Defective linear and circular RNAs biogenesis in Huntington's disease: CAG repeat expansion hijacks neuronal splicing

Dilara Ayyildiz[1,2,(*)], Alan Monziani[3,(*)], Takshashila Tripathi[3,(*)], Jessica Döring[3], Guendalina Bergonzoni[3], Emanuela Kerschbamer[3], Francesca Di Leva[3], Elia Pennati[3], Luisa Donini[3], Marina Kovalenko[4], Jacopo Zasso[5], Luciano Conti[5], Vanessa C. Wheeler[4,6], Christoph Dieterich[7], Silvano Piazza[1,(#)], Erik Dassi[8,(#)] and Marta Biagioli[3,(#)]

[1]Bioinformatic facility, Department of Cellular, Computational and Integrative Biology - CIBIO, University of Trento (ITALY)

[2]Biomedical Sciences and Biotechnology, University of Udine, ITALY

[3]NeuroEpigenetics laboratory, Department of Cellular, Computational and Integrative Biology - CIBIO, University of Trento (ITALY)

[4]Molecular Neurogenetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, U.S.A.

[5]Laboratory of Stem Cell Biology, Department of Cellular, Computational and Integrative Biology - CIBIO, University of Trento (ITALY)

[6]Department of Neurology Harvard Medical School, Boston, MA 02115, USA

[7]Section of Bioinformatics and Systems Cardiology, University Hospital Heidelberg, Germany

[8]Laboratory of RNA Regulatory Networks, Department of Cellular, Computational and Integrative Biology - CIBIO, University of Trento (ITALY)

26   (*) Joint first authors.

27   (#)   *Correspondence   to: silvano.piazza@icgeb.org;   erik.dassi@unitn.it;*

28   *marta.biagioli@unitn.it*

29

30

31   Running title: CAG repeat expansion hijacks neuronal splicing

ABSTRACT

Alternative splicing (AS) appears to be altered in Huntington's disease (HD), but its significance for early, pre-symptomatic disease stages has not been inspected.

Here, taking advantage of *Htt* CAG knock-in mouse *in vitro* and *in vivo* models, we demonstrate a strong correlation between *Htt* CAG repeat length and increased aberrant linear AS, specifically affecting neural progenitors and, *in vivo,* the striatum prior to overt behavioral phenotypes stages. Remarkably, expanded *Htt* CAG repeats reflect on a previously neglected, global impairment of back-splicing, leading to decreased circular RNAs production in neural progenitors.

Though the mechanisms of this dysregulation remain uncertain, our study unveils network of transcriptionally altered micro-RNAs and RNA-binding proteins (CELF, hnRNPS, PTBP, SRSF) which, in turn, might influence the AS machinery, primarily in neural cells.

We suggest that this unbalanced expression of linear and circular RNAs might result in altered neural fitness, contributing to HD striatal vulnerability.

INTRODUCTION

Huntington's disease (HD) is a hereditary, fatal neurodegenerative disorder caused by a CAG trinucleotide expansion within exon 1 of the *HTT* gene (1993). Explicit clinical onset typically occurs in mid-life and leads to an inexorable decline to death after 10-15 years (Perutz et al. 1999). A polymorphic CAG tract up to 35 repeats is found in unaffected individuals, whereas alleles bearing 36 or more repeats lead to HD symptoms. Since the *HTT* gene is ubiquitously expressed during human development and in all body districts, the effects of the mutation is strongly pleiotropic (Bassi et al. 2017). The central nervous system, however, remains the main region affected by mutant *HTT*, with a prominent loss of GABAergic medium-sized spiny neurons of the striatum, constituting the major contributor to movement, cognitive and behavioral dysfunctions (Vonsattel et al. 1985; Rosas et al. 2003; Han et al. 2010). So far, the rate-limiting mechanism (s) of neurodegeneration remains elusive although chromatin, transcription, and RNA processing dysregulations are emerging as fundamental features (Bassi et al. 2017; Kerschbamer and Biagioli 2016). In particular, RNA processing and alternative splicing (AS) alterations might affect the level and composition of a broad repertoire of proteins, thus contributing to HD striatal vulnerability and pathogenesis. The spliceosomal activity can be directly modulated by the expression level and/or sequestration of various proteins that bind to nascent mRNA. Interestingly, huntingtin can associate with the WW-containing proteins HYPA and HYPC (*Htt* Yeast two-hybrid Protein A and C), also known as FBP11/PRPF40A, and PRPF40B, respectively (Faber et al. 1998; Jiang et al. 2011; Passani et al. 2000), participating in early spliceosomal assembly and 5' site recognition. On the other hand, mis-splicing events in individuals with highly expanded *HTT* CAG repeats have been shown to produce the small, highly pathogenic, polyadenylated exon 1-intron 1 *HTT* transcript (Neueder et al. 2017). Moreover, tissue-restricted trans-splicing regulators, binding auxiliary exonic and intronic cis-regulatory

4

72 signals, modulate splice-site choice by interacting with components of the splicing machinery

73 (Wahl et al. 2009; Ule and Blencowe 2019). Recent evidence supports a role for huntingtin in

74 regulating the expression of four RNA-binding proteins (PTBP1, SFRS4, RBM4, SREK1) in

75 HD *post-mortem* brains, thus, in turn, affecting the relative abundance of specific target

76 isoforms (Lin et al. 2016). A third layer of complexity in AS regulation is added by broad

77 chromatin conformation and transcriptional kinetics (Andersson et al. 2009; Spies et al. 2009;

78 Schwartz et al. 2009). In fact, chromatin relaxation accelerates RNA Pol II processing,

79 correlating with alternative exon skipping. Conversely, packed nucleosomes slow down RNA

80 Pol II progression, causing a pause in transcription and not-constitutive weak exon inclusion.

81 Interestingly, altered chromatin remodeling and histone modifications enrichment emerge as

82 key features in HD. The work from Vashishtha et al., 2013 shows changes in H3K4me3 in

83 HD mice and *post-mortem* tissues (Vashishtha et al. 2013), whereas we recently described an

84 effect of wild-type and mutant huntingtin on PRC2 and Mll-containing complexes (Biagioli

85 et al. 2015; Seong et al. 2010). In addition, huntingtin can directly interact with the

86 HypB/SetD2 H3K36me3 methyltransferase, implicated in RNA splicing and Pol II

87 elongation of transcription (Faber et al. 1998; Passani et al. 2000; Simon et al. 2014; Zhu et

88 al. 2017). As experimental evidence points toward dysregulated AS as a key feature in HD

89 pathology, it is tempting to hypothesize that nuclear huntingtin might regulate AS outcome

90 through different direct and/or indirect modes, thus contributing to HD pathogenesis.

91   Importantly, AS regulation is crucial not only to the establishment of a repertoire of

92 protein-coding isoforms extremely relevant for the proper physiology of the nervous system,

93 but also to the biogenesis of circular RNAs (circRNAs), unusually stable non-coding RNAs

94 produced by the circularization of exons (Jeck et al. 2013; Memczak et al. 2013; Salzman et

95 al. 2012). CircRNAs are highly enriched in neurons and have been implicated in a wide

96 variety of pathological conditions, including neurological diseases (Holdt et al. 2018).

97 CircRNA biogenesis seems to correlate with exon-skipping events, but the exact mechanisms

98 of spliced exon-exon circularization remains unclear (Kelly et al. 2015; Barrett et al. 2015).

99 The vast majority of circular RNAs localize in the cytoplasm and their exonic sequences

100 might encompass micro RNAs (miRNAs) and RNA Binding Proteins (RBPs) binding sites

101 (Memczak et al. 2013; Hansen et al. 2013), thus acting as competing-endogenous RNA

102 (Salmena et al. 2011). In this scenario, a circRNA could efficiently engage ("sponge") either

103 miRNAs or RBPs, or both, eventually relieving their canonical mRNA targets from post-

104 transcriptional regulation (Memczak et al. 2013; Piwecka et al. 2017; Hansen et al. 2011;

105 Abdelmohsen et al. 2017; Chen et al. 2019). Whether mutant huntingtin may modulate

106 circular RNA expression levels remains unexplored.

107 Here, we investigated whether linear and back-splicing processes, producing linear

108 mRNAs and circular isoforms, might be altered by the HD mutation in a CAG-dependent and

109 tissue-specific manner. We unveiled new mechanisms dysregulated in HD at the genome-

110 wide level and possible regulatory cross talks with miRNAs and RBPs.

111

112

113

114 **RESULTS**

115 **Aberrant linear alternative splicing shows *Htt*-CAG length and age dependency,**

116 **specifically in the mouse striatum**

117 In order to gauge evidence of AS alterations in the brain regions ofneural HD animal models,

118 we examined publicly available RNA sequencing (RNAseq) datasets (striatum (GSE65774),

119 cortex (GSE65770) and liver (GSE65772), available through HDinHD, (see also Materials

120 and Methods) (Langfelder et al. 2016). The AS events in striatum, cortex and liver of 6

121 knock-in (KI) mouse models (Q20, Q80, Q92, Q111, Q140 and Q175) carrying different *Htt*

122 CAG repeat lengths were determined using rMATS (Shen et al. 2012, 2014; Park et al. 2013).

123 Q20 mice, with the lowest *Htt* CAG repeat length, were used as controls. The total number of

124 genes showing one of the five major types of AS events [alternative 5' splice site (A5SS),

125 alternative 3' splice site (A3SS), mutually exclusive exons (MXE), retained intron (RI) and

126 skipped exon (SE)] was evaluated using a collection of both reads spanning splicing junctions

127 (JC) and reads spanning splicing junctions plus on target (JCEC) (Supplemental Table_S1).

128 The striatum, the most severely affected brain region in HD and HD model mice,

129 exhibited the highest number of aberrantly spliced transcripts, while cortex and liver did not

130 show overt aberrant splicing phenotypes (Fig. 1a, Supplemental_Fig_S1, Supplemental

131 Table_S1). Strikingly, the alteration in the total number of mis-spliced events in the striatum

132 strongly correlated (0.8 to 0.97 R^2, Supplemental_Fig_S2) with *Htt* CAG length

133 (Supplemental_Fig_S2), with Q175 mice showing the greatest number of abnormal splicing

134 events at all three time points analyzed (2, 6 and 10 months). Cortical mis-splicing showed

135 some, more variable, degree of CAG correlation at 2 and 6 months (~ 0.6 to 0.9 R^2,

136 Supplemental_Fig_S2), while no significant association was detected in the liver.

137 Interestingly, a *Htt* CAG correlated increase in aberrant striatal AS was already visible at

7

138   early stages (2 months), becoming more significant as the pathologic process progressed (6-

139   10 months) (Fig. 1a, Supplemental Table_S1).

Fig. 1 - D. Ayyildiz et al., 2021

140
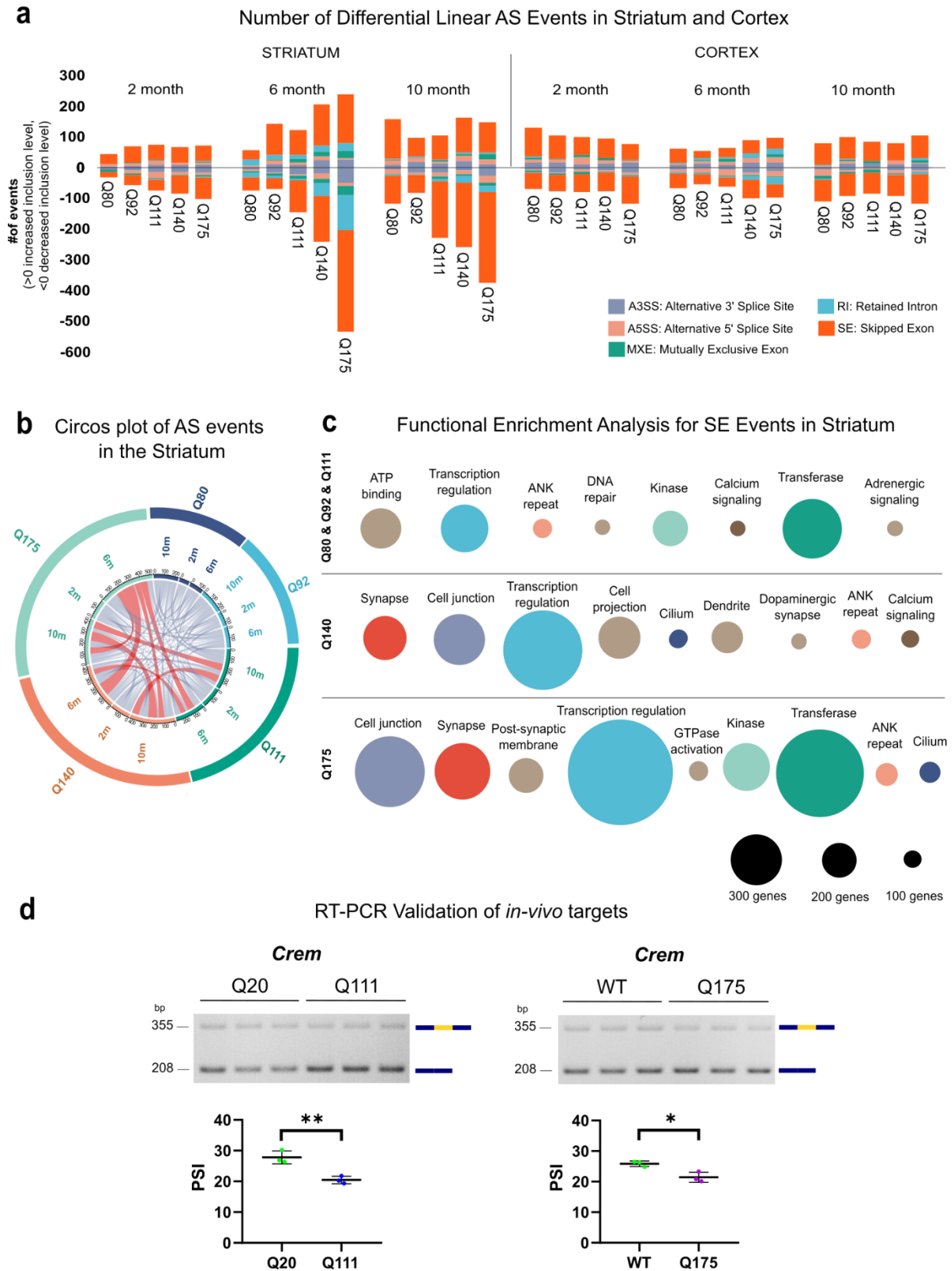
Figure 1. Aberrant linear alternative splicing in the striatum of KI animal models of HD.

a) Bar graphs show the number of differential AS events in the striatum and cortex from 6 mouse KI models (Q80, Q92, Q111, Q140 and Q175) of HD, presenting different *Htt*-CAG repeat lengths and 3 ages (2, 6, and 10 months). The number of events is shown for each genotype, time point and brain region. The inclusion level is calculated in comparison to Q20 controls and the positive or negative values are plotted. Source data by Langfelder P. *et al* (2016). Further details can be found in the Methods section. Each color of the bar chart represents a different AS event. **b)** Circos plot represents the number of transcripts within the striatum - showing differential AS events - shared between different genotypes (Q80, Q92, Q111, Q140 and Q175) and time points (2, 6, and 10 months). Conditions (genotypes and/or time points) sharing more than 50 transcripts are depicted in red. **c)** Weighted nodes graphical representation shows the functional enrichment analysis for transcripts showing significant skipped exon (SE) events in the striatum. Highly expanded *Htt* CAG sizes (Q140 and Q175), the major contributors to aberrant SE in the striatum, are shown separated. Nodes' size legend is depicted at the bottom. Nodes are ordered by adjusted p-values using Benjamini Hochberg (<0.05). The same enrichment terms shared between genotypes are presented with nodes of the same color, while unique terms are colored in light brown. **d)** Representative agarose gel images (top) and dot-plots (bottom) show the RT-PCR results of AS validation for *Crem*, a selected transcript target. RT-PCR assay and quantification were performed on RNA from striata from an independent set of wild-type (WT), Q20 and Q111/Q175 mice. *Crem* isoforms with inclusion or exclusion of the variable exon (in yellow) are visualized and quantified. Dot plots report the PSI, percent-spliced-in. *$P < 0.05$, **$P < 0.01$ (Student's unpaired t-test; n = 3), error bars indicate standard error of the mean.

9

166   While RI, MXE and A3SS showed some association with CAG length, especially in

167   the striatum at 6 months of age, the SE splicing subtype showed the strongest repeat length

168   association (see R^2 in Supplemental_Fig_S3). Importantly, the overlap, between genes

169   exhibiting AS alteration and expression differences at each time point and with any genotype

170   tested, was negligible (1.2 %) (Supplemental_Fig_S4), indicating that the AS alterations were

171   not caused by concomitant transcriptional changes or *vice versa*.

172   Because the striatum exhibited a more prominent alteration in aberrant AS events

173   during the progression of the pathologic process, we then asked whether aberrant AS events

174   were shared among a common set of striatal genes regardless of *time points* and genotype. As

175   depicted by the circos plot in Fig. 1b, the vast majority of genes with differential AS were

176   unique to a specific genotype and time point, though involved in redundant GO terms such as

177   'synapse', 'cell junction', 'transcription' (Supplemental_Table_S1). However, some common

178   genes could be found among the lines with the most expanded CAG tracts (Q140 and Q175),

179   mainly at 6 months of age (Fig. 1b). Due to the prevalence of aberrant SE events compared to

180   other subtypes, we performed functional enrichment analysis on genes exhibiting CAG-

181   dependent SE mis-splicing events. We analyzed the highest CAG repeat length lines (Q140,

182   Q175) separately, while the "lower" repeat length lines (Q80, Q92, and Q111), with fewer

183   altered isoforms, were grouped together (Fig. 1c). Functional enrichment revealed that

184   'transcriptional regulation', 'cell junction', and 'synapse' were recurrent among genotypes

185   and the most significantly enriched terms (Fig. 1c, Supplemental_Table_S1).

186   Given the predominance of aberrant SE in the mouse striatum, we wanted to experimentally

187   validate the differential exon inclusion observed using two separate lines (Q111 and Q175) in

188   striatal samples dissected from 6 months old mice. We prioritized a list of transcripts from the

189   AS splicing analysis involved in 'transcriptional regulation', 'synapses' processes, known

190   pathways altered in HD, and showing a significant change in percent-spliced-in ($|\Delta PSI| >$

191 0.25) of alternative exons between Q20 controls and Q111 or Q175

192 (Supplemental_Table_S1). *Crem*, a cAMP-responsive element modulator (Nagamori et al.

193 2006), was an interesting target showing aberrant SE (ΔPSI (JCEC) = - 0.3, SE increase in

194 *Htt* CAG expanded alleles) in both Q111 and Q175 striata. Replicate RT-PCR experiments,

195 using an independent set of Q20, Q111 and Q175 mouse striatal samples, confirmed the

196 altered *Crem* SE event (Fig. 1d).

197

198 **Aberrant linear alternative splicing correlates with *Htt* CAG size in murine neural**

199 **progenitor cells**

200 In order to follow how linear AS behaved during the transition from pluripotency to neuronal

201 commitment in presence of the *Htt* CAG expansion mutation, we took advantage of a panel

202 of mouse embryonic stem cells (mESCs) (Q20, Q50, Q92 and Q111) harboring different

203 CAGs within a single *Htt* allele (Auerbach et al. 2001; Wheeler et al. 1999; Jacobsen et al.

204 2011a), closely mimicking the human HD mutations. We also included wild-type (WT) and

205 *Htt* null alleles in which the endogenous *Htt* locus was unaltered or inactivated by the

206 insertion of a neomycin cassette between exon 4 and 5, respectively (White et al. 1997).

207 mESC with different genotypes were pushed toward neural differentiation following a

208 previously reported protocol (Conti et al. 2005). Differentiated mouse neural progenitor cells

209 (mNPC) were initially characterized by immunofluorescence and RT-qPCR, to confirm the

210 correct reduction of pluripotency markers (i.e. *Pou5f1, Nanog*) and the induction of neural

211 progenitor markers (i.e. *Nes, Vim*, *Msi1* and *Sox2*) (Supplemental_Fig_S5 a-b).

212 Following RNAseq, linear AS events for mESCs and mNPCs were analyzed by the same

213 pipeline employed for the *in vivo* analysis, comparing the different expanded *Htt* CAG repeat

214 lengths (Q50, Q92, and Q111) to the Q20 *Htt* CAG repeat (control). Similar to the previous

215   results (Fig. 1a), a direct correlation between the total number of genes with aberrant AS

216   events and CAG repeat length was observed in mNPCs but not in mESCs (Fig. 2a).

Fig. 2 - D. Ayyildiz et al., 2021

217

218
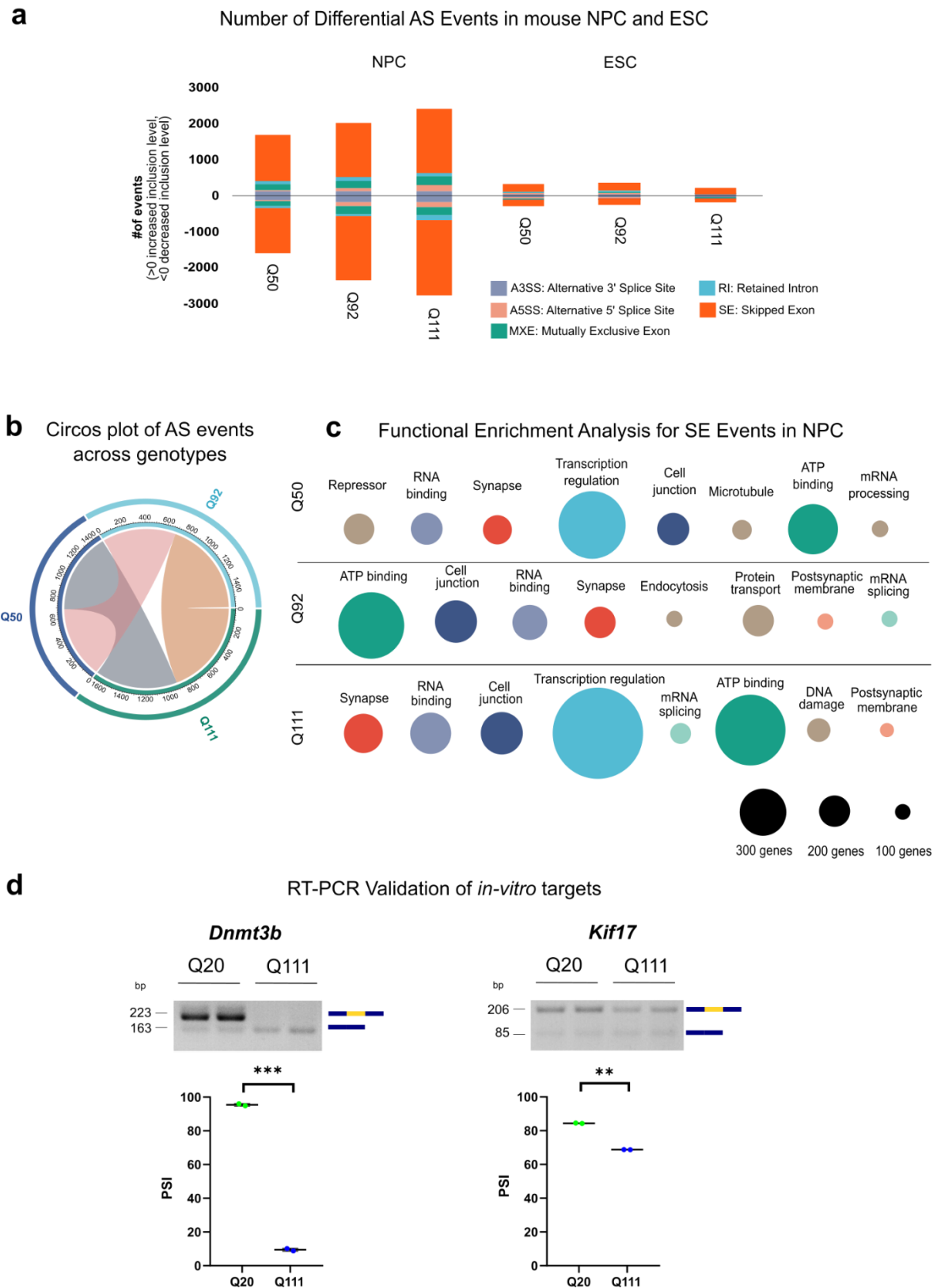
219 **Figure 2: Aberrant Linear Alternative Splicing correlates with *Htt* CAG size in mNPC.**

220 **a)** The bar graph reports the number of differential AS events in the mNPC and mESC from

221 KI models of HD, presenting 3 different *Htt*-CAG repeat lengths (Q50, Q92, Q111). The

222 number of events is reported for each genotype and differentiation stage. The inclusion level

223 is calculated in comparison to Q20 controls and the positive or negative values plotted in the

224 graph. Further details in the Methods section. Each color of the bar chart represents a

225 different AS event. **b)** Circos plot reports the number of transcripts in mNPC - presenting

226 differential AS events - shared between different genotypes (Q80, Q92, Q111). Conditions

227 sharing more than 50 transcripts are depicted in colors. **c)** Weighted nodes graphical

228 representation shows the functional enrichment analysis for transcripts presenting significant

229 skipped exon (SE) events in mNPC. Each genotype is shown separately. Nodes' size legend

230 is depicted at the bottom. Nodes are ordered by adjusted p-values using Benjamini Hochberg

231 (<0.05). The same enrichment terms shared between genotypes are presented with nodes of

232 the same color, while unique terms are colored with light brown. **d)** Representative agarose

233 gel images (top) and dot-plots (bottom) report the RT-PCR results of *in vitro* AS validation

234 for *Dnmt3b* and *Kif17*, selected transcript targets. RT-PCR assay and quantification were

235 performed on RNA from Q20 and Q111 mNPCs. *Dnmt3b* and *Kif17* isoforms presenting

236 inclusion or exclusion of the variable exons (in yellow) are visualized and quantified. Dot

237 plots report the PSI, percent-spliced-in. **P < 0.01, ***P < 0.001 (Student's unpaired t-test; n

238 = 3), error bars indicate standard error of the mean.

239

240   Interestingly, similar results were obtained from the analysis of AS events in *Htt* null mNPCs

241   but the affected genes were mostly different (only 11% were overlapping) and unique to KO

242   or KI condition (Supplemental_Fig_S6). Similar to the *in vivo* models, the highest proportion

243   of aberrant AS events was the SE subtype and increasing *Htt* CAG size was more strongly

244   linked to decreased rather than increased inclusion levels (Fig. 2a). Importantly, also in the *in*

245   *vitro* models, the AS alterations were vastly not associated with concomitant transcriptional

246   changes, with minor overlaps (20.8%) detected between differentially expressed (DE) and

247   AS-affected genes among genotypes and differentiation time points

248   (Supplemental_Table_S2). Consistent with a more homogeneous cellular population, mNPCs

249   showed a clear overlap between genes showing/exhibiting aberrant splicing (any subtype)

250   across different genotypes (Fig. 2b). Moreover, functional enrichment of genes

251   showing/exhibiting mis-splicing events showed that 'RNA-binding', 'mRNA processing',

252   'splicing', 'kinase' and 'synapse' were the most significantly enriched terms

253   (Supplemental_Table_S2). Due to the prevalence of SE subtype among the mis-spliced

254   events, a dedicated functional enrichment analysis was performed for each genotype

255   separately. The analysis highlighted terms such as 'transcriptional regulation', 'ATP

256   binding', 'RNA binding' and 'synapse' as recurrent among the genotypes (Fig. 2d,

257   Supplemental_Table_S2) and mirroring the *in vivo* data (Fig. 1c). Similarly, we prioritized

258   transcripts involved in the indicated biological pathways with statistically significant SE

259   events with $|\Delta PSI| > 0.5$ between Q20 and Q111 NPC for RT-PCR experimental

260   validation. *Dnmt3b* (SE (Q111_JCEC) $\Delta PSI = -0.86$, SE increase in *Htt* CAG expanded

261   alleles), involved in maintenance of DNA methylation and transcriptional regulation

262   processes (Okano et al. 1999) and *Kif17* (Kinesin Family Member 17) (SE (Q111_JCEC)

263     $\Delta PSI = -1.0$, SE increase in *Htt* CAG expanded alleles) with ATPase and microtubule motor

15

264   activities (Wong-Riley and Besharse 2012) were prioritized. Replicate RT-PCR experiments

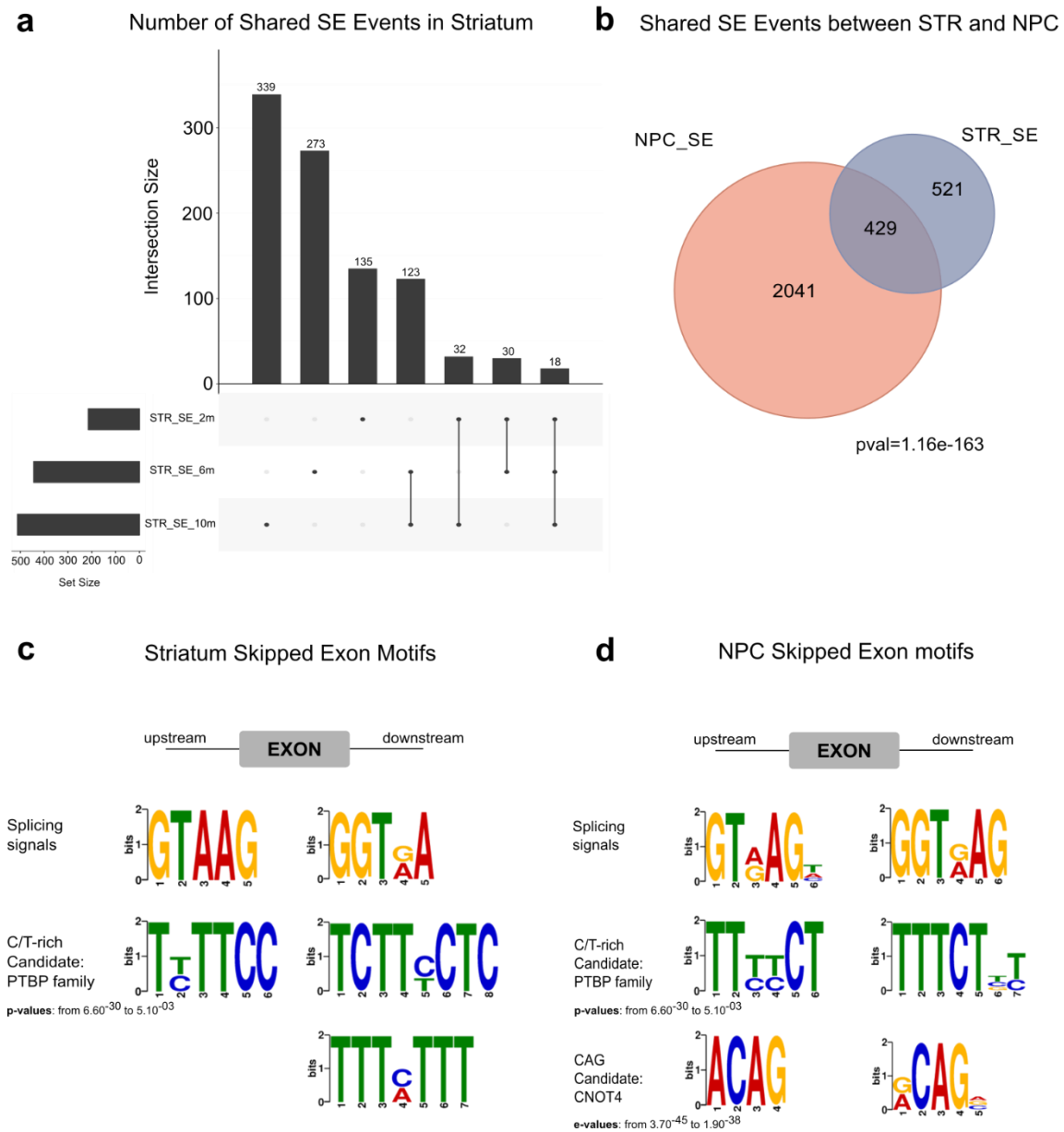265   validated the altered SE events (Fig. 2d).

266   **_Htt_ CAG size dictates a specific linear AS signature in neural cells.**

267   Because of the similarities in the AS results between the _in vivo_ and _in vitro_ HD

268   models, we sought to directly intersect the genes exhibiting SE in the mouse striatum and

269   mNPCs. While the majority of SE events in the striatum were age-specific (Fig. 3a), some

270   more overlap was observed at 6 and 10 months of age (123 genes shared, 13% of total).

271   Strikingly, however, _Htt_ CAG expansion caused a specific AS disruption in genes common in

272   the _in vivo_ and _in vitro_ HD models (45% of the total striatal SE events, p-value: 1.16^{-163},

273   Fisher's exact test) (Fig. 3b), which were mainly involved in 'cell junction' and 'synapse'

274   enriched GO terms (Supplemental_Table_S1-2).

275   Importantly, by inspecting the regions adjacent to the differentially skipped exons (+/-

276   100bp) in both mouse striatum and mNPCs, we detected a specific and significant enrichment

277   for C/T rich motifs (p-values from 6.60^{-30} to 5.10^{-03}), typically bound by polypyrimidine

278   tract binding proteins (PTBP) (Ray et al. 2013) (Fig. 3c, d and Supplemental_Table_S3).

279   Furthermore, a shorter, CAG-rich motif was found (p-values from 3.70^{-45} to 1.90^{-38}),

280   which could possibly be bound by CNOT4 (Lau et al. 2009).

281

Fig. 3 - D. Ayyildiz et al., 2021

a Number of Shared SE Events in Striatum

b Shared SE Events between STR and NPC

c Striatum Skipped Exon Motifs

d NPC Skipped Exon motifs

282

283   **Figure 3. *Htt*-CAG size dictates a specific linear AS signature in neural cells.**
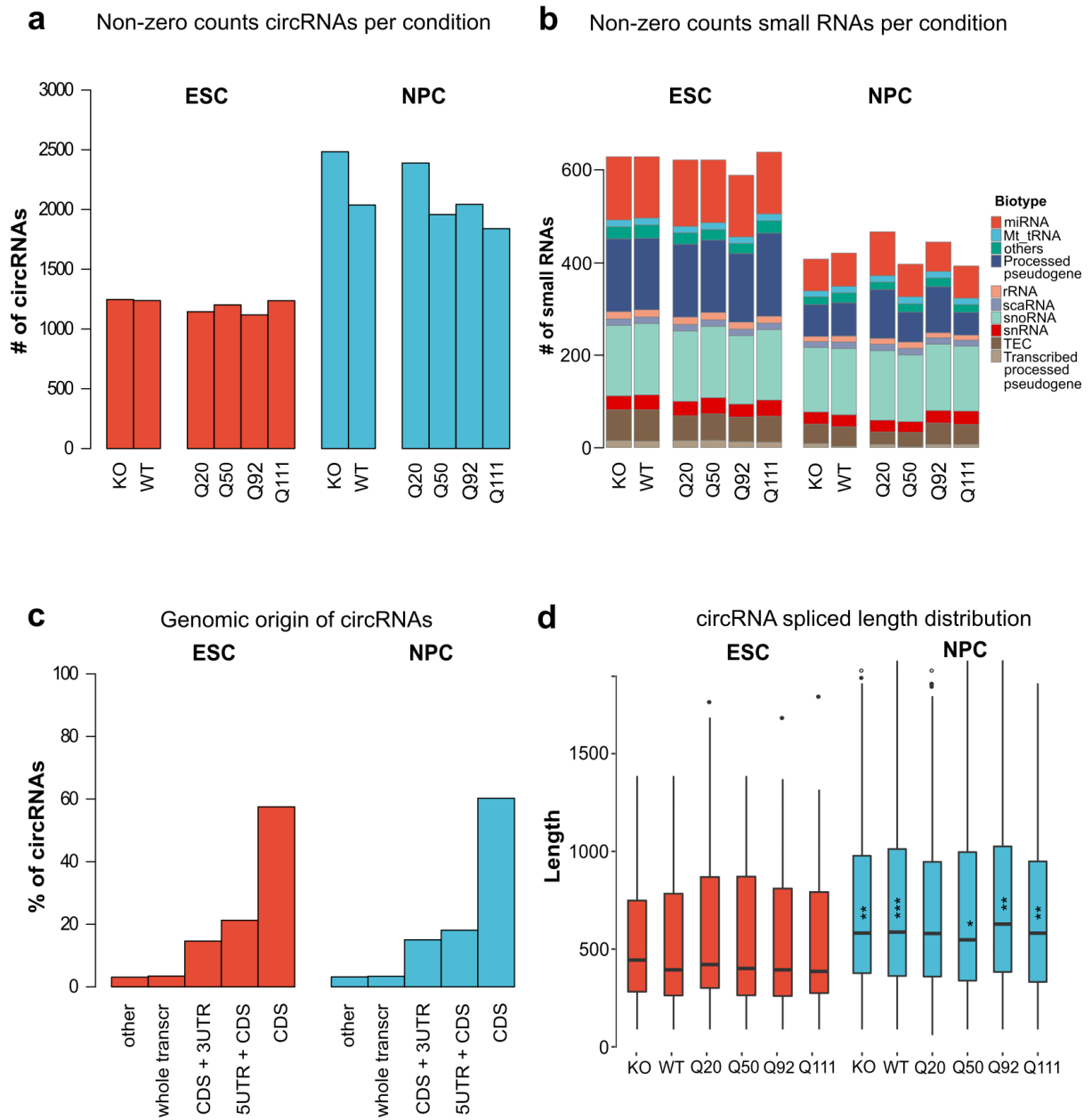
284   **a)** The upset plot displays the number of SE events shared among different time points (2, 6

285   and 10 months) within the mouse striatum. The number of events within each intersection is

286   presented in the vertical bars. Intersection groups (lines) or single time points (dots) are

287   shown in the lower panel. Sample set size is indicated at the bottom left of the panel. **b)** The

288   Venn diagram reports the comparison between striatal and mNPC SE events. All genotypes

289   (Q80, Q92, Q111, Q140 and Q175) and time points (2, 6 and 10 months) for striatal districts

290   and all genotypes (Q50, Q92, Q111) for mNPC were combined. Shared SE events (45% of

291   the total striatal SE events, p-value: 1.16^{-163}, Fisher's exact test) are indicated in the

292   intersection. **c-d)** Motif analysis identifies the splicing factors and/or RBP binding sites in the

293   ± 100bp upstream and downstream adjacent regions to the alternatively spliced exons for the

294   striatum (c) and for mNPC (d). The P-value of enrichment testing for individual motifs in

295   each data set is indicated. The candidate binding splicing factor and/or RBP family is shown.

296   *Htt* **CAG mutation impacts circRNA biogenesis in mouse neural progenitors.**

297   Considering that AS is fundamental to the biogenesis of circRNAs, we asked whether

298   the *Htt*-CAG expansion could be also associated with defective back-splicing in the *in vitro*

299   system. First, the analysis of the percentage of circRNAs originating from each chromosome

300   generally complied with the expected ratio relative to transcript mass in both mESCs and

301   mNPCs (Supplemental_Fig_S7 a-b). CircRNA reads (average reads/sample = 3825) were

302   then quantified in all samples by means of the DCC platform (Cheng et al. 2016)

303   (Supplemental_Table_S4-5). As shown in Fig. 4a, we observed, for all genotypes, a higher

304   number of expressed circRNAs in mNPC than in mESC samples with a corresponding higher

305   circRNA fraction of total transcript mass (circRNA reads/circRNA + host transcript reads,

306   Supplemental_Fig_S7c). This trend was specific for circRNAs, given that when evaluating

307   small RNAs [miRNAs, mitochondrial transfer RNAs (Mt-tRNAs), processed pseudogenes,

308   ribosomal RNAs (rRNAs), small nucleolar RNAs (snRNAs), small Cajal body-specific

309   RNAs (scaRNAs), To be Experimentally Confirmed (TEC), transcribed processed

310   pseudogenes] (average aligned reads/sample = 1.04M, see methods, Supplemental_Table_4-

311   5), we observed a general reduction in mNPCs compared to mESCs (Fig. 4b). Specifically,

312   miRNAs and processed pseudogenes subtypes appeared to be strongly affected by neural

313   differentiation, considerably reduced in mNPCs (Fig. 4b and Supplemental_Fig_S8).

314   While more than half of circRNA species derived from coding sequences (CDSs) and 5'-

315   untranslated regions (5'-UTRs) in both mESCs and in mNPCs (Fig. 4c), their spliced lengths

316   were significantly longer in mNPCs (Fig. 4d).

317

Fig. 4 - D. Ayyildiz et al., 2021

a Non-zero counts circRNAs per condition

b Non-zero counts small RNAs per condition

c Genomic origin of circRNAs

d circRNA spliced length distribution

318

319

320 **Figure 4. Neural differentiation impacts on circRNAs and small-RNAs biogenesis.**

321 **a)** The bar chart shows the number of detected circRNAs (>1 count/sample, see also

322 Methods), comparing pluripotent (mESC) and neural committed progenitors (mNPC).

323 Different *Htt* genotypes are presented: *Htt* double knock-out (KO) and a series of 4 *Htt*-CAG

324 expansion alleles (Q20, Q50, Q92 and Q111). Average circRNA count from two biological

325 replicate experiments is plotted. **b)** The color bar graphs report the number of small RNAs

326 (>1 count/sample, see also Methods) comparing mESC and mNPC conditions. *Htt* genotypes

327 as in a). Different classes of small RNAs are examined. Abbreviations as follows:

328 microRNAs (miRNAs), Mitochondrial transfer RNAs (Mt-tRNAs), processed pseudogenes,

329 ribosomal RNAs (rRNAs), small nucleolar RNAs (snRNAs), small Cajal body-specific

330 RNAs (scaRNAs), To be Experimentally Confirmed (TEC), transcribed processed

331 pseudogenes. **c)** The bar chart reports the percentage of circRNAs derived from specific

332 transcripts areas: transcript coding sequence (CDS), 3' or 5' untranslated regions

333 (3UTR/5UTR), whole transcript or other. The data compare mESC and mNPC conditions.

334 All *Htt* genotypes were combined. **d)** The box plot displays the spliced length distribution

335 [average ± standard deviation (SD)] for circRNAs in the various conditions of mESC and

336 mNPC cells. *Htt* genotypes as in a). Wilcoxon test p-values of the difference between

337 corresponding conditions of mESC and mNPC are shown as stars (* < 0.05, ** < 0.01, *** <
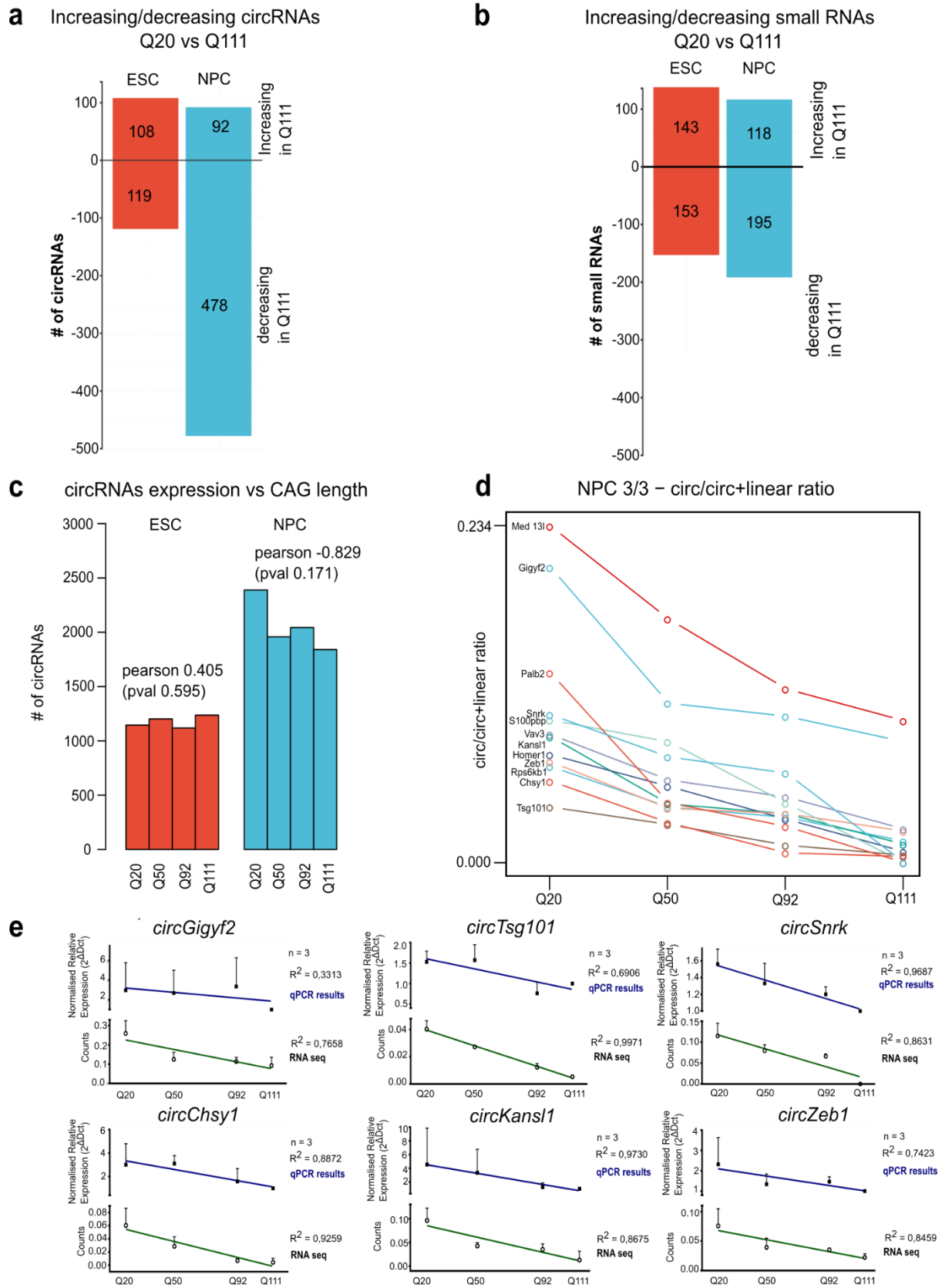
338 0.001).

339

340

21

341 We then analyzed the changes in circRNA expression between samples of increasing CAG

342 length. As shown in Fig. 5a, while in mESC the number of increasing and decreasing

343 circRNAs was almost equal, circRNA expression in mNPCs predominantly decreased in

344 Q111 vs Q20 samples (92 increasing vs 478 decreasing). Genes harboring differentially

345 expressed circRNAs were over-represented in pathways and functions such as 'cGMP-PKG

346 signaling pathway', 'Lysosome' and 'DNA binding' (Supplemental_Table_S4). Interestingly,

347 the complete absence of huntingtin in mNPCs elicited an opposite effect on circRNA

348 production with the majority of circular RNA molecules turned on in *Htt* KO mNPCs

349 (Supplemental_Fig_S9 a-b).

350 To uncover a possible link between *Htt* CAG expansion and decreased circRNA expression,

351 we counted the number of expressed circRNAs in each *Htt* CAG genotype. Although not

352 nominally statistically significant (p = 0.171), possibly because of the reduced number of

353 genotypes examined (n = 4), a high inverse correlation (Pearson's coefficient = -0.829) was

354 observed in mNPCs (Fig. 5c and Supplemental_Fig_S7c). Such a trend was absent in mESCs.

355 We then selected a stringently defined subset of differentially expressed circRNAs whose

356 expression was (i) monotonically decreased with increasing CAG lengths; (ii) significantly (p

357 < 0.05) changed according to Pearson's correlation coefficient and (iii) significantly different

358 as defined by CircTest algorithm (Cheng et al. 2016). 12 circRNAs fulfilling these criteria

359 and showing a negative trend of expression with increasing *Htt* CAG length (Fig. 5d), were

360 used as circRNA candidates for experimental RT-qPCR validation. Biological triplicates of

361 ribo-depleted RNAs for the 4 *Htt* CAG repeat genotypes were used for quantification of

362 linear and circular isoforms (circRNA/circRNA+linear ratio). We experimentally confirmed

363 the expected expression changes, i.e. circRNA expression linearly decreasing from Q20 to

364 Q111, in 6 of 10 targets (Fig. 5e). Significantly, circ*Snrk* and circ*Kansl1,* previously

365 characterized for their function in regulating apoptosis (Zhu et al. 2021) and as miRNA

366   sponges (Wang et al. 2020), showed the highest CAG correlation Pearson's R^2 and most

367   significant p-values (circ*Snrk* R^2 = 0.9687, p = 0.0313 and circ *Kansl1* R^2 = 0.9730,

368   p = 0.0270) (Fig. 5e).

369

23

Fig. 5 - D. Ayyildiz et al., 2021

a Increasing/decreasing circRNAs Q20 vs Q111

b Increasing/decreasing small RNAs Q20 vs Q111

c circRNAs expression vs CAG length

d NPC 3/3 − circ/circ+linear ratio

e circGigyf2, circTsg101, circSnrk, circChsy1, circKansl1, circZeb1

370

371

372

Figure 5. *Htt*-CAG dependent reduction of circRNAs in mNPC.

a) The bar graph reports the number of circRNAs differentially expressed between *Htt* Q111 versus Q20 genotypes. The comparison is presented for pluripotent (mESC) and neural committed progenitors (mNPC). The number of circRNAs increasing (*Increasing in Q111* - upper part of the plot), and decreasing their expression in Q111 versus Q20 (*Decreasing in Q111* - lower part of the plot) is depicted. **b)** The bar chart shows the number of small-RNAs differentially expressed between *Htt* Q111 versus Q20 genotypes. The comparison is presented for pluripotent (mESC) and neural committed progenitors (mNPC). The number of small-RNAs increasing (*Increasing in Q111* - upper part of the plot), and decreasing their expression in Q111 versus Q20 (*Decreasing in Q111* - lower part of the plot) is depicted. **c)** The bar chart reports the total number of circRNAs expressed at each differentiation stage (mESC and mNPC) and for each *Htt*-CAG genotypes (Q20, Q50, Q92 and Q111). A negative correlation between the number of expressed circRNAs and *Htt*-CAG length is observed (although not nominally significant) in mNPC, but not in mESC cells. Pearson's correlation R = -0829, p-value = 0.171. d) Line plot describes the expression pattern of the 12 circRNAs of mNPC selected by three criteria (decreasing expression and negative correlation with *Htt*-CAG, significantly different expression by circTest- see Methods for further details). *Htt*-CAG expansion alleles (Q20, Q50, Q92 and Q111) are presented. **e)** The line charts report the ratio (circ/linear) of the expression levels for the selected circRNAs candidates in mNPC across different *Htt*-CAG genotypes. RT-qPCRs results of 3 biological replicates are plotted as normalized circ/linear expression ($2^{\wedge\Delta\Delta CT}$) and average ± standard deviation (SD) is presented. *Htt*-Q111 genotype – with lower circ/linear ratios - was used as relativizing condition. The linear trend lines are presented for the qPCR validation (blue line) and the corresponding RNAseq data (green line) (qPCR, filled symbols; RNAseq, empty symbols).

397   RNAseq data are presented as normalized transcript counts (counts). Pearson's correlation R

398   square values are indicated for the two regressions (qPCR and RNAseq).

399

400   **Linear and back-splicing alterations correlate with direct and indirect mis-regulation of**

401   **RNA-binding proteins and splicing factors.**

402   In order to deepen our understanding of the effects of *Htt* CAG expansion on linear and back-

403   splicing regulation and to gather new mechanistic insights, we focused on miRNAs, small

404   RNAs subtype, which are very sensitive to neural differentiation (Fig. 4b) and [already]

405   reported to control the stability of RNA-binding proteins and splicing factors in neurons

406   (Weiss et al. 2015; Fukao et al. 2015; Gardiner et al. 2015). Although miRNA expression

407   strongly decreased following neural differentiation (Fig. 4b and Supplemental_Fig_S8), we

408   asked whether we could identify differentially expressed miRNAs affected by the expression

409   of *Htt* CAG expansion in mNPC. Thus, we compared Q20 versus Q111 mNPCs and

410   identified 9 significantly differentially expressed miRNAs (Supplemental_Table_S5). By

411   exploiting the mirDB database (Chen and Wang 2020), we inferred their mRNA targets and

412   intersected them with mouse splicing factors and RNA-binding proteins (RBPs) lists (see

413   methods and Suppl. Table 5). We identified 73 transcripts (Fig. 6a), known RNA-binding

414   proteins and splicing factors (of the *Celf, Elav, HnRNPs, Mbnl, Ptbp, Srsf* families). These

415   were also targets of miRNAs dysregulated by mutant huntingtin and showing expression

416   levels changes with $|Log2FC| > 0.5$ and a significant p-value ($p < 0.05$) (Fig. 6a, Table).

417   Importantly, significant overlap and enrichment was also identified between miRNA targets

418   and transcripts showing differential alternative splicing or back-splicing in mNPC

419   (Supplemental_Fig_S10 and Supplemental_Table_S7), suggesting a possible role for

420   miRNAs in the regulation of AS and back-splicing events. Strikingly, *Ptbp3*, usually

421   expressed in thymus, lymph nodes and digestive system (Uhlén et al. 2015) whose binding

422   motifs were significantly enriched in the ± 100bp adjacent to the differentially skipped exons

423   in both the *in vivo* and *in vitro* systems (Fig. 3c,d), was significantly transcriptionally

424   upregulated (Fig. 6a, Table).

425   Taken together, while we cannot exclude direct transcriptional alterations of RBPs

426   and splicing factors by mutant huntingtin, we identified a pool of dysregulated miRNAs,

427   sensitive to the *Htt* mutation and possibly targeting linear and back-splicing factors. Thus, our

428   data support a complex direct and indirect mode of mutant huntingtin regulation of canonical

429   linear and circular RNA-producing splicing (Fig. 6b).

430

Fig. 6 - D. Ayyildiz et al., 2021

a

NPC miR Targets
Q111 vs Q20

Splicing Factors

RNA Binding Proteins

122

36

160

73

4172

224

683

p-value color scale

1.00e-10 1.00e-05 1.00e+00

Gene	Log2FC	p-value	Targeting miRNA(s)
Celf1	1,03	$1,31^{E-13}$	miR-293-5p; miR-106a-3p; miR-292a-5p; miR-877-3p;
Celf5	-1,19	$8,90^{E-04}$	let-7c-1-3p;
Elavl4	-1,40	$1,60^{E-06}$	miR-293-5p; miR-302b-5p; miR-409-3p; miR-292a-5p;
Hnrnpa3	1,07	$5,36^{E-03}$	miR-106a-3p;
Hnrnph3	-0,50	$3,21^{E-04}$	miR-302b-3p;
Hnrnpll	0,65	$1,16^{E-04}$	miR-293-5p; miR-292a-5p;
Hnrnpu	0,85	$6,70^{E-08}$	let-7c-1-3p;
Mbnl1	1,35	$2,26^{E-06}$	miR-293-5p; miR-106a-3p; miR-292a-5p; let-7c-1-3p;
Mbnl3	2,49	$4,01^{E-09}$	miR-293-5p; miR-302b-3p; miR-302b-5p; let-7c-1-3p; miR-106a-3p; miR-292a-3p; miR-292a-5p;
Ptbp3	0,80	$2,02^{E-05}$	miR-293-5p; miR-106a-3p; miR-292a-5p; let-7c-1-3p;
Rbfox2	-0,49	$8,37^{E-03}$	miR-877-3p;
Srsf1	0,99	$8,51^{E-03}$	miR-302b-5p; miR-106a-3p;
Srsf2	0,99	$2,28^{E-04}$	miR-293-5p; miR-409-5p; miR-292a-5p; miR-877-3p;
Srsf3	0,51	$4,19^{E-05}$	miR-293-5p; miR-292a-5p;

b

431

432

28

433 **Figure 6. Linear and back-splicing alterations correlate with direct and indirect mis-**

434 **regulation of RNA-binding proteins and splicing factors.**

435 **a)** The Venn diagrams reports the overlap between mutant huntingtin's dysregulated miRNA

436 targets (4505 transcripts targets of 9 dysregulated miRNAs comparing Q20 versus Q111 *Htt*-

437 CAG genotypes in mNPC, see Methods and Supplemental_Table_S5), the list of splicing

438 factors (391) and RNA binding proteins (1140) (see Methods and Supplemental_Table_S5).

439 The enrichment scores observed for the different lists of splicing regulators is represented as

440 color-coded within the intersection areas. Color-coded scale bar on the bottom left corner of

441 the figure. A subset of transcripts with splicing regulator or RBP functions which are targets

442 of mutant huntingtin's dysregulated miRNAs (smaller intersection, # 73) are highlighted in

443 the left table. Log fold change (LogFC) transcripts difference in the reference conditions

444 (mNPC Q20 versus Q111), associated p-value and targeting miRNAs are shown. **b)** The

445 scheme summarizes the main findings and the proposed mechanistic hypotheses of the

446 current study, supporting a complex, direct and/or indirect, mode of mutant huntingtin

447 regulation of canonical linear and circular RNA-producing splicing. The summary scheme

448 was created by using BioRender.com.

449

450

DISCUSSION

Alterations in the choice of splice sites resulting in proteins with different structures and functions, altered mRNA localization, translation or decay is crucial for the complexity of the mammalian nervous system (Li et al. 2007; Zheng and Black 2013; Raj and Blencowe 2015; Darnell 2013). Splicing defects impacting the functionality of mature neurons are increasingly implicated in neurological and neurodegenerative diseases. Thus, there is an increasing need to better understand these regulatory processes.

Since the discovery of the causal genetic mutation underlying HD, great efforts have been made to uncover the functions of both wild-type and mutant huntingtin, which is now thought to act in a truly pleiotropic manner. The challenge nowadays is to reveal which of these pathways might exert a crucial early role in the onset and progression of HD pathology.

Here, we investigated how RNA processing and, specifically, alternative splicing is affected by mutant huntingtin. By resourcing to publicly available RNA-seq data from *in vivo* HD KI mice (Q20, 80, 92, 111, 140 and 175) with different CAG lengths (Langfelder et al. 2016) and our newly generated RNA-seq from an isogenic panel of mouse ESCs and NPCs, we uncovered a neural specific, CAG-length dependent alteration in alternative splicing. Exon skipping and, to a lesser extent, intron retention events were primarily altered. Notably, the increased total number of AS events was significantly correlated with CAG length (and to a lesser extent with age) in the striatum, but not within the cortex and liver, hence supporting preferential striatal vulnerability. Interestingly, significant changes in the AS pattern could already be detected prior to overt behavioral phenotypes stages (2 months). While previous reports correlated mutant huntingtin expression to mis-splicing events, locally affecting the *Htt* locus (Neueder et al. 2017; Sathasivam et al. 2013; Schilling et al. 2019), and also more generally altering the whole brain's transcriptome (Lin et al. 2016; Elorza et al. 2021), this is the first evidence of a direct correlation between the number of *Htt* CAG repeats and the

476 degree of AS, extending the repertoire of molecular phenotypes directly linked to the CAG

477 expansion. The concept of CAG-dependency is currently extensively investigated to gain

478 insights into how *Htt* CAG repeat length might modify HD pathogenesis (Langfelder et al.

479 2016; Galkina et al. 2014; Reis et al. 2011; Seong et al. 2005). Notably, our findings

480 supporting a primary CAG length-dependent effect in striatum – with less involvement of

481 cortex [and liver] – is well in line with previous observations that highlighted CAG length-

482 dependent modules of co-expressed genes (Langfelder et al. 2016).

483   Similarly, the results obtained from the AS analyses, on the isogenic *in vitro* KI

484 (Auerbach et al. 2001; Wheeler et al. 1999; White et al. 1997; Duyao et al. 1995; Jacobsen et

485 al. 2011b) system in the transition from pluripotency to neural committed progenitors,

486 provided an independent, genome-wide validation of the AS *Htt* CAG correlation. Further,

487 we confirmed an increase of AS and SE events with increasing CAG length. However, this

488 phenotype was limited to mNPCs, which also exhibited a greater number of total AS events

489 relative to mESCs. These observations globally suggest that (i) neural progenitors activate

490 prominent neural splicing processes (Weyn-Vanhentenryck et al. 2018), and (ii) the

491 correlation with CAG length likely requires factors expressed in the neural lineage.

492 Functional enrichment analyses for transcripts experiencing SE events, *in vivo* and *in vitro*,

493 revealed GO and pathways already associated with HD (Labbadia and Morimoto 2013;

494 Hodges et al. 2006; Kuhn et al. 2007), with prevalence of transcriptional and neural-related

495 functions. Taken together, these findings suggest that linear splicing becomes progressively

496 dysregulated as CAG length increases, specifically in neuronally-committed cells and adult

497 striatum, thus possibly contributing to HD pathogenesis.

498   It has been ascertained that AS regulation is also crucial to the biogenesis of

499 circRNAs, stable, neuronally-enriched, circular non-coding RNAs (Ashwal-Fluss et al. 2014;

500 Ivanov et al. 2015), with important roles during neural development (You et al. 2015; Venø et

501 al. 2015a) and in brain function (Piwecka et al. 2017; Lu et al. 2019). While their contribution

502 to brain pathologies and specifically to neurodegenerative diseases is becoming recognized

503 (Jia et al. 2020; Zhang et al. 2020; Dube et al. 2019), circRNAs remain largely neglected in

504 HD. We aimed to provide a first characterization of back-splicing regulation in the presence

505 of mutant huntingtin expression, exploiting the isogenic *in vitro* KI system. In contrast to

506 what was observed for linear splicing, we detected a general decrease in circRNA abundance

507 with increasing CAG length. While the impact of mutant huntingtin on circRNAs biogenesis

508 is remarkable when comparing the two extreme genotypes (Q20 versus Q111), the genes

509 affected by AS and back-splicing are largely different, and also enriched in different GO and

510 pathways terms. However, once again, the more striking phenotype appeared to be confined

511 to neural progenitors with minimal changes in mESC.s It is unclear at present whether these

512 circRNAs play known or novel regulatory roles (Wilusz 2018), and whether their depletion

513 might contribute to HD pathogenesis.

514 Taken together, our observations demonstrate an opposite correlation of linear AS and back-

515 splicing with the number of CAG repeats, thus suggesting a possible link between the two

516 splicing types, with contrary/reversed regulation. Previous studies reported exon skipping as

517 a promoter of skipped-exon circularization (Kelly et al. 2015). However, other data suggested

518 that alternative splicing is likely in competition with back-splicing (Holdt et al. 2018;

519 Ashwal-Fluss et al. 2014). Our results support the latter idea, where a global increase of SE

520 events downregulates back-splicing efficiency although it remains unclear whether the *Htt*

521 CAG mutation primarily affect linear splicing, exon circularization, or both simultaneously.

522 The effect exerted by the *Htt* CAG expansion on circRNA biogenesis in mNPCs, might be

523 exacerbated by the higher number of circRNAs detected at this developmental stage, already

524 reported in a number of studies (Izuogu et al. 2018; Szabo et al. 2015; Venø et al. 2015b). On

525 the other hand, it might suggest that the *Htt* CAG expansion could, directly or indirectly,

526 regulate some neural splicing factors also implicated in back-splicing control. the *Htt* CAG

527 expansion, thus, might induce an abnormal interaction and/or alter the expression levels of

528 small RNAs, splicing factors or RBP, specifically in neural-lineages (Murthy et al. 2019), in

529 turn affecting the two types of splicing.

530 Here we explored a network of miRNAs and their targets to expose some mechanistic

531 insights. miRNAs revealed to be altered by neural differentiation and, to a lesser extent, by

532 the *Htt* CAG expansion. Interestingly, a fraction of the targets of mutant huntingtin's-

533 dysregulated miRNAs, were significantly enriched for RBP and splicing factors, with

534 members of the PTBP, CELF, SRSF and hnRNPs families, also showing transcriptional

535 alterations (Fig. 6a, Table). This suggests that these AS regulatory elements might be

536 indirectly (via miRNAs alteration) or directly dysregulated by *Htt* CAG expansion (Fig. 6 b).

537 Importantly, targets of mutant huntingtin's-dysregulated miRNAs also revealed a significant

538 enrichment for transcripts exhibiting SE and back-splicing events in mNPCs, supporting the

539 existence of a complex RNA regulatory network by mutant huntingtin (Fig. 6 b).

540 In conclusion, we identified splicing and back-splicing as highly impacted molecular

541 alterations in HD model systems. Although previous reports associated mutant huntingtin

542 expression to linear splicing dysregulation (Lin et al. 2016; Schilling et al. 2019; Elorza et al.

543 2021), this is the first study reporting the relationship between magnitude of AS alterations

544 and *Htt* CAG length and the very first evidence of circRNA deregulation in HD. Moreover,

545 the identification of aberrantly regulated miRNAs, RBPs, and splicing factors, specifically in

546 neural committed cells and in combination with *Htt* mutations, proposes novel molecular

547 players contributing to HD pathogenesis and delineating new targets of therapeutic

548 intervention.

549

550   **METHODS**

551   **Generation and characterization of mNPC with different *Htt* CAG sizes**

552   The isogenic panel of wild-type, *Htt* null *Hdh*[ex4/5/ex4/5] and heterozygous *Htt* CAG knock-in

553   *Hdh*[Q20/7], *Hdh*[Q50/7], *Hdh*[Q92/7] and *Hdh*[Q111/7] mESCs, kindly provided by Dr. Marcy E.

554   MacDonald (Massachusetts General Hospital and Harvard Medical School, Boston, USA),

555   were cultured as described previously (Auerbach et al. 2001; Wheeler et al. 1999; White et al.

556   1997; Duyao et al. 1995; Jacobsen et al. 2011b). Pluripotent cells were maintained in

557   KnockOut DMEM (Gibco), supplemented with 15% of ESC-grade FBS (Gibco), 2 mM L-

558   glutamine (Gibco), 100 U/ml Penicillin/Streptomycin (Gibco), 1% Non-essential Amino

559   Acids (Gibco), 0.1 mM 2-mercaptoethanol (Sigma) and 1000 U/ml of leukemia inhibitory

560   factor (LIF) (Voden), on plates coated with 0.1% gelatin (Millipore) or on a feeder layer of

561   CF-1 IRR mouse embryonic fibroblast (TebuBio).

562   Neural differentiation was performed as previously described (Conti et al. 2005). Briefly,

563   self-renewing mESCs were dissociated and plated onto 0.1% gelatin-coated plates at a

564   density of 0.5–1.5 10^4cells/cm^2 in N2-B27 medium (Ying et al., 2003). After 7 days, cells

565   were detached using Accutase (Thermo Fisher Scientific) and plated on 3 μg/ml laminin-

566   coated dishes in neural stem (NS) expansion medium [composed by Euromed-N (Euroclone)

567   supplemented with 20ng/ml FGF (R&D) and EGF (Sigma), 1% N-2 Supplement (Gibco), 2

568   mM L-glutamine (Gibco) and 100 U/ml Penicillin/Streptomycin (Gibco)] (Conti et al. 2005).

569   Mouse neural progenitor cells were routinely passaged 1:2-1:4 every 3-5 days using Accutase

570   and maintained in NS expansion medium on laminin-coated plates (Sigma, 3 μg/ml). Both

571   mESCs and mNPCs were incubated at 37 °C and 5% CO2. The pluripotency of mESCs and

572   their differentiation to mNPCs was evaluated by RT-qPCR using stage specific markers (full

573   list of primers used in Supplemental_Fig_S6).

574

Immunofluorescence

Cells were seeded on a 24-well the day before the experiment and then fixed with 4% PFA for 15 minutes at room temperature, permeabilized with 0.5% Triton X-100 in 1X PBS and incubated with blocking solution (0.3% Triton X-100, 5% FBS in 1X PBS) for 1 hour. Primary antibody incubation was carried out at 4 °C overnight, followed by three washes with 1X PBS. Proper secondary antibodies were eventually used and images acquired using a confocal microscope (Leica TCS SP5). The following antibodies were used: rat anti-Nestin (Santa Cruz Biotechnology), rabbit anti-Sox2 (GeneTex), goat Alexa Fluor 546 and 647 (Life Technologies). Nuclei were stained using Hoechst 33342 (Life Technologies), diluted 1:20000 in 1X PBS.

Tissues dissection and isolation

All animal experiments were conducted to minimize pain and discomfort, under approved Institutional Animal Care and Use Committee (IACUC) protocol of the Italian Ministry of Health (project authorization n. 781/2016-PR) and the Massachusetts General Hospital. 3 *Htt* KI mouse lines with an *Htt* CAG repeat knock-in allele (Htt^{Q20}, Htt^{Q111} and zQ175) (C57BL/6 J inbred background) have been described previously (Wheeler et al. 1999; Lee et al. 2011; Grima et al. 2017; Menalled et al. 2003). Mice were maintained as heterozygotes and genotyped according to previously published protocols (Wheeler et al. 1999; Lee et al. 2011; Grima et al. 2017; Menalled et al. 2003).

Mice were sacrificed by CO_2 asphyxiation followed by cervical dislocation. The brain regions of interest (striatum and cortex) were dissected on ice, rapidly removed, snap frozen and stored at −80 °C for further use. Each group/genotype included 3 males and 3 females to avoid sex-dependent bias, were collected at 6 months of age.

RNA extraction and reverse transcription

Isolation of total RNA from tissues and cells was performed using TRIzol-based extraction method according to the manufacturer's protocol (Thermo Scientific). RNA was resuspended in nuclease-free water and quantified by a Nanodrop 2000 spectrophotometer (Thermo Scientific). The quality of RNA was estimated by using the RNA 6000 Nano or Pico Bioanalyzer 2100 Assay (Agilent). RNA samples with RNA integrity number between 6.8 and 9.5 were used. Unless otherwise noted, 1 μg of total RNA was reverse transcribed using SensiFAST cDNA Synthesis Kit following the manufacturer's protocol (Bioline).

Primer design

Transcripts were searched on Ensembl genome browser 95 (https://www.ensembl.org/Mus_musculus/Info/Index) and primers were designed on the coding sequence of each gene. Primers for end-point PCR were generated using the default setting on Primer3web (http://primer3.ut.ee/) and those for quantitative PCR by using Roche Universal ProbeLibrary for Mouse (https://lifescience.roche.com/). The list of primers used in this study is presented in Supplemental_Table_S6.

Quantitative PCR analysis

Quantitative polymerase chain reaction (qPCR) was performed using SensiFAST™ SYBR No-ROX Kit (Bioline, BIO-98020) according to the manufacturer's instructions. Data was normalized with RNA Polymerase II subunit A (*Polr2a*) as housekeeping gene (ΔCT) and analyzed with the 40-ΔCT method. The 40-ΔCT method was used to distinguish between down-regulation and overexpression of markers of mESCs and mNPCs

Validation of linear AS candidates' transcripts

625   Candidate transcripts presenting differential linear AS among the different conditions were

626   filtered on the basis of two parameters: (i) the percentage difference between the length of the

627   target exon – included or excluded - and the invariant adjacent exons should be higher than

628   25% and (ii) the percentage of target exon inclusion/exclusion difference between Htt^{Q111} and

629   Htt^{Q20} or Htt^{Q175} and Htt^{Q20} mice should be higher than 25%. For the selected candidates,

630   primers were designed targeting the invariant adjacent exons to amplify the two splicing

631   isoforms. RT-PCR was then performed using Phusion Green Hot Start II High-Fidelity PCR

632   Master Mix (Thermo Scientific) and 20-100 ng cDNA. PCR amplicons were resolved on 2%

633   agarose gels pre-stained with Xpert Green DNA Stain (Grisp) in $1\times$ TBE buffer (89 mM Tris-

634   borate and 2 mM EDTA, pH 8.3). The intensity of the bands was quantified using ImageJ

635   software (NIH, Bethesda, MD) and the percent spliced-in value was calculated as (intensity

636   of band with exon inclusion divided by the sum of the intensity of bands with both exon

637   inclusion and exclusion) x 100. Replicate experiments (2 or 3) were carried out and statistical

638   analysis was conducted using Student's unpaired t-test. $P \leq 0.05$ (*) or $P \leq 0.01$ (**)

639

640   **Validation of expression changes for selected circular RNA candidates**

641   For circular RNA validation experiments, 10 µg of total RNA was depleted of rRNAs by

642   using the Human/Mouse RiboMinus Transcriptome Isolation Kit (Invitrogen). Ribosomal

643   RNA depletion was confirmed by an Agilent 2100 Bioanalyzer (Agilent). Ribo-depleted

644   RNA was reverse-transcribed and quantitative PCR was performed by employing iTaq

645   Universal SYBR® Green Supermix (Biorad). To test the correlation between circular RNAs

646   candidates and the number of CAG repeats to faithfully recapitulate the bioinformatic

647   analysis, we considered the expression of both linear and circular isoforms of the selected

648   transcripts. We designed convergent and divergent primer sets for each candidate

649   (Supplemental_Table_S6). To specifically target the circular isoform, the backspliced

650 junction was selected. Expression values were obtained employing the $2^{\wedge -\Delta\Delta Ct}$ method,

651 normalizing with *Polr2a* and *Actb* housekeeping genes (HKG) and relativizing to Q111. The

652 selected HKGs were chosen among a panel of 5, showing a more stable expression in our *Htt*

653 KI samples. PCR amplicons were loaded on a 2% agarose gel, bands were carefully excised,

654 purified using PureLink Quick Gel Extraction and PCR Purification Combo Kit" (Bioline)

655 and eventually Sanger sequenced to finally prove their identity and the back-splice junction.

656 GraphPad Prism (GraphPad Softwares) was used to establish a regression line between

657 circular to linear ratios and the CAGs serie (20, 50, 89, 111). Finally, R^2 values and

658 corresponding p-values were calculated.

659

660 **Library preparation and RNA sequencing**

661 For total RNA, samples were subjected to DNAse I treatment (Ambion) according to

662 manufacturer instructions. Ribodepletion and barcoded stranded RNA-seq libraries

663 preparation was performed by the European Molecular Biology Laboratory Genomic Core

664 facility, following the standard Illumina protocols. Libraries were then sequenced with

665 NextSeq 500 paired-end, 75bp reads, obtaining ~60M reads/library.

666 Similarly, miRNA libraries were generated by the European Molecular Biology Laboratory

667 Genomic Core facility according to standard procedures and were sequenced on the Illumina

668 HiSeq 2500 using single-end, 50bp reads obtaining ~15M reads/sample.

669

670 **RNA-seq data, miRNA-seq and circRNA data analysis**

671 Data from RNA-seq and miRNA-seq on mESC and mNPC allelic series were processed as

672 follows. Reads filtering (minimum base quality of Q30) and adapter trimming was performed

673 with Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

674 Remaining reads were then aligned with STAR v2.5.3a (Dobin et al. 2013) to the mm10

675   assembly of the mouse genome coupled to the M14 Gencode genes annotation

676   (10.1093/nar/gky955). STAR parameters were set as recommended by the DCC pipeline

677   (doi: 10.1093/bioinformatics/btv656) to allow for the subsequent detection of circRNAs

678   through chimeric junctions. Reads derived from circular RNA species were extracted from

679   the WT, KO, and Q20 - Q111 samples of mESC and NPC cells by the DCC pipeline v0.4.4

680   (Cheng et al. 2016), run with default parameters. Repeats annotation provided to DCC was

681   obtained from the UCSC Genome Browser (doi: 10.1093/nar/gkx1020), by combining the

682   RepeatMasker and Simple Repeats track for the mm10 mouse assembly. Reads from

683   circRNA host genes were also quantified by DCC in the same run. Total transcript mass was

684   computed for a given transcript as the sum of its linear and corresponding circular RNA reads

685   count. The relative abundance of the circular form of the transcripts was then computed as

686   circular RNA read counts / circular + linear RNA read counts.

687   All p-values were computed by a Wilcoxon rank-signed test unless explicitly stated.

688   Exons/introns annotation and the number of protein-coding genes per chromosome were

689   obtained from UCSC (doi: 10.1093/nar/gkx1020) for the mm10 mouse genome assembly.

690   These annotations were then used to compute circRNAs spliced length and the enrichment of

691   expressed circRNAs in chromosomes (by a Fisher test).

692   Gene Ontology and pathways enrichment were computed with the topGO   (Alexa and

693   Rahnenfuhrer 2021) and clusterProfiler (Yu et al. 2012) R packages, using a BH adjusted p-

694   value threshold of 0.05.

695   Differential expression of circRNAs was computed by the CircTest R package (Cheng et al.

696   2016). circRNAs were filtered by requiring at least five reads in at least one sample and at

697   least 5% of total transcript mass in at least one condition. circRNAs having an FDR lower or

698   equal to 0.05 were considered to be differentially expressed.

699

Motif analysis

700

701 The sequence of the 100nts upstream and downstream of significant skipped exons events

702 were extracted with the biomaRt R package (Durinck et al. 2009). A motif search was

703 performed with DREME v4.10 (Bailey 2011), considering only same-strand matches, using a

704 background sequence set generated by DREME via shuffling of input sequences and a 1.0E-

705 03 threshold on the E-value to select significant motifs.

706

Differential Expression and Differential Linear Alternative Splicing Analysis

707

708 For the *in vivo* differential splicing analysis, the raw data from the study of Langfelder P et al

709 (2016) was used. Striatum (GSE65774), cortex (GSE65770) and liver (GSE65772) mRNA

710 expression profile datasets were retrieved for the analysis through the online database

711 HDinHD portal (https:// www.hdinhd.org/). At each of 3 time points (2, 6, 10 months), 8

712 heterozygous knock-in mice from each of the 6 Htt CAG repeat lengths (Q20, Q80, Q92,

713 Q111, Q140, and Q175) were used, resulting in 48 samples from each tissue and each time

714 point. Raw reads were subjected to sequence quality control using FastQC

715 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Removal of low-quality reads

716 and trimming of the adapter sequences were achieved by Trim Galore

717 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). In order to eliminate

718 variability between sequencing runs the reads were trimmed to 45 bp by Trim Galore,

719 resulting in the removal of only 4% of the reads. Raw sequences were aligned to mm10

720 mouse genome assembly (UCSC) with STAR RNA-seq aligner version 2.5.3a (Dobin et al.

721 2013) using standard settings. The differential alternative splicing (AS) events between each

722 of 5 samples for *Htt* CAG repeat lengths (Q80, Q92, Q111, Q140 and Q175) and Q20 (as

723 control) were identified by rMATS v4.0.1 (Shen et al. 2014) (http://rnaseq-

724 mats.sourceforge.net) that detects five major types of AS events from RNA-Seq data with

725 replicates (Shen et al. 2012, 2014). Analyses results were then further processed with

726 R/Bioconductor.

727 For the *in-vitro* differential splicing analysis, both mESC and mNPC samples were subjected

728 to the same pipelines used for *in-vivo* differential splicing analysis with same parameters and

729 tools. For the analysis of the differential alternative splicing (AS) events of 3 samples for *Htt*

730 CAG repeat lengths (Q50, Q92, and Q111) Q20 samples were used as control. For the

731 analysis of the differential alternative splicing (AS) events in knockout cells, wild type

732 samples were used as control.

733 Differential expression analysis was performed on striatum samples for *in vivo* with read

734 counts output of the alignment which was performed in previous step. edgeR (v3.24.3) was

735 used within the Bioconductor environment in R and p-values were adjusted for multiple

736 comparisons using the Benjamini–Hochberg method within each contrast and genes with

737 FDR-adjusted p-value < 0.05 were considered significantly differentially expressed. To

738 identify the genes both differentially expressed and alternatively spliced (all the different AS

739 subtypes were included), differential expression and alternative splicing results were

740 overlapped.

741 The functional enrichment analysis of the genes with AS events was performed using the

742 DAVID functional enrichment tool v6.8 mainly based on GO terms (biological process,

743 cellular compartment and molecular function), KEGG pathway, InterPro and UniProtKB

744 keywords. The enriched terms were filtered according to FDR adjusted P-value <0.05.

745

746 **DATA ACCESS**

747 The RNA-sequencing runs for mESC and NPC lines were deposited in GEO with ID

748 GSE175658.

749

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

V.C.W. is a scientific advisory board member of Triplet Therapeutics, a company developing new the- rapeutic approaches to address triplet repeat disorders such Huntington's disease and Myotonic Dystrophy. Her financial interests in Triplet Therapeutics were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. She is a scientific advisory board member of LoQus23 Therapeutics and has provided paid consulting services to Alnylam and Acadia Pharmaceuticals.

770 **AUTHORS' CONTRIBUTION**

771 D.A. was responsible for differential expression and linear alternative splicing (AS) analyses.

772 A.M. and T.T. conducted most of the wet lab work, including mESCs differentiation to

773 mNPCs, RNA collection and extraction, reverse transcription and PCR. J.D. and E.D.

774 provided experimental validation to linear AS differences. L.D. and A.M. validated circRNAs

775 expression differences. G.B. and M.K. contributed to mouse tissues collection and animal

776 handling and caring. F.D.L. helped with RNA collection and initial RNA libraries

777 preparation. J.Z. and L.C. supervised the differentiation of mESCs to mNPCs. V.C.W.

778 provided mouse striatal tissues for AS validation. C.D. provided computational methods and

779 suggestions for circRNAs data analysis. G.B. contributed to figures' design. E.K. assisted

780 with initial quality controls of the sequencing data. S.P. supervised D.A. during RNA-seq and

781 AS analyses. E.D. conducted miRNA-seq and circRNAs data analysis. D.A., E.D., S.P.,

782 provided critical advice on statistical analyses. M.B., E.D., S.P. conceptualized and designed

783 the study. D.A., J.D., E.D., G.B. and M.B. wrote the manuscript. M.B., E.D., S.P. supervised

784 the project. All authors revised, read and approved the submitted version.

785

786

REFERENCES

Abdelmohsen K, Panda AC, Munk R, Grammatikakis I, Dudekula DB, De S, Kim J, Noh JH, Kim KM, Martindale JL, et al. 2017. Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1. *RNA Biol* **14**: 361–369.

Alexa A, Rahnenfuhrer J. 2021. *topGO: Enrichment Analysis for Gene Ontology*. Bioconductor version: Release (3.13) https://bioconductor.org/packages/topGO/ (Accessed August 12, 2021).

Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. 2009. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res* **19**: 1732–1741.

Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S. 2014. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* **56**: 55–66.

Auerbach W, Hurlbert MS, Hilditch-Maguire P, Wadghiri YZ, Wheeler VC, Cohen SI, Joyner AL, MacDonald ME, Turnbull DH. 2001. The HD mutation causes progressive lethal neurological disease in mice expressing reduced levels of huntingtin. *Hum Mol Genet* **10**: 2515–2523.

Bailey TL. 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**: 1653–1659.

Barrett SP, Wang PL, Salzman J. 2015. Circular RNA biogenesis can proceed through an exon-containing lariat precursor. *Elife* **4**: e07540.

809   Bassi S, Tripathi T, Monziani A, Di Leva F, Biagioli M. 2017. Epigenetics of Huntington's

810   Disease. *Adv Exp Med Biol* **978**: 277–299.

811   Biagioli M, Ferrari F, Mendenhall EM, Zhang Y, Erdin S, Vijayvargia R, Vallabh SM,

812   Solomos N, Manavalan P, Ragavendran A, et al. 2015. Htt CAG repeat expansion

813   confers pleiotropic gains of mutant huntingtin function in chromatin regulation. *Hum*

814   *Mol Genet* **24**: 2442–2457.

815   Chen Y, Wang X. 2020. miRDB: an online database for prediction of functional microRNA

816   targets. *Nucleic Acids Research* **48**: D127–D131.

817   Chen Y, Yang F, Fang E, Xiao W, Mei H, Li H, Li D, Song H, Wang J, Hong M, et al. 2019.

818   Circular RNA circAGO2 drives cancer progression through facilitating HuR-

819   repressed functions of AGO2-miRNA complexes. *Cell Death Differ* **26**: 1346–1364.

820   Cheng J, Metge F, Dieterich C. 2016. Specific identification and quantification of circular

821   RNAs from sequencing data. *Bioinformatics* **32**: 1094–1096.

822   Conti L, Pollard SM, Gorba T, Reitano E, Toselli M, Biella G, Sun Y, Sanzone S, Ying Q-L,

823   Cattaneo E, et al. 2005. Niche-Independent Symmetrical Self-Renewal of a

824   Mammalian Tissue Stem Cell. *PLOS Biology* **3**: e283.

825   Darnell RB. 2013. RNA protein interaction in neurons. *Annu Rev Neurosci* **36**: 243–270.

826   Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,

827   Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:

828   15–21.

829   Dube U, Del-Aguila JL, Li Z, Budde JP, Jiang S, Hsu S, Ibanez L, Fernandez MV, Farias F,

830   Norton J, et al. 2019. An atlas of cortical circular RNA expression in Alzheimer

831   disease brains demonstrates clinical and pathological associations. *Nat Neurosci* **22**:

832   1903–1912.

833   Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of

834   genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**: 1184–

835   1191.

836   Duyao MP, Auerbach AB, Ryan A, Persichetti F, Barnes GT, McNeil SM, Ge P, Vonsattel

837   JP, Gusella JF, Joyner AL. 1995. Inactivation of the mouse Huntington's disease gene

838   homolog Hdh. *Science* **269**: 407–410.

839   Elorza A, Márquez Y, Cabrera JR, Sánchez-Trincado JL, Santos-Galindo M, Hernández IH,

840   Picó S, Díaz-Hernández JI, García-Escudero R, Irimia M, et al. 2021. Huntington's

841   disease-specific mis-splicing unveils key effector genes and altered splicing factors.

842   *Brain*. https://doi.org/10.1093/brain/awab087 (Accessed August 12, 2021).

843   Faber PW, Barnes GT, Srinidhi J, Chen J, Gusella JF, MacDonald ME. 1998. Huntingtin

844   interacts with a family of WW domain proteins. *Hum Mol Genet* **7**: 1463–1474.

845   Fukao A, Aoyama T, Fujiwara T. 2015. The molecular mechanism of translational control via

846   the communication between the microRNA pathway and RNA-binding proteins. *RNA*

847   *Biol* **12**: 922–926.

848   Galkina EI, Shin A, Coser KR, Shioda T, Kohane IS, Seong IS, Wheeler VC, Gusella JF,

849   Macdonald ME, Lee J-M. 2014. HD CAGnome: a search tool for huntingtin CAG

850   repeat length-correlated genes. *PLoS One* **9**: e95556.

851   Gardiner AS, Twiss JL, Perrone-Bizzozero NI. 2015. Competing Interactions of RNA-
852   Binding Proteins, MicroRNAs, and Their Targets Control Neuronal Development and
853   Function. *Biomolecules* **5**: 2903–2918.

854   Grima JC, Daigle JG, Arbez N, Cunningham KC, Zhang K, Ochaba J, Geater C, Morozko E,
855   Stocksdale J, Glatzer JC, et al. 2017. Mutant Huntingtin Disrupts the Nuclear Pore
856   Complex. *Neuron* **94**: 93-107.e6.

857   Han I, You Y, Kordower JH, Brady ST, Morfini GA. 2010. Differential vulnerability of
858   neurons in Huntington's disease: the role of cell type-specific features. *J Neurochem*
859   **113**: 1073–1091.

860   Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013.
861   Natural RNA circles function as efficient microRNA sponges. *Nature* **495**: 384–388.

862   Hansen TB, Wiklund ED, Bramsen JB, Villadsen SB, Statham AL, Clark SJ, Kjems J. 2011.
863   miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular
864   antisense RNA. *EMBO J* **30**: 4414–4422.

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

868   Holdt LM, Kohlmaier A, Teupser D. 2018. Molecular roles and function of circular RNAs in
869   eukaryotic cells. *Cell Mol Life Sci* **75**: 1071–1098.

870   Ivanov A, Memczak S, Wyler E, Torti F, Porath HT, Orejuela MR, Piechotta M, Levanon
871   EY, Landthaler M, Dieterich C, et al. 2015. Analysis of intron sequences reveals
872   hallmarks of circular RNA biogenesis in animals. *Cell Rep* **10**: 170–177.

873   Izuogu OG, Alhasan AA, Mellough C, Collin J, Gallon R, Hyslop J, Mastrorosa FK,

874   Ehrmann I, Lako M, Elliott DJ, et al. 2018. Analysis of human ES cell differentiation

875   establishes that the dominant isoforms of the lncRNAs RMST and FIRRE are

876   circular. *BMC Genomics* **19**: 276.

877   Jacobsen JC, Gregory GC, Woda JM, Thompson MN, Coser KR, Murthy V, Kohane IS,

878   Gusella JF, Seong IS, MacDonald ME, et al. 2011a. HD CAG-correlated gene

879   expression changes support a simple dominant gain of function. *Human Molecular*

880   *Genetics* **20**: 2846–2860.

881   Jacobsen JC, Gregory GC, Woda JM, Thompson MN, Coser KR, Murthy V, Kohane IS,

882   Gusella JF, Seong IS, MacDonald ME, et al. 2011b. HD CAG-correlated gene

883   expression changes support a simple dominant gain of function. *Hum Mol Genet* **20**:

884   2846–2860.

885   Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE.

886   2013. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*

887   **19**: 141–157.

888   Jia E, Zhou Y, Liu Z, Wang L, Ouyang T, Pan M, Bai Y, Ge Q. 2020. Transcriptomic

889   Profiling of Circular RNA in Different Brain Regions of Parkinson's Disease in a

890   Mouse Model. *Int J Mol Sci* **21**: 3006.

891   Jiang Y-J, Che M-X, Yuan J-Q, Xie Y-Y, Yan X-Z, Hu H-Y. 2011. Interaction with

892   polyglutamine-expanded huntingtin alters cellular distribution and RNA processing of

893   huntingtin yeast two-hybrid protein A (HYPA). *J Biol Chem* **286**: 25236–25245.

894   Kelly S, Greenman C, Cook PR, Papantonis A. 2015. Exon Skipping Is Correlated with Exon

895   Circularization. *J Mol Biol* **427**: 2414–2417.

896   Kerschbamer E, Biagioli M. 2016. Huntington's Disease as Neurodevelopmental Disorder:

897   Altered Chromatin Regulation, Coding, and Non-Coding RNA Transcription. *Front*

898   *Neurosci* **9**. http://journal.frontiersin.org/Article/10.3389/fnins.2015.00509/abstract

899   (Accessed August 19, 2021).

900   Kuhn A, Goldstein DR, Hodges A, Strand AD, Sengstag T, Kooperberg C, Becanovic K,

901   Pouladi MA, Sathasivam K, Cha J-HJ, et al. 2007. Mutant huntingtin's effects on

902   striatal gene expression in mice recapitulate changes observed in human Huntington's

903   disease brain and do not differ with mutant huntingtin length or wild-type huntingtin

904   dosage. *Hum Mol Genet* **16**: 1845–1861.

905   Labbadia J, Morimoto RI. 2013. Huntington's disease: underlying molecular mechanisms and

906   emerging concepts. *Trends Biochem Sci* **38**: 378–385.

907   Langfelder P, Cantle JP, Chatzopoulou D, Wang N, Gao F, Al-Ramahi I, Lu X-H, Ramos

908   EM, El-Zein K, Zhao Y, et al. 2016. Integrated genomics and proteomics to define

909   huntingtin CAG length-dependent networks in HD Mice. *Nat Neurosci* **19**: 623–633.

910   Lau N-C, Kolkman A, van Schaik FMA, Mulder KW, Pijnappel WWMP, Heck AJR,

911   Timmers HTM. 2009. Human Ccr4-Not complexes contain variable deadenylase

912   subunits. *Biochem J* **422**: 443–453.

913   Lee J-M, Pinto RM, Gillis T, St Claire JC, Wheeler VC. 2011. Quantification of age-

914   dependent somatic CAG repeat instability in Hdh CAG knock-in mice reveals

915   different expansion dynamics in striatum and liver. *PLoS One* **6**: e23647.

916   Li Q, Lee J-A, Black DL. 2007. Neuronal regulation of alternative pre-mRNA splicing. *Nat*

917   *Rev Neurosci* **8**: 819–831.

918 Lin L, Park JW, Ramachandran S, Zhang Y, Tseng Y-T, Shen S, Waldvogel HJ, Curtis MA,
919   Faull RLM, Troncoso JC, et al. 2016. Transcriptome sequencing reveals aberrant
920   alternative splicing in Huntington's disease. *Hum Mol Genet* **25**: 3454–3466.

921 Lu Y, Tan L, Wang X. 2019. Circular HDAC9/microRNA-138/Sirtuin-1 Pathway Mediates
922   Synaptic and Amyloid Precursor Protein Processing Deficits in Alzheimer's Disease.
923   *Neurosci Bull* **35**: 877–888.

924 Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD,
925   Gregersen LH, Munschauer M, et al. 2013. Circular RNAs are a large class of animal
926   RNAs with regulatory potency. *Nature* **495**: 333–338.

927 Menalled LB, Sison JD, Dragatsis I, Zeitlin S, Chesselet M-F. 2003. Time course of early
928   motor and neuropathological anomalies in a knock-in mouse model of Huntington's
929   disease with 140 CAG repeats. *J Comp Neurol* **465**: 11–26.

930 Murthy V, Tebaldi T, Yoshida T, Erdin S, Calzonetti T, Vijayvargia R, Tripathi T,
931   Kerschbamer E, Seong IS, Quattrone A, et al. 2019. Hypomorphic mutation of the
932   mouse Huntington's disease gene orthologue. *PLoS Genet* **15**: e1007765.

933 Nagamori I, Yomogida K, Adams PD, Sassone-Corsi P, Nojima H. 2006. Transcription
934   factors, cAMP-responsive element modulator (CREM) and Tisp40, act in concert in
935   postmeiotic transcriptional regulation. *J Biol Chem* **281**: 15073–15081.

936 Neueder A, Landles C, Ghosh R, Howland D, Myers RH, Faull RLM, Tabrizi SJ, Bates GP.
937   2017. The pathogenic exon 1 HTT protein is produced by incomplete splicing in
938   Huntington's disease patients. *Sci Rep* **7**: 1307.

939   Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b

940   are essential for de novo methylation and mammalian development. *Cell* **99**: 247–257.

941   Park JW, Tokheim C, Shen S, Xing Y. 2013. Identifying differential alternative splicing

942   events from RNA sequencing data using RNASeq-MATS. *Methods Mol Biol* **1038**:

943   171–179.

944   Passani LA, Bedford MT, Faber PW, McGinnis KM, Sharp AH, Gusella JF, Vonsattel JP,

945   MacDonald ME. 2000. Huntingtin's WW domain partners in Huntington's disease

946   post-mortem brain fulfill genetic criteria for direct involvement in Huntington's

947   disease pathogenesis. *Hum Mol Genet* **9**: 2175–2182.

948   Perutz MF, Harper PS, Ferguson-Smith MA. 1999. Introduction to Glutamine repeats and

949   neurodegenerative diseases: molecular aspects. A Discussion Meeting held at the

950   Royal Society on 7 and 8 October 1998. *Philosophical Transactions of the Royal*

951   *Society of London Series B: Biological Sciences* **354**: 957–961.

952   Piwecka M, Glažar P, Hernandez-Miranda LR, Memczak S, Wolf SA, Rybak-Wolf A,

953   Filipchyk A, Klironomos F, Cerda Jara CA, Fenske P, et al. 2017. Loss of a

954   mammalian circular RNA locus causes miRNA deregulation and affects brain

955   function. *Science* **357**: eaam8526.

956   Raj B, Blencowe BJ. 2015. Alternative Splicing in the Mammalian Nervous System: Recent

957   Insights into Mechanisms and Functional Roles. *Neuron* **87**: 14–27.

958   Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M,

959   Zheng H, Yang A, et al. 2013. A compendium of RNA-binding motifs for decoding

960   gene regulation. *Nature* **499**: 172–177.

961 Reis SA, Thompson MN, Lee J-M, Fossale E, Kim H-H, Liao JK, Moskowitz MA, Shaw SY,
962   Dong L, Haggarty SJ, et al. 2011. Striatal neurons expressing full-length mutant
963   huntingtin exhibit decreased N-cadherin and altered neuritogenesis. *Hum Mol Genet*
964   **20**: 2344–2355.

965 Rosas HD, Koroshetz WJ, Chen YI, Skeuse C, Vangel M, Cudkowicz ME, Caplan K, Marek
966   K, Seidman LJ, Makris N, et al. 2003. Evidence for more widespread cerebral
967   pathology in early HD: an MRI-based morphometric analysis. *Neurology* **60**: 1615–
968   1620.

969 Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. 2011. A ceRNA hypothesis: the Rosetta
970   Stone of a hidden RNA language? *Cell* **146**: 353–358.

971 Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. 2012. Circular RNAs are the
972   predominant transcript isoform from hundreds of human genes in diverse cell types.
973   *PLoS One* **7**: e30733.

974 Sathasivam K, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, Smith DL,
975   Faull RLM, Roos RAC, Howland D, et al. 2013. Aberrant splicing of HTT generates
976   the pathogenic exon 1 protein in Huntington disease. *PNAS* **110**: 2366–2370.

977 Schilling J, Broemer M, Atanassov I, Duernberger Y, Vorberg I, Dieterich C, Dagane A,
978   Dittmar G, Wanker E, van Roon-Mom W, et al. 2019. Deregulated Splicing Is a
979   Major Mechanism of RNA-Induced Toxicity in Huntington's Disease. *J Mol Biol*
980   **431**: 1869–1877.

981 Schwartz S, Meshorer E, Ast G. 2009. Chromatin organization marks exon-intron structure.
982   *Nat Struct Mol Biol* **16**: 990–995.

983   Seong IS, Ivanova E, Lee J-M, Choo YS, Fossale E, Anderson M, Gusella JF, Laramie JM,

984   Myers RH, Lesort M, et al. 2005. HD CAG repeat implicates a dominant property of

985   huntingtin in mitochondrial energy metabolism. *Hum Mol Genet* **14**: 2871–2880.

986   Seong IS, Woda JM, Song J-J, Lloret A, Abeyrathne PD, Woo CJ, Gregory G, Lee J-M,

987   Wheeler VC, Walz T, et al. 2010. Huntingtin facilitates polycomb repressive complex

988   2. *Hum Mol Genet* **19**: 573–583.

989   Shen S, Park JW, Huang J, Dittmar KA, Lu Z, Zhou Q, Carstens RP, Xing Y. 2012. MATS: a

990   Bayesian framework for flexible detection of differential alternative splicing from

991   RNA-Seq data. *Nucleic Acids Res* **40**: e61.

992   Shen S, Park JW, Lu Z, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. 2014. rMATS: robust

993   and flexible detection of differential alternative splicing from replicate RNA-Seq data.

994   *Proc Natl Acad Sci U S A* **111**: E5593-5601.

995   Simon JM, Hacker KE, Singh D, Brannon AR, Parker JS, Weiser M, Ho TH, Kuan P-F,

996   Jonasch E, Furey TS, et al. 2014. Variation in chromatin accessibility in human

997   kidney cancer links H3K36 methyltransferase loss with widespread RNA processing

998   defects. *Genome Res* **24**: 241–250.

999   Spies N, Nielsen CB, Padgett RA, Burge CB. 2009. Biased chromatin signatures around

1000   polyadenylation sites and exons. *Mol Cell* **36**: 245–254.

1001   Szabo L, Morey R, Palpant NJ, Wang PL, Afari N, Jiang C, Parast MM, Murry CE, Laurent

1002   LC, Salzman J. 2015. Statistically based splicing detection reveals neural enrichment

1003   and tissue-specific induction of circular RNA during human fetal development.

1004   *Genome Biol* **16**: 126.

1005 Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å,
1006   Kampf C, Sjöstedt E, Asplund A, et al. 2015. Proteomics. Tissue-based map of the
1007   human proteome. *Science* **347**: 1260419.

1008 Ule J, Blencowe BJ. 2019. Alternative Splicing Regulatory Networks: Functions,
1009   Mechanisms, and Evolution. *Mol Cell* **76**: 329–345.

1010 Vashishtha M, Ng CW, Yildirim F, Gipson TA, Kratter IH, Bodai L, Song W, Lau A,
1011   Labadorf A, Vogel-Ciernia A, et al. 2013. Targeting H3K4 trimethylation in
1012   Huntington disease. *Proc Natl Acad Sci U S A* **110**: E3027-3036.

1013 Venø MT, Hansen TB, Venø ST, Clausen BH, Grebing M, Finsen B, Holm IE, Kjems J.
1014   2015a. Spatio-temporal regulation of circular RNA expression during porcine
1015   embryonic brain development. *Genome Biology* **16**: 245.

1016 Venø MT, Hansen TB, Venø ST, Clausen BH, Grebing M, Finsen B, Holm IE, Kjems J.
1017   2015b. Spatio-temporal regulation of circular RNA expression during porcine
1018   embryonic brain development. *Genome Biol* **16**: 245.

1019 Vonsattel JP, Myers RH, Stevens TJ, Ferrante RJ, Bird ED, Richardson EP. 1985.
1020   Neuropathological classification of Huntington's disease. *J Neuropathol Exp Neurol*
1021   **44**: 559–577.

1022 Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic
1023   RNP machine. *Cell* **136**: 701–718.

1024 Wang D, Chen Z, Zhuang X, Luo J, Chen T, Xi Q, Zhang Y, Sun J. 2020. Identification of
1025   circRNA-Associated-ceRNA Networks Involved in Milk Fat Metabolism under Heat
1026   Stress. *Int J Mol Sci* **21**: 4162.

1027   Weiss K, Antoniou A, Schratt G. 2015. Non-coding mechanisms of local mRNA translation
1028   in neuronal dendrites. *Eur J Cell Biol* **94**: 363–367.

1029   Weyn-Vanhentenryck SM, Feng H, Ustianenko D, Duffié R, Yan Q, Jacko M, Martinez JC,
1030   Goodwin M, Zhang X, Hengst U, et al. 2018. Precise temporal regulation of
1031   alternative splicing during neural development. *Nat Commun* **9**: 2189.

1032   Wheeler VC, Auerbach W, White JK, Srinidhi J, Auerbach A, Ryan A, Duyao MP, Vrbanac
1033   V, Weaver M, Gusella JF, et al. 1999. Length-dependent gametic CAG repeat
1034   instability in the Huntington's disease knock-in mouse. *Hum Mol Genet* **8**: 115–122.

1035   White JK, Auerbach W, Duyao MP, Vonsattel JP, Gusella JF, Joyner AL, MacDonald ME.
1036   1997. Huntingtin is required for neurogenesis and is not impaired by the Huntington's
1037   disease CAG expansion. *Nat Genet* **17**: 404–410.

1038   Wilusz JE. 2018. A 360° view of circular RNAs: From biogenesis to functions. *Wiley*
1039   *Interdiscip Rev RNA* **9**: e1478.

1040   Wong-Riley MTT, Besharse JC. 2012. The kinesin superfamily protein KIF17: one protein
1041   with many functions. *BioMolecular Concepts* **3**: 267–282.

1042   You X, Vlatkovic I, Babic A, Will T, Epstein I, Tushev G, Akbalik G, Wang M, Glock C,
1043   Quedenau C, et al. 2015. Neural circular RNAs are derived from synaptic genes and
1044   regulated by development and plasticity. *Nat Neurosci* **18**: 603–610.

1045   Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing
1046   biological themes among gene clusters. *OMICS* **16**: 284–287.

1047   Zhang Y, Zhao Y, Liu Y, Wang M, Yu W, Zhang L. 2020. Exploring the regulatory roles of
1048   circular RNAs in Alzheimer's disease. *Transl Neurodegener* **9**: 35.

1049   Zheng S, Black DL. 2013. Alternative pre-mRNA splicing in neurons: growing up and
1050   extending its reach. *Trends Genet* **29**: 442–448.

1051   Zhu K, Lei P-J, Ju L-G, Wang X, Huang K, Yang B, Shao C, Zhu Y, Wei G, Fu X-D, et al.
1052   2017. SPOP-containing complex regulates SETD2 stability and H3K36me3-coupled
1053   alternative splicing. *Nucleic Acids Res* **45**: 92–105.

1054   Zhu Y, Zhao P, Sun L, Lu Y, Zhu W, Zhang J, Xiang C, Mao Y, Chen Q, Zhang F. 2021.
1055   Overexpression of circRNA SNRK targets miR-103-3p to reduce apoptosis and
1056   promote cardiac repair through GSK3β/β-catenin pathway in rats with myocardial
1057   infarction. *Cell Death Discov* **7**: 1–16.

1058   1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on
1059   Huntington's disease chromosomes. The Huntington's Disease Collaborative
1060   Research Group. *Cell* **72**: 971–983.

1061

1062

1063

1064

1065