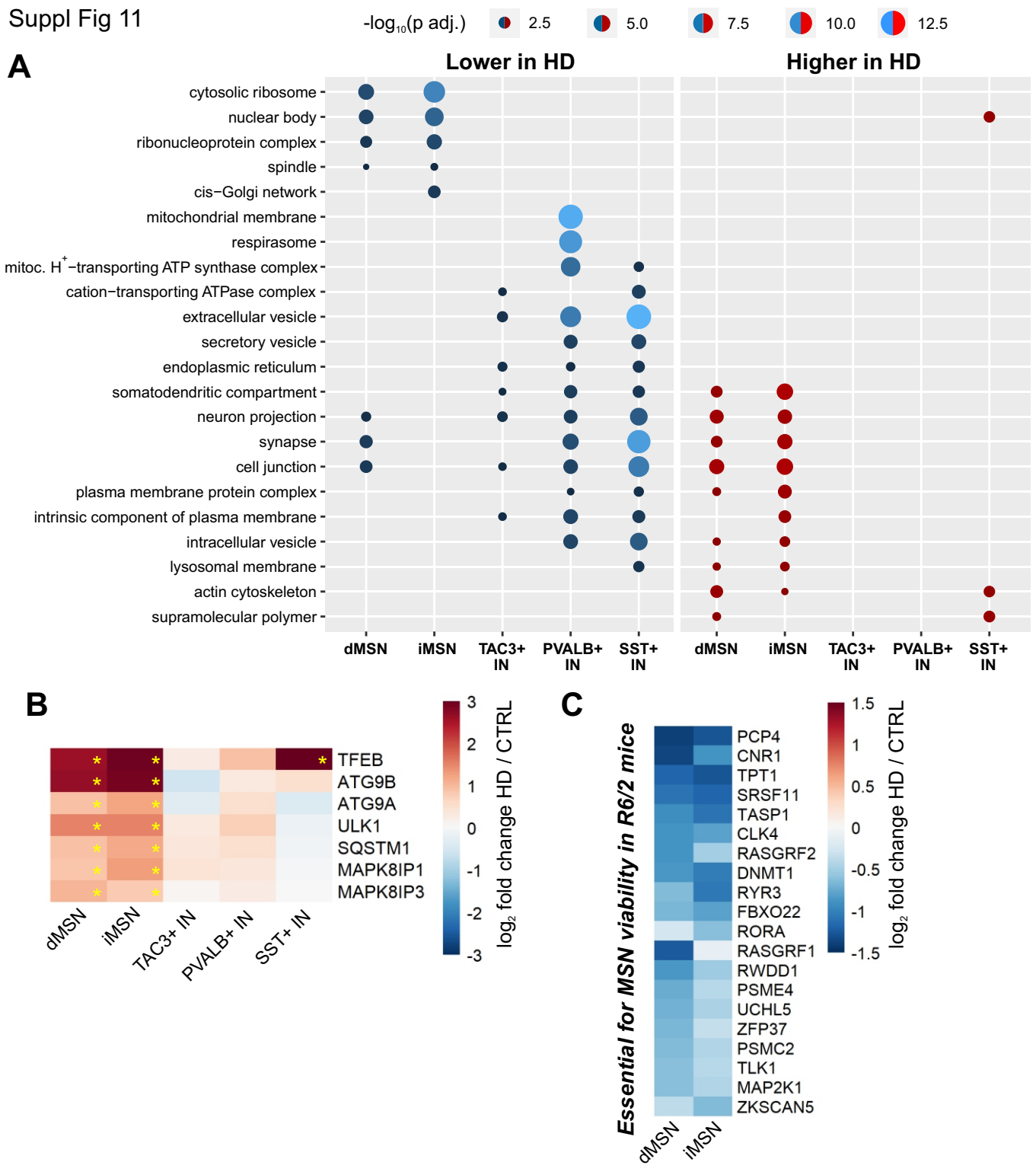
Suppl Figure 9. HTT interactome genes with low expression in human striatal projection neurons. (**A**) Heatmap shows gene expression in the putamen of human control donors and includes genes for which P adj. <0.05 and log2 of fold change < -1 (by DESeq2) when comparing expression in dMSN or iMSN to any interneuron type in both putamen and caudate nucleus. (**B-C**) Genes for which lower expression in human MSNs was not predicted by (**B**) snRNA-seq data from mouse striatum (Saunders, Macosko et al. 2018) or (**C**) by TRAP RNA-seq data from mouse cholinergic neurons (Chat-TRAP mouse line) and striatal projection neurons (Drd1-TRAP and Drd2-TRAP lines) (Doyle, Dougherty et al. 2008). N=3 for all TRAP lines. Heatmaps show log2-transformed relative expression in each cell type, calculated based on DESeq2-normalized counts (for **A** and **C**) or 'transcripts per 100,000' values (for **B**). (**D**) Distribution of TRAP-RNAseq reads mapped to *Hap1*, showing its expression in projection neurons and cholinergic interneurons of mouse striatum. (**E**) Distribution of human FANSseq and ATAC-seq reads mapped to *HAP1*. Arrows mark the position of annotated transcriptional start sites. The data are from a 41-year-old male donor.

Suppl Fig 10


A
Expression level of genes included in analysis
dMSN
iMSN
TAC3
PVALB
SST
number of genes
30,000
25,000
20,000
15,000
10,000
5,000
0
- acc - acc - acc - acc - acc
>10 TPM
1-10 TPM
0-1 TPM
B
PCA of top 500 variable genes with accessible TSS
P.C.2: 24.3% variance
P.C.1: 31.9% variance
CTRL dMSN
CTRL iMSN
HD dMSN
HD iMSN
C
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
correlation coefficient

Suppl Figure 10. (**A**) The number of genes with accessible transcriptional start sites (acc.), defined based on the overlap of TSS and consensus peak of ATACseq data from HD donor MSNs and control donor interneurons. Genes have been grouped according to their average expression level in FANSseq datasets from HD donors (TPM - transcripts per million). **(B**) Principal component analysis of FANSseq data from putamen and caudate nucleus MSNs from HD and control donors, performed after the exclusion of genes with inaccessible TSSs. **(C)** Correlation analysis of disease-associated expression changes of genes expressed in striatal neuronal types studied.

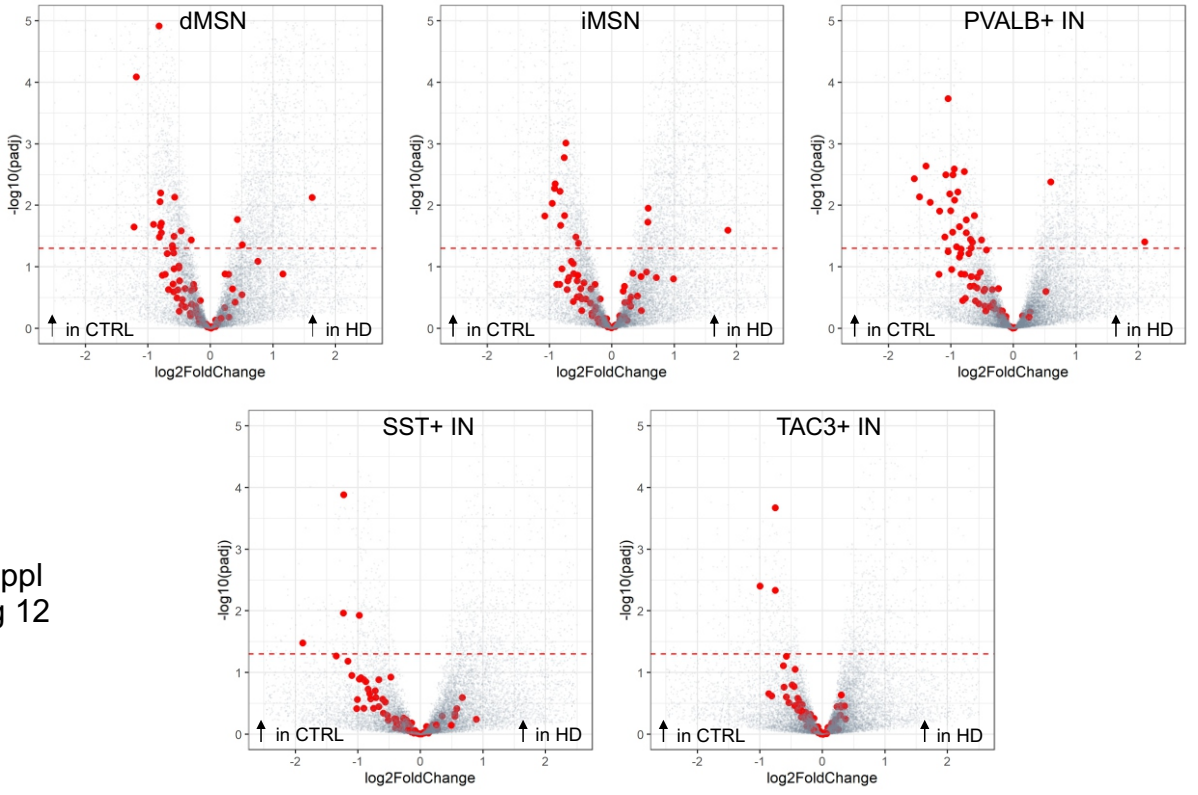
Suppl Fig 11


-log10(p adj.)
2.5
5.0
7.5
10.0
12.5
A
Lower in HD
Higher in HD
cytosolic ribosome
nuclear body
ribonucleoprotein complex
spindle
cis−Golgi network
mitochondrial membrane
respirasome
mitoc. H+−transporting ATP synthase complex
cation−transporting ATPase complex
extracellular vesicle
secretory vesicle
endoplasmic reticulum
somatodendritic compartment
neuron projection
synapse
cell junction
plasma membrane protein complex
intrinsic component of plasma membrane
intracellular vesicle
lysosomal membrane
actin cytoskeleton
supramolecular polymer
dMSN
iMSN
TAC3+ IN
PVALB+ IN
SST+ IN
B
TFEB
ATG9B
ATG9A
ULK1
SQSTM1
MAPK8IP1
MAPK8IP3
log2 fold change HD / CTRL
C
Essential for MSN viability in R6/2 mice
PCP4
CNR1
TPT1
SRSF11
TASP1
CLK4
RASGRF2
DNMT1
RYR3
FBXO22
RORA
RASGRF1
RWDD1
PSME4
UCHL5
ZFP37
PSMC2
TLK1
MAP2K1
ZKSCAN5

Suppl Figure 11. (A) Selected non-redundant Gene Ontology Cellular Component terms from enrichment analysis of genes with disease-associated expression changes in each striatal neuron type (p adj < 0.05). (**B**) Heatmap depicting disease-associated changes in transcript levels of selected genes regulating autophagosome formation and transport. All changes were significant only in MSNs. Asterisks denote p adj < 0.05. (**C**) Genes essential for MSN viability in the R6/2 mouse model of HD and that are also downregulated in HD by more than a third in either iMSNs or dMSNs (log2FC<-0.6, p<0.01).

A

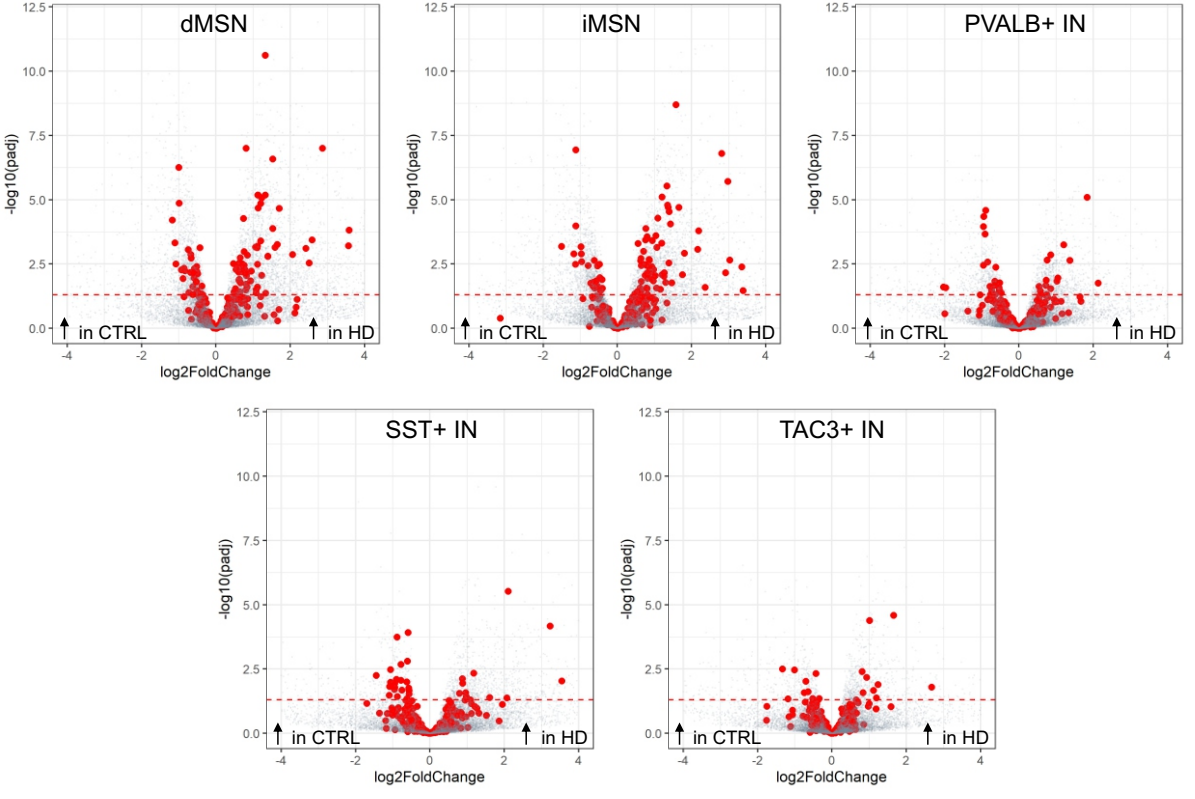
Expression of genes in GOCC gene set '*Respirasome*'



Suppl Fig 12

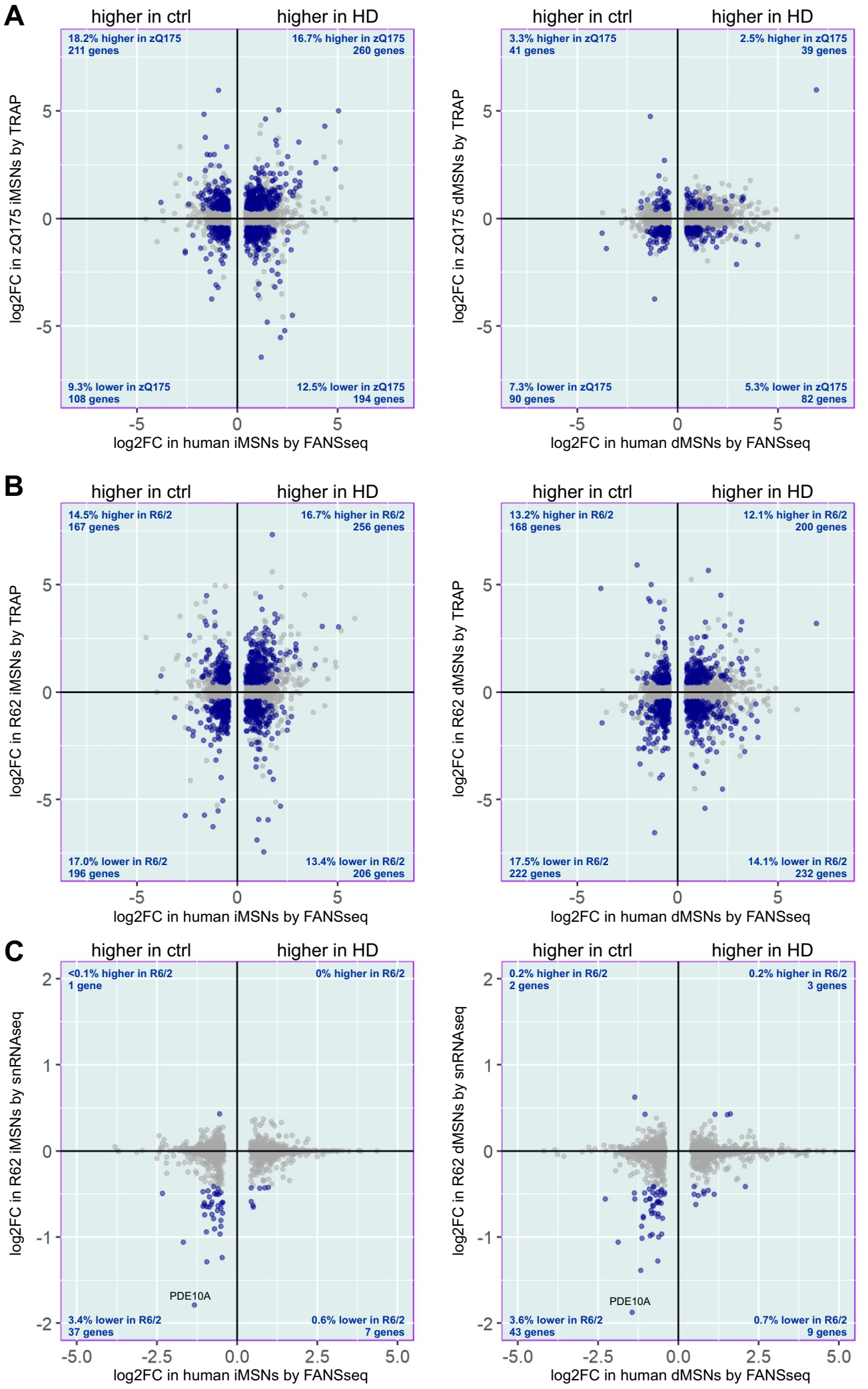
B

Expression of genes in GOCC gene set '*Lysosomal membrane*'



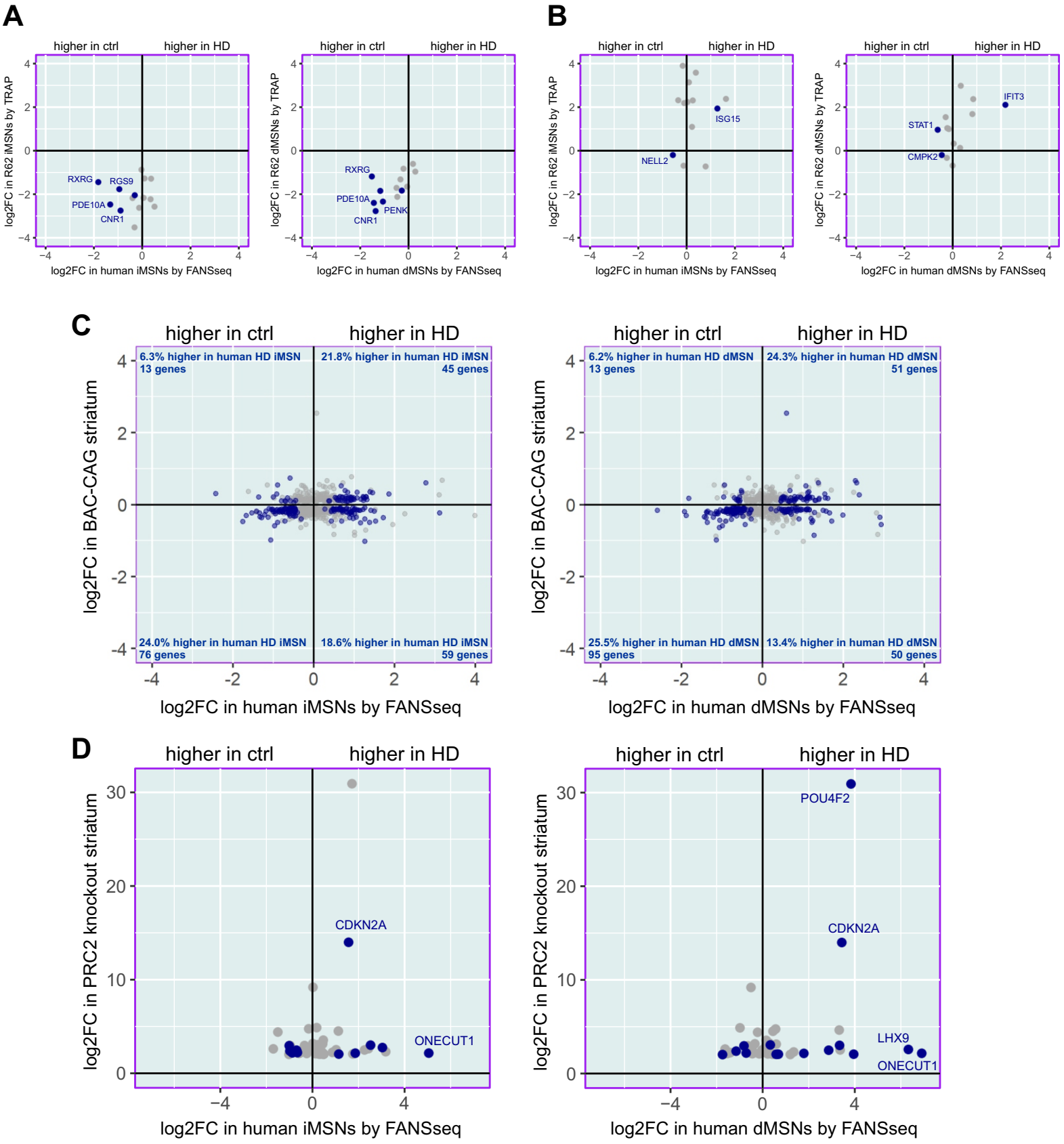
Suppl Figure 12. Disease-associated changes in genes associated with GO Cellular Component terms (**A**) ‘Respirasome’ (GO:0070469) and (**B**) ‘Lysosomal membrane’ (GO:0005765). Dashed line denotes the threshold of statistical significance (p adj = 0.05).

Suppl Fig 13


A
higher in ctrl
higher in HD
18.2% higher in zQ175
211 genes
16.7% higher in zQ175
260 genes
9.3% lower in zQ175
108 genes
12.5% lower in zQ175
194 genes
log2FC in zQ175 iMSNs by TRAP
log2FC in human iMSNs by FANSseq
3.3% higher in zQ175
41 genes
2.5% higher in zQ175
39 genes
7.3% lower in zQ175
90 genes
5.3% lower in zQ175
82 genes
log2FC in zQ175 dMSNs by TRAP
log2FC in human dMSNs by FANSseq
B
14.5% higher in R6/2
167 genes
16.7% higher in R6/2
256 genes
17.0% lower in R6/2
196 genes
13.4% lower in R6/2
206 genes
log2FC in R62 iMSNs by TRAP
log2FC in human iMSNs by FANSseq
13.2% higher in R6/2
168 genes
12.1% higher in R6/2
200 genes
17.5% lower in R6/2
222 genes
14.1% lower in R6/2
232 genes
log2FC in R62 dMSNs by TRAP
log2FC in human dMSNs by FANSseq
C
<0.1% higher in R6/2
1 gene
0% higher in R6/2
PDE10A
3.4% lower in R6/2
37 genes
0.6% lower in R6/2
7 genes
log2FC in R62 iMSNs by snRNAseq
log2FC in human iMSNs by FANSseq
0.2% higher in R6/2
2 genes
0.2% higher in R6/2
3 genes
PDE10A
3.6% lower in R6/2
43 genes
0.7% lower in R6/2
9 genes
log2FC in R62 dMSNs by snRNAseq
log2FC in human dMSNs by FANSseq

Suppl Figure 13. Genes with disease-associated transcript level change (p adj. < 0.05 and |log2FC|>0.415) in FANSseq data from human iMSN or dMSN were plotted against the mouse orthologue transcript level change in (**A**) TRAP data from iMSNs or dMSNs of zQ175 mice (compared to Q20 mice) and (**B**) TRAP data or (**C**) snRNAseq data from iMSNs or dMSNs of R6/2 mice (compared to wt mice) (Lee, Fenster et al. 2020). Genes with a significant transcript level change in the HD mouse model (p adj. < 0.05 and |log2FC|>0.415) are marked with blue color. Note that for most of these genes in mouse there is no significant transcript level change according to these criteria.

Suppl Fig 14


A
higher in ctrl
higher in HD
log2FC in R62 iMSNs by TRAP
log2FC in human iMSNs by FANSseq
RXRG
RGS9
PDE10A
CNR1
log2FC in R62 dMSNs by TRAP
log2FC in human dMSNs by FANSseq
RXRG
PDE10A
PENK
CNR1
B
log2FC in R62 iMSNs by TRAP
log2FC in human iMSNs by FANSseq
ISG15
NELL2
log2FC in R62 dMSNs by TRAP
log2FC in human dMSNs by FANSseq
IFIT3
STAT1
CMPK2
C
higher in ctrl
higher in HD
log2FC in BAC-CAG striatum
log2FC in human iMSNs by FANSseq
6.3% higher in human HD iMSN
13 genes
21.8% higher in human HD iMSN
45 genes
24.0% higher in human HD iMSN
76 genes
18.6% higher in human HD iMSN
59 genes
log2FC in human dMSNs by FANSseq
6.2% higher in human HD dMSN
13 genes
24.3% higher in human HD dMSN
51 genes
25.5% higher in human HD dMSN
95 genes
13.4% higher in human HD dMSN
50 genes
D
higher in ctrl
higher in HD
log2FC in PRC2 knockout striatum
log2FC in human iMSNs by FANSseq
CDKN2A
ONECUT1
log2FC in human dMSNs by FANSseq
POU4F2
CDKN2A
LHX9
ONECUT1

Suppl Figure 14. HD-associated transcript level change in human orthologues of (**A**) MSN marker genes downregulated in R6/2 mice, (**B**) innate immunity genes induced in R6/2 mice, (**C**) genes altered in the striatum of BAC-CAG mice (FDR < 0.1), and (**D**) genes induced in the striatum of MSN-specific Ezh1/Ezh2 double knockout mice. Genes with a significant transcript level change in human FANSseq data (p adj. < 0.05) are marked with blue color. Note that for most of these genes there is no significant transcript level change in the nuclei of human MSNs.