

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Fan1 and Huntingtin Uninterrupted CAG Repeats in Huntington's Disease

**Permalink**

<https://escholarship.org/uc/item/4r752080>

**Author**

Richman, Jeffrey Bryan

**Publication Date**

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Fan1 and Huntingtin Uninterrupted CAG Repeats in Huntington's Disease

A dissertation submitted in partial satisfaction of the requirements  
for the degree of Doctor of Philosophy in Neuroscience

by

Jeffrey Bryan Richman

2020



## ABSTRACT OF THE DISSERTATION

### Fan1 and Huntingtin Uninterrupted CAG Repeats in Huntington's Disease

by

Jeffrey Bryan Richman

Doctor of Philosophy in Neuroscience

University of California, Los Angeles, 2020

Professor Xiangdong William Yang, Chair

Huntington's disease (HD) is the most common monogenic, neurodegenerative disease. Despite identifying the causal mutation over 20 years ago, there are no disease-modifying treatments currently available for HD patients. A recent genome wide association study has identified several potential genetic modifiers of HD, including FAN1, but validation and further mechanistic investigation is needed. To determine whether Fan1 is a modifier of HD and to elucidate the mechanisms involving Fan1 that hasten or delay HD onset, I crossed a Fan1 knockdown (Fan1-KD) mouse model with a knock-in model of HD with 140 CAG repeats. I found that Fan1-KD caused minimal effect on gene transcription, but a significant enrichment of oligodendrocyte and microglia genes was observed. Fan1-KD led to increased mutant huntingtin (mHtt) aggregates in the striatum of HD mice. Lastly, Fan1-KD led to increased somatic mHtt CAG repeat instability in the liver at 6m, but not 12m, but showed minimal impact on somatic CAG repeat instability in the brain of HD mice. Our findings suggest that Fan1 may act as a modifier in HD by protecting oligodendrocytes from mHtt-induced toxicity and preventing a non-cell autonomous exacerbation of striatal pathology. I also helped characterize

a novel full-length human huntingtin genomic mouse model, known as BAC-CAG, with roughly 120 uninterrupted CAG repeats. BAC-CAG mice show robust behavioral phenotypes, age-dependent striatal transcriptionopathy, nuclear accumulation of mHTT and age-dependent increase in somatic mHTT CAG repeat instability. We also demonstrate that indices of somatic mHTT CAG repeat instability in the cortex and striatum correlate to activity/sleep-related behavioral deficits in BAC-CAG mice. Together, my results provide novel mechanistic insights into recently identified modifiers of HD age of onset.

The dissertation of Jeffrey Bryan Richman is approved.

Christopher S. Colwell

Steve Hortvath

Douglas L. Black

Xiangdong William Yang, Committee Chair

University of California, Los Angeles

2020

*This dissertation is dedicated to my wife, Tina,  
my parents, Norman and Lydia,  
my grandparents, Lydia, Robert, and Esther,  
And my brothers, Andrew and Brandon, for their love and support.*

## TABLE OF CONTENTS

<b>Abstract</b> .....	i
<b>Dedication</b> .....	v
<b>Table of Content</b> .....	vi
<b>List of Figures and Tables</b> .....	ix
<b>Acknowledgements</b> .....	xi
<b>Vita</b> .....	xiv
<b>Chapter 1: Introduction</b> .....	1
1.1.1 Huntington’s Disease.....	1
1.1.2 Cellular pathology of Huntington’s disease.....	2
1.2.1 Modifiers of HD age of onset.....	5
1.2.2 DNA Mismatch Repair in Huntington’s Disease.....	11
1.2.3 The role of FAN1 in Huntington’s Disease.....	13
1.3.1 Mouse models used to study Huntington’s disease.....	16
1.3.2 N-Terminal HTT fragment mouse models of Huntington’s Disease.....	17
1.3.3 Full-length Huntingtin transgenic mouse models of Huntington’s disease.....	18
1.3.4 Full-length huntingtin knock-in mouse models of Huntington’s disease.....	20
<b>Chapter 2: Fan1’s Protective role in Huntington’s disease</b> .....	22
2.1 Significance of studying <i>Fan1</i> in Huntington’s Disease.....	22
2.2 Fan1-KD causes age-related dysregulation of HD- and glial-related transcripts in wildtype and HD mice.....	23
2.3. Using a novel GFP-Fan1 KI mice to study cellular and subcellular expression of Fan1 wildtype.....	30

2.4 Preliminary studies suggest Fan1-KD exacerbates neuropathology in HD mice.....	32
2.5 Fan1-KD causes minimal behavioral changes in normal and HD mice at 12m.....	34
2.6 Age- and Tissue-specific effects of Fan1-KD on somatic mHTT CAG repeat instability.....	35
2.7 Discussion.....	38
<b>Chapter 3: Uninterrupted Huntingtin CAG-Repeat Length Predicts Striatal Nuclear Inclusions and Transcriptionopathy In Vivo.....</b>	<b>44</b>
3.1 Significance of developing novel human HTT BAC Transgenic mouse model with uninterrupted CAG repeats.....	45
3.2 Generation and initial characterizations of BAC-CAG mice.....	46
3.3 Neuropathology in BAC-CAG mice.....	46
3.4 Striatal transcriptionopathy in BAC-CAG mice is highly concordant with transcriptional changes seen in allelic series KI mice Striatum.....	50
3.5 Uninterrupted mHTT/mHtt CAG-repeat length predicts transcriptional dysregulation in striatum of full-length huntingtin models of HD.....	52
3.6 BAC-CAG mice exhibit activity and sleep deficits.....	54
3.7 BAC-CAG mice exhibit age-dependent somatic instability of mHTT CAG repeats, which correlates with activity/sleep deficits.....	56
3.8 Discussion.....	58
<b>Chapter 4: Conclusions and future directions.....</b>	<b>61</b>
4.1 Fan1's role in the brain.....	61
4.2 4.2 Uninterrupted mHTT CAG repeats modify HD-related behavior deficits.....	62
4.3 Novel mouse models to aid investigations of Fan1's role in HD.....	63
4.4 Concluding thoughts.....	65
<b>Appendix: Mouse models to aid in investigations of Fan1.....</b>	<b>66</b>

A.1 Generation and validation of Fan1-R510H and GFP-Fan1 mice.....	66
A.2 Generation of BAC-FAN1 mice.....	71
A.3 Fan1-KD does not impact the expression of alternatively spliced mHTT transcript with intron1 retention.....	74
<b>Materials and Methods</b> .....	76
<b>References</b> .....	84

## LIST OF FIGURES

Figure 1: Schematic of human FAN1 Protein Domains.....	15
Figure 2: Generation of Fan1-KD mice crossed to Q140 HD mouse model.....	24
Figure 3: Fan1-KD causes minimal transcriptional change in 6m striatum.....	25
Figure 4: Fan1-KD causes minimal transcriptional change in 12m striatum.....	26
Figure 5: Fan1-KD causes minimal transcriptional change in 6m cortex.....	27
Figure 6: Fan1-KD causes minimal transcriptional change in 12m cortex.....	28
Figure 7: Fan1-KD causes upregulation of striatal microglia genes and downregulation of cortical oligodendrocyte genes in the 12m mouse brain.....	29
Figure 8: GFP-Fan1 is expressed in cortical and striatal neurons and glia.....	31
Figure 9: Fan1-KD mice exhibit increased mHTT aggregates and decreased synaptic markers.....	33
Figure 10: Fan1-KD is correlated with reduced anxiety, but not motor deficits.....	34
Figure 11: Fan1-KD causes tissue- and age-dependent changes in somatic mHTT CAG repeat instability.....	36
Figure 12: Generation and initial characterizations of BAC-CAG mice.....	47
Figure 13: Age-dependent mHTT nuclear accumulation and aggregation in BAC-CAG mice...	48
Figure 14: Age-dependent cortical and striatal transcriptionopathy in BAC-CAG mice.....	51
Figure 15: Striatal transcriptional changes better predicted by uninterrupted CAG repeat length than PolyQ length.....	53
Figure 16: Locomotor, sleep, and psychiatric phenotypes in BAC-CAG mice.....	55
Figure 17: Age-dependent somatic CAG repeat instability in BAC-CAG mice.....	57
Figure 18: Somatic mHTT CAG repeat instability correlate with activity and sleep deficits.....	58
Figure 19: Generation of Fan1-R510H KI mice.....	67
Figure 20: Generation of GFP-Fan1 KI mice.....	68

Figure 21: Validation of Fan1-R510H and GFP-Fan1 KI mice.....	69
Figure 22: Modifying human FAN1 BAC.....	72
Figure 23: Generation of BAC-FAN1 mice.....	73
Figure 24: Expression of alternatively spliced mHTT transcript with intron1 retention.....	75

## ACKNOWLEDGMENTS

I would like to acknowledge individuals who have supported me throughout my PhD training. Foremost, I would like to thank my mentor, William Yang, who entrusted me with an important and highly impactful project for our field. I am very proud of the work I've accomplished while in William's lab and grateful for the patience he's shown during my training, which allowed me the opportunity to grow as a scientist. William has always encouraged me to approach my work with ambition and rigor, and I greatly appreciate the experience I've gained in his lab. I would also like to thank my committee members, both past and present, including Doug Black, Steve Horvath, Chris Colwell, and Giovanni Coppola. Steve has always been supportive of my progress, both in our JSC projects and of my thesis. I also want to thank Chris Colwell for being a late addition to my committee and for providing advice on how I should organize the topics in my thesis.

I would also like to thank several post-docs in William's lab who played a critical role in my growth as a scientist. When I first joined the lab, Matt Veldman taught me everything from cloning to managing mouse colonies. His mentorship was instrumental to my early development as scientist and his support as a friend encouraged me to push forward in the face of many failures. Nan Wang also played a crucial role as a mentor throughout my PhD. Nan is a great friend and I want to thank her for putting up with me constantly asking for her advice on my thesis projects. Her guidance helped me learn how to manage mouse colonies and I am proud of what we've been able to accomplish as part of the JSC project, which she supervises. I would also like to thank Xiaofeng Gu, who's advice I sought on almost all cloning-related projects and who taught me how to engineer BACs. His tremendous work ethic and positive attitude served as a source of inspiration. Daniel Lee also served as one of my scientific mentors. He always

encouraged me to consider the larger scientific question when framing how I approach my project. I also want to thank Daniel for his advice on my written thesis. Shasha Zhang played an important role in advising my mouse fibroblast experiments. I also want to thank her for putting up with my cluttered desk as my lab bench neighbor. Lastly, I want to thank Peter Langfelder, who not only analyzes all of our labs RNAseq data, but also taught me a lot about statistics and data analysis. I greatly appreciated the time and effort Peter invested to introduce me to R.

Next, I would like to thank the members of the lab, who's friendship made working in the lab a joy. Lalini is great friend and is always willing to go the extra mile to help someone. She helped proofread nearly every word of my thesis, to which I am extremely grateful. Discussing science with Lalini everyday made the lab environment very collegial. Plus, we would share tips on where to find free food on campus. I also want to thank Matt Stricos and Ray Vaca, who helped the success of my thesis project and who I shared lunch with on most days. Surprisingly, I also feel the urge to thank Chris Park. Despite his sarcasm, or maybe because of it, I consider him a good friend, and I've relied on his advice on where to find the best food in LA on numerous occasions. I'm also inspired by the surgical level of precision he brings to his work. I also want to thank Jason Zhang. The opportunity to mentor Jason came at an important time in my development as a grad student. Although I will not claim to be a good scientific mentor, I am glad Jason was able to find a home working as Chris' slave. Lastly, I want to thank Linna Deng, who will be taking over the Fan1 project. I greatly respect Linna's persistence and ambition and trust she will do a great job finishing up this exciting project.

I want to thank William's previous grad students, Jeff Cattle and Anthony Daggett. Jeff taught me how to run qRT-PCR and extract RNA, which played a big role in enabling my ability to contribute to the JSC. Both Anthony and Jeff have shared salient advice on how to navigate my journey through the Yang lab and grad school in general, to which I am very grateful.

I would also like to thank the CHDI and the ARCS foundation, who's financial support has enabled me to complete my graduate studies.

I want to thank all my friends, including Daniel Lopez, Daniel Tracy, Daniel Kinsley, Nick Fuller, Chris Jones, Kevin Lanzo, Aizen Malki, Chris Cartaya and Steve Liu. Their friendship and support during my PhD was necessary to keep me (relatively) sane. I also want to thank the NSIDP PhD students in my incoming cohort, including Ian Heimbuch, Cassie Meyer, Catherine Schweppe, Martin Safrin, Karen Cheng, Lauren MacIntyre, Kathleen Wang, Wendy Herbst, and Kadidia Adula. Their friendship was essential to helping me feel settled when I first move to Los Angeles and sharing stories about graduate school has been cathartic.

Lastly, I would like to thank my family for being a solid foundation which has allowed me to grow. My grandparents, Lydia Montejo and Esther and Bob Richman, have helped raise me, with constant love and support, since I was a child. My parents, Lydia and Norman Richman, have always been there to pick up when I called nearly every single day. I know it was difficult for my mom to let her eldest son move across the country. But without both of their support, it would not have been possible for me to complete this PhD. I would also like to thank my brothers, Andrew and Brandon Richman, who are my best friends. Finally, and most importantly, I would like to thank my wife, Tina Wang. Her support, friendship, mentorship, guidance, and unconditional love have been essential for me to grow and enabled me to find the strength to complete my graduate studies.

## VITA

### EDUCATION

Bachelor of Science  
Neuroscience  
Florida Atlantic University, Wilkes Honors College  
Jupiter, Florida 2014

Graduate Student Researcher  
Neuroscience  
University of California, Los Angeles  
Los Angeles, California 2014 - 2020

### HONORS AND AWARDS

ARCS scholars fellowship 2016 - 2019

### PRESENTATIONS

Xiaofeng Gu, Jeff Richman, Peter Langfelder, Nan Wang, Lucia Yang, X. William Yang (2018).  
Development and Characterization of a Novel BAC Transgenic Mouse Model of Huntington's  
Disease with Elongated Pure CAG Repeats. Society for Neuroscience annual meeting, San  
Diego. Nov 3-7, 2018.

Jeff Richman, Peter Langfelder, Matthew Stricos, Fuying Gao, Nan Wang, Matthew B. Veldman,  
Giovanni Coppola and X. William Yang (2018). Transcriptomic Analysis of Fan1 Genetic  
Reduction in the Q140 Huntingtin Knockin Mouse Model. Hereditary Disease Foundation. Aug  
7-11, 2018.

### PUBLICATIONS

Beena Kadakkuzha, Timothy Spicer, Peter Chase, Jeff Richman, Peter Hodder, and  
Sathyanarayanan V. Puthanveetil (2014). High-throughput screening for small molecule  
modulators of motor protein Kinesin. Assay Drug Dev Technol, 12(8), 470-480.

Valerio Rizzo, Jeff Richman and Sathyanarayanan V. Puthanveetil (2014). Dissecting  
mechanisms of brain aging by studying the intrinsic excitability of neurons. Front Aging  
Neurosci, 6, 337.

## CHAPTER 1:

### Introduction

#### 1.1.1. Huntington's disease

Huntington disease (HD) is the most common monogenic, neurodegenerative disorder in the developed world, with roughly 1 in 7,300 individuals in Western populations affected (Evans et al., 2013; Fisher and Hayden, 2014; Morrison et al., 2011). The disease-causing mutation is an expansion of a CAG repeat in exon1 of the huntingtin (*HTT*) gene (Group, 1993). Most unaffected individuals possess *HTT* alleles with <20 CAG repeat lengths (Kay et al., 2016), while individuals with >40 huntingtin CAG repeats will develop the disease and people with 35-40 CAG repeats show variable penetrance (McNeil et al., 1997; Orr and Zoghbi, 2007). Roughly 5-8% of diagnosed patients are de novo cases (Squitieri et al., 1994; Warby et al., 2009), which inherit a mutant huntingtin (*mHTT*) allele from a parent without known diagnosis of the disease. In de novo cases, a parent typically possesses a huntingtin allele of 27-35 CAG repeats, which expands to disease-causing lengths during meiosis (Semaka et al., 2013a; Semaka et al., 2013b).

An inverse relationship between CAG repeat length and age of onset of HD symptoms exists, in which longer CAG repeats lead to earlier onset of symptoms (Bates et al., 2015). Typically, onset begins around 45 years of age, although individuals with >60 CAG repeats can develop symptoms before the age of 20, known as Juvenile HD (Andresen et al., 2007). HD patients develop a progressive movement disorder. Early in the disease, patients commonly develop "chorea," an involuntary, dance-like movement (Bates et al., 2015). Parkinsonian symptoms, such as dystonia, bradykinesia, and rigidity, are often seen in late-onset and juvenile HD patients and are progressive with disease (Reilmann, 2019). Cognitive impairments are also

seen in HD patients, including problems with attention, mental flexibility, planning, emotion recognition, and cognitive slowing (Stout et al., 2016; Stout et al., 2011). Psychiatric symptoms such as depression, irritability, and apathy are also seen in HD patients, but prevalence of these symptoms is more variable (Thompson et al., 2012). Lastly, HD patients often experience sleep disturbances, with many patients having difficulty sleeping at night and staying awake during the day (Morton, 2013).

### **1.1.2 Cellular pathology of Huntington's disease**

Massive degeneration of neurons in the basal ganglia and, to a lesser extent, cortex is a hallmark of HD pathology (Mattsson et al., 1974). The caudate and putamen appear to be especially vulnerable, with neuronal loss preceding that of cortex (Dumas et al., 2012; Paulsen et al., 2010). Late in disease other brain regions, including the cerebellum, also show evidence of degeneration, although not to the same degree as the caudate and cortex (Jeste et al., 1984; Kremer et al., 1991; Schmitz et al., 1999; Vonsattel and DiFiglia, 1998).

Before overt neuronal loss manifests in the HD brain, deficits in cortico-striatal communication emerge. Morphometric and histological studies in postmortem HD brains show myelin breakdown and loss of white matter volume (Bartzokis et al., 2007; de la Monte et al., 1988; Mann et al., 1993). Decreased trophic support from cortex to striatum (Plotkin et al., 2014; Zuccato et al., 2001; Zuccato et al., 2005) and aberrant cortico-striatal signaling likely contributes to the shift from vulnerable neuronal populations to the outright death of susceptible cells (Cepeda et al., 2007; Raymond et al., 2011). As well, loss of connectivity might underlie many of the early clinical decrements (Tabrizi et al., 2009).

To distinguish the importance of mHTT-dependent pathology in the cortex and striatum on overall HD phenotypes, Wang et al. (Wang et al., 2014a) used a full-length human mHTT genomic transgenic mouse model of HD, which has two loxp sites flanking *HTT* exon1

(BACHD), to selectively reduce mHTT expression in either the cortex , striatum, or both regions. Interestingly, selectively reducing mHTT expression in the cortex, but not the striatum, lead to a partial rescue of psychiatric-like behavioral symptoms and motor deficits in BACHD mice, and the greatest rescue of deficits occurred when mHTT was removed from both brain regions. This study highlights the important role of mHTT-related cortical pathology in striatal pathology and HD overall. It also laid out a framework where other labs could use BACHD mice to study the role of mHTT expression in specific cell-types and brain regions to study HD pathogenesis.

The GABAergic medium-spiny neurons (MSNs) in the striatum are the most vulnerable cell-types to neurodegeneration in HD (Vonsattel et al., 1985). The D2 dopamine receptor expressing MSNs in the “indirect” pathway appear to be affected earlier than D1 expressing MSNs (Aizman et al., 2000; Richfield et al., 1995; Sapp et al., 1995; Surmeier et al., 1996) in the “direct” pathway. This may be due to intrinsic cell-autonomous differences (Heiman et al., 2008; Lobo et al., 2006) or cell-nonautonomous difference between these cell types, such as differential cortical innervation of these subpopulations of MSNs (Lei et al., 2004). The second most vulnerable neurons in HD are the glutamatergic pyramidal neurons in middle and deep cortical layers (Cudkowicz and Kowall, 1990; Hedreen et al., 1991; MacDonald and Halliday, 2002; Wall et al., 2013)

While dysfunction and degeneration of the vulnerable striatal and cortical neurons tend to receive the most attention in HD research, glial cells of the brain also develop a number of mHTT-induced pathologies (Benraiss et al., 2016). HD post-mortem brains show astrocytosis and the number of GFAP-positive reactive astrocytes is correlated with the gradient of striatal neurodegeneration (Lee et al., 2013; Vonsattel et al., 1985). Astrocytes play a critical role in normal brain functions, including glutamate uptake (Bezzi et al., 2004), modulating synapses (Fiacco and McCarthy, 2004), neuronal metabolism (Brown and Ransom, 2007; Pfrieger and Ungerer, 2011), and neuroinflammatory responses (Carson et al., 2006; Glass et al., 2010).

Interestingly, reactive astrocytes are not commonly seen in the HD cortex, even in regions with neuronal loss and microglia activation (Palpagama et al., 2019; Sotrel et al., 1991). BACHD mice with mHTT depleted in astrocytes show a rescue of motor, synaptic marker, and electrophysiological deficits (Wood et al., 2019), demonstrating the crucial role astrocytes play in the disease.

Postmortem HD brains are also marked by accumulation of reactive microglia (Tai et al., 2007). Microglia are the resident immune cells of the brain (Frost and Schafer, 2016). At baseline, microglia are in a surveilling state but following an inflammatory stimulus, microglia become activated to respond and take part in an inflammatory response. Microglial density and activation can be detected up to 15 years prior to symptom onset and increased microglial activation is correlated with greater disease severity and striatal neuron loss (Pavese et al., 2006). Mutant huntingtin expression in microglia has been shown to dysregulate transcription and contribute to the disease in mouse models (Crotti et al., 2014). Interestingly, depletion of mHTT in microglia shows no significant rescue of behavioral performance in BACHD mice (Petkau et al., 2019). These conflicting findings, in part due to the lack of full neurodegenerative and neuroinflammatory pathology of these models compared to the patients, suggests further studies are needed to delineate the precise pathogenic and therapeutic roles of microglia in HD.

Another cell-type which is affected early in HD is oligodendrocytes (OLs), which play an important role in ensheathing the axons of projection neurons and cortico-striatal connectivity. White matter abnormalities appear very early in the disease course, many years before neurological onset in patients (Paulsen et al., 2008; Rosas et al., 2018; Tabrizi et al., 2009) and before any neuronal loss in animal models of HD (Teo et al., 2016; Xiang et al., 2011). Deficits in OLs or their precursors can lead to axonal pathology and neurodegeneration (Nave, 2010). BACHD mice with mHTT depleted in OLs show rescued behavioral deficits (rotarod, open field, forced swim), highlighting their important contribution to HD neuropathogenesis (Ferrari Bardile

et al., 2019). As well, (Osipovitch et al., 2019) showed that human glial progenitor cells (hGPCs) from HD patient embryonic stem cells showed downregulation in OL-specific transcription factors (e.g. OLIG2, MYRF) and showed deficits in glial differentiation causing reduced myelination in the mouse brain. Thus, it is likely that mHTT-induced OL-pathology may play an important role in early HD deficits and diverse array of symptoms.

### **1.1.3 Molecular Pathology of Huntington's disease**

Within the cells which are most affected in HD, a myriad of molecular dysfunctions and pathologies are at play. The normal function of HTT is not fully understood, but it is known to play diverse functions (Cattaneo et al., 2005; Saudou and Humbert, 2016), including but not limited to development of the nervous system, transcriptional regulation, vesicular transport of cargos such as brain-derived neurotrophic factor (BDNF), cell adhesion, mitotic spindle orientation, synaptogenesis and neuronal survival (Burrus et al., 2020; Dragatsis et al., 2000; Duyao et al., 1995; Godin et al., 2010; McKinstry et al., 2014; Zeitlin et al., 1995; Zuccato and Cattaneo, 2014). The poly-glutamine (polyQ) expansion in the huntingtin protein is believed to be mostly a toxic gain-of-function mutation, but one cannot fully rule out a role of partial HTT loss-of-function in HD pathogenesis (Landles and Bates, 2004; Ross and Tabrizi, 2011). Mutant huntingtin expression is known disrupt a number of pathways including protein homeostasis and degradation and cause transcriptional dysregulation and mitochondrial dysfunction (Bates et al., 2015; Labbadia and Morimoto, 2013; Ross and Tabrizi, 2011).

A pathological hallmark of HD is the presence of mHTT aggregates in the form of large intranuclear inclusions (Nis) and cytoplasmic aggregates (DiFiglia et al., 1997). These inclusions occur predominantly in the cortical and striatal neurons, but could also be in other cells that have entered mitotic arrest, which suggests that cell division acts to delay the pathogenetic process (Moffitt et al., 2009). Several reports have provided evidence that large mHTT-

containing inclusions are not correlated with cytotoxicity (Kim et al., 1999) and might even be protective (Arrasate et al., 2004). How *mHTT* aggregation impacts HD pathology remains unclear but understanding how different species of aggregates impact cellular physiology may be critical to understanding their role in HD pathogenesis (Arrasate and Finkbeiner, 2012; Ratovitski et al., 2009).

Studies of transcriptional changes in post-mortem HD patient brains (Agus et al., 2019; Hodges et al., 2006; Labadorf et al., 2015) and mouse models of HD (Langfelder et al., 2016) have revealed 1000s of gene transcripts which show an age-dependent dysregulation in the striatum and relatively fewer genes changing in the cortex. The greatest number and magnitude of differentially expressed mRNAs are detected in the caudate nucleus/striatum (Hodges et al., 2006; Langfelder et al., 2016). By analyzing an allelic series of mutant huntingtin knock-in mice with 20, 80, 92, 111, 140, and 175 CAG repeat lengths, our lab was able to identify 13 striatal and 5 cortical modules of coexpressed genes, defined by Weighted Gene Coexpression Network (WGCNA) analysis, that are significantly correlated with CAG length and age (Langfelder et al., 2016). Top striatal modules implicate *mHTT* CAG repeat expansion and age in graded impairment of striatal MSN identity gene expression, dysregulation of cAMP signaling, and increase in cell death and protocadherin gene expression. Interestingly, several modules were enriched with glial genes (M11, M20, and M45), highlighting the role these cells play in HD. As well, some unpublished data from our lab that will be discussed in Chapter 3 shows that transcriptional changes in the striatum of HD mouse models is predicted by length of uninterrupted CAG, but not polyQ length.

Besides the impact that changes in mRNA expression have on altering protein expression and activity, RNAs themselves can cause molecular pathologies as well. For example, an antisense transcript to huntingtin that contains a CUG repeat was identified in human brain tissue (Chung et al., 2011) and is able to repress *HTT* transcription in a reporter

cell line, in a CAG repeat length-dependent manner. In myotonic dystrophy, expanded CUG repeats form RNA foci that can sequester various RNA binding proteins (Conlon et al., 2016; DeJesus-Hernandez et al., 2011; Mankodi et al., 2000; Taneja et al., 1995; Timchenko et al., 2001; Wojciechowska and Krzyzosiak, 2011). This may result in a potential disruption of cellular homeostasis or act as sites of nucleation for protein aggregation (Todd and Paulson, 2010). (Banez-Coronel et al., 2012) showed that mHTT mRNA can be cleaved by Dicer to create small CAG-repeat RNAs which decrease neuronal viability by inhibiting expression of CAG repeat containing RNAs. Lastly, incomplete splicing of mHTT exon 1, leading to the retention of intron1, has been found in full-length *HTT* mouse models of HD as well as HD patient post-mortem brains (Neueder et al., 2017; Sathasivam et al., 2013). Expression of this truncated mHTT mRNA is proportional to CAG-repeat length and that encodes the highly pathogenic mHTT exon 1 protein. Interestingly, the partial mHTT exon 1-intron 1 RNA transcript can also be found in BACHD and YAC128 models, suggesting it is not dependent on uninterrupted CAG repeats.

Oxidative stress and DNA damage have also been shown to contribute to HD molecular pathology in HD patients, animals, and cells (Browne and Beal, 2006). The high oxygen demand of the brain exposes its cell to numerous reactive oxygen species (ROS), which can lead to DNA damage (Cooke et al., 2003). Further oxidative damage can arise as a result of inflammation. DNA damage in neurons is predominantly single stranded and CAG repeats might be especially susceptible to damage such as hydrolytic depurination (i.e. loss of A or G) and cytosine deamination (Lindahl, 1993). "Programmed" DNA strand breakage can also be a source of DNA damage. Topoisomerases, which can introduce temporary single- or double-strand breaks in DNA to regulate supercoiling, are essential for the expression of long genes (e.g. *HTT*) and are particularly important in the brain (King et al., 2013; Zylka et al., 2015).

The DNA damage response involves over 450 genes in humans and subsets of these DNA repair genes are deployed depending on the type of DNA damage, type of cell, and stage

of cell cycle (Pearl et al., 2015). For DNA double strand break repair, dividing progenitors in S/G2 phase use accurate homologous recombination (HR), while mature postmitotic neurons in G0/G1 rely upon less accurate nonhomologous end joining (NHEJ) (Massey and Jones, 2018). Single strand breaks (SSBs), on the other hand, are usually repaired by processes known as base excision repair (BER), nucleotide excision repair (NER) or mismatch repair (MMR). BER is responsible for most oxidative or alkylation DNA damage. In mitotic cells, genomic errors formed during DNA replication are repaired by MMR. These pathways can also be used for non-canonical repair. For example, MMR enzymes have been shown to recognize and repair mis-paired bases in nondividing yeast cells (Rodriguez et al., 2012).

An area of HD research which has gained a lot of interest recently is the instability of *mHTT* CAG repeats. Uninterrupted trinucleotide repeats show a length-dependent instability in both dividing cells, such as during meiosis (Kremer et al., 1995; Ranen et al., 1995), and in post-mitotic cells, which is known as “somatic” instability (Kennedy et al., 2003). In HD mouse models with repeat instability, most tissues (e.g. cerebellum, tail) show relatively stable CAG repeat lengths. However, the liver, striatum, and sometimes the cortex, show robust instability that is biased towards repeat expansion and is progressive with age (Larson et al., 2015; Lee et al., 2011; Lee et al., 2010) (unpublished data from our lab). *mHTT* CAG repeat expansion has also been found in post-mortem human patients’ brains (Kennedy et al., 2003), however the exact role instability plays in HD is not fully understood. HD mouse models with CAA interrupted CAG repeats (both encode glutamine) and lack repeat instability, such as YAC128 (Hodgson et al., 1999; Slow et al., 2003) and BACHD (Gray et al., 2008), still develop neuropathology suggesting the somatic instability is not required for certain aspects of HD pathology manifested in these models. However, somatic instability may still act as a modifier of HD since CAG repeat instability in the post-mortem human cortex is correlated with disease onset in HD patients (Swami et al., 2009).

How tissue-specific CAG instability emerges is unclear but some evidence suggests that differences in stoichiometric expression levels of repair proteins are associated with variable levels of CAG repeat instability (Goula et al., 2009; Goula et al., 2012). Another question being considered by investigators is, how do CAG repeats become unstable in post-mitotic cells? A parsimonious theory is that error prone repair of DNA lesions near CAG repeats may cause contraction or expansion of the repeats (Massey and Jones, 2018; Schmidt and Pearson, 2016). Indeed, tandem repeats can readily adopt noncanonical conformations in DNA, such as slipped strands, hairpin loops, G-quadruplexes and R-loops (Mirkin, 2007; Neil et al., 2017), and processing of lesions at repeats may lead to instability. For example, formation and processing of slipped-DNAs and bidirectional transcription in a human cell model (such as occurring at *HTT*) have been shown to increase CAG repeat instability (Haeusler et al., 2014; Kadyrova et al., 2020; Lin et al., 2010; Lin et al., 2009; Nakamori et al., 2011; Pluciennik et al., 2013; Reddy et al., 2014; Schmidt and Pearson, 2016; Zhang et al., 2012).

### **1.2.1 Modifiers of Huntington's Disease age of onset**

Although the disease-causing mutation of HD was discovered more than 20 years ago (Group, 1993), there are no disease-modifying treatments currently available for HD patients. Recent clinical trials for HTT lowering therapies using anti-sense oligonucleotides (ASOs) or adeno-associated viruses (AAVs) have shown promising pre-clinical and clinical data (Leavitt and Tabrizi, 2020; Tabrizi et al., 2019b). But evidence suggesting HTT plays a role in survival of post-mitotic cells (De Souza and Leavitt, 2015) has led to concerns over the negative effects of lowering wildtype huntingtin expression in the adult brain. This has encouraged research into allele-specific *mHTT* lowering and research into therapeutics which do not target HTT. As well, treatments targeting other genes/proteins can be used synergistically as an adjunctive therapy

with HTT lowering drugs, if shown to be safe. Thus, clinicians and investigators have sought to identify modifiers of HD pathogenesis.

*mHTT* CAG repeat length explains ~60% of individual variation in HD age of onset (Djousse et al., 2003) and the remaining variance is largely heritable. This suggests that genetic modifiers of the disease may exist. To identify novel HD modifiers, the Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (GeM-HD, 2015, 2019) led a genome-wide association (GWA) study, which now includes over 9,000 HD patients. Briefly, the residual age of onset (i.e. the difference in 'predicted' age of onset based on CAG repeat length and 'actual' age of onset) for each patient was calculated. Patients were then genotyped using microarrays and patient haplotypes were imputed using the Haplotype Reference Consortium (HRC; (McCarthy et al., 2016)). Using this approach, GeM-HD was able to identify genetic loci/haplotypes which are significantly associated with residual age of onset of HD motor symptoms. Near the significant genetic loci are a number of genes which have the potential to hasten or delay onset and may represent therapeutic targets for HD.

One of the most significant discoveries of this study was that *HTT* uninterrupted CAG repeat length is a better predictor of HD age of onset than HTT polyglutamine length (GeM-HD, 2019). The vast majority of the Western population carry a canonical CAG repeat region with an uninterrupted CAG repeat followed by a terminal CAACAG sequence (both encoding glutamine). Thus, the encoded polyglutamine (polyQ) repeat length is consistently 2 residues longer than the uninterrupted CAG repeat. Individuals lacking the penultimate CAA have a hastened age of onset than what would be predicted by their polyQ length. Similarly, individuals with a duplication of the CAACAG have a delayed onset, compared to their predicted onset based on polyQ length. This suggests that there may be some molecular pathologies which depend on uninterrupted CAG repeats, separate from encoding glutamine, that affect timing of HD onset.

The GWA studies also highlighted the importance of DNA repair genes in modifying HD age of onset. Many of the genetic loci identified contain genes associated with DNA repair (*PMS1*, *MLH1*, *MSH3*, *PMS2*, *FAN1*, *LIG1*), while others (*DHFR*, *TCERG1*, *RRM2B*, *UBR5*, *CCDC82*, *SYT9*) appear to be involve other cellular processes including mitochondrial regulation.

### 1.2.2 DNA Mismatch Repair in Huntington's Disease

By assigning significant single nucleotide polymorphisms (SNPs) to the nearest gene, the GeM-HD consortium was able to carry out a pathway enrichment analysis, which highlighted the mismatch repair (MMR) pathway as an important modifier of HD onset. Taken together with the rate-determining property of the uninterrupted CAG repeats, this suggests that MMR genes may influence HD onset through a DNA-level effect on somatic expansion of the *HTT* CAG repeat. Indeed, previous studies have shown that MMR proteins affect trinucleotide repeat instability (Dragileva et al., 2009; Manley et al., 1999; Massey and Jones, 2018; Pinto et al., 2013) and somatic CAG expansion in the human cortex is a significantly correlated with HD age of onset (Swami et al., 2009).

In *Escherichia coli*, MutS and MutL are required to initiate MMR and act as homo dimers (Modrich and Lahue, 1996). MutS is involved in DNA mismatch recognition and MutL physically interacts with MutS to enhance mismatch recognition and act as a molecular matchmaker to facilitates assembly of a functional MMR complex (Sancar and Hearst, 1993). MMR is a highly conserved biological pathway and several human MMR genes have been identified based on their homology to MMR genes in *E. coli*. In humans, there are two heterodimer MutS complexes: MutS $\alpha$  (MSH2-MSH6) and MutS $\beta$  (MSH2-MSH3) (Drummond et al., 1995; Fishel et al., 1993; Leach et al., 1993; Palombo et al., 1995; Reenan and Kolodner, 1992). MutS $\alpha$  recognizes base-base mismatches and mispairs of 1 or 2 nucleotides, while MutS $\beta$

preferentially recognizes larger mispairs. There are also three MutL heterodimer complexes: MutL $\alpha$  (MLH1-PMS2), MutL $\beta$  (MLH1-PMS1) and MutL $\gamma$  (MLH1-MLH3) (Bronner et al., 1994; Nicolaides et al., 1994; Papadopoulos et al., 1994). Interestingly, MutL $\alpha$ , possesses endonuclease activity that is required for mammalian MMR (Kadyrov et al., 2006).

The first study to show that MMR plays a role in *mHTT* repeats instability was (Manley et al., 1999), who showed that Msh2 KO prevents CAG repeat instability. This was later expanded upon by (Kovalenko et al., 2012) who showed that Msh2 KO in MSNs was sufficient to prevent the majority of striatal somatic CAG repeat instability. Interestingly, Msh2 KO in MSNs also delayed nuclear accumulation and aggregation of *mHTT*. A role for Msh complex proteins in CAG repeat instability was further elucidated by (Dragileva et al., 2009), who showed that Msh3 KO, but not Msh6 KO, prevented CAG repeat instability in the striatum of Q111 knock-in mouse model of HD. Interestingly, Msh3 KO also showed decreased nuclear mutant huntingtin staining in Q111 mice, Taken together, these studies suggest that the MutS $\beta$  (Msh2/Msh3) may play a more important role in somatic CAG instability, compared to MutS $\alpha$  (Msh2/Msh6), that striatal CAG repeat instability is mostly caused by instability within MSNs, and that instability plays a role in enhancing/modifying HD striatal pathology.

A role for MutL complex proteins (MLH1 and MLH3) in *mHTT* CAG repeat instability was later established in a study by (Pinto et al., 2013). This study found differences in somatic instability of striatal CAG repeats in mice from different inbred backgrounds. By genotyping these mice using a panel of 117 SNPs and correlating mouse genotype with somatic instability, they were able to identify a single quantitative trait locus (QTL) which was near the *Mlh1* gene. They then showed that *Mlh1* KO prevents striatal and liver CAG repeat instability in Q111 mice. They also showed that *Mlh3* KO (a binding partner of *Mlh1*) prevented CAG repeat instability, suggesting MutL $\gamma$  plays a critical role in somatic CAG repeat instability. KO of either *Mlh1* or *Mlh3* led to decreased *mHTT* nuclear accumulation, suggesting a rescue or delay of HD

pathology. A parallel study looking at the effects of MutL complex proteins in myotonic dystrophy type 1 found that Pms2 KO eliminated somatic instability in kidney (Gomes-Pereira et al., 2004), suggesting MutL $\alpha$  (Mlh1/Pms2) is also involved in somatic repeat instability.

While several of studies have shown that MutS and MutL complex proteins are involved in somatic instability of mHTT CAG repeats, a mechanistic model of how this occurs in post-mitotic neurons has not been established. A review by (Schmidt and Pearson, 2016) summarizes the present literature on how these proteins may work in concert to cause trinucleotide repeat instability. MutS is required for repair of isolated, short slip-outs (1–3 excess repeat units) (Panigrahi et al., 2010), while slip-outs with >3 excess repeat units are processed independently of MMR (Panigrahi et al., 2005; Zhang et al., 2012). It has been proposed that MutS $\beta$ , but not MutS $\alpha$ , promotes trinucleotide repeat expansion through a non-canonic MMR pathway (Gomes-Pereira et al., 2004; Keogh et al., 2017; Pinto et al., 2013; Tome et al., 2009), and a role for slip-out size-dependent recruitment of MutL complexes has also been seen, with both MutL $\alpha$  (Mlh1/Pms2) and MutL $\gamma$  (Mlh1/Mlh3) being required for processing of short CAG.CTG slip-outs (Panigrahi et al., 2012; Pluciennik et al., 2013).

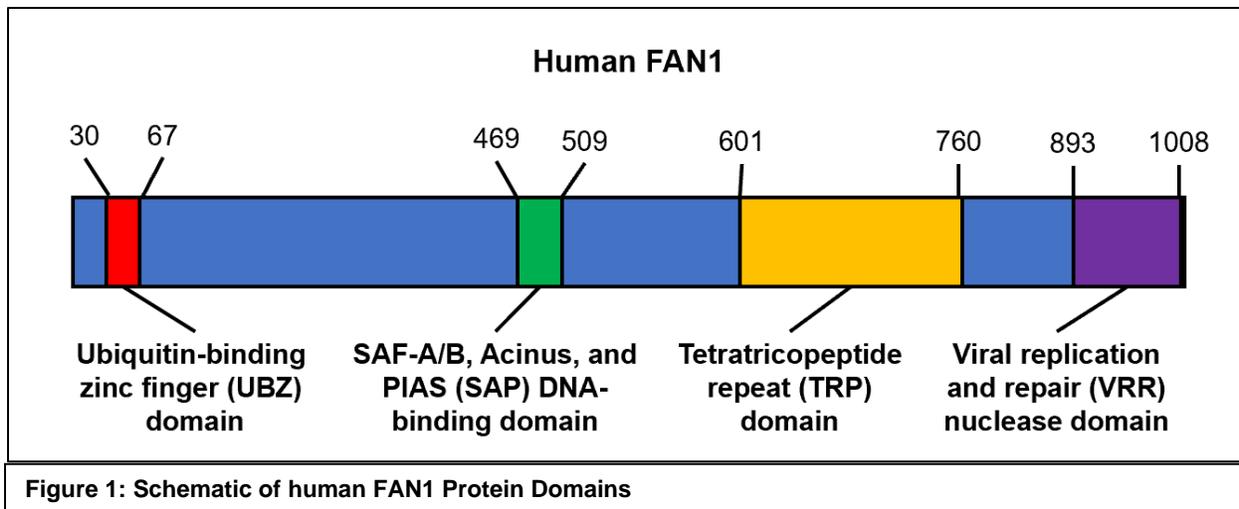
### **1.2.3 The role of *FAN1* in Huntington's Disease**

The genetic locus that was most significantly associated with modifying HD age of onset was on Chr15, near a gene called *FAN1* (GeM-HD, 2019). In this region, two rare, onset hastening haplotypes (15am1 and 15am3) and two common, onset-delaying haplotypes (15am2 and 15am4) were identified. Interestingly, tag SNPs for 15am2 and 15am4 haplotypes are significantly associated with cis-eQTLs for increased *FAN1* mRNA expression in the cortex, while 15am1 and 15am3 are tagged by SNP alleles that cause missense mutations (R507H and R377W, respectively), which are predicted to be deleterious by the Sorting Intolerant From Tolerant (SIFT) algorithm. SIFT predicts whether a SNP will be deleterious based off

conservation of the amino acid affected across species, position at which the change occurred, and the type of amino acid change (Ng and Henikoff, 2003). Taken together, this suggests that *FAN1* may play a protective role in HD.

*FAN1* was first identified in a genetic screen for genes involved in the Fanconi Anemia (FA) DNA repair pathway (Smogorzewska et al., 2010). The FA pathway has evolved to repair DNA interstrand crosslinks (ICLs), which are life-threatening lesions that create a bidirectional polymerase blockade. Human patients homozygous recessive for *FAN1* mutations develop a form of chronic kidney disease known as Karyomegalic Interstitial Nephritis (KIN) (Zhou et al., 2012). These patients develop renal pathologies including widespread fibrosis and enlarged nuclei known as karyomegaly. As well, cells from individuals with *FAN1* mutations exhibit sensitivity to the ICL-inducing agent, mitomycin C (MMC). *Fan1* KO mice also show ICL repair deficits and karyomegaly in kidney and liver (Lachaud et al., 2016; Thongthip et al., 2016). SNPs in *Fan1* have also been found to modify the onset of other polyglutamine diseases (Bettencourt et al., 2016). *FAN1* resides in the 15q13.3 chromosomal locus where microdeletions have been significantly associated with autism spectrum disorder (ASD), mental retardation and psychiatric disorders (Ben-Shachar et al., 2009) and certain *FAN1* variants are significantly associated with ASD and schizophrenia (Ionita-Laza et al., 2014).

*FAN1* possesses 5' flap endonuclease and 5'-3' exonuclease activities (Kratz et al., 2010). The *FAN1* protein contains an N-terminal ubiquitin binding zinc finger (UBZ) domain, a SAF-A/B, Acinus, and PIAS (SAP) DNA-binding domain, a TPR domain, and a C-terminal virus-type replication repair nuclease (NUC) domain (Gwon et al., 2014; Kratz et al., 2010; Pennell et al., 2014; Smogorzewska et al., 2010) (Figure 1). *FAN1* has two modes of recruitment to ICLs, one is very fast and depends on the SAP domain and the other is a slower accumulation that depends on the UBZ domain (Thongthip et al., 2016). Recent crystallographic data revealed that the SAP domain interacts extensively with the DNA (Gwon et al., 2014; Wang et al., 2014b;



Zhang and Walter, 2014) and recognizes the sugar-phosphate backbone of the duplex region downstream of the 5' flap (Zhao et al., 2014). The SAP domain is also involved in FAN1 dimerization, in which the TPR and NUC domains of the first FAN1 molecule interacts with the SAP domain of a second FAN1.

The mechanism of how *FAN1* acts as a modifier of HD age of onset is not yet understood, but a parsimonious explanation may be that *FAN1* protects against somatic *mHTT* CAG repeat instability and that loss of *FAN1* leads to CAG repeat expansion in vulnerable striatal MSNs and cortical pyramidal neurons, which results in an earlier onset of HD pathology. Evidence for this hypothetical model comes from a study of Fragile X mice in which *Fan1* KO lead to an increase in CGG repeats in mutant *Fmr1* mice (Zhao and Usdin, 2018). Another study found that level of *Fan1* expression determines the rate of CAG repeat expansion in vitro (Goold et al., 2019). Interestingly, their results suggest *Fan1*'s nuclease activity is not required for protection against CAG repeat expansion. As well, while *Fan1* was found to interact with *Htt* CAG repeats, it was not significantly enriched at CAG repeats throughout the genome. This suggests that *Fan1*'s role is not specific to CAG repeat repair. Finally, unpublished study by Vanessa Wheeler's group showed viral-mediated, CRISPR/Cas9-based deletion of murine *Fan1*

in an Htt knockin model (Q111) leads to increased mHtt CAG repeat instability in the liver and striatum (CHDI HD Therapeutic Conference, 2019).

Another clue towards Fan1's potential role in HD may come from its expression in single-cell RNAseq datasets from human and mouse brains. Allen brain atlas has compiled a well-curated bioinformatic tool to browse their human and mouse single-cell RNAseq data (Tasic et al., 2018). From the human dataset of multiple cortical areas and mouse dataset of cortex and hippocampus, Fan1 expression is seen in both inhibitory and excitatory neurons, including upper- and lower-layer cortical projection neurons. Interestingly, Fan1 is also expressed in cortical glia including astrocytes, microglia, and OLs. A single-cell RNAseq study in the mouse striatum, from Steve Quake's lab found Fan1 expression in striatal astrocytes, OLs, ependymal cells, and D1- and D2-MSNs (Gokce et al., 2016b). Lastly, single-cell RNAseq study from Steve McCarroll's lab looking at 9 different regions of the adult mouse brain found that striatal and cortical neurons and glia express Fan1, with OLs having the highest level of expression in both tissues (Saunders et al., 2018).

### **1.3.1 Mouse models used to study Huntington's disease**

An important question to consider when designing an HD study is which model is best suited to test my hypothesis? Should I use a human cell-line or a mouse model? What level of disease severity do need? Thankfully, in the HD field there are a wide variety of models for investigators to choose from. A common method of characterizing a disease model is based on face, construct, and predictive validity (Nestler and Hyman, 2010). Face validity asks how well the model recapitulates human patient molecular and behavioral phenotypes. Construct validity addresses how similar the model's disease-causing mutation is to the mutation seen in patients. Lastly, predictive validity is a measure of how well treatments that are efficacious in the model translate to treat human patients.

Mouse models of HD can be grouped into three broad categories: (1) N-terminal *HTT* fragment models, (2) full-length human *HTT* genomic models, and (3) *HTT* knock-in models. Each type of model offers specific advantages and limitations, and proper consideration is necessary to choose the appropriate model for a given study (Crook and Housman, 2011; William Yang and Gray, 2011). For an excellent review of HD models and best practices, please see “A Field Guide to Working with Mouse Models of Huntington’s Disease” which is a joint publication by CHDI, Jackson Laboratories, and PsychoGenics, Inc. (Menalled et al., 2014).

### **1.3.2 N-Terminal *HTT* fragment mouse models of Huntington’s Disease**

N-terminal *HTT* fragment lines carry a small portion of the 5’ end of the human *HTT* gene, including exon 1 which encodes the polyQ repeat region. The most well characterized N-terminal fragment models are the R6/1 and R6/2 lines, which were the first HD mouse models to be developed (Mangiarini et al., 1996). Many investigators choose to study these mouse models for their robust and progressive neurological abnormalities, brain atrophy, and their accelerated disease course including premature death, relative to other genetically engineered lines (Menalled et al., 2009). As well, these models were useful in demonstrating that an N-terminal mutant *HTT* fragment is sufficient to elicit disease-like phenotypes, including mHTT aggregates. N-terminal *HTT* fragment models also show somatic CAG repeat instability in the cortex, striatum, and liver (Larson et al., 2015).

While these models show strong face validity in terms of their advanced pathology and behavior impairments, they are limited by a lack of overt cell loss, are prone to seizures, display more widespread neuropathology than that seen in HD patients and premature death (Menalled et al., 2009). Another limitation of the N-terminal fragment models is that they have relatively low construct validity. While there is some evidence that impaired mHTT mRNA splicing can lead to expression of mHTT-exon1 fragments, human patients also express the full-length *HTT* protein.

Thus, HD-pathology driven by the full-length mHTT protein cannot be studied in these models. As well, the N-terminal fragment mice are engineered via pronuclear injection, with each transgene integrating into the genome randomly with multiple copies of the transgene. Such transgene insertion and copy number related phenotypes could also be a confounding factor.

### **1.3.3 Full-length Huntingtin transgenic mouse models of Huntington's disease**

Full-length HTT genomic mouse models use bacterial artificial chromosomes (BACs) or yeast artificial chromosomes (YACs) to insert large pieces of human genomic DNA into the mouse genome. The main advantage of this approach is that it allows the entire human huntingtin locus (including promoter, enhancers, intronic regions, and other flanking genomic regulatory regions) to be integrated along with the intact exons and introns (together encompass the open reading frame or ORF) of the gene, which leads to more physiologically relevant expression pattern and level (Yang and Gong, 2005). Other advantages of this method are that the larger size of the transgene makes it more likely to integrate with low copy numbers (1-10 transgene copies) and are more insulated from the positional effects that are often seen in small cDNA transgenic models.

The first full-length *HTT* genomic mouse models were the YAC18, YAC46, and YAC72 models (Hodgson et al., 1999). The YAC128 transgenic mouse line is the most characterized YAC model. It expresses multiple copies of full-length human *HTT* with 125 glutamine repeats encoded by a tract that is composed primarily of CAG codons with 9 interspersed CAA codons (Pouladi et al., 2012), at roughly the same level (~75%) as the endogenous mouse *HTT* (Slow et al., 2003). The BACHD mouse model, developed by our laboratory, contains full-length human HTT with a floxed exon 1 containing 97 codons of a mixed CAG/CAA repeat tract, and expresses human HTT at levels 2 to 3 times that of the endogenous mouse Htt (Gray et al., 2008).

In some respects, these models offer better face validity compared to the fragment models. Genomic mouse models show a slower progressing disease course, with phenotypes developing gradually over many months and showing relatively normal survival (Gray et al., 2008; Menalled et al., 2009; Van Raamsdonk et al., 2005). As well, they show progressive motor, cognitive and psychiatric-like behavioral deficits, and selective cortical and striatal atrophy including a reduction in synaptic markers (synaptophysin, actn2, PSD95) at 12m (Gray et al., 2008; Menalled et al., 2012; Slow et al., 2003; Wang et al., 2014a). BACHD mice also show a robust sleep disruption phenotype with progressively declining amplitude of sleep rhythms with age (Kudo et al., 2011).

Compared to the N-terminal HTT fragment models, the human *HTT* genomic mice have higher construct validity because they contain the full-length human *HTT* gene in the context of its endogenous genomic regulatory elements. This feature has made human genomic mice the model of choice for translational research into HTT lowering therapies (Carroll et al., 2011; Kordasiewicz et al., 2012; Monteys et al., 2017; Skotte et al., 2014; Wild and Tabrizi, 2017).

However, human genomic models also have several limitations, including minimal transcriptional dysregulation, delayed mHTT aggregate formation and mostly lack of nuclear inclusions, and an age-related weight gain (Gray et al., 2008; Van Raamsdonk et al., 2007) due to overexpression of full-length human HTT (Pouladi et al., 2010). The latter is not seen in patients and can partially affect behavioral endpoints (Gray et al., 2008; Menalled et al., 2009). Another limitation of the human genomic models' construct validity are the CAA codons that interrupt the CAG repeat region, preventing the study of germline and somatic mHTT CAG repeat instability and the potential toxicity of CAG and CUG repeat containing RNAs. Human genomic HD models also offer improved predictive validity compared to the fragment models since they can be used to test therapies targeting the human HTT gene, RNA, or protein, but cannot be used to study therapies targeting somatic repeat instability.

#### 1.3.4 Full-length huntingtin knock-in mouse models of Huntington's disease

The knock-in (KI) HD models were generated by homologous recombination in mouse embryonic stem cells to introduce an expanded CAG repeat into mouse *Htt*. Some KI models had expanded CAG repeats inserted into mouse *Htt* exon1 (Lin et al., 2001; Sathasivam et al., 2013), while others were generated by humanizing *HTT* exon1, creating a chimeric mouse-human *HTT* (Levine et al., 1999; Menalled et al., 2003; Shelbourne et al., 1999; Wheeler et al., 1999).

These models offer high construct validity, with physiological *Htt* expression. The HD knock-in mice genocopy human patients, with one wildtype and one CAG-expanded *Htt* allele (Aziz et al., 2010; Cattaneo et al., 2005; Djousse et al., 2003; Farrer et al., 1993; Lee et al., 2012; Wexler et al., 1987). However, there are several limitations to their construct validity. The size of the repeats most studied (e.g., 111, 150, 175, 200) are well above the normal range for adult- or even juvenile-onset HD. As well, the mostly murine *Htt* context differs from the human *HTT* in >300 amino acids, suggesting that the full-length transgenic models are more useful for studying the human *HTT* DNA, RNA and protein contexts.

In considering the face validity of the KI models, they display a robust age- and tissue-dependent transcriptionopathy (Langfelder et al., 2016) and somatic instability (Lee et al., 2011; Lee et al., 2010), unlike the existing full-length human genomic models. The Q140 and Q175 KI models show mild striatal atrophy, wide-spread mHTT aggregates (Heikkinen et al., 2012; Menalled et al., 2012; Menalled et al., 2003) and heterozygous Q140 mice show an age-dependent reduction in synaptic marker VGLUT1 (Deng et al., 2013) and robust striatal neuronal electrophysiological deficits (Cepeda et al., 2003; Cummings et al., 2010; Hickey et al., 2012; Raymond et al., 2011). However, one limitation of these models is that they have milder and

less progressive behavioral abnormalities than the fragment and genomic models (Brooks et al., 2012a, b; Brooks et al., 2006; Heng et al., 2010; Hickey et al., 2008; Lin et al., 2001; Menalled et al., 2012).

When considering the use of KI models of HD studies, most HD investigators have focused on using high Q length (92 or greater) knock-in mice since lower Q length mice do not appear to demonstrate robust measurable behavioral phenotypes. Unlike human genomic models, KI mice can be used to study CAG repeat instability, sense and antisense HTT mRNA, noncanonical repeat associated non-ATG (RAN) translation (Banez-Coronel et al., 2015) and molecular, cellular and physiological deficits in the striatal and cortical neurons in HD mice (Liu et al., 2016; Menalled, 2005; Peng et al., 2016). In terms of predictive validity, KI models allow for testing of mHTT CAG repeat targeting therapeutics (Zeitler et al., 2019), but offer limited utility for testing HTT lowering therapies targeting human-specific sequences at DNA, RNA or protein levels (Tabrizi et al., 2019a).

## CHAPTER 2:

### *Fan1*'s Protective role in Huntington's disease

#### 2.1 Significance of studying *Fan1* in Huntington's disease

There remains a need for novel therapeutics to modify the disease course of HD. Currently, the most exciting novel HD therapies are trying to lower expression of HTT (e.g. Zinc-finger nuclease, ASO, etc.), with many clinical trials showing promising preclinical (Kordasiewicz et al., 2012; Southwell et al., 2018) and Phase1a/2b data (Tabrizi et al., 2019b). Another class of promising therapeutics may come from the genetic modifiers identified by the GeM-HD GWAS (GeM-HD, 2019). Interestingly, many of the loci were adjacent to DNA repair genes and the most significant locus identified was in Chr15 near a DNA repair nuclease, *FAN1*.

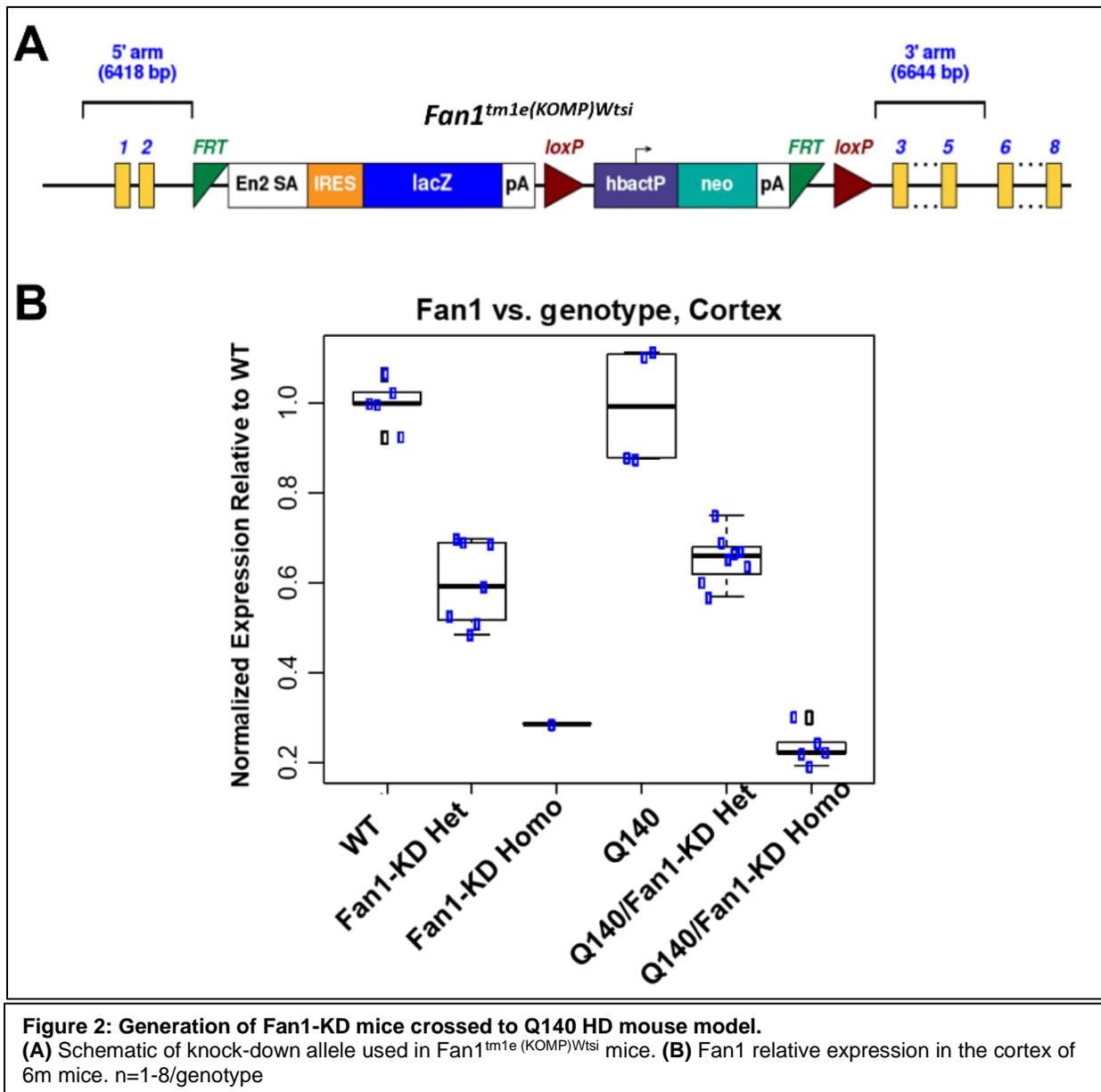
Several lines of evidence suggest *FAN1* is the genetic modifier of HD at the Chr15 locus. (1) The most significant SNP in the GeM-HD GWAS (rs150393409) is within the *FAN1* ORF and causes a missense mutation (R507H), which is predicted by SIFT to be a loss of function mutation (GeM-HD, 2019). (2) Transcriptome-wide association studies (TWAS) revealed that increased *FAN1* expression, but not expression of any nearby genes, is significantly associated with delayed onset and slower progression of HD (GeM-HD, 2019; Goold et al., 2019). (3) *Fan1* has been shown to be protective against somatic instability of CGG repeats in a mouse model of Fragile X (Zhao and Usdin, 2018) and in *mHTT* CAG repeats in-vitro (Goold et al., 2019). (4) *FAN1* protein directly interacts with MMR proteins including, *MLH1* and *PMS2* (Goold et al., 2019; Rikitake et al., 2020; Smogorzewska et al., 2010), which are were also highlighted as potential modifiers of HD age of onset and affect CAG repeat instability. (5) One of the

haplotypes identified at the Chr15 locus, which is associated with delayed onset of HD, is tagged by SNPs which are correlated with increased *FAN1* expression in the cortex.

To test the hypothesis that *Fan1* is a genetic modifier of mHtt-induced toxicities in vivo and to explore the underlying mechanisms, we analyzed the effects of *Fan1* genetic reduction in one of the allelic series KI HD mouse models with 140 CAG repeats (i.e. Q140) (Menalled et al., 2003). We analyzed the transcriptome changes, neuropathology, behavioral deficits and somatic CAG repeat instability of these mice at 6m and 12m of age. Our results thus far suggest that *Fan1* plays a protective role in HD and represents a novel therapeutic target to delay onset in HD patients.

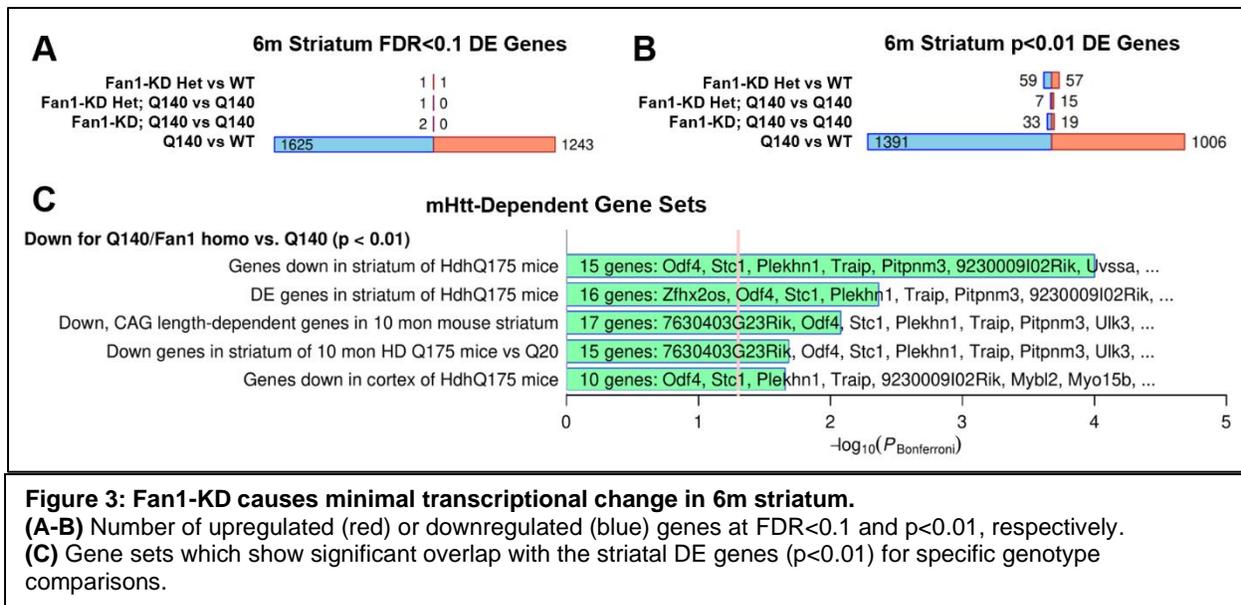
## **2.2 *Fan1*-KD causes age-related dysregulation of HD- and glial-related transcripts in wildtype and HD mice**

We obtained a *Fan1* knock-down (KD) mouse model from KOMP (*Fan1*<sup>tm1e(KOMP)Wtsi</sup>), which uses a gene-trap cassette targeted to *Fan1* intron 2 (Figure 2A). The cassette uses an En2 splice acceptor to interrupt the *Fan1* mRNA with an IRES, LacZ, and polyA sequence after *Fan1* exon2. I crossed heterozygous *Fan1*-KD (*Fan1*-KD het) mice with heterozygous Q140 KI HD (Q140) mice to obtain *Fan1*-KD het;Q140 mice. *Fan1*-KD het; Q10 mice were then crossed with *Fan1*-KD het mice to obtain wildtype, *Fan1*-KD het, homozygous *Fan1*-KD (*Fan1*-KD homo), Q140, *Fan1*-KD het;Q140, and *Fan1*-KD homo;Q140 mice. All Q140 mice analyzed were heterozygous and will be referred to as Q140. Homozygous *Fan1*-KD mice will be referred to as *Fan1*-KD thereafter. I measured the effect of the *Fan1*-KD allele by measuring *Fan1* mRNA expression level in 6m old mice using RNAseq (Figure 2B) and qRT-PCR (not shown). I found a gene dose-dependent reduction in mRNA expression, with *Fan1*-KD Het mice showing



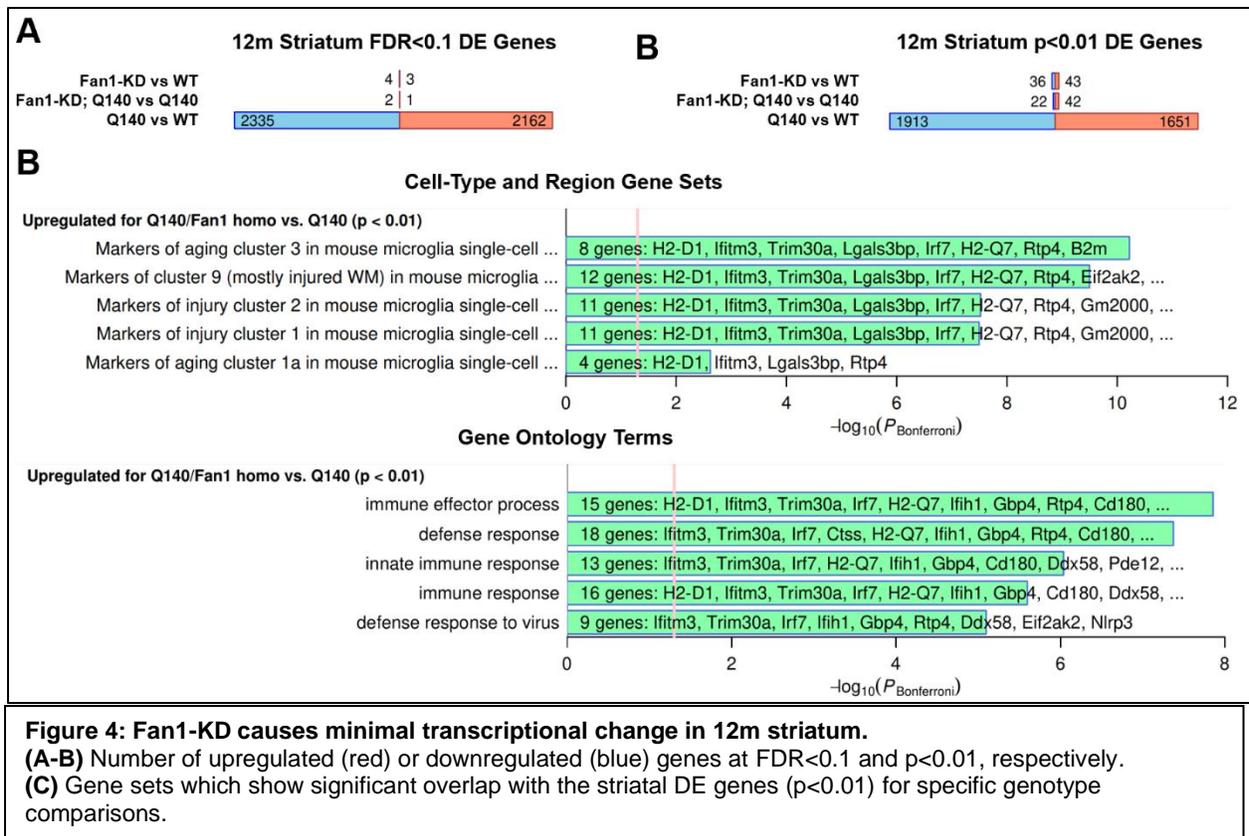
a ~35% reduction and Fan1-KD mice showing ~70% reduction of Fan1 transcripts in both wildtype and Q140 backgrounds.

To gain a better understanding of role of Fan1 in the brain in both normal and HD context, I performed RNAseq on brain tissues which are most affected in HD (i.e. cortical and striatal tissue) at two ages. Peter Langfelder helped me perform all differential expression and gene enrichment analyses, as well as generate figures. At 6m old, Fan1 reduction (both heterozygous and homozygous) caused very few genes to be differentially expressed (DE) at



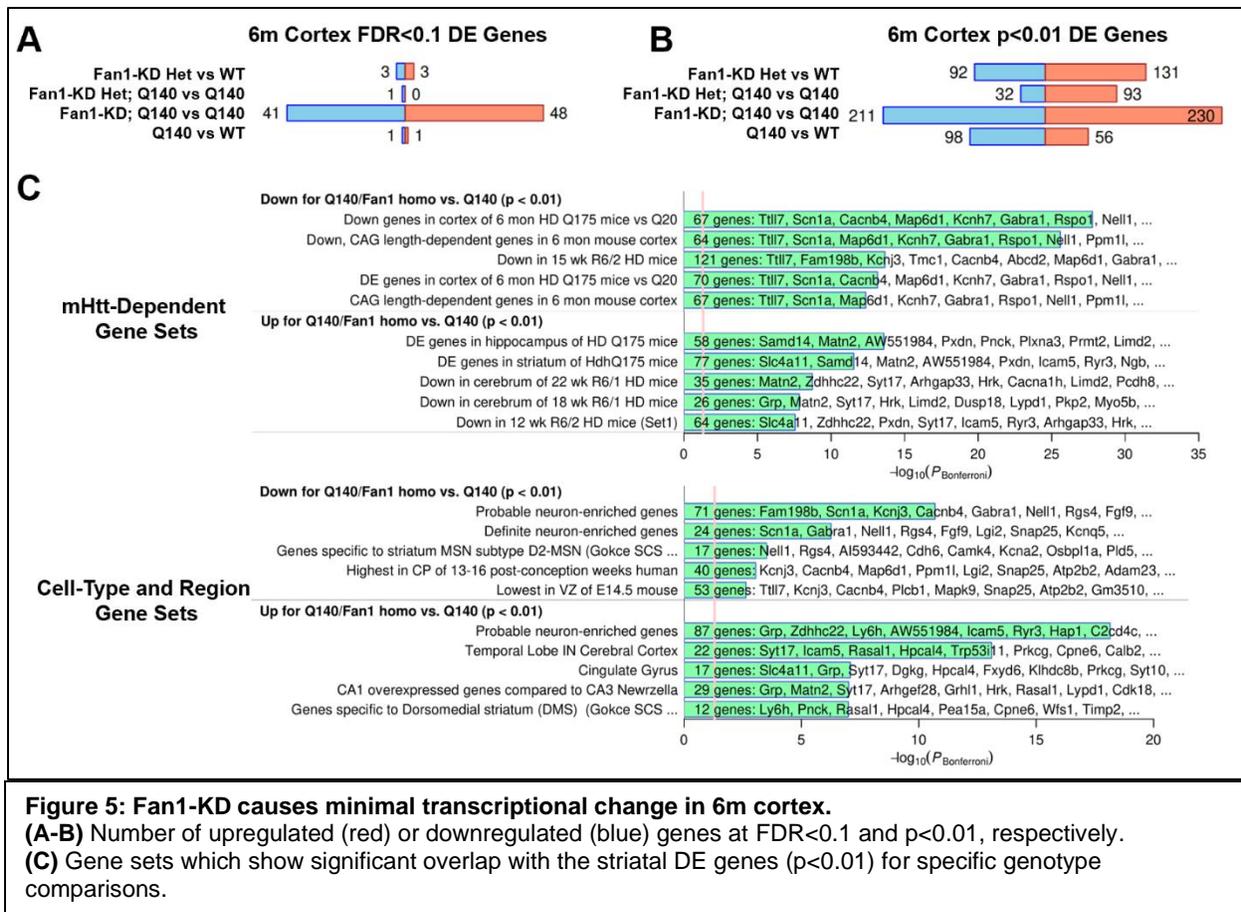
FDR<0.01 in the striatum (Figure 3A). To gain a better understanding of how these transcriptional changes might impact the striatum, we performed a gene set enrichment analysis using anRichtman (Langfelder et al., 2016) on the significant DE gene (p<0.01; Figure 3B) among the different genotype comparisons. Previous studies in our lab have shown that using a lower DE gene significance threshold for the purpose of enrichment analyses can yield find meaningful gene enrichment terms. We found a significant enrichment of mHTT-dependent genes among the Fan1-KD; Q140 vs Q140 DE genes (Figure 3C). Specifically, there was overlap between the striatal genes down-regulated in Fan1-KD; Q140 vs Q140 and in Q175 mouse striatum vs WT mice, suggesting Fan1-KD has an exacerbating effect, in terms of further downregulating these HD-related transcripts.

Because Fan1-KD het mice showed minimal effect on the HD transcriptionopathy at 6m, and because Fan1-KD mice only show a reduction of 70% on transcript expression, we did not include Fan1-KD Het and Fan1-KD het;Q140 mice our 12m RNAseq study. Fan1-KD caused very few genes to significantly change at FDR<0.01 in 12m mice (Figure 4A). However, there were 79 and 64 DE gene (p<0.01) between Fan1-KD vs WT and Fan1-KD; Q140 vs Q140, respectively, in the 12m striatum (Figure 4B). Gene enrichment analysis showed an



upregulation of microglia markers (including in the context of white matter injury) and immune-related genes in the Fan1-KD; Q140 vs Q140 comparison (Figure 4C). This suggests Fan1-KD may exacerbate an immune or inflammation response in the HD mouse brain.

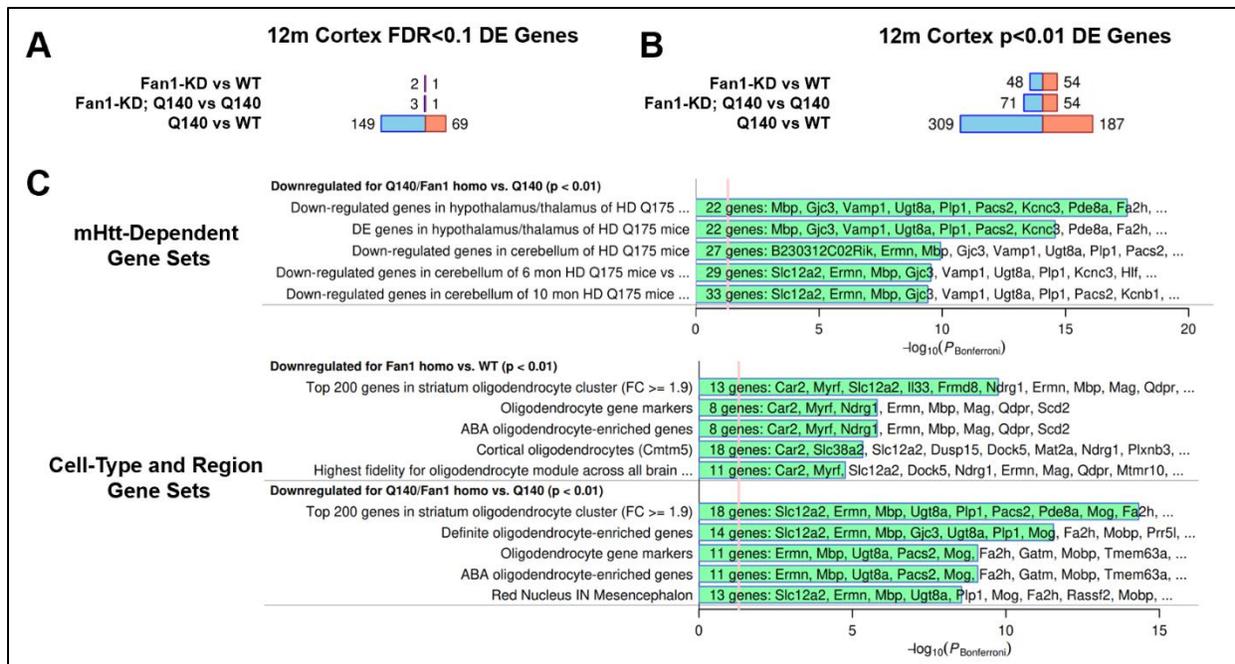
Based on results from the GeM-HD GWAS (GeM-HD, 2019), Fan1 expression in the cortex may be important to its role in modifying HD age of onset. To gain a better understanding of Fan1's role in the cortex, we analyzed the effect of Fan1 reduction on the cortical transcriptome in WT and Q140 mice. Similar to the striatum, very few genes were dysregulated by Fan-KD het in the WT and Q140 mice at FDR<0.1. But surprisingly, there were ~90 DE genes in 6m cortical Fan1-KD;Q140 vs Q140 comparison (Figure 5A). Lowering the significance threshold to p<0.01, we found ~ 200 DE genes between Fan1-KD vs WT, 120 DE genes between Fan1-KD Het; Q140 vs Q140, and over 400 DE genes between Fan1-KD; Q140 vs Q140 in the cortex at 6m (Figure 5B). An enrichment analysis found significant overlap between



*mHTT*-dependent gene sets and neuron-enriched gene sets among the 6m Fan1-KD; Q140 vs Q140 cortex DE genes (Figure 5C).

In the 12m cortex, minimal Fan1-KD related DE genes at FDR<0.1 (Figure 6A). However, we found ~100 DE genes (p<0.01) between Fan1-KD vs WT and Fan1-KD;Q140 vs Q140 (Figure 6B). Interestingly, enrichment analysis of the 12m cortex DE genes found significant overlap between *mHTT*-dependent gene sets and OL-enriched genes (Figure 6C) and DE genes in both Fan1-KD vs WT and Fan1-KD; Q140 vs Q140 comparison. This suggests that Fan1-KD by itself is sufficient to dysregulated certain *mHTT*-related genes and that Fan1-KD may play a significant role in OLs in both WT and Q140 mice at 12m.

Two glial cell types were highlighted in the 12m cortex and striatum RNAseq data, striatal microglia and cortical OLs. To get a better understanding of how these glial-related



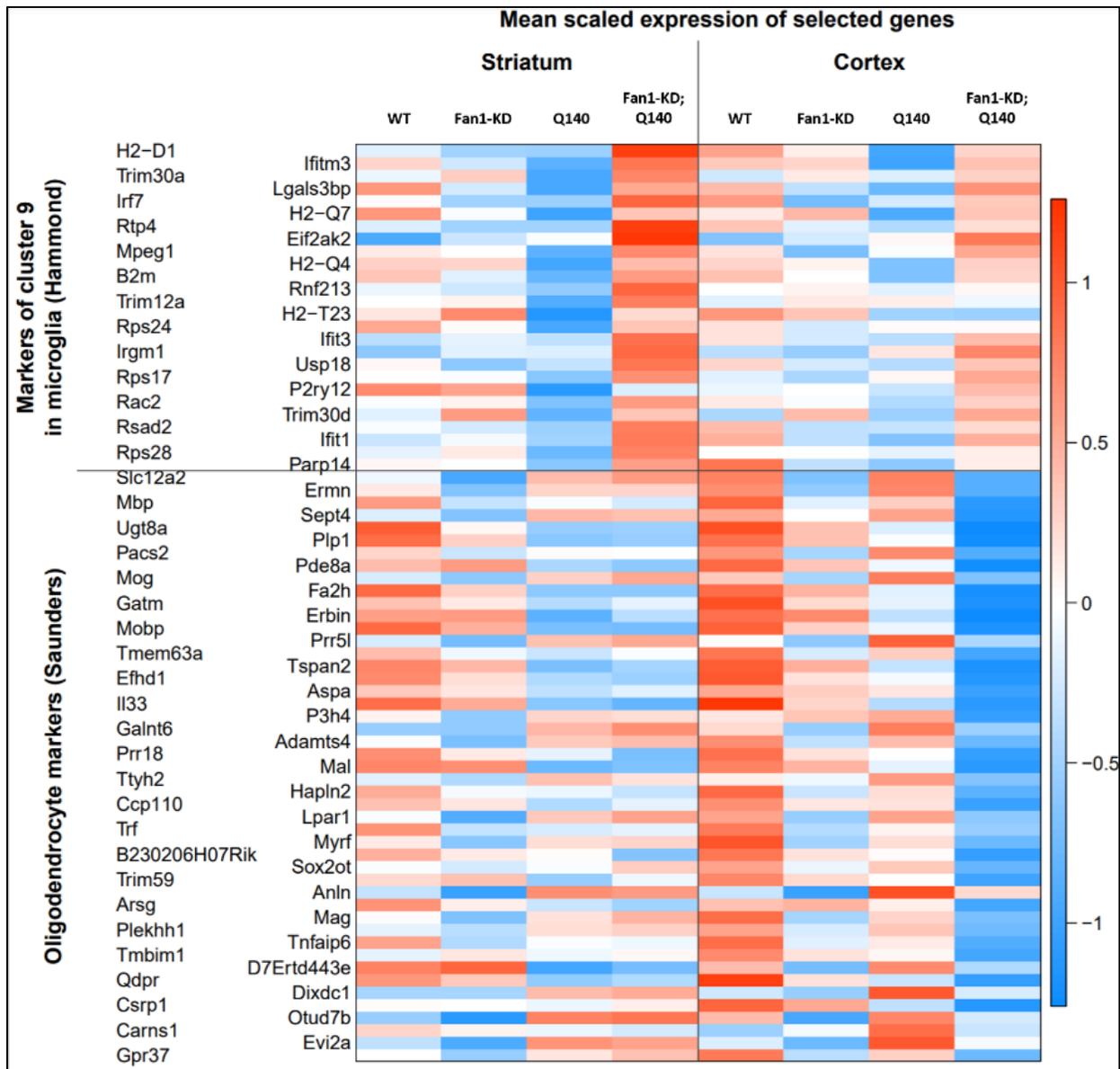
**Figure 6: Fan1-KD causes minimal transcriptional change in 12m cortex.**

(A-B) Number of upregulated (red) or downregulated (blue) genes at FDR<0.1 and p<0.01, respectively.

(C) Gene sets which show significant overlap with the striatal DE genes (p<0.01) for specific genotype comparisons.

genes were changing with Fan1-KD in WT and Q140 mice, we generated a heatmap to visualize their expression level (Figure 7). The microglia gene set comes from Beth Steven's lab, in which they performed single-cell RNAseq profiling of mouse microglia during development, in old age, and after brain injury (Hammond et al., 2019). From the heatmap, we can see that Q140 mice show a downregulation of microglial genes and Fan1-KD; Q140 have a marked upregulation of microglial genes in the striatum, compared to WT. Interestingly, a similar trend can also be seen for microglial genes changing expression in the cortex, although it is not as striking as it is in the striatum.

The OL gene set comes from Steve McCarroll's lab in which they performed single-cell RNAseq of 9 regions of the adult mouse brain including frontal cortex and striatum (Saunders et al., 2018). We find that Fan1-KD; Q140 mice have a reduction of OL genes in the cortex, compared to WT or Q140, and a modest reduction in OL genes in Fan1-KD cortex. This



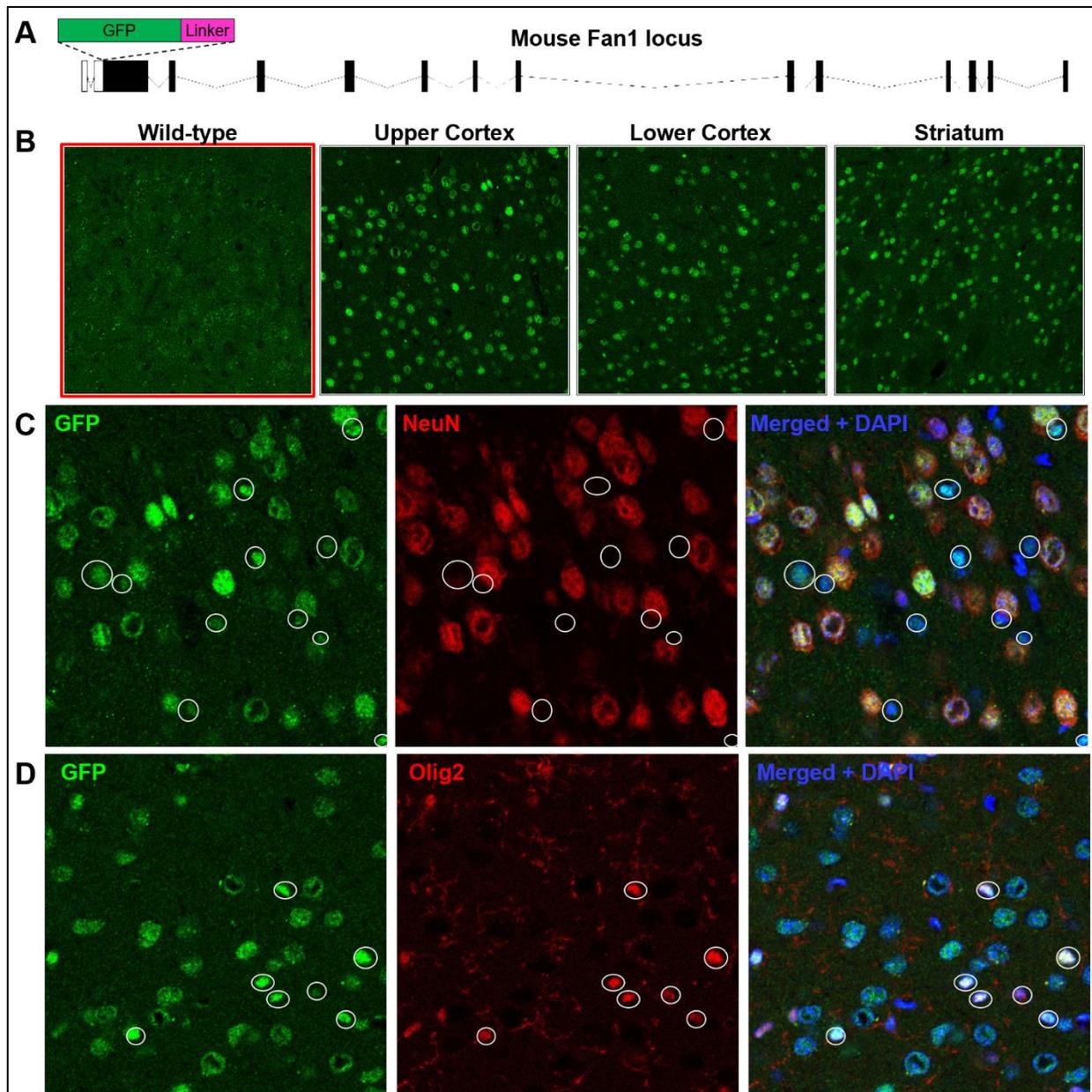
**Figure 7: Fan1-KD causes upregulation of striatal microglia genes and downregulation of cortical oligodendrocyte genes in the 12m mouse brain.**

heatmap highlights the list of microglial and oligodendrocyte genes that are differentially affected in the striatum and cortex of Fan1-KD in Q140 and WT mice.

### **2.3. Using a novel GFP-Fan1 KI mice to study cellular and subcellular expression of Fan1 wildtype**

Previous single-cell RNAseq studies in human (Tasic et al., 2018) and mouse (Gokce et al., 2016a; Loo et al., 2019; Saunders et al., 2018; Tasic et al., 2018) brains have reported Fan1 expression in a wide range of neuronal and glial cell-types. To confirm which cell-types express Fan1 in the brain, we generated a novel knock-in mouse model with GFP fused to the N-terminus of the Fan1 protein, (Figure 8A). A short chain of amino acids was added in between the GFP and Fan1 amino acid sequences to act as a spacer and was previously shown by Agata Smogorzewska's lab to minimize the effect of GFP on Fan1's protein conformation. A description of how the GFP-Fan1 mice were generated can be found in the appendix (Figure 20). GFP-Fan1 heterozygous (Het) mice were bred to obtain GFP-Fan1 homozygous (henceforth referred to as GFP-Fan1) mice. We stained brain section from 6m WT and GFP-Fan1 mice with a GFP antibody. We found minimal background GFP antibody staining in the wildtype mouse brain while the GFP-Fan1 brains showed GFP-positive cells in the upper and lower layers of the cortex and the striatum (Figure 8B). GFP-Fan1 was mostly localized to the nucleus. We next wanted to determine whether *Fan1* is expressed in either neurons, glia, or both cell-types. We co-stained GFP-Fan1 brain sections from 6m old mice with a neuronal marker, NeuN, GFP-antibody, and DAPI which labels nuclei. We found DAPI labeled nuclei in the cortex which show colocalization of GFP and NeuN, suggesting that these are neuronal cells expressing GFP-Fan1 (Figure 8C). We also found DAPI-labeled nuclei which show with GFP, but not NeuN, signal, suggesting there are non-neuronal cells which also express GFP-Fan1.

Since our 12m RNAseq data suggested that Fan1-KD causes a downregulation of OL genes in the cortex, we wanted to confirm whether cortical OLs express Fan1. We co-stained GFP-Fan1 brain sections from 6m mice with DAPI, a GFP-antibody and an OL-specific



**Figure 8: GFP-Fan1 is expressed in cortical and striatal neurons and glia.**

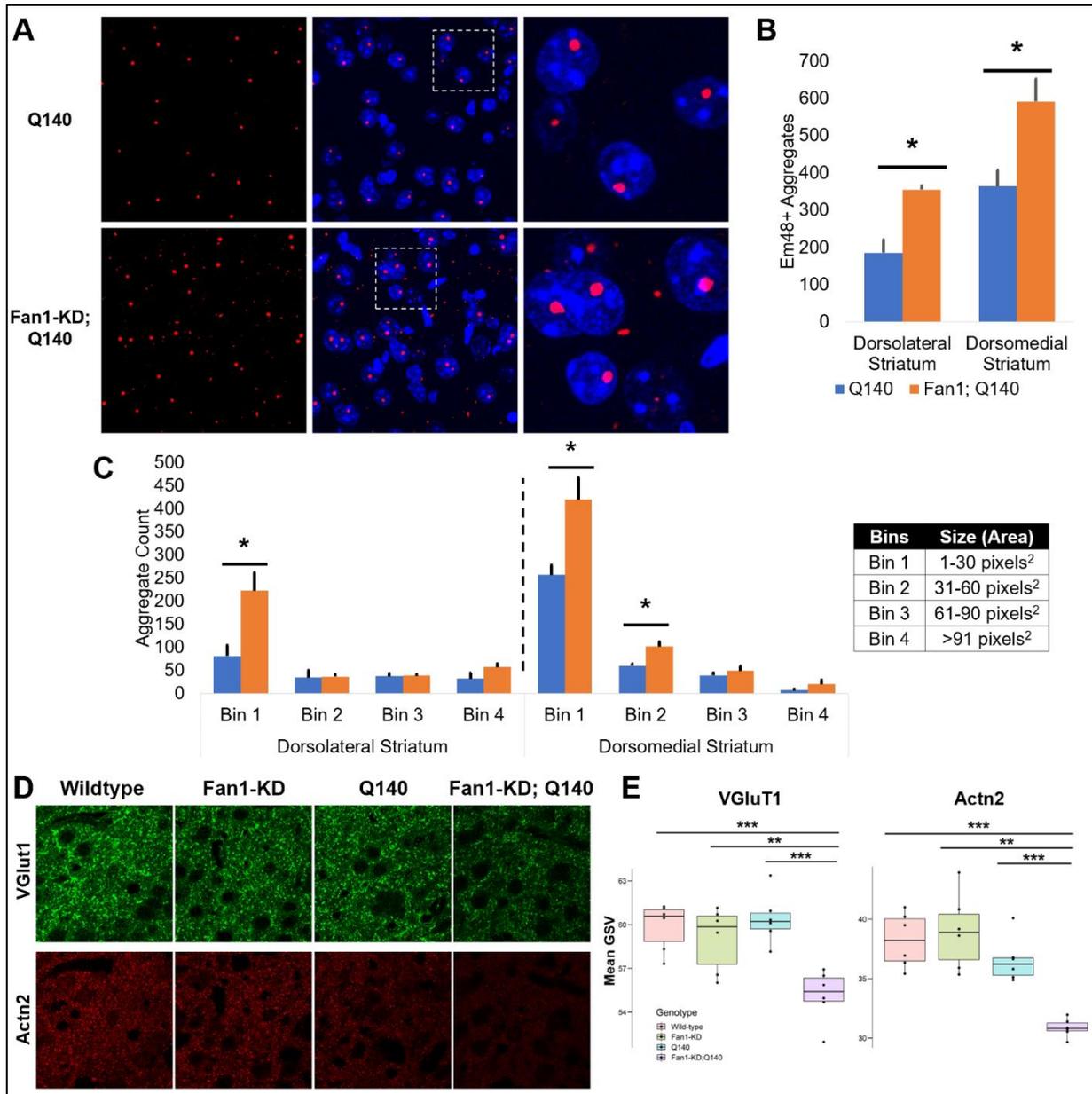
**(A)** Schematic of GFP knock-in created in 5' end of mouse Fan1. **(B)** Comparing anti-GFP staining in wildtype mice (highlighted in red) compared to GFP-Fan1 mice in the upper cortex, lower cortex, and striatum. **(C)** Colocalization of GFP and Neuronal Marker (NeuN) signal in mouse cortex of 6m GFP-Fan1 mice. White circles highlight glia cells with GFP and DAPI but lacking NeuN signal. **(D)** Colocalization of GFP and oligodendrocyte marker (Olig2). White circle highlight OLs which show colocalization of GFP, Olig2, and DAPI.

antibody, Olig2. We found cortical nuclei which show colocalization of the GFP, Olig2, and DAPI (Figure 5D), suggesting that GFP-Fan1 is expressed in cortical OLs. In combination with our 12m cortex RNAseq data, this result supports the hypothesis that Fan1 is playing an important role in cortical oligodendrocytes in both WT and Q140 mice.

## **2.4 Preliminary studies suggest Fan1-KD exacerbates neuropathology in HD mice**

We hypothesize that *Fan1* is a protective modifier in HD and loss of *Fan1* expression should exacerbate HD-related neuropathology. An age-related nuclear accumulation and aggregation of mHTT is seen in HD KI mice, with older mice showing larger and more numerous mHTT aggregates (Carty et al., 2015; Menalled et al., 2003). We stained the brains of 12m old Q140 and Fan1-KD; Q140 mice with EM48 antibody, which selectively recognizes the aggregated form of mHTT protein (Gutekunst et al., 1999; Li et al., 1999; Li et al., 2000). Fan1-KD; Q140 mice showed larger mHTT inclusions and had more cytoplasmic aggregates, compared to Q140 mice (Figure 9A). We quantified the number of EM48+ aggregates using ImageJ and found that Fan1-KD; Q140 mice show increased number of EM48+ aggregates in both the dorsolateral striatum (DLS) and dorsomedial striatum (DMS), compared to Q140 mice (Figure 9B). We next wanted to test whether large nuclear inclusions or smaller cytoplasmic aggregates were changing with Fan1-KD. Aggregates were divided into 4 bins based on their size: 1-30, 31-80, 61-90, 91+ pixels<sup>2</sup>. We found that 12m Fan1-KD; Q140 mice showed a significant increase in number aggregates in Bin 1 in the DLS and in Bin 1 and Bin 2 in DMS (Figure 9C).

Because our 12m RNAseq data showed that Fan1-KD leads to a reduction in cortical OL genes, we hypothesized that Fan1-KD may impact cortico-striatal synaptic communication. To test this hypothesis, we stained 12m brain sections from WT, Fan1-KD, Q140, and Fan1-KD; Q140 mice with a presynaptic marker (VGlut1) and postsynaptic marker (Actn2). The intensity of the VGlut1 and Actn2 staining was quantified using ImageJ to measure the mean gray scale value (GSV) of each image (n=4-6/genotype). We found preliminary data suggesting 12m Fan1-KD; Q140 mice showed significantly decreased VGlut1 and Actn2 signal compared with WT, Fan1-KD, and Q140 mice (Figure 9D-E). However, repeat staining by Linna Deng for Actn2 and



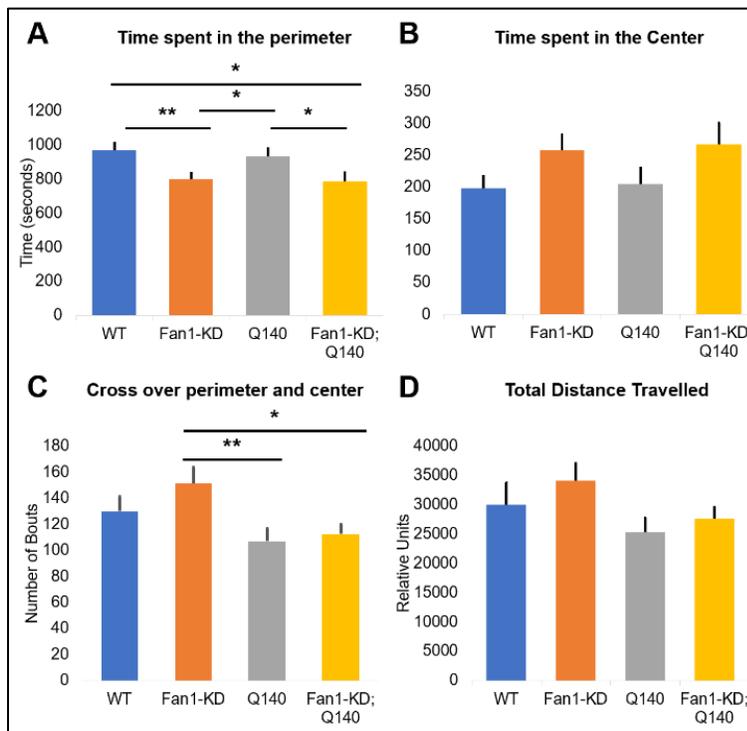
**Figure 9: Fan1-KD mice exhibit increased mHTT aggregates and decreased synaptic markers.** (A) Dorsolateral striatum of 12m Q140 and Fan1-KD; Q140 mice showing mHTT nuclear inclusions and cytoplasmic aggregates. (B) Quantification of number of mHTT aggregates in each visual field. (C) Binning of Em48-positive aggregates by size. (D) Presynaptic (VGlut1) and Postsynaptic markers (Actn2) staining in striatum from 12m old mice. (E) Quantification of signal intensity from synaptic marker. N=3-5/genotype/region, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, 2-tail student t-test, error bars = S.E.M.

VGlut1 showed no difference between Q140 and Fan1-KD; Q140 (data not shown). Taken together, these preliminary results suggest Fan1-KD may lead to reduced cortico-striatal

synaptic markers, but an independent validation study with larger sample size is needed to rigorously examine the pathology in these mouse models.

## 2.5 Fan1-KD causes minimal behavioral changes in normal and HD mice at 12m

We hypothesized that if Fan1-KD exacerbates cortical pathology or cortico-striatal communication, it may also exacerbate motor and psychiatric-like deficits in Q140 mice. We characterized a cohort of WT, Fan1-KD, Q140, and Fan1-KD; Q140 mice at 12m using openfield (locomotion) tests. Q140 mice show minimal behavioral deficits, but openfield phenotypes have been observed at various different ages in homozygous mice (Menalled et al., 2003). We found that Fan1-KD and Fan1-KD; Q140 mice spent less time in the perimeter of the



**Figure 10: Fan1-KD is correlated with reduced anxiety, but not motor deficits.**

(A) Time spent within perimeter of the chamber during 15min trial. (B) Time spent within the center of chamber. (C) Number of bouts crossing over the center and perimeter. (D) Total distance traveled within chamber. N=11-18 mice/genotype, \*p<0.05, \*\*p<0.01, 2-tail student's t-test, error bars = S.E.M.

openfield box, compared to WT and Q140 mice (Figure 10A). Time spent in the perimeter is presumed to reflect an anxious behavior, suggesting that the Fan1-KD and Fan1-KD; Q140 mice are showing reduced anxiety compared to WT and Q140 mice. Interestingly, time spent in the center did not reach significance by single-factor ANOVA (p=0.140, Figure 10B). Fan1-KD mice showed increased bouts of crossing from center to perimeter (Figure 10C), but it is not statistically significant compared to

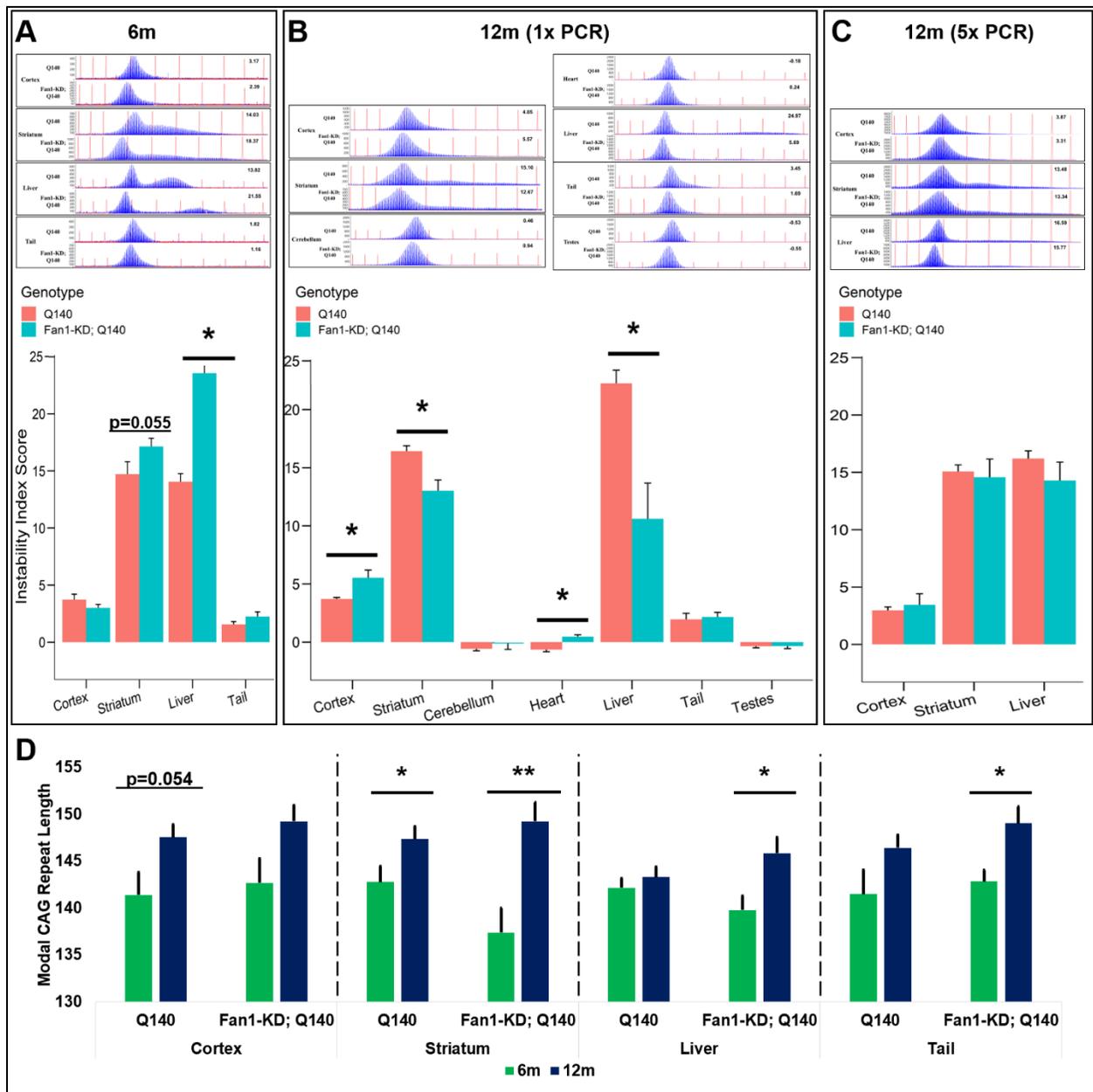
wildtype littermates. Lastly, there was no significant difference in distance traveled ( $p=0.152$ , single-factor ANOVA, Figure 10D). Overall, openfield results do not suggest that Fan-KD exacerbates motor deficits, and only moderately reduces anxiety-like behavior compared to WT and Q140 mice.

## **2.6 Age- and Tissue-specific effects of Fan1-KD on somatic mHTT CAG repeat instability**

With previous studies finding Fan1 plays a protective role in stabilizing Fmr1 CGG repeats in a Fragile X mouse model (Zhao and Usdin, 2018) and mHTT CAG repeats in HD cell lines (Goold et al., 2019), we wanted to test the hypothesis that Fan1-KD would cause expansion of mHTT CAG repeats in vivo. To analyze CAG repeat instability, we adopted a widely used PCR-based protocol, established in Vanessa Wheeler's lab, to visualize repeat instability and quantify instability by calculating a repeat instability index score (IIS) (Lee et al., 2010). A positive IIS indicates a bias towards expansion, while a negative IIS indicate a bias towards contraction. A full description of how nucleotide repeat instability can be measured and how the IIS is calculated can be found in the Material and Methods and in (Lee et al., 2010).

We assayed a cohort of Q140 and Fan1-KD; Q140 mice at 6m and 12m for somatic CAG repeat instability in tissues that were previously reported to have stable (e.g. heart, tail cerebellum) and unstable (e.g. striatum, liver) mHTT CAG repeats. We found that 6m Fan1-KD; Q140 mice showed significantly increased expansion of mHTT CAG repeats in the liver, compared to Q140 mice (Figure 11A). Interestingly, while the striatum showed a trend towards increased IIS in 6m Fan1-KD; Q140 mice, it was not significant ( $p=0.055$ , 2-tail student's t-test). Cortex and tail mHTT CAG repeat instability were also unchanged in 6m Fan1-KD; Q140 mice.

We next wanted to test whether the impact of Fan1-KD on somatic CAG repeat expansion was progressive with age. We assayed somatic CAG repeat instability in a different cohort of mice at 12m. We found that Fan1-KD; Q140 mice showed increased cortex IIS,



**Figure 11: Fan1-KD causes tissue- and age-dependent changes in somatic mHTT CAG repeat instability.** (A-C) Example GeneMapper traces of mHTT exon1 CAG repeat PCR products and bar plots showing CAG repeat instability index score (Lee et al., 2010) for specific tissues at either 6m or 12m of age. (D) Bar plots comparing modal CAG repeat lengths (tallest peak) for specific tissues between Q140 and Fan1-KD; Q140 at either 6m or 12m old. N=4-8 mice/genotype/tissue/age, \* $p < 0.05$ , \*\* $p < 0.01$ , 2-tail student's t-test, error bars = S.E.M.

compared to Q140 mice (Figure 11B). This was surprising because the cortex was previously reported to be relatively stable in Q111 KI mouse models (Lee et al., 2010), although cortical CAG repeat instability has been shown in fragment HD models (Larson et al., 2015) and post-mortem human cortex (Kennedy et al., 2003; Swami et al., 2009). Fan1-KD; Q140 mice also

showed increased IIS score in the heart, another tissue which was assumed to have relatively stable trinucleotide repeats (Zhao and Usdin, 2018).

Because 5-8% of HD patients are de novo cases (Squitieri et al., 1994; Warby et al., 2009) which likely come from *mHTT* CAG repeat expansion during meiosis (Semaka et al., 2013a; Semaka et al., 2013b), we also analyzed CAG repeat instability in the testes. We found no significant difference in the IIS of Fan1-KD; Q140 vs Q140 mice at 12m. We also found no effect of Fan1-KD on CAG repeat instability in the cerebellum or tail in 12m mice.

Surprisingly, Fan1-KD; Q140 mice showed decreased IIS in the striatum and liver, compared to Q140. PCR through long CAG repeats is inefficient, leading to reduced PCR product from highly expanded repeats. The GeneMapper software excludes peaks which are below 50 relative fluorescence units (RFUs) in an attempt to minimize PCR generated noise. We hypothesized that our IIS calculation were underrepresenting highly expanded alleles in 12m Fan1-KD; Q140 mouse striatum and liver sample due to their peak heights being below 50 RFUs, leading to the reduced IIS in these tissues. To test this hypothesis, we re-assayed cortex, striatum, and liver tissues from 12m Q140 and Fan1-KD; Q140 mice using 5 PCR reactions for each tissue sample and combining the PCR products to yield taller peak heights. The reanalyzed tissues now showed there is no significant difference in the CAG repeat instability in the cortex, striatum, or liver at 12m (Figure 11C). This result casts doubt on the hypothesis that Fan1's modifying role in HD is through the mechanism of somatic expansion of *mHTT* CAG repeats in the brain.

Another method of measuring somatic CAG repeat instability is to measure the growth of the modal CAG repeat length (i.e. the repeat size of the tallest peak in the GeneMapper traces), which is the *mHTT* CAG repeat length most prevalent in the assayed tissue. We found no significant difference between Fan1-KD; Q140 and Q140 mice at the same age, in any tissue compared. However, we did find that the modal CAG repeat lengths were significantly expanded

(2-10 CAGs) for certain tissues when comparing the 6m to 12m time points for the same genotype (Figure 11D). Specifically, the striatum, showed an increase in modal CAG repeat length from 6m to 12m in both Q140 and Fan1-KD; Q140 mice. Interestingly, the growth in modal CAG repeat length was larger in Fan1-KD; Q140 striatum, with an expansion of 11.9 CAGs from 6m to 12m, compared to 4.6 CAGs in Q140 mice. Fan1-KD; Q140 mice also showed a significant age-related expansion in modal CAG repeats in the liver and tail, while Q140 did not. Cortex modal CAG repeats did not significantly expand between 6m and 12m in either genotype. This suggests that Fan1 may play a role in growth of modal CAG repeat length over time in certain tissues.

## 2.7 Discussion

The aim of our current study was to test the hypothesis that *Fan1* is the genetic modifier in linkage disequilibrium with the Chr15 locus that is significantly associated with modifying HD age of onset. We provided several pieces of evidence which support this hypothesis.

The results of our RNAseq study suggested that at 6m, Fan1-KD may have a modest exacerbating effect on transcripts which are being downregulated in the cortex and striatum in HD mouse models. More importantly, the results from our 12m RNAseq study highlighted *Fan1*'s potential role in of two glia cell-types, striatal microglia and cortical OLs. Interestingly, Fan1-KD in WT mice also lead to downregulation of cortical mature OL genes, suggesting Fan1 may plays a modest protective role in these cell-types in the wildtype background but more robust role in the HD context. This finding needs to be followed up with more detailed analyses of oligodendrocyte pathology in the Fan1-KD; Q140 mice.

The increased microglia and inflammation-related transcripts seen in the striatum of Fan1-KD; Q140 mice can be caused by several mechanism. One possible explanation is that it is caused by a cell-autonomous mechanism in which loss of Fan1 directly leads to inflammation

signaling in the striatum. Fan1 may play a protective role in MSNs or microglia and Fan1 has been shown to be expressed in both these striatal cell-types (Gokce et al., 2016a; Saunders et al., 2018).

An alternative explanation is that the increased striatal inflammation may be caused by a non-cell autonomous mechanism. Loss of Fan1 in the cortex may impair cortico-striatal communication, possibly due to effects on oligodendrocytes, leading to downstream microglial inflammatory response in the striatum. Oligodendrocytes exhibit high metabolic rate due to their role in myelin maintenance and are vulnerable to oxidative stress (Smith et al., 1999). Fan1 may play an important role in protecting cortical OLs from ROS over-production in HD brains. Indeed, our 12m RNAseq found that Fan1-KD; Q140 and Fan1-KD show a downregulation of MYRF ( $p < 0.01$ ) in the cortex, which regulates oligodendrocyte maturation and is essential for proper myelination (Koenning et al., 2012). Reduction of MYRF transcriptional activity has been associated with OL dysfunction and myelin impairment in Q140 mice (Huang et al., 2015). Fan1-KD related impairment of OL or myelination may lead to earlier appearance of white matter lesions, which are seen many years before symptom onset in HD patients (Paulsen et al., 2008; Rosas et al., 2018; Tabrizi et al., 2009).

We next sought to test whether loss of *Fan1* exacerbated HD-related neuropathology in Q140 mice. We found that Fan1-KD increased aggregation in the striatum, especially small aggregates. This may be explained by Fan1-KD leading to increased CAG repeat instability, which yields *mHTT* alleles with longer CAG repeats and are more prone to *mHTT* accumulation. However, results from our CAG repeat instability analysis did not find a significant effect of Fan1-KD on CAG repeat instability in the striatum at 6m or 12m.

Another hypothesis is that Fan1-KD lead to increased instability in the cortex which then has a downstream effect of increasing aggregation in the striatum. Indeed, 12m Fan1-KD; Q140 mice showed increased cortex IIS compared to Q140 in our 1x PCR data. However, the 5x PCR

data for 12m cortex did not show a significant effect. The minor effect of Fan1-KD on cortical IIS may be due to the fact that cortex is a more heterogenous tissue compared to the striatum, and only a small proportion of cells show a Fan1-KD induced increased IIS.

A third hypothesis may be related to a previously unidentified role of *Fan1* as an RNA nuclease and regulating expression of RNAs. In human cells, RNA foci form by phase separation of the repeat-containing RNA (Jain and Vale, 2017). mHTT RNAs with long CAG or AS-HTT RNAs with long CUG repeats may phase separate and act as site of mHTT nucleation. Fan1 may play a role in controlling the turnover of these long CAG/CUG repeat mRNAs. Thus, loss of Fan1 might cause increased expression of HTT or AS-HTT RNA, leading to increasing mHTT aggregation in the striatum, a hypothesis that can be readily examined in our mouse models.

Lastly, we analyzed the impact of reduced *Fan1* expression on somatic CAG repeat instability. We analyzed the somatic instability of CAG repeats using two methods: (1) calculating an IIS and (2) measuring the growth in modal CAG repeat length from 6m to 12m. One caveat of the age-related expansion in modal CAG repeats analysis we performed is that the 6m and 12m data were generated using different cohorts of mice. Thus, mHTT alleles with different CAG repeat length could have been inherited from the distinct breeders used to generate the 6m and 12m cohorts of mice, and the baseline repeat sizes of these mice at 2m of age were not assayed. Thus, what was observed (Figure 11D) may be an effect of mouse cohort rather than age. However, if this was the case, we would expect to see this cohort effect in all tissues assayed. Thus, the fact that we do not see an age-related increase in modal CAG repeats in the cortex suggests it is not simply a cohort effect. Rather, there may be a real interaction between genotype, age and tissue on modal CAG repeat length. We intend to use additional mouse cohorts to validate these findings in our future studies.

At 6m, *Fan1*-KD lead to an increase in liver IIS, a finding I have independently verified using a second cohort of mice (Appendix, Figure 20). This result agrees with previous studies which found *Fan1* plays a role in stabilizing trinucleotide repeats (Goold et al., 2019; Zhao and Usdin, 2018). *Fan1*'s interacts with several MMR genes (*Mlh1*, *Pms2*) which have also been shown to play a role in CAG repeat instability (Goold et al., 2019; Rikitake et al., 2020; Smogorzewska et al., 2010). Specifically, loss of MSH $\beta$  and MLH complex proteins leads to reduced instability of CAG repeats (Dragileva et al., 2009; Manley et al., 1999; Pinto et al., 2013). *Fan1* may interact with *Mlh1* or other *Mlh* complex proteins to stabilize repeats in the liver. Interestingly, we did not observe an increase in liver IIS at 12m, suggesting that the exacerbating effect of *Fan1*-KD on liver *mHTT* CAG repeat expansion may normalizes with age.

A more surprising result was that we did not detect a significant impact of *Fan1*-KD on striatal *mHTT* CAG repeat instability. The most parsimonious theory of *Fan1*'s role as a modifier, is that loss of *Fan1* function increases the CAG repeat instability in striatal MSNs and the expanded *mHTT* alleles lead to an earlier onset of HD symptoms in human patients. However, we observed a very modest increase in striatal IIS in 6m *Fan1*-KD; Q140 mice, which did not reach significance. Moreover, 12m *Fan1*-KD; Q140 mice either show reduced instability in the striatum (1x PCR), or no change in instability (5xPCR), compared to Q140 mice.

To independently verify this result, we could use small-pool PCR, in which a small number of DNA molecules are used for PCR amplification and is followed by a southern blot (Dandelot and Gourdon, 2018; Gomes-Pereira et al., 2004; Swami et al., 2009). Small-pool PCR has been used to reveal the presence of dramatic somatic instability in HD patient brains that was undetected or underestimated using standard PCR amplification of 'bulk' genomic DNA.

As corroborative evidence for this result, we looked at the expression of the alternatively spliced *mHTT* RNA with intron1 retained. RNA polymerase read-through of exon1 and

expression of truncated *HTT* exno1-intron1 transcripts is proportional to CAG-repeat length (Neueder et al., 2017; Sathasivam et al., 2013). Thus, increased striatal CAG repeat instability should cause increased expression of *mHTT* splice variants with intron1-retention. However, we did not see an increase in *HTT* intron1 read count in Fan1-KD; Q140 mice compared to Q140 at 12m (Appendix, Figure 24), suggesting Fan1-KD does not increase *mHTT* CAG repeat instability in the striatum.

While Fan1 may play an important role in stabilizing CAG repeats in the liver at 6m, our current evidence does not show similar effects of in stabilizing *mHTT* CAG repeats in the striatum at 6m and 12m of age. Since our study only used Fan1 KD mice with 70% reduction of Fan1 transcripts, and our analysis is up to 12m of age, we cannot rule out Fan1 effects on striatal *mHTT* repeat instability could be observed with more complete depletion of Fan1 gene expression or with our current mouse models at a much older age. However, we consider our study still informative as the disease modifying alleles of FAN1 in HD are not known to be null alleles. We would like to interpret the possible new insights on Fan1's role in HD is its effects on cortical oligodendrocytes, based on our transcriptomic findings. It is possible that Fan1 functions to delay the onset of oligodendrocyte dysfunction, which in turn leads to cortico-striatal miscommunication and non-cell autonomous striatal pathogenesis (Teo et al., 2016; Veldman and Yang, 2018; Wang et al., 2014a). This hypothesis is also supported by prior literature suggesting GWAS FAN1 alleles may alter the expression of FAN1 in the cortex but not in the caudate and putamen (GeM-HD, 2019; Goold et al., 2019). (GeM-HD, 2019; Goold et al., 2019).

Overall, our findings support the hypothesis that *Fan1* is a modifier of HD. Fan1-KD led to altered transcription of HD-related transcripts, caused a modest increase in striatal inflammation-related transcripts, reduced cortical OL-related gene expression, and increased *mHTT* aggregation in the striatum. While the exact mechanism of Fan1's modifying influence on HD is not fully understood, we provide evidence suggesting further investigation of our Fan1-

KD;Q140 mouse model systems, particularly cellular and molecular changes in the cortex, is warranted.

## CHAPTER 3:

### **Uninterrupted Huntingtin CAG-Repeat Length Predicts Striatal Nuclear Inclusions and Transcriptionopathy In Vivo**

Chapter 3 contains work from a collaborative project and the data I generated that will be part of a manuscript that I will be 2<sup>nd</sup> author on.

Xiaofeng Gu, Jeff Richman, Peter Langfelder, Huei Bin Wang, Nan Wang, Jeff Cattle, Lucia Yang, Lalini Ramanathan, Chang Chris Park, Giovanni Coppola, Steve Horvath, Gillian Bates, Christopher Colwell and X. William Yang

My specific research efforts for this project include analyzing somatic mHTT CAG repeat instability, in collaboration with Laragen. I also quantified several human HTT protein species using MSD assay, in collaboration with Evotec. Lastly, I performed night-time open field behavioral testing, extracted RNA samples for RNAseq experiments and helped perform all tissue dissections. Xiaofeng Gu engineered the human HTT BAC to contain an expanded, uninterrupted CAG repeat. Dr. Gu then generated the BAC-CAG mice, managed the mouse colony, measured body weight, and performed immunostaining with S830, GFAP, and Actn2 antibodies. Nan Wang and Lucia Yang performed western blotting of full-length huntingtin. Huei Bin Wang performed the activity/sleep study as well as grip strength testing. Peter Langfelder performed differential expression, enrichment, and concordance analyses on RNAseq data and generated figures describing the results. He also performed correlational analyses between CAG repeat indices and sleep/activity behavioral measures.

### **3.1 Significance of developing novel human *HTT* BAC Transgenic mouse model with uninterrupted CAG repeats**

For decades, the poly-glutamine (polyQ) repeats in the mHTT protein have been considered the main driver of HD pathogenesis. However, recent mounting evidence suggests that the expanded, uninterrupted CAG repeats in mHTT may directly contribute to HD pathology (Lieberman et al., 2019). This was recently highlighted by GeM-HD (2019), in which they showed that, if polyQ length is kept constant, then presence, absence, or duplication of terminal CAACAG sequence (both encoding for Q) HTT CAG repeats significantly correlates with modifying HD age of onset. Importantly, this suggests that uninterrupted CAG repeat length is a better predictor HD onset than polyQ length.

The GWAS also identified genomic loci near MMR genes that were significantly associated with modifying onset of HD. MMR genes have been studied in HD for their role in somatic mHTT CAG repeat instability (Manley et al., 1999; Dragileva et al., 2009; Flower et al., 2019; Gomes-Pereira et al., 2004; Kovalenko et al., 2012; Manley et al., 1999; Pinto et al., 2013), and somatic instability in the human cortex is significantly correlated with HD onset (Swami et al., 2009). Thus, MMR proteins are believed to influence HD progression by effecting somatic instability of mHTT CAG repeats in the brain.

The current study describes the generation and characterization of a novel human BAC transgenic mouse model expressing full-length *HTT* with approximately 120 uninterrupted CAG repeats (BAC-CAG). We provide insight into the direct contribution of mHTT uninterrupted CAG repeats to HD pathogenesis by comparing phenotypes seen in BAC-CAG mice with previous full-length HD mouse models. Interestingly, despite expressing a human HTT protein which is similar to that expressed by our previous BACHD model (Gray et al., 2008), BAC-CAG mice show phenotypic differences which better recapitulate the phenotypes seen in the allelic series KI mice (e.g. Q140 and Q175). This includes robust mHTT nuclear accumulation and nuclear

inclusion, age-dependent striatal transcriptionopathy, and somatic CAG repeat instability. Our current study provides the first evidence that phenotypic differences between allelic series KI models and full-length genomic HD models is the presence/absence of long patient-like, uninterrupted CAG repeats in HTT exon1.

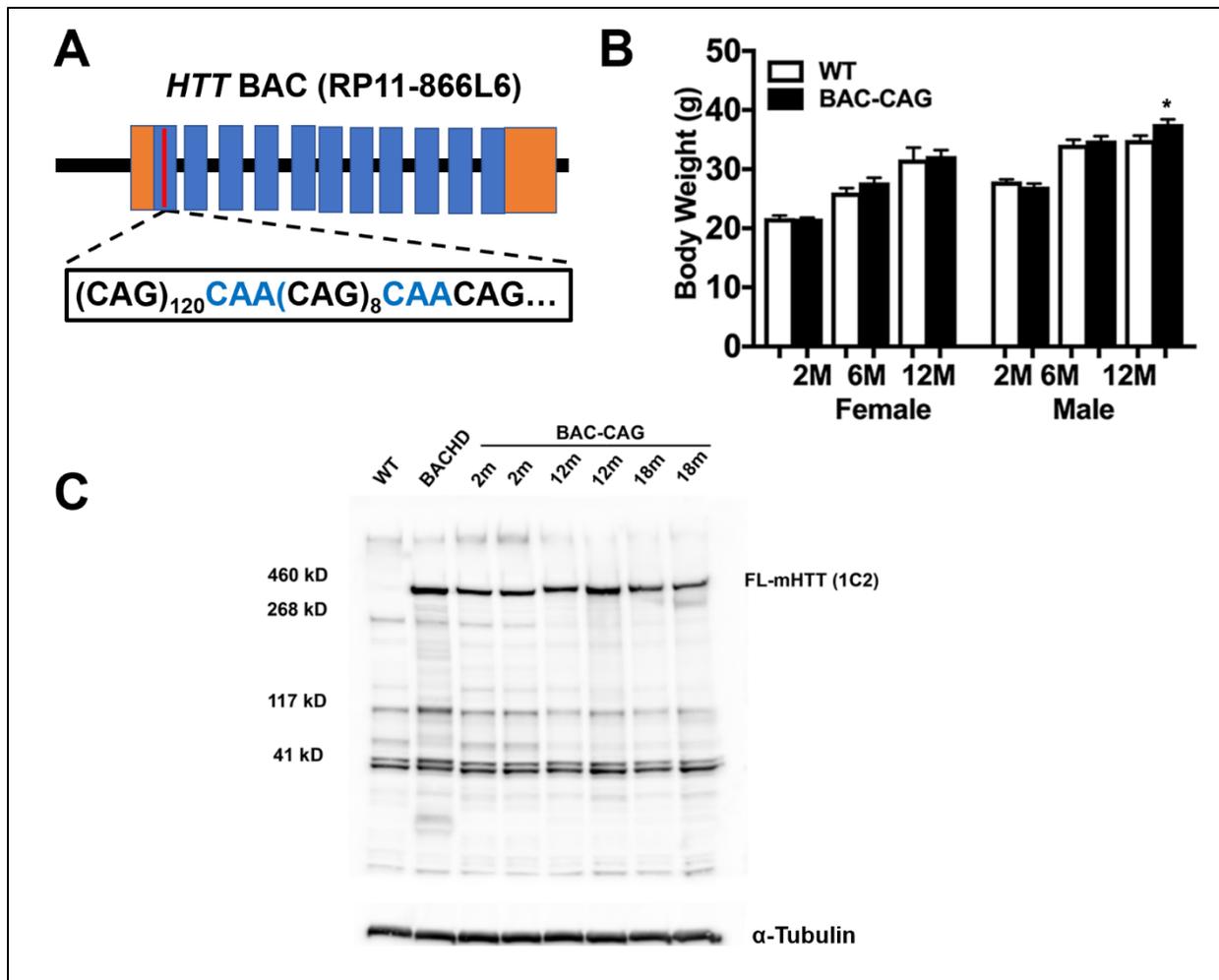
### **3.2 Generation and initial characterizations of BAC-CAG mice**

To test the hypothesis that phenotypic differences between prior human genomic models and the KI models of HD are due to the difference in the genomic DNA encoding the polyQ repeat in mHTT, Dr. Xiaofeng Gu generated a novel HD mouse model which contains a full-length human *HTT* BAC with uninterrupted CAG repeats integrated into the mouse genome. The human *HTT* BAC was engineered to contain roughly 120 pure CAG repeats followed by one CAA, 8 CAGs and another CAA-CAG, to encode about 141 glutamine repeats (Figure 12A). Using qRT-PCR, we determined that BAC-CAG mice contain 2 copies of the BAC transgene (data not shown).

Previous human genomic models of HD showed progressive weight gain (Gray et al., 2008). However, BAC-CAG mice show either no weight gain, or minimal (about 5%) weight gain at 12m, and normal weight at 2m and 6m (Figure 12b). As well, BAC-CAG mice express intact mHTT in the striatum at 2, 12 and 18 months of age with minimal soluble mHTT fragments at these ages (Figure 12C).

### **3.3 Neuropathology in BAC-CAG mice**

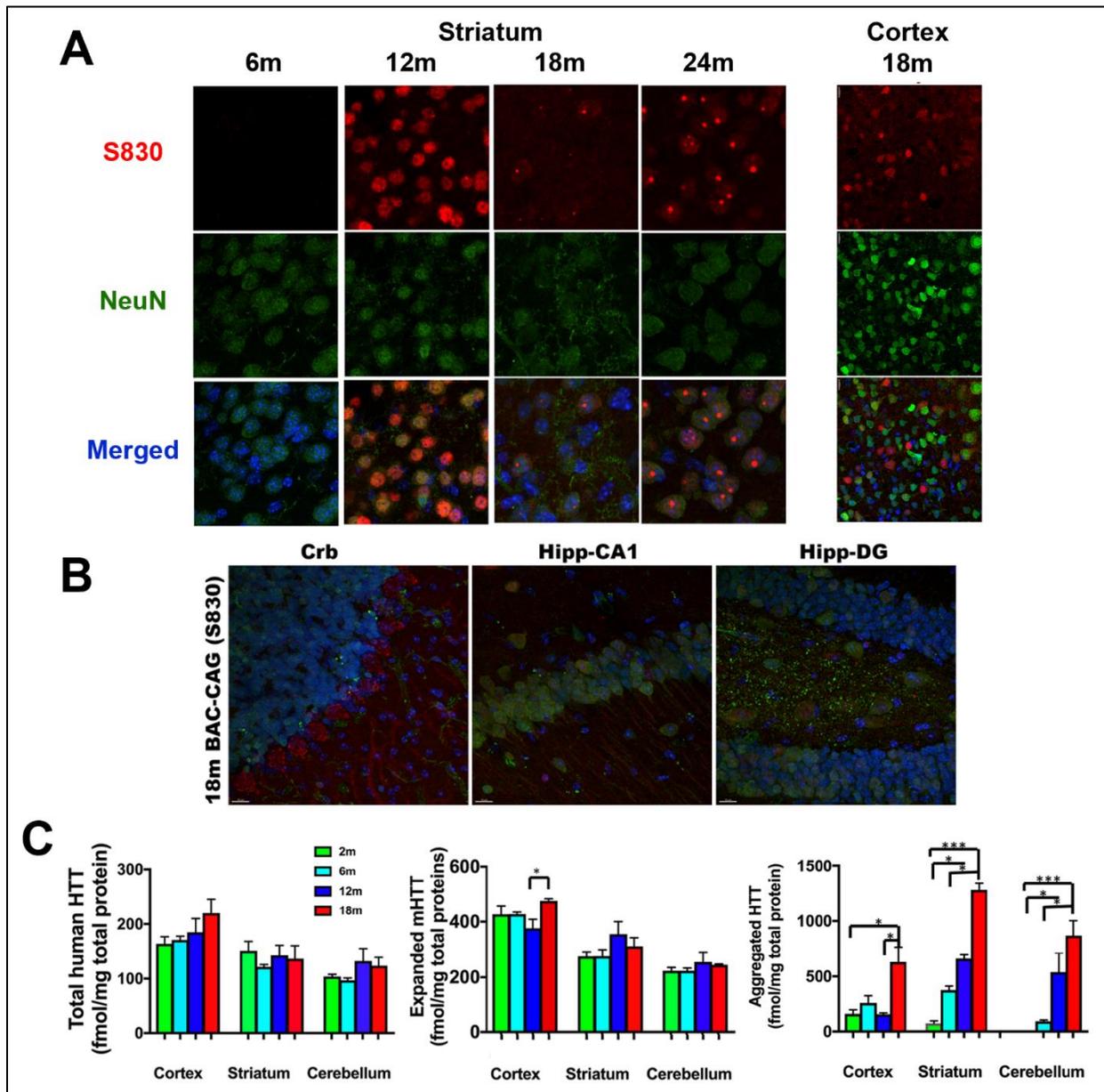
mHTT-containing nuclear inclusions (NIs) are a hallmark of HD (DiFiglia et al., 1997). While robust mHTT aggregates are prominently displayed in the striatum and cortex of *mHtt* KI mice (Carty et al., 2015), NIs only emerge at very advanced age in human genomic models (Bayram-Weston et al., 2015; Gray et al., 2008). To determine whether BAC-CAG mice display



**Figure 12: Generation and initial characterizations of BAC-CAG mice.**

(A) Schematic representing the modified human *HTT* BAC (RP11-866L6) with expanded, uninterrupted CAG repeats. (B) Bar plot showing the mean body weight of male and female mice at different ages. (C) Western blot with anti-polyQ antibody (1C2) shows intact mHTT in the striatum at different ages. \* $p < 0.05$ , 2-tail student's t-test, error bars = S.E.M.

robust mHTT aggregation, Dr. Gu stained brain sections from BAC-CAG mice at different ages (6m, 12m, 18m and 24m) with antibodies against aggregated mHTT. We did not find nuclear mHTT accumulation or NIs in 6m mice using either EM48 (not shown) or another mHTT aggregate-sensitive antibody, S830 (Sathasivam et al., 2001) (Figure 13A). However, we found diffuse nuclear aggregation in 12m, 18m and 24m striatum and in 18m cortex (Figure 13A). We also observed progressive NI formation in 18m and 24m striatum. Nuclear accumulation of mHTT was also seen in the cerebellum of 18m BAC-CAG mice, but not in hippocampus CA1 and



**Figure 13: Age-dependent mHTT nuclear accumulation and aggregation in BAC-CAG mice.**

(A-B) S830 and NeuN staining in different brain regions in BAC-CAG mice at various ages. (C) Bar plots of mean quantified mHTT protein in cortex, striatum, and cerebellum tissues from BAC-CAG mice at different ages. n=6 mice/age/tissue

dentate gyrus (Figure 13B). This suggests that the distribution of mHTT accumulation in BAC-CAG mice is much more restricted to the striatum and deep cortical layers, compared to KI mice (e.g. Q140 and Q175).

To independently and quantitatively evaluate total and aggregated mHTT levels in BAC-CAG mice, we used Meso Scale Discovery (MSD) (Macdonald et al., 2014; Reindl et al., 2019)

on protein samples from multiple brain regions (cortex, striatum, cerebellum) and at different ages (2m, 6m, 12m, 18m). MSD uses antibodies conjugated to electrochemiluminescent labels which generate light when stimulated by electricity. This approach offers lower background and improved signal amplification, sensitivity, and dynamic range compared to traditional ELISA.

KI models of HD (Q140 and HdhQ150) have previously been reported to show a decrease in mHTT expression with age (Franich et al., 2018). To determine whether the expression of human HTT changes over time in BAC-CAG mice, we used MSD to measure total HTT and mHTT, which should yield similar results given that BAC-CAG only express expanded HTT (Figure 13C). MSD for total human HTT, using antibodies against N-terminus (HTT 7-13aa) and polyproline regions (HTT 51-71aa), showed no significant change in the cortex, striatum and cerebellum of BAC-CAG mice from 2m to 18m. MSD assay for mHTT, using antibodies against N-terminus (HTT 7-13aa) and polyQ regions, showed only the cortex is significantly increased in mHTT from 12m to 18m, while striatum and cerebellum remained unchanged. Overall, this suggests that human HTT expression in BAC-CAG mice remains mostly unchanged with age.

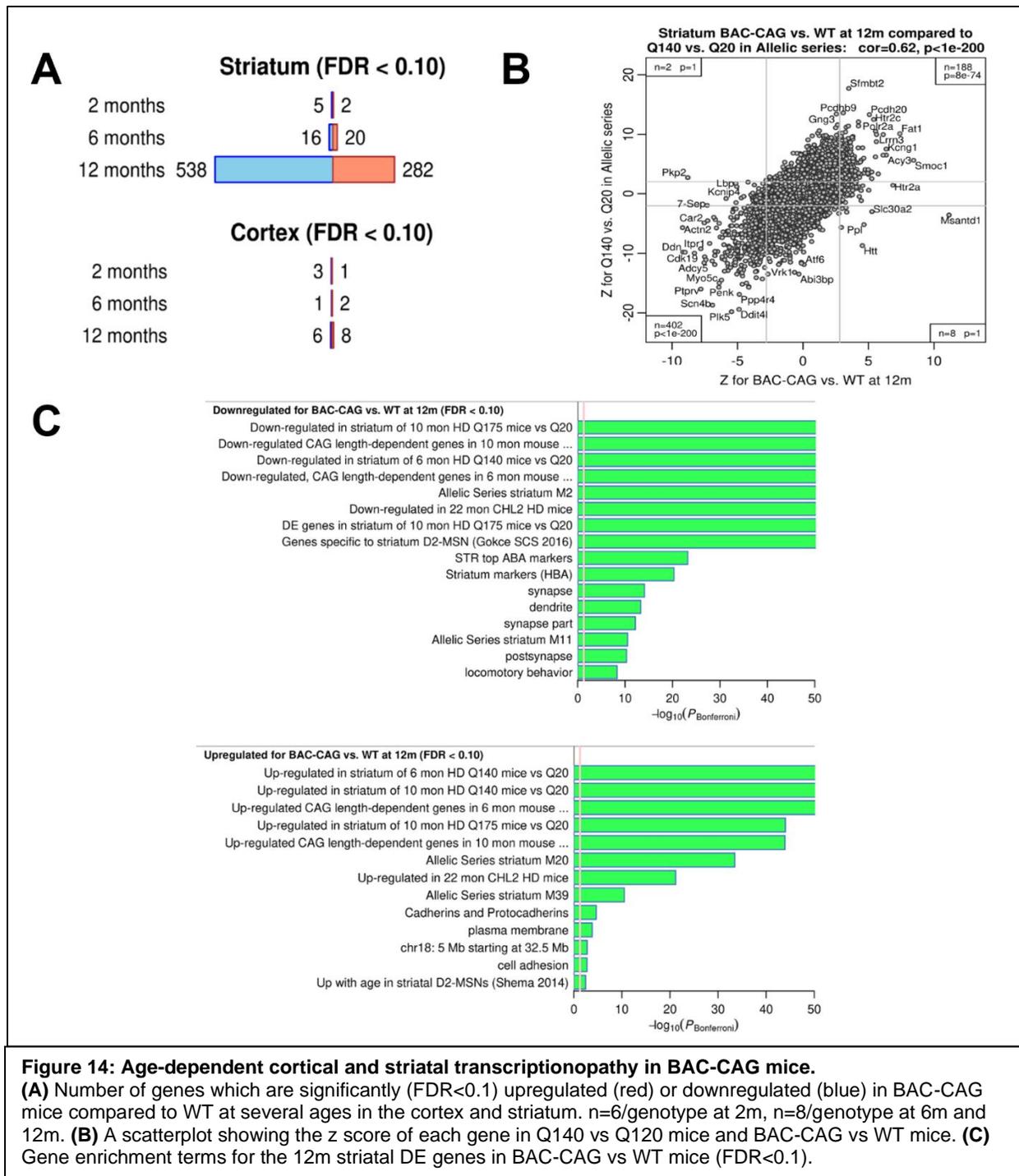
We next used MSD to measure aggregated mHTT, using antibodies against polyproline (HTT 51-71aa) and an immediately adjacent region (HTT 82-90aa) region, to see if it would confirm our findings from S830 immunostaining. We found a significant increase with age in aggregated HTT in all 3 tissues (cortex, striatum, and cerebellum) with age (Figure 13C). The cortex showed no change in aggregated mHTT level from 2m to 12m, but became significantly increased at 18m. The striatum, on the other hand, showed progressive increase in mHTT aggregation from 2m to 18m, and reached significance at 18m. Finally, the cerebellum showed a similar trend as the striatum, with progressive increase in mHTT aggregation starting at 6m and reaching a significant increase by 18m. Thus, the MSD assay confirmed our findings with immunostaining and demonstrates that BAC-CAG mice develop robust and progressive mHTT

aggregation in multiple brain regions with age. Thus, 12m BAC-CAG mice display neuropathological features including early mHTT aggregation and NI formation that were not seen in previous human genomic HD mouse models.

### **3.4 Striatal transcriptionopathy in BAC-CAG mice is highly concordant with transcriptional changes seen in allelic series KI mice Striatum.**

HD patients and KI mouse models develop age-related transcriptional changes in the striatum (Kuhn et al., 2007; Langfelder et al., 2016). However, human genomic HD models do not recapitulate this phenotype well (Bayram-Weston et al., 2015). To determine whether long uninterrupted CAG repeats enables BAC-CAG mice to recapitulate the transcriptional changes seen in HD patients and KI mice, we performed RNAseq analysis of the striatum and cortex at 2m, 6m and 12m of age. We observed relatively few significant DE genes (FDR<0.1) in the striatum at 2m and 6m (7 and 36 DE genes, respectively). However, 820 genes showed differential expression in the striatum at 12m of age (Figure 14A). By contrast, the cortex showed relatively few significant DE genes (FDR<0.1) at all three ages.

We next performed concordance and enrichment analyses to gain a better understanding of which cellular and molecular pathways are being dysregulated in the 12m BAC-CAG striatum. We found high concordance between 12 striatal DE genes in BAC-CAG vs WT and Q140 KI vs WT (Figure 14B) and enrichment analyses of the 12m BAC-CAG vs WT DE genes (FDR<0.1) revealed that up- and down-regulated genes were strongly enriched in the same direction as allelic series KI mice (Q175, Q140, Q150) striatal DE genes (Figure 14C). This suggests that the differentially expressed transcripts in BAC-CAG striatum are related to HD pathology, and not merely an effect of the transgene integration. Enrichment analyses also find that BAC-CAG vs WT striatal DE genes are enriched for *HTT* CAG-length dependent modules (M2, M11, M20 and M39) (Langfelder et al., 2016) and showed selective



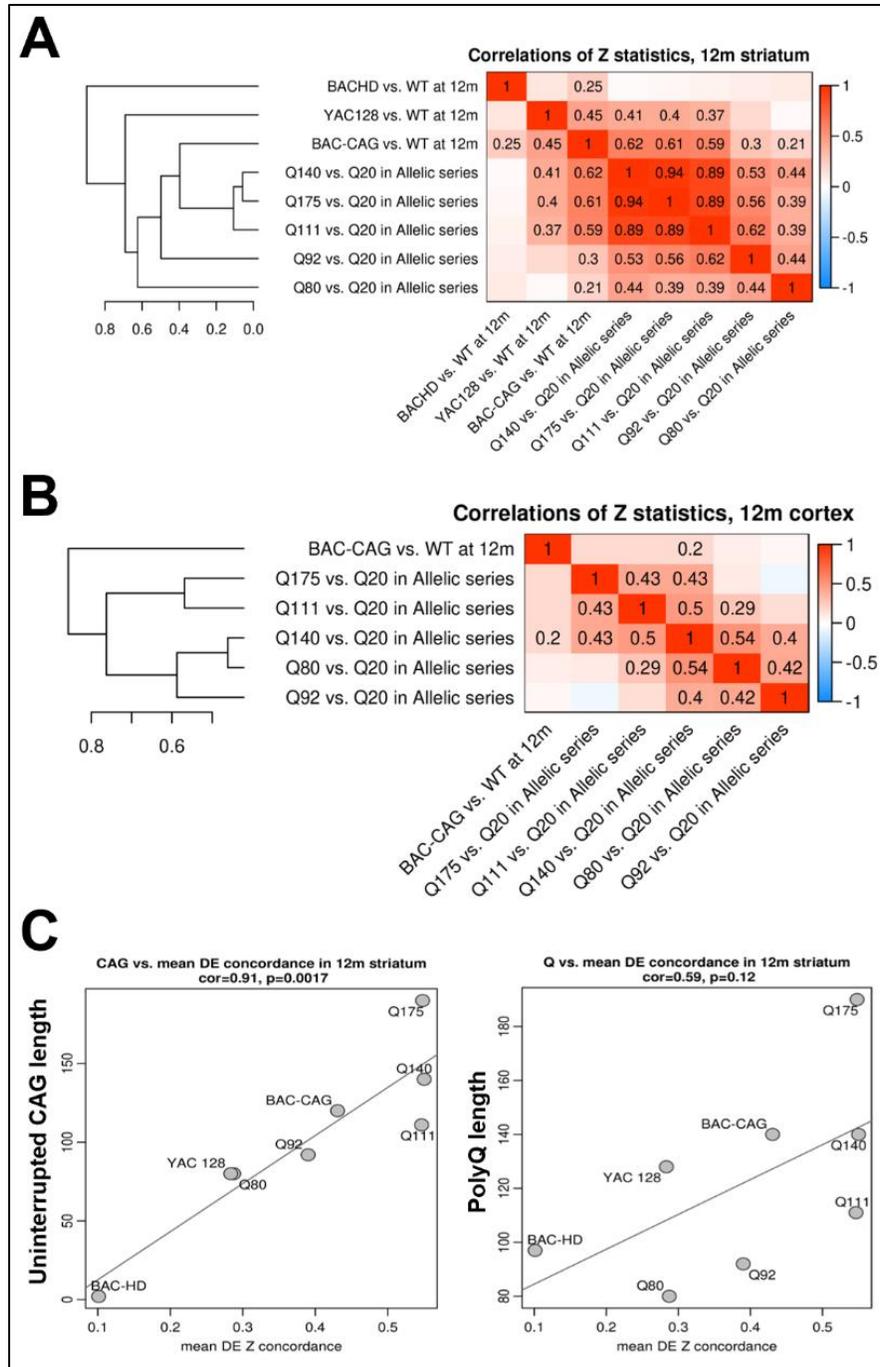
**Figure 14: Age-dependent cortical and striatal transcriptionopathy in BAC-CAG mice.**  
**(A)** Number of genes which are significantly (FDR<0.1) upregulated (red) or downregulated (blue) in BAC-CAG mice compared to WT at several ages in the cortex and striatum. n=6/genotype at 2m, n=8/genotype at 6m and 12m. **(B)** A scatterplot showing the z score of each gene in Q140 vs Q120 mice and BAC-CAG vs WT mice. **(C)** Gene enrichment terms for the 12m striatal DE genes in BAC-CAG vs WT mice (FDR<0.1).

downregulation of MSN and synapses markers, and upregulation of clustered protocadherins, similar to the allelic series KI mice. Thus, BAC-CAG mice display robust age-dependent, striatal-selective transcriptionopathy, unlike previous human genomic models of HD.

### **3.5 Uninterrupted *mHTT/mHtt* CAG-repeat length predicts transcriptional dysregulation in striatum of full-length huntingtin models of HD.**

To better understand how the BAC-CAG striatal transcriptome compares to previous full-length huntingtin models, Peter Langfelder performed hierarchical clustering of the striatal transcriptome of allelic series KI mice at 6m and human genomic models (BAC-CAG, BACHD and YAC128) at 12m. These ages were chosen because they are the earliest timepoint where robust transcriptional dysregulation is observed in each model (Bayram-Weston et al., 2015; Langfelder et al., 2016). Interestingly, the BAC-CAG striatal transcriptome, but not BACHD or YAC128, clustered closely with the allelic series KI mice (Figure 15A). This suggests that of the human genomic models of HD, BAC-CAG mice best recapitulate the striatal transcriptionopathy seen in the KI HD models (Q140 and Q175). To see if BAC-CAG cortical transcriptome is also concordant with the allelic series KI mice, we repeated this analysis using the BAC-CAG and allelic series cortex transcriptomic data. We found that the BAC-CAG cortical transcriptome shows minimal concordance with Q140 mice and no concordance with the other allelic series models (Figure 15B). These results suggest that a human genomic model of *mHTT* with uninterrupted CAG repeat and murine KI *mHtt* mice share common transcriptomic signatures in the striatum, but not the cortex.

The GeM-HD GWAS recently identified that uninterrupted CAG repeats are a better predictor of HD age of onset than polyQ length (GeM-HD, 2019). We therefore asked whether uninterrupted CAG repeat length would also predict transcriptionopathy in HD mouse models. We found that uninterrupted CAG repeat length, but not polyQ length, showed a significant correlation with transcriptional dysregulation in the striatum of human genomic and KI models of



**Figure 15: Striatal transcriptional changes better predicted by uninterrupted CAG repeat length than PolyQ length.**

(A-B) Tombstone plots and clustering of HD mouse models based on correlation of transcriptional changes in a given tissue at 12m in BAC-CAG mice or 10m in allelic series mice. (C) Scatter plots highlighting the robust correlation between uninterrupted CAG repeat length or polyQ length and concordance of striatal DE genes.

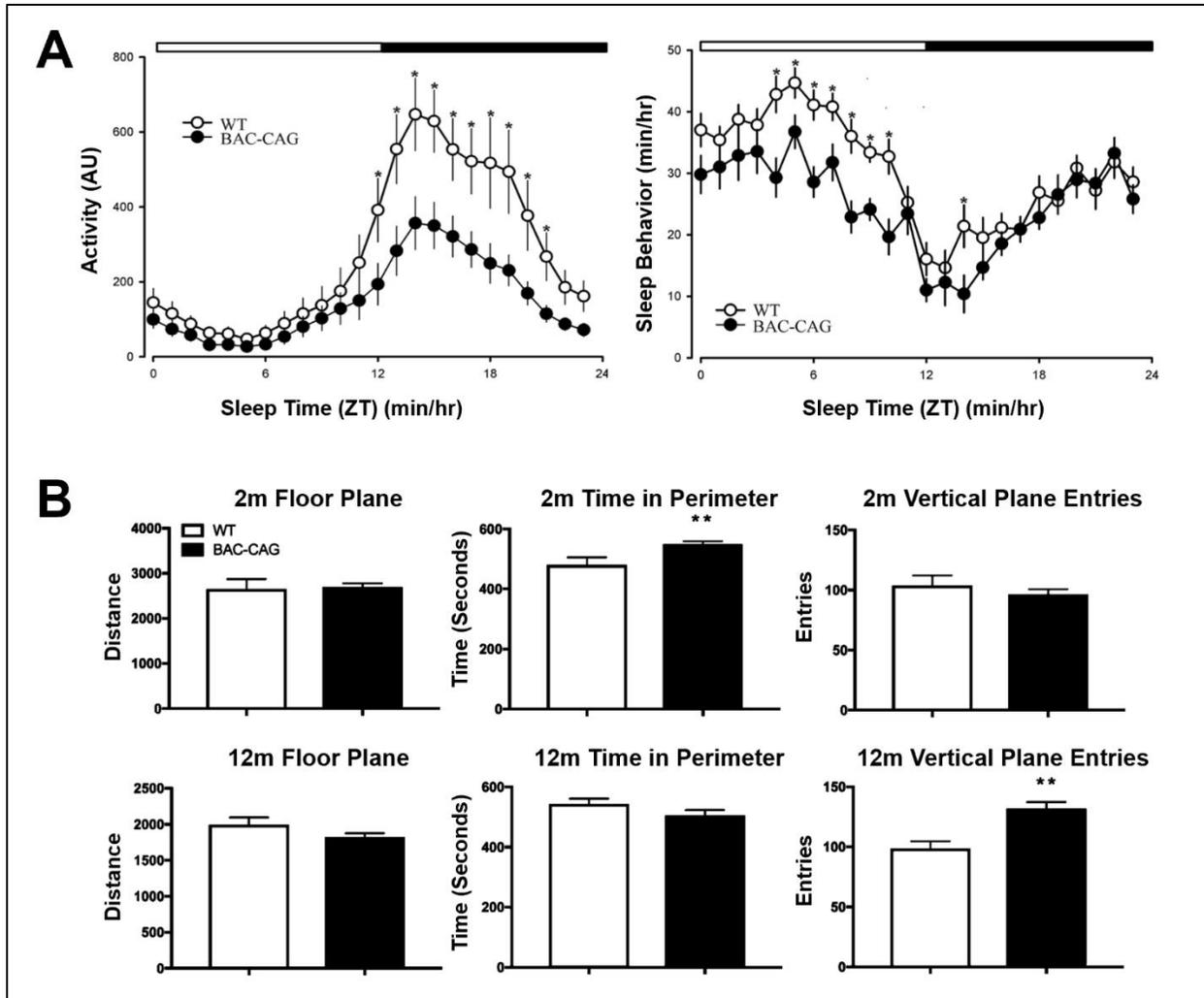
HD (Figure 15C). This suggests that uninterrupted mHTT CAG repeats may directly contribute to transcriptional dysregulation in the striatum in full-length HTT models.

Together, our RNAseq analyses reveal that BAC-CAG mice showed age-dependent, striatum-selective transcriptional dysregulation that closely resemble the allelic series *mHtt* KI mice. Moreover, our study demonstrates that among the full-length models, the length of the uninterrupted *mHTT* CAG repeat length, but not polyQ length, can predict the striatal DE gene concordance.

### **3.6 BAC-CAG mice exhibit activity and sleep deficits**

One of the earliest behavioral deficits seen in HD patients is sleep disturbance (Morton, 2013), and altered sleep/wake rhythms have been reported in previous HD models (Kudo et al., 2011; Loh et al., 2013; Morton et al., 2005). To determine whether BAC-CAG mice also exhibit sleep disruptions, we collaborated with Dr. Chris Colwell and Huei Bin Wang to evaluate locomotor activity and sleep behavior of these mice (Figure 16A). 12m BAC-CAG mice display diurnal rhythms of activity. However, BAC-CAG mice were less active than WT mice during the night (normally their most active period). BAC-CAG mice also slept significantly less than WT mice during the day (inactive period). Taken together, this suggests BAC-CAG develop deficits in sleep/wake cycles, similar to previous human genomic models.

Psychiatric symptoms such as anxiety and depression are commonly seen in HD patients (Thompson et al., 2012), as well as motor impairment (Bates et al., 2015; Reilmann, 2019). Open field testing can be used to measure locomotor and psychiatric behaviors and our lab has previously reported that 12m BACHD mice show reduced floor plate distance traveled during open field testing (Wang et al., 2014a). To test if BAC-CAG mice exhibit reduced locomotion or psychiatric-like phenotype, we performed open field testing on BAC-CAG and WT mice at 2m and 12m of age. Open field testing was performed at night, when mice are most active. We found that 2m and 12m BAC-CAG mice do not exhibit any change in floor plate



**Figure 16: Locomotor, sleep, and psychiatric phenotypes in BAC-CAG mice.**

**(A)** Daily average activity or sleep behavior over a 10-day period. N=13 male mice/genotype. **(B)** Bar plots showing mean of various open field measurements. N=5-17 mice/genotype/age. \*p<0.05, \*\*p<0.005, 2-tail student's t-test, error bars = S.E.M.

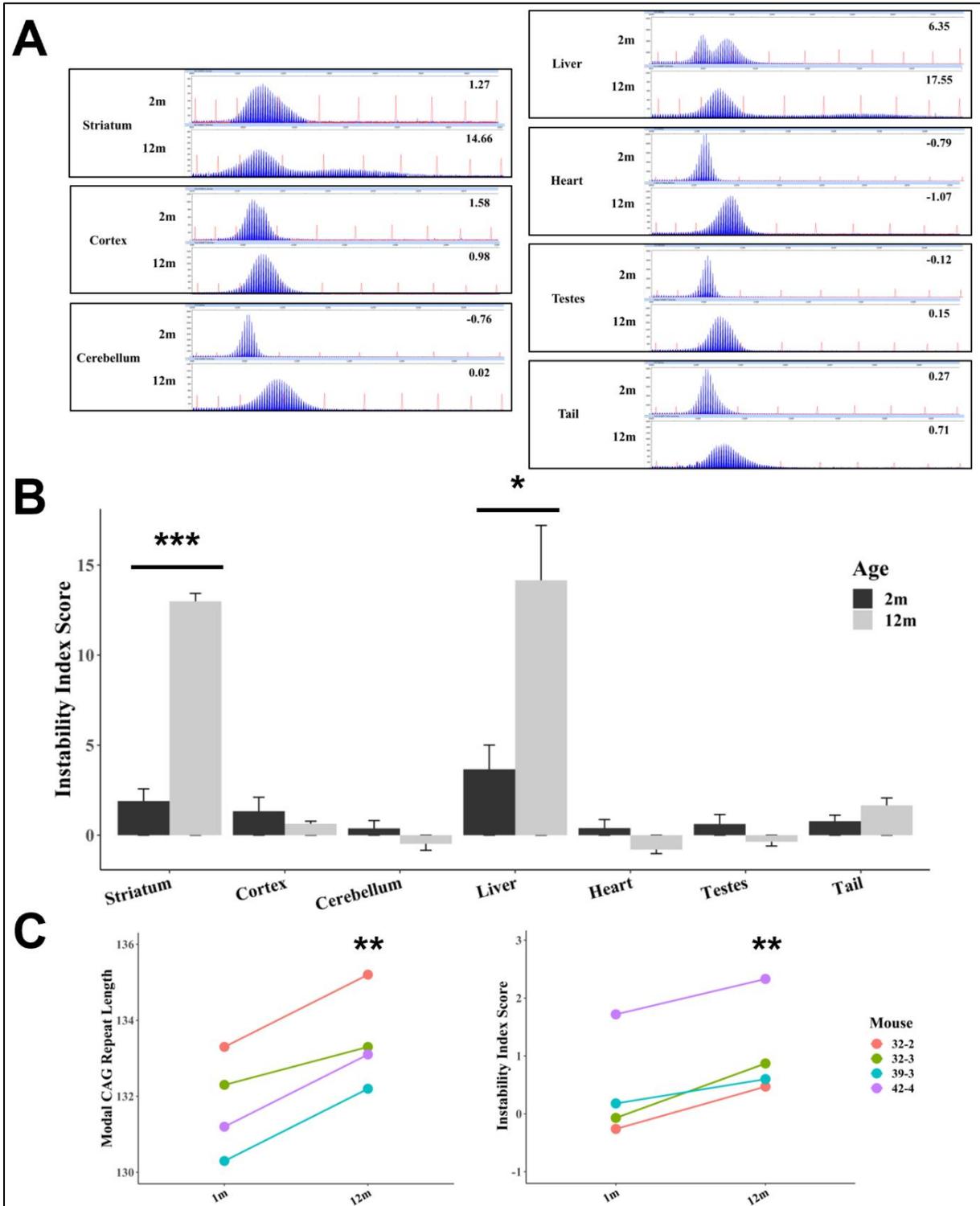
distance traveled (Figure 16B). 2m BAC-CAG mice show increased time spent in the perimeter of the open field box, which is an anxiety-like behavior. However, this phenotype was not seen at 12m. Lastly, 12m BAC-CAG mice showed an increase in number vertical plane entries (i.e. rearing), which is an exploratory behavior, and was not seen at 2m. Overall, BAC-CAG mice did not exhibit robust phenotypes using night-time openfield testing.

### **3.7 BAC-CAG mice exhibit age-dependent somatic instability of *mHTT* CAG repeats, which correlates with activity/sleep deficits**

Somatic *mHTT* CAG repeat instability has become a major focus of HD research, due to its potential role in modifying HD age of onset (GeM-HD, 2015, 2019; Swami et al., 2009). Similar to our study in Fan1-KD; Q140 and Q140 mice in chapter 2, we used a PCR-based protocol established in Vanessa Wheeler's lab to visualize and quantify repeat instability (Lee et al., 2010). We found that BAC-CAG mice show an age-dependent increase in somatic *mHTT* CAG repeat instability in the striatum and liver, but not cortex, cerebellum, heart, testes, or tail (Figure 17A-B). This replicates what is previously reported in Q111 KI mice (Lee et al., 2011; Lee et al., 2010).

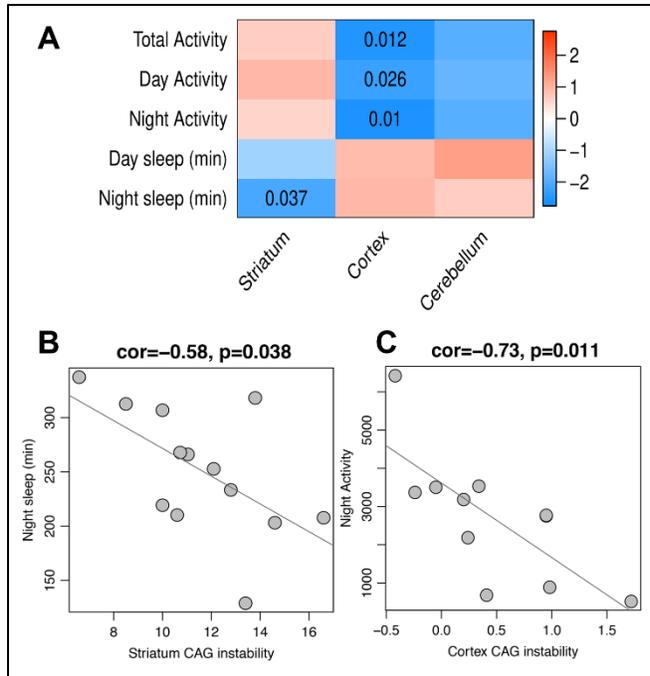
To confirm that the increase in CAG repeat instability is an effect of mouse age, and not an effect of the cohort of mice assayed, we measured the tail CAG repeat instability and modal CAG repeat length from the same mice at 1m and 12m of age. Mouse tail was used due to its relative ease of access and minimal impact on the health of the mice. We found that both tail modal CAG repeat length and IIS increased from 1m to 12m, in each of the BAC-CAG mice tested (Figure 17C). Thus, BAC-CAG mice show age-dependent increase in CAG repeat instability and growth of modal CAG repeat length.

CAG repeat instability in post-mortem human cortex correlates with onset of HD (Swami et al., 2009). We therefore hypothesized that instability indices in brain tissues should correlated with behavioral deficits in HD mice. To test this hypothesis, we correlated the IIS of cortex, striatum, and cerebellum from 12m BAC-CAG mice with their activity/sleep behavior. We found a significant negative correlation between striatum IIS and night-time sleep (Figure 18A-B). This suggests that increased striatal *mHTT* CAG repeat expansion impairs night-time sleep in BAC-



**Figure 17: Age-dependent somatic CAG repeat instability in BAC-CAG mice.**

(A) Representative GeneMapper traces of various tissues showing somatic CAG repeat instability in 2m and 12m BAC-CAG mice. (B) Bar plot showing mean IIS in various tissues from 2m and 12m BAC-CAG mice.  $n=4-28$  mice/tissue/age. Student's T test. (C) Tail modal CAG repeat length and IIS from the same BAC-CAG mice at two different ages. Paired T-test  $*p<0.05$ ,  $**p<0.01$ ,  $***p<0.001$ , error bars = S.E.M.



**Figure 18: Somatic mHTT CAG repeat instability correlate with activity and sleep deficits.** (A) Tombstone plot showing positive correlations (red) and negative correlations (blue) between the IIS of various brain tissues from 12m male BAC-CAG mice and activity/sleep-related behaviors. Only boxes with p-values shown are significant. (B-C) Scatter plots showing the correlation between striatal and cortical IIS and activity/sleep-related behavior. n=13 male mice.

CAG mice. Mice are nocturnal and night-time sleep can be thought of as a “day-time nap-like” behavior in humans.

The cortex and cerebellum IISs in BAC-CAG mice is relatively low and both are generally thought of as a “stable” mHTT CAG repeat tissues in Q111 KI models of HD (Lee et al., 2010). Surprisingly, cortex IIS, but not cerebellum, was correlated with several locomotor activity measures including day, night, and total activity (Figure 18A and 18C). These results suggest that somatic *mHTT* CAG repeat instability in the cortex and striatum may act as a modifier of activity/sleep deficits

and possibly other behavioral deficits as well.

### 3.8 Discussion

The current study helps uncover how mHTT uninterrupted CAG repeats directly contribute to HD pathology. We’ve shown that patient-like, pure CAG repeats cause earlier mHTT accumulation and NI formation, compared to BACHD and YAC128 mouse models. It also led of a robust striatal transcriptionopathy that is highly concordant with that in the allelic series KI models and HD patients. We also established that striatal transcriptionopathy in full-length HD models is predicted by uninterrupted CAG repeat length, and not polyQ length. Thus, our current study provides definitive evidence that a human genomic transgenic model of HD with

long uninterrupted CAG repeats can produce many of the phenotypes missing in the BACHD and YAC128 models. Our results provide insight into the role of uninterrupted mHTT CAG repeats in directly contributing to HD pathogenesis.

Our study is the first to show that CAG repeat instability, or more specifically IIS in brain tissues, correlate with behavioral impairments in HD mouse models. Vanessa Wheeler's Lab has previously shown that somatic instability correlates with mHTT nuclear accumulation (Wheeler, 2003; Kovalenko, 2012; Pinto, 2013), whether mHTT aggregates play a toxic or protective role remains unclear (Finkbeiner, 2011). We provided evidence that somatic instability correlates with activity/sleep-related behavior in an HD mouse model. The BAC-CAG model will be a useful resource in future investigations to understand how somatic CAG repeat instability, and MMR genes, modify HD pathogenesis.

BAC-CAG mice represent an important advance in HD mouse modeling. BAC-CAG mice share the strength of both allelic series KI mice and the previous genomic HD models, in terms of expressing a full-length human HTT protein, displaying robust behavioral phenotypes, and developing progressive mHTT nuclear accumulation and age-dependent transcriptional dysregulation in the striatum. BAC-CAG mice are uniquely suited to study novel therapeutic approaches to delaying HD onset. Previous genomic HD models are ideal for testing HTT lowering therapies, as they contain the entire human *HTT* genomic locus, but the presence of CAA interruption limited their utility in testing therapeutics to reduce CAG repeat instability. Fragment and KI HD models, on the other hand, are ideal for testing therapeutics to dampen CAG repeat instability but have limited utility in testing HTT lowering drugs due to the mostly murine genomic context. Thus, BAC-CAG mice are the first HD model to allow both HTT lowering and CAG repeat instability dampening therapeutics to be tested within the same model. Moreover, BAC-CAG mice allow for testing of the synergistic potential of these novel

therapeutic approaches as an adjunct therapy which, in theory, would increase the potential to delay HD onset in human patients.

## CHAPTER 4:

### Conclusions and future directions

#### 4.1 *Fan1*'s role in the brain

Previously, the most parsimonious explanation of how *Fan1* acts as a modifier in HD is that *Fan1* protects vulnerable cell-types (e.g. striatal MSNs) from somatic mHTT CAG repeat instability. Thus, loss of *Fan1* may cause increase expansion of CAG repeats, leading to an earlier onset of HD symptoms. However, the results presented in Chapter 2, do not support this hypothesis. While, *Fan1*-KD led to increased instability of liver CAG repeats at 6m, it had had minimal impact on somatic instability in the brain, even at 12m old.

Instead, *Fan1* may act as a modifier through its role in the cortex and more specifically through OLs. There are several lines of evidence suggesting cortical somatic CAG repeat instability is an important driver of HD pathology and that *Fan1* may be involved in the process. (1) Highly expanded mHTT alleles have been found in the cortex of R6/2 and postmortem human brains (Kennedy et al., 2003; Larson et al., 2015). (2) Somatic mHTT CAG repeat instability in the cortex of HD patients correlated with age of onset (Swami et al., 2009). (3) In chapter 3, we found that cortical IIS is correlated with locomotor behavior in 12m BAC-CAG mice. (4) (GeM-HD, 2019) identified a haplotype which is significantly associated with a delayed onset of HD and is correlated with increased expression of *Fan1* in the cortex. (5) In chapter 2, we found that 12m *Fan1*-KD; Q140 mice showed significantly increased cortical IIS using our 1x PCR protocol, although this difference became non-significant using 5x PCR (Figure 11B-C).

Having shown that *Fan1*-KD in WT and Q140 mice leads to a decrease in cortical OL gene expression and that the *Fan1*-KD; Q140 striatum shows upregulation microglia/immune-related gene are increased mHTT aggregation, this suggests *Fan1* plays a protective role in

cortical OLs, possibly through protecting against somatic instability, and has a non-cell autonomous effect in the striatum. Thus, loss of *Fan1* may lead to earlier mHtt-induced white matter lesions, deficits in cortico-striatal communication and trophic support and exacerbation of striatal pathology. Future studies should verify some of our preliminary findings in chapter 2 and begin testing the updated model of *Fan1*'s role in the brain. It will be important to show that *Fan1*-KD; Q140 mice have increased white matter lesions and impaired cortico-striatal communication. As well, it will also be important to show that specific cortical cell-types, such as CPNs or OLs, have increased somatic mHtt CAG repeat instability. Additionally, one could also test if *Fan1* plays a role in the cortex by crossing a conditional *Fan1* KO lines with cell-type or cortex-specific Cre line to see if selectively knocking-down *Fan1* in these cell-types recapitulate the phenotypes seen in our *Fan1*-KD mice.

#### **4.2 Uninterrupted mHTT CAG repeats modify HD-related behavior deficits**

In chapter 3, we described the generation and characterization of a novel human genomic model of HD which contains mHTT with long uninterrupted CAG repeats. We showed these mice develop (1) robust behavioral phenotypes, (2) nuclear accumulation of mHTT, (3) age-dependent striatal transcriptionopathy which is highly concordant with DE genes seen in the allelic series KI mice, and (4) age-dependent CAG repeat instability. We also demonstrated that somatic mHTT CAG repeat instability measurements in the cortex and striatum, but not cerebellum, correlate with behavioral deficits in 12m BAC-CAG mice.

An important validation of the potential of somatic instability dampening therapeutics would be to show that preventing somatic instability in the striatum or cortex improves HD symptoms. This could be tested by in BAC-CAG by showing that MMR lowering drug or crossing BAC-CAG mice to MMR genes (*Mlh1*, *Msh2*, etc) knocked-out lines improves activity/sleep-related deficits. Alternatively, we have shown that *Fan1*-KD may increase cortical

IIS (Figure 11B), and thus crossing BAC-CAG mice to Fan1-KD mice may increase cortical CAG repeat instability and exacerbate activity-related behavior deficits.

#### **4.3 Novel mouse models to aid investigations of *Fan1*'s role in HD**

Fan1's role in ICL repair has been studied extensively and loss of function mutations in *Fan1* are known to cause, KIN, a chronic kidney disease (Airik et al., 2016; Kratz et al., 2010; Law et al., 2020; Smogorzewska et al., 2010; Thongthip et al., 2016; Zhou et al., 2012). But Fan1's role in brain cells has not yet been comprehensively studied. An exhaustive profiling of direct Fan1 interactions, including protein-protein interactors and protein-nucleotide interactors (DNA/RNA), would greatly aid in understanding the cellular pathways that *Fan1* is directly involved in. As well, a brain-wide mapping of Fan1's expression in both health and disease would aid in understanding Fan1's cell-specific roles in the brain. Unfortunately, no validated antibodies for Fan1 exist to date. Thus, we developed a GFP-tagged *Fan1* KI (GFP-Fan1) mouse model to aid our investigation of *Fan1*'s role in the HD brain. As part of my thesis, I have worked with the University of California, Irvine (UCI) to develop the GFP-Fan1 model (described in the appendix; Figure 20) and have used these mice to confirm *Fan1* expression in neurons and glia, including OLs (Figure 8D).

Our lab will cross GFP-Fan1 mice with two mouse models of HD (Q140 and BAC-CAG) to study the following 3 aims. (1) Determine brain region and cell-type specific expression of Fan1 in wildtype and HD mice at different ages. (2) Compile an exhaustive list of *Fan1* protein-protein interactors using co-immunoprecipitation (coIP) and mass spectrometry and apply systems biology approaches to gain insight into the cellular pathways involving *Fan1*. (3) Identify which regions of the genome *Fan1* directly interacts with by performing ChIPseq. By completing these three aims, we hope to better understand the expression of Fan1 in different brain regions and cell-types in normal and HD mice, gain deeper insight into the molecular

pathways involving *Fan1*, and determine whether *Fan1* is selectively recruited in vivo to repetitive regions of the genome, such as trinucleotide repeats.

A key question in understanding *Fan1*'s role as a modifier of HD onset is whether the *Fan1* R507H missense mutation, which is the most significantly associated SNP (rs150393409) in the GeM-HD GWAS (GeM-HD, 2019), causes a loss of *Fan1* function. Sarah Tabrizi's lab has previously characterized this mutation in vitro and found that it does not impair *Fan1*'s role in ICL repair or in protecting against CAG repeat instability (Goold et al., 2019). To better characterize the impact of this human SNP on *Fan1* function in vivo, we developed a KI mouse model with an equivalent SNP to the one found in humans, called *Fan1*-R510H mice. A description of how these mice were generated can be found in the appendix (Figure 19).

Our goal is to cross *Fan1*-R510H mice with Q140 mice to determine whether *Fan1*-R510H; Q140 mice recapitulate the phenotypes seen in *Fan1*-KD; Q140 mice, including increased aggregation in the striatum and increased liver instability at 6m of age. If *Fan1*-R510H can recapitulate one or all of these phenotypes in Q140 mice, it would suggest that R507H causes a loss of *Fan1* function. It would also suggest that *Fan1* is a protective modifier in HD.

GeM-HD identified several *FAN1* haplotypes which were significantly associated with modifying HD age of onset. One haplotype which delays HD onset, is significantly correlated with increased expression of *FAN1* mRNA in the cortex, but not striatum. To test the hypothesis that a physiological increase in *FAN1* expression is protective in HD, we generated a novel BAC transgenic mouse model which overexpresses human *FAN1*, known as BAC-*FAN1* mice. A description of how we generated this mouse model can be found in the appendix (Figure 22 and Figure 23).

We plan to cross these mice with two mouse models of HD (Q140 and BAC-CAG) to test whether overexpression of human *FAN1* can rescue HD pathology. These mice will be tested for transcriptional changes, pathological accumulation of mHTT, behavioral (sleep/activity)

changes, and somatic *mHTT* CAG repeat instability. Completion of this study will provide compelling evidence to determine whether *FAN1* overexpression is protective in HD and would highlight the potential of *FAN1* as being a viable therapeutic target to delay the onset of HD in human patients.

#### **4.4 Concluding thoughts**

We also showed how uninterrupted CAG repeat in *mHTT* are directly involved in molecular pathologies, separate from encoding glutamine. While HD models with interrupted CAG repeats (e.g. BACHD and YAC128) recapitulate many HD phenotypes, *mHTT* nuclear accumulation, NIs, and transcriptional changes in the striatum seem to be related to length of uninterrupted CAG repeats. The BAC-CAG model will be an invaluable for investigations of pure CAG repeat-dependent pathological mechanism of HD. As well, they will serve as a pivotal resource for development of novel HD therapeutic to delay onset of HD.

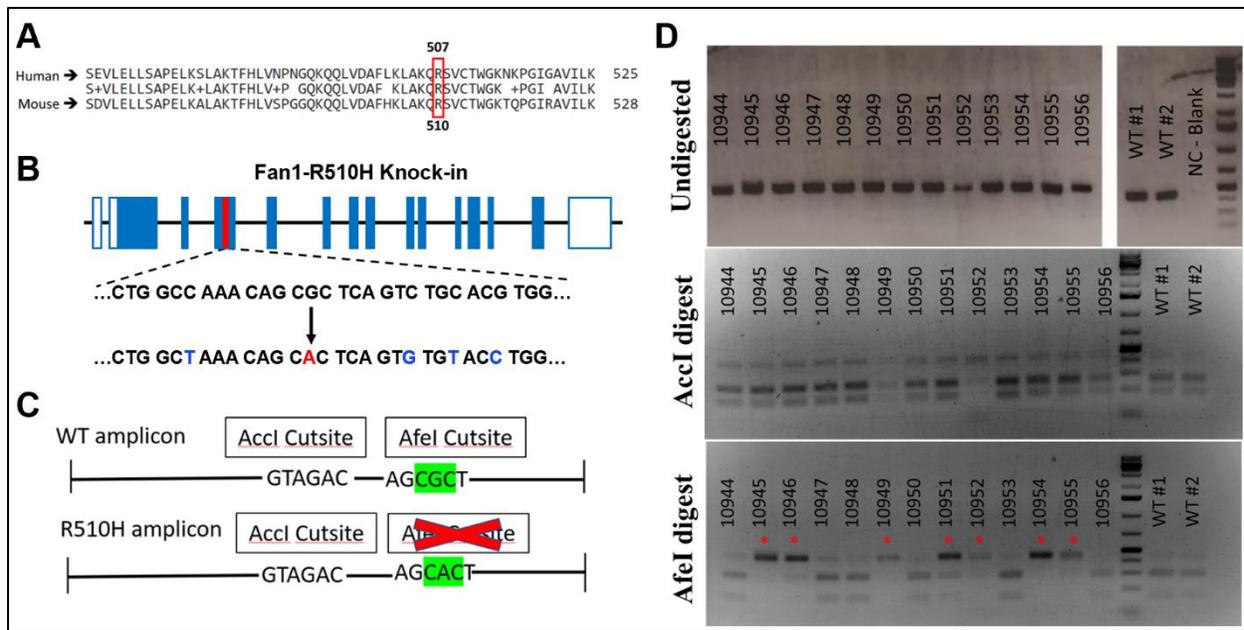
Occam's razor would suggest that *FAN1*, as a DNA repair nuclease that interacts with MMR proteins, would likely act as a HD modifier through protection against CAG repeat instability in striatal MSNs. But our current data does not support this hypothesis. Rather, data from GeM-HD GWAS and our RNAseq results suggest *Fan1* plays a significant HD-related role in the cortex, potentially through stabilizing somatic CAG repeat instability. And the increase in striatal *mHTT* aggregation we observed in 12m *Fan1*-KD; Q140 mice suggests *Fan1*-KD also contributes to striatal HD pathology, either through cell autonomous or non-cell autonomous mechanisms. More fully elucidating the brain regions and cell-types expressing *Fan1* and the molecular mechanisms involving *Fan1* in the brain will undoubtedly yield deeper insights, not only into how *Fan1* acts as a modifier, but also into fundamental HD biology.

## APPENDIX

### Mouse models to aid in investigations of *Fan1*

#### A.1 Generation and validation of *Fan1*-R510H and GFP-*Fan1* mice

The human missense R507H variant of *FAN1* (rs150393409) is significantly associated with an earlier onset of HD motor symptoms (GeM-HD, 2015, 2019). To better understand the functional significance of this variant, we generated a murine KI mouse line with an equivalent to the human *FAN1* R507H variant, which in the mouse is corresponding to the murine R510 residue (Figure 19A). We used CRISPR-Cas9 directed homology-dependent repair to introduce a targeted mutation in exon 4 of the murine *Fan1* gene. Working with the UC core facility at UC Irvine (i.e. UCI Transgenic Mouse Facility or UCI-TMF), we injected two guide-RNAs along with Cas9 and a single-stranded oligonucleotide (ssOligo) in >200 mouse oocytes and implant into pseudo pregnant dams. The ssOligo contained silent mutations to destroy the CRISPR cutsite after KI, which prevents re-cleavage by Cas9 and increases the efficiency of generating the targeted mutation (Figure 19B). 13 live pups were born and screened for the R510H knock-in using a PCR- and restriction digest-based assay (Figure 19C). Briefly, KI of the R510H variant destroys an *AfeI* restriction enzyme cut site. Primers flanking the R510H knock-in site were used to PCR amplify KI region. PCR products were then digested with either *AfeI* (which should only digest PCR product from WT alleles) or *AccI* (which should digest PCR products from both WT an R510H KI alleles). Using this assay, we identified 7 potential founders of *Fan1*-R510H allele (Figure 19D). Sanger sequencing of the mutated allele confirmed the successful KI of the R510H point mutations without unintended mutations occurring. We crossed the mosaic

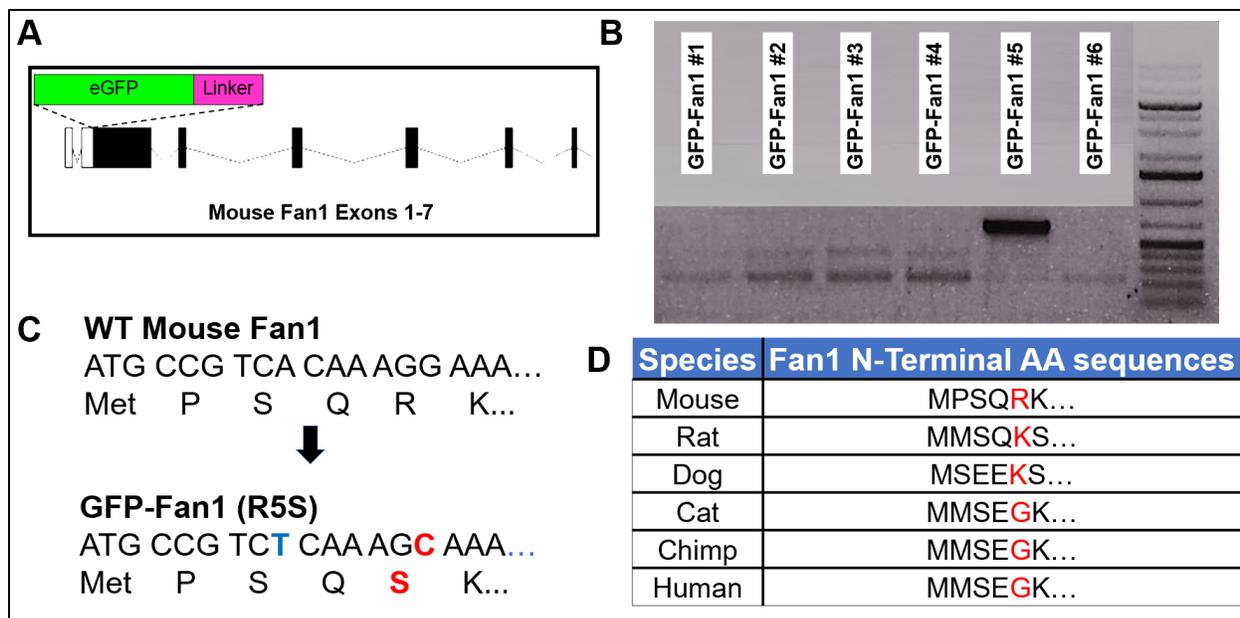


**Figure 19: Generation of Fan1-R510H KI mice.**

**(A)** Alignment of Human and mouse Fan1 amino acid sequence showing that human FAN1 amino acid 507 is equivalent to mouse Fan1 amino acid 510. **(B)** Schematic showing silent (blue letters) and missense (red letter) mutations created by 510H knock-in. **(C)** Diagram showing how R510H mutation destroys Afel cut site which was used to genotype mosaic founders. **(D)** Results of genotyping mosaic founders by PCR and restriction enzyme digest. Red asterisk highlights pups which have a mutation at R510.

founders with WT mice to establish F1 mice and confirmed the successful germline transmission of the *Fan1*-R510H KI allele.

Using a similar principle as described above, we worked with UCI-TMF to develop a *Fan1* KI mouse encoding an N-terminal GFP fused to the endogenous *Fan1* protein by a 21 amino acid linker sequence (GFP-Fan1; Figure 20A). The linker amino acid sequence used (“ARGYQTLISLYKKAGSAAAPFT”) was suggested by a postdoc in Agata Smogorzewska lab, Ryan White PhD. To generate the GFP-Fan1 model, two gRNAs were used to create double-strand breaks near the N-terminus of murine *Fan1*, in exon2 of the murine *Fan1* gene. The gRNAs were co-injected into C57/BL6J zygotes with Cas9 protein and a long synthesized single strand DNA containing GFP and linker sequences flanked by *Fan1* genomic DNA sequence (i.e. homology arms). The embryos were implanted into pseudo-pregnant surrogate females and resulting pups were screened by PCR using one primer 5' of the knock-in site and one primer

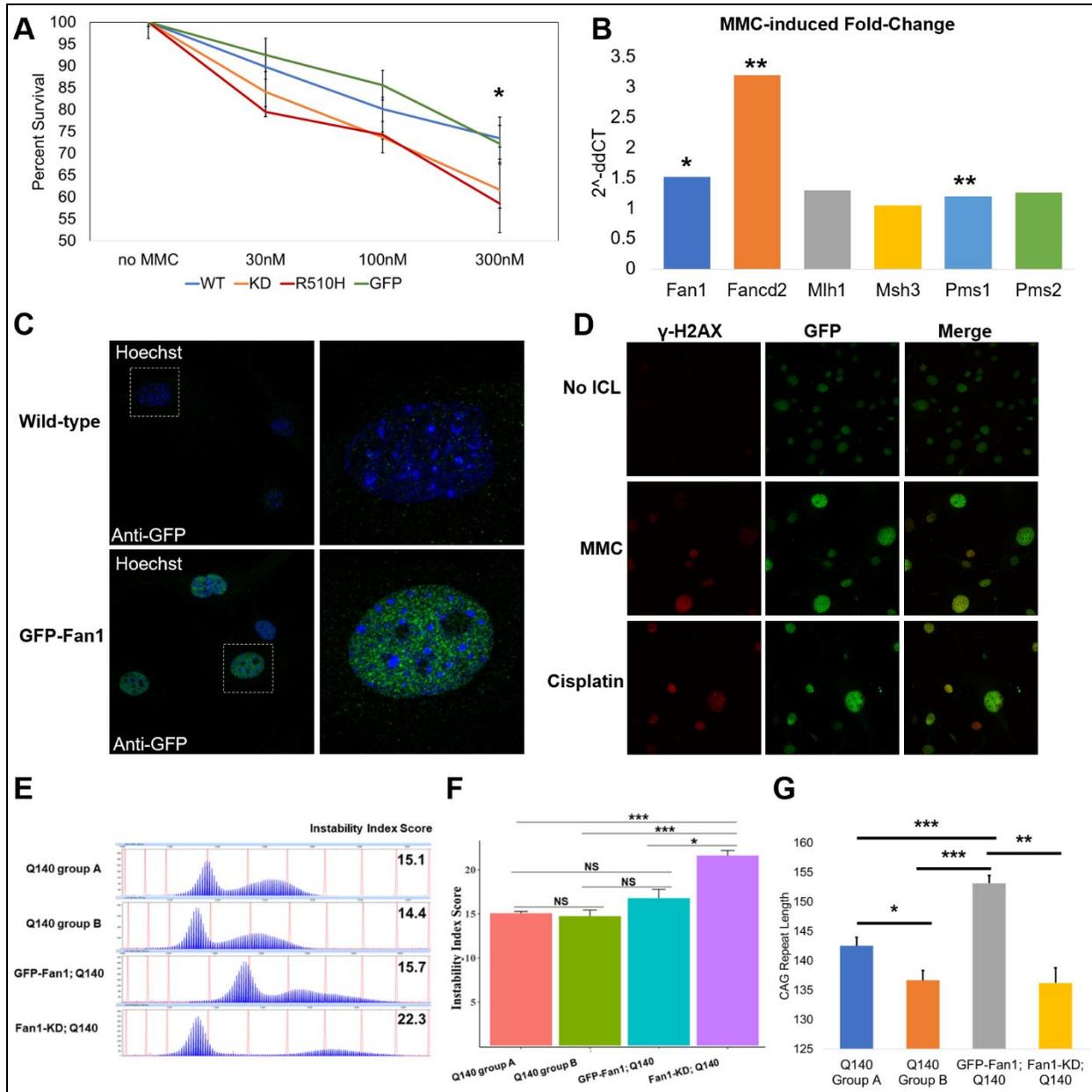


**Figure 20: Generation of GFP-Fan1 KI mice.**

**(A)** Schematic showing GFP and linker sequencing being knocked-in to mouse *Fan1* exon 2. **(B)** Agarose gel showing one pup with correct GFP knock-in **(C)** Diagram showing how in the process of creating the GFP knock-in, on silent mutation (blue) and one missense mutation (red) were created in the N-terminus of *Fan1*. **(D)** Comparison of several species showing that the R5S missense mutation is not in a highly conserved amino acid.

within the GFP-knock-in (Figure 20B). We obtained one correctly targeted GFP-Fan1 transgenic line that conferred germline transmission of the GFP-Fan1 allele. We performed sanger sequencing of the N-terminal *Fan1* region (not shown) to confirm the successful knock-in of the GFP and linker sequences. From sequencing, we found one non-conserved amino acid residue change at position 5 of *Fan1* (R5S; Figure 20C). This residue is R and K in rodents, but G in cat, chimp and human, suggesting it is not highly conserved in mammals (Figure 20D). This suggests that the R5S variant in GFP-Fan1 KI mice is unlikely to alter *Fan1* function.

To determine whether the R510H or GFP-Fan1 KI mutations impact the function of *Fan1*, we generated fibroblast cell lines from wildtype, *Fan1*-KD, GFP-Fan1 and *Fan1*-KD mice. These fibroblast cell lines were acutely exposed to increasing concentrations of MMC and cell survival was measured using Cell Counting Kit-8 (Dojindo). We found that acute exposure of high concentrations of MMC (300nM) lead to a significant increase in cell death in *Fan1*-KD and



**Figure 21: Validation of Fan1-R510H and GFP-Fan1 KI mice.**

(A) Assay for Fan1 function by testing the sensitivity of different fibroblast lines acute exposure of MMC. \* $p < 0.05$  comparing R510H and WT. (B) Fold change of genes in the mismatch repair and Fanconi anemia pathway in fibroblasts that were acutely exposed to MMC. (C) 64x confocal image of GFP-Fan1 subcellular localization in fibroblast after acute exposure to MMC. Cells were fixed and stained with GFP antibodies to amplify signal. (D) Comparing GFP-Fan1 fluorescent intensity at baseline to after acute exposure to ICL-inducing agents MMC or cisplatin. (E) Representative GeneMapper traces from 6m mouse liver with the IIS score of each trace shown on the right. (F) Mean instability index score of 6m mouse liver ( $n=6$ /genotype). (G) Mean modal CAG repeat length from 6m mouse liver ( $n=6$ /genotype). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , student's t-test.

*Fan1*-R510H fibroblasts lines, compared to WT and GFP-*Fan1* cell lines ( $p < 0.05$ , single-factor ANOVA and student's t-test; Figure 21A). This suggests that *Fan1*-R510H fibroblasts are hypofunctional and that GFP-*Fan1* mice possess normal *Fan1* function.

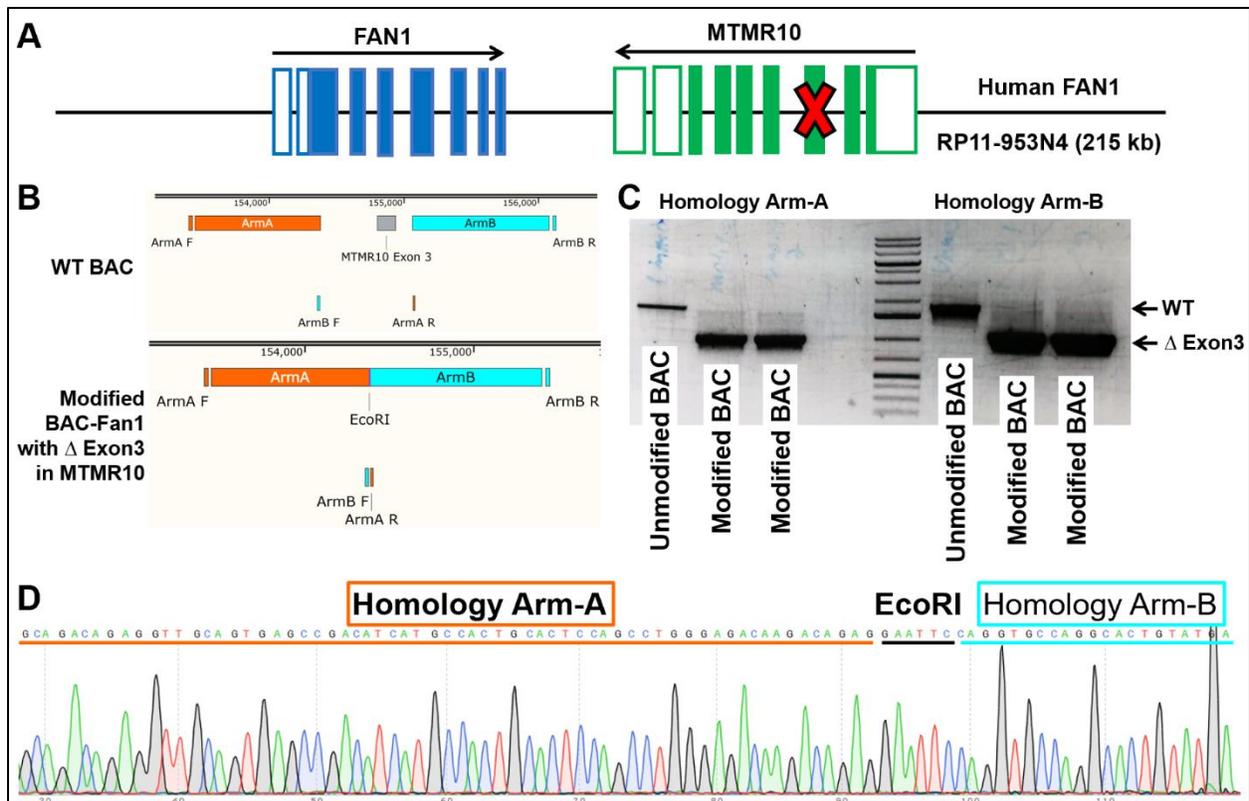
To determine whether the N-terminal GFP KI impacts *Fan1*'s ability to prevent somatic expansion of liver mHTT CAG repeats, we crossed GFP-*Fan1* mice to Q140 mice. We measured the CAG repeat instability of the liver from two independent cohorts of Q140 mice, GFP-*Fan1*; Q140 mice, and *Fan1*-KD; Q140 mice at 6m old (Figure 21E-F). We found that the two cohorts of Q140 mice showed no difference in IIS, confirming the replicability of results using this assay. We also found that there was a significant increase in the *Fan1*-KD; Q140 liver, compared to Q140 or GFP-*Fan1*; Q140, confirming our results in chapter 2 (Figure 8A). Lastly, we found no significant difference between Q140 and GFP-*Fan1*; Q140 mice, suggesting that the N-terminal GFP KI and R5S mutations do not impact *Fan1*'s function in somatic instability of CAG repeats.

We noticed that the modal CAG repeat lengths (i.e. the tallest peak in the GeneMapper traces) were significantly different between the different cohort of mice (Figure 21G). Group B Q140 mice showed a significantly shorter modal CAG repeat compared to Group A Q140 mice. Similarly, GFP-*Fan1*; Q140 mice had a significantly longer modal CAG repeat length than the other 3 cohorts. We believe this is an effect of the mHTT allele inherited from different Q140 breeders used to generate these different cohort of mice, rather than effect of the N-terminal GFP KI or R5S mutations. This is because the two groups of Q140 mice showed a significant difference, even without a mutation in *Fan1*. As well, *Fan1*-KD; Q140 mice showed no difference from Q140 mice, suggesting loss of *Fan1* function does not increase modal CAG repeat length in the liver. Thus, it is most likely that GFP-*Fan1*; Q140 cohort of mice inherited a mHTT allele with a longer modal CAG repeat length than the other 3 genotypes.

We next characterized GFP-Fan1's subcellular localization. We used WT mouse fibroblast cells to test if acute exposure to MMC would increase the mRNA expression of *Fan1* or other DNA repair genes using qRT-PCR. We found that *Fan1*, and two direct interactors of *Fan1* (*Fancd2* and *Pms1*) showed a significant increase in mRNA expression in response to acute exposure to MMC. (Figure 21B). We then exposed WT and GFP-Fan1 fibroblasts to MMC before fixing the cells and staining with GFP antibodies. We found that wildtype cell showed minor background GFP-antibody staining, while GFP-Fan1 fibroblasts showed robust nuclear punctate expression of GFP-Fan1 (Figure 21C). We next wanted to see if acute exposure of ICL-inducing agents would increase GFP-Fan1 protein expression level, like it induced an increase mRNA transcript expression. We acutely exposed GFP-Fan1 cells to either MMC or Cisplatin, before fixing and staining with GFP antibodies. We found that without ICL-inducing agents, GFP-Fan1 showed moderate GFP-signal in the nucleus (Figure 21D). But after acute exposure to either MMC or cisplatin, the GFP fluorescence increase dramatically. This suggests that acute exposure of ICL-inducing agents leads to an increase in *Fan1* mRNA and protein expression, and the *Fan1* is localized mostly to the nucleus.

## **A.2 Generation of BAC-FAN1 mice**

One of the *FAN1* haplotypes identified by the GeM-HD (GeM-HD, 2019) GWAS was significantly associated with delayed onset of HD and was correlated with increased expression of *Fan1* in the cortex. This suggests that overexpression of *FAN1* in HD may be protective. To test this hypothesis, we develop a human *FAN1* BAC transgenic mouse line. We obtained a human BAC containing the *FAN1* gene (RP11-953N4) from BACPAC Resources at Children's Hospital Oakland Research Institute. This BAC contained two genes, *FAN1* and *MTMR10* which were encoded on opposite strands. In order to selectively overexpress *FAN1*, we needed to KO *MTMR10* by delete one of the exons on the BAC (Figure 22A). We choose to delete *MTMR10*

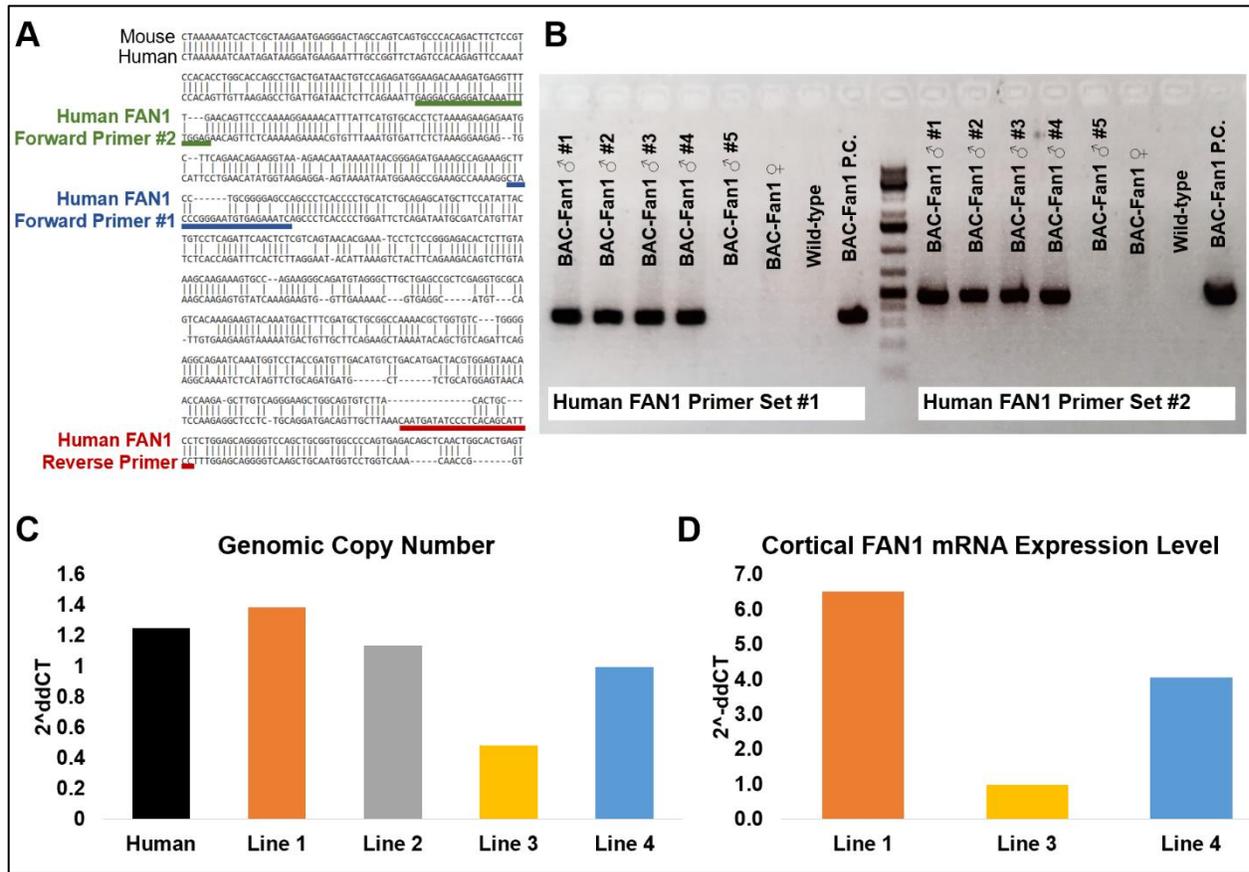


**Figure 22: Modifying human FAN1 BAC**

(A) Schematic of the genes contained within Human BAC RP11-953N4. (B) Diagram of the BAC modification to delete MTMR10 exon 3 and replace with EcoRI restriction enzyme cut site. (C) PCR to confirm MTMR10 exon 3 deletion by using PCR primers that will amplify through either homology arm A or homology arm B. (D) Representative sanger sequencing result showing that MTMR10 exon 3 has been replaced by an EcoRI site.

exon3 because it would cause a frameshift mutation and should lead to nonsense mediated decay of the transcripts. We used homologous recombination to replace the MTMR10 exon3 with an EcoRI site (Figure 22B). We then confirmed the deletion of MTMR10 exon3 using PCR with two pairs primers (i.e. ArmA and ArmB primer sets; Figure 22B). The ArmA forward primer was 5' homology Arm A and the reverse primer was on the 5' end of ArmB. Similarly, the ArmB forward primer was on the 3' end of ArmA and the reverse primer was just 3' of Arm B. In this way, the unmodified BAC would yield a larger amplicon and the modified BAC with MTMR10 exon3 deleted would yield a smaller band. Using both primer sets, identified several modified BACs which had (Figure 22C), which were subsequently sequenced to confirm that MTMR10 exon 3 was replaced with an EcoRI site (Figure 22D).

With MTMR10 successfully knocked-out, we linearized and purified our modified *FAN1* BAC. Working with the Cedars Sinai Transgenic Core, we microinjected the modified *FAN1* BAC into FvB/NJ fertilized oocytes and implant into pseudo-pregnant dams. 6 mosaic pups were born (5 male and 1 female), which were screened for the *FAN1* BAC transgene by PCR with human *FAN1* specific primers (Figure 23A-B). 4 of the 6 pups were positive for the *FAN1* BAC transgene and are referred to as BAC-*FAN1* lines 1-4. All 4 BAC-*FAN1* lines were able to germline transmit the BAC-*FAN1* transgene.



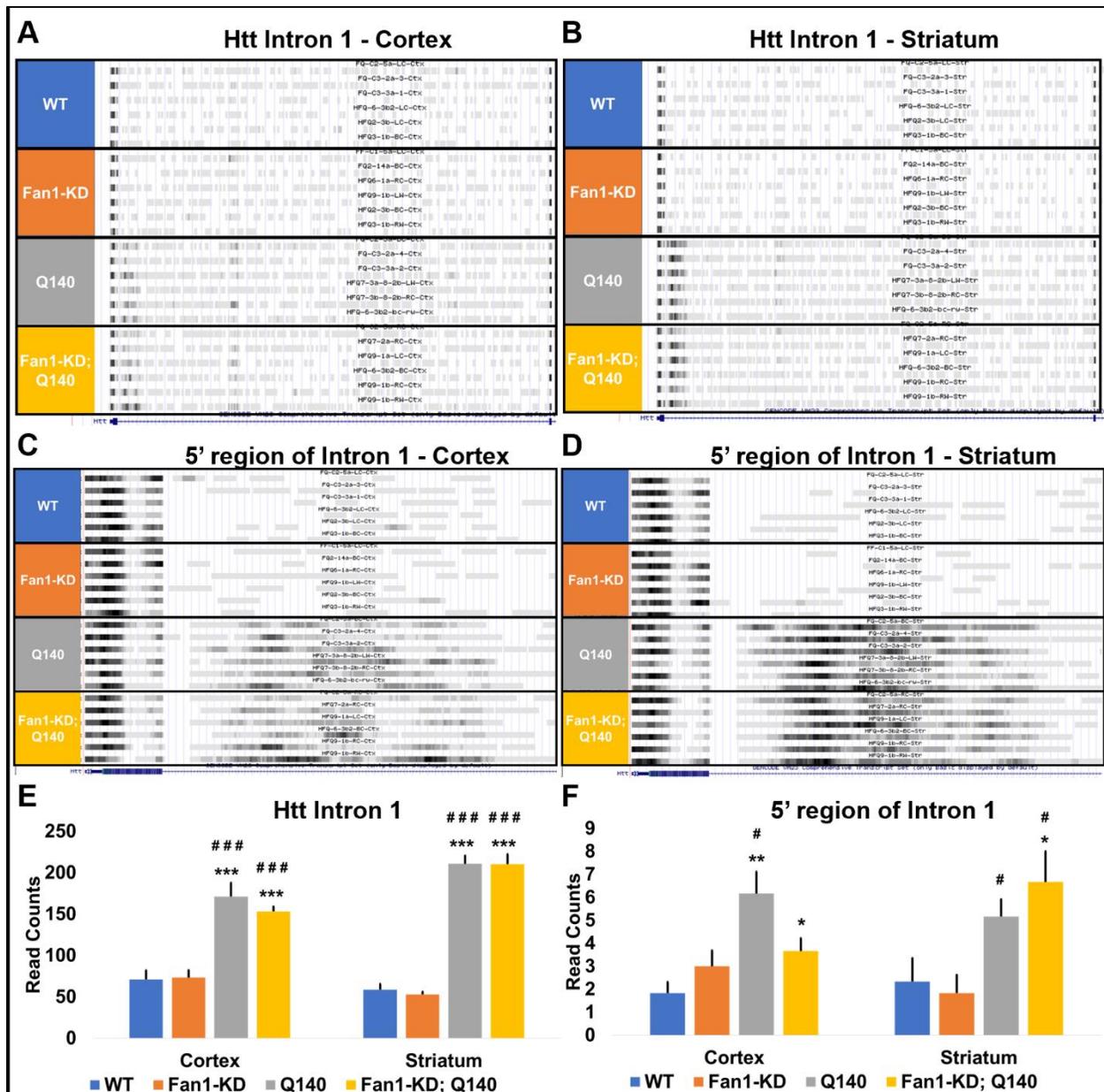
**Figure 23: Generation of BAC-*FAN1* mice**

(A) Schematic showing how primers were designed to amplify specifically from the human *FAN1* gene. (B) Genotyping results identifying 4 transgenic pups which contain integrated BAC-Fan1. BAC-Fan1 P.C. was amplified using the modified *FAN1* BAC DNA as a positive control (C) qRT-PCR result showing the genomic copy number of human *Fan1* BACs integrated in the 4 lines which were generated. (D) qRT-PCR result comparing *FAN1* mRNA expression in 3 of the BAC-*FAN1* lines which were generated.

We then performed qRT-PCR to determine the copy number of BAC-*FAN1* transgenes integrated into the mouse genome in each BAC-*FAN1* line. By comparing the  $2^{\Delta\Delta Ct}$  between the BAC-*FAN1* lines and a human genomic sample (which has 2 copies of *FAN1*), we were able to determine that BAC-*Fan1* lines 1, 2, and 4 have two copies of BAC-*FAN1* integrated, and line 3 has one copy. Next compared the expression of human *FAN1* mRNA expression in the cortex between three of the BAC-*FAN1* lines. We found that BAC-*FAN1* had the highest expression of *FAN1* in the cortex and will be used as the main BAC-*FAN1* line in future studies.

### **A.3 Fan1-KD does not impact the expression of alternatively spliced mHTT transcript with intron1 retention**

Incompletely splicing of *HTT* mRNA is seen in HD mouse models and human HD patients and generates a truncated mRNA comprising *HTT* exon1 and the 5' region of intron1 (Neueder et al., 2017; Sathasivam et al., 2013). The expression of this read-through product is proportional to CAG repeat length and leads to the production of the exon 1 *HTT* protein fragments, which are the most pathogenic *HTT* fragments. To test whether Fan1-KD affects the read-through of exon1 and the expression of *HTT* exon1-intron1 mRNA transcript, we mapped the reads from our 12m RNAseq study in the striatum (Figure 4) and cortex (Figure 6) of WT, Fan1-KD, Q140, and Fan1-KD; Q140 mice (Figures 23A-B). We found Q140 and Fan1-KD; Q140 mice show increased read counts in *HTT* intron1 in both the cortex and striatum, compared to WT and Fan1-KD mice (Figure 24E-F). The increased number of reads seemed to be enriched in the 5' region of intron1 (Figure 24), supporting the idea that this reflects the truncated *HTT* exon1-intron1 transcript. However, we did not see a significant difference between Q140 and Fan1-KD; Q140 mice in either the cortex or striatum (Figure 23E-F). This



**Figure 24: Expression of alternatively spliced mHTT transcript with intron1 retention**

(A-B) Mapping of RNAseq reads in Htt intron1 from 12m striatum and cortex. (C-D) Mapping of RNAseq reads in 5' region of Htt intron1 from 12m striatum and cortex. (E-F) Bar graphs show the mean read counts within Htt intron1 and 5' region of intron1. n=6 mice/genotype, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001 compared to WT, #p<0.05, ##p<0.01, ###p<0.001 compared to Fan1-KD; Q140, student's t-test.

suggests that *Fan1* does not play a role in alternative splicing of this truncated *HTT* exon1-intron1 mRNA or the regulation of this transcript.

## MATERIALS AND METHODS

### Mouse Breeding, Maintenance, and Genotyping

Fan1-KD, GFP-Fan1, and Q140 mice were bred, maintained in the C57BL/6J background. Fan1-KD Hets were crossed with Q140 mice obtain Fan1-KD Het; Q140 mice. Fan1-KD Het; Q140 mice were then bred with Fan1-KD Het mice to obtain the following 4 genotypes: WT, Fan1-KD, Q140, and Fan1-KD; Q140. GFP-Fan1 Het mice were crossed with Q140 mice obtain GFP-Fan1 Het; Q140 mice. GFP-Fan1 Het; Q140 mice were then bred with GFP-Fan1 Het mice to obtain to obtain genotypes: WT, GFP-Fan1, Q140, and GFP-Fan1; Q140 mice. Although only WT and GFP-Fan1 mice were used for the studies described. Mice were bred and housed under standard conditions, consistent with NIH guidelines and approved by the UCLA Chancellor's Animal Research Committee. Mice were genotyped with the following primer sequences (all sequences are shown in the 5' to 3' direction):

Allele	Forward Primer	Reverse Primer
Fan1-KD WT	GCTTTCAGAGCTTCCTCGTCTGG	CAGAGTGAGTTCTAGGACAGGCAGG
Fan1-KD KD	GGGATCTCATGCTGGAGTTCTTCG	CAGAGTGAGTTCTAGGACAGGCAGG
GFP-Fan1 WT	GGTCGCCCACTTTCTCTAATAG	GGTGGTGATAATTTCTGAGGTGTC
GFP-Fan1 KI	GGTCGCCCACTTTCTCTAATAG	GAAGTCGATGCCCTTCAGC
Q140	CTGCACCGACCGTGAGTCC	GAAGGCACTGGAGTCGTGAC
BAC-CAG	GCAACGTGCGTGTCTCTG	TGTTCCCAAAGCCTGCTC
Fan1-R510H WT	CAAACAGCGCTCAGTCTGCACG	CTGTCTGGCTGAGAGCAGTG
Fan1-R510H KI	CTAAACAGCACTCAGTGTGTACC	CTGTCTGGCTGAGAGCAGTG
BAC-FAN1	CTACCCGGGAATGTGAGAAATC	GGAATGCTGTGAGGGATATCATTG

### Transgene construct and generation of BAC-CAG mice

*mHTT* exon1 containing expanded CAG repeats was PCR amplified from DNA template of *mHTT*-Exon1 plasmid (gift from Dr. G. Bates) and cloned into shuttle vector pLD53-SCAB.

Multiple clones were picked up and CAG repeat sizes were sequenced at Laragen ( Los Angeles, CA). We selected the clone with the longest CAG repeat size [(CAG)<sub>130</sub>-CAA-(CAG)<sub>10</sub>-CAACAG] and cultured in LB medium in 30 °C incubator. Purified plasmid DNAs were used to modify human huntingtin BAC (RP11-866L6) clone as reported previously (Yang et al., 1997; Gu et al., 2005). The modified BAC DNAs were confirmed by sequencing before subjected to microinject into FvB/NJ fertilized oocytes at the Cedars Sinai Transgenic Core to generate BAC-CAG mice. Ten founders were identified by PCR amplification and sequencing. We selected two lines with relatively higher transgene expression levels in Western blot assay for further analysis. Transgenic mice were bred and maintained under standard conditions, consistent with NIH guidelines. All procedures complied with regulations and were approved by the UCLA Institutional Animal Care and Use Committee.

#### **Quantitation of mHTT expression by Western blot and MSD assays**

Three mice each from 10 founder lines and BACHD positive control were used for mHTT quantification by Western blot assay probed with MAB5492 antibody (anti-human HTT polyproline domain; EMD Millipore, Billerica, MA), 1C2, and  $\alpha$ -tubulin as reported previously (Gray et al 2008; Gu et al. 2009). Briefly, brains from 6-week old mice of each line were dissected. 40  $\mu$ g protein from each forebrain lysate was mixed with NuPAGE LDS loading buffer (Invitrogen) and heated for 10 min at 70°C, then resolved on 3-8% Tris-Acetate NuPAGE gel (Invitrogen). After protein transfer to PVDF membrane, blots were probed with respective antibodies. We used human huntingtin specific and polyQ-independent MAB5492 antibody for detection and quantification. Densitometric values from scanned Western blot films were obtained using ImageQuantTL software (GE Healthcare).

Total, expanded and aggregated forms of mHTT protein levels in cortex, striatum and cerebellum of 2, 6 and 12 month old of ages were measured using MSD method based on Tr-

FRET Immunoassay platforms carried out at Evotec® (www.evotec.com). Briefly, about 50 µg tissues of cortex, striatum and cerebellum were dissected out from 4 mice of each groups of 2, 6 and 12 month old of ages of BAC-CAG and its WT littermates. Frozen tissues were sent to Evotec and total *mHTT* proteins were measured using 2B7/4C9 antibody pair; expanded *mHTT* proteins 2B7/MW1 and aggregated *mHTT* MW8/4C9 antibody pairs.

### **Behavioral Tests**

Accelerating Rotarod test is used to evaluate the motor coordination and motor skill learning of the mouse. Briefly, mice of 2 month old of age were trained to stay on the bars for 2 continuous days at 4 rpm steady speed; then tested for 3 days at rotate speed from 4 to 40 rpm within 5 minutes. Three trials per day were carried out for each mouse. The time of the testing mouse remind on the bar was recorded. Mice were tested at 2, 6 and 12 months of age (Gray et al. 2008; Gu et al. 2009). The mice were coded, and the investigator was blind to the genotype of the mice during testing and analysis.

The grip strength test is used to measure the neuromuscular function. Remove a mouse from its home cage, Lower the mouse over the grid keeping the torso horizontal and allowing only its forepaws to attach to the grid. Gently pull the mouse back by its tail ensuring the mouse grips the top portion of the grid and the torso remains horizontal and record the maximal grip strength value. Repeat this procedure twice more. Average of the three readings as the final result.

Open field test is used to measure locomotion and psychiatric-like phenotypes. Automated open field analysis was used to assess locomotor activity. Mice were tested in the automated open field (Coburn) at 12m. Sessions lasted 15 minutes.

### **Fresh Frozen Tissue collection**

Tissue was dissected from WT, Fan1-KD, Q140, and Fan1-KD; Q140 mice at 6m and 12m. Brain (cortex, striatum, cerebellum) and peripheral tissue (heart, liver, gonads, tail) were dissected simultaneously and immediately flash-frozen on dry ice.

### **Analysis of CAG repeat instability**

About 100 µg tissues of cortex, striatum, cerebellum, heart, liver, gonad and tail dissected from 4 mice for each groups of 6 and 12 month old mice, and were digested with proteinase K overnight at 55°C water bath. Genomic DNAs were isolated using phenol: chloroform extraction and precipitated with sodium acetate /ethanol, washed extensively with 70% ethanol. The *mHTT* CAG repeat was amplified using a human-specific PCR assay from sample DNAs. The forward primer was fluorescently labeled with 6-FAM (Applied Biosystems) and products were resolved using the ABI 3730xl DNA analyzer (Applied Biosystems) with GeneScan 500 LIZ as internal size standard (Applied Biosystems). GeneMapper v5 (Applied Biosystems) was used to generate CAG repeat size distribution traces. Expansion and contraction indices were quantified from the GeneMapper CAG repeat distributions as reported (Lee, et al. 2010). Briefly, peaks below 50 relative fluorescent units (RFUs) were excluded from analysis. Peak heights were then normalized to the sum of all peak heights. Normalized peak heights were multiplied by the change in CAG repeat length of each peak relative to the highest peak and summed to generate an instability index score representing the mean CAG length change from the main allele.

### **RNA Extraction and mRNA sequencing**

Four male and female littermate mice of each genotype (WT, Fan1-KD Het, Fan1-KD, Q140, Fan1-KD Het; Q140, Fan1-KD; Q140) were used for 6m RNAseq study. Three male and female littermate mice of each genotype (WT, Fan1-KD, Q140, Fan1-KD; Q140) were used for 12m

RNAseq study. Animals were deeply anesthetized with pentobarbital prior to transcardial perfusion with 0.1 mM phosphate buffered saline (PBS), pH7.4, made with DEPC-treated water. Brains were dissected in ice cold PBS, and flash-frozen on dry ice. Tissue was stored at -80°C until all tissues were ready for RNA preparation. Tissues were disrupted and homogenized in Trizol (Invitrogen; Carlsbad, CA) before total RNA was extracted using Qiagen (Valencia, CA) RNeasy kit with QIAshredder columns, per the manufacturers' recommendations, including on-column DNase digestion. Total RNAs were stored at -80°C until use, avoiding freeze-thaw cycles. RNA quality was measured using the Agilent Bioanalyzer system according to the manufacturer's recommendations and cDNA libraries were simultaneously generated and sequenced using an Illumina HiSeq 2000 sequence using strand-specific, paired-end, 50-mer sequencing protocols to a minimum read depth of 40 million reads per sample, as previously described (Langfelder P., et al. 2016). Clipped reads were aligned to mouse genome mm9 using the STAR aligner using default settings. Read counts for individual genes were obtained using HTSeq.

### **mRNA expression data preprocessing**

Expression data from each age were analyzed as a separate set. We retained only genes with at least 1 count per million aligned reads in at least  $\frac{1}{4}$  of the total number of samples. The rationale is to only include genes that are likely to be expressed in at least half of the samples of each genotype group. We used a modified version of the sample network methodology originally described in (Oldham et al., 2012). Specifically, to quantify inter-sample connectivity, we first transformed the raw counts using variance stabilization (R function `varianceStabilizingTransformation`) and then used Euclidean inter-sample distance based on the scaled profiles of the 8000 genes with highest mean expression. The intersample connectivities  $k$  were transformed to Z scores using robust standardization,

$$Z_a = \frac{k_a - \text{median}(k)}{1.4826 \times \text{MAD}(k)},$$

where index  $a$  labels samples, MAD is the median absolute deviation, a robust analog of standard deviation, and the constant 1.4826 ensures asymptotic consistency (approximate equality of MAD and standard deviation for large, normally distributed samples). Finally, samples with  $Z_a < -6$  were flagged as outliers.

### Differential expression analysis

To make DE analysis more robust against potential outlier measurements (counts) that may remain even after outlier sample removal, we defined individual observation weights designed to downweigh potential outliers. The weights are constructed separately for each gene. First, Tukey bi-square-like weights  $\lambda$  (Wilcox, 2012) are calculated for each (variance-stabilized) observation  $x_a$  (index  $a$  labels samples) as

$$\lambda_a = (1 - u_a^2)^2,$$

where

$$u_a = \min \left\{ 1, \frac{|x_a - \text{median}(x)|}{9 \text{MMAD}(x)} \right\}.$$

The median is calculated separately for each gene across all samples. MMAD stands for modified MAD, calculated as follows. For each gene, we first set  $\text{MMAD} = \text{MAD}$ . The following conditions are then checked separately for each gene: (1) 10th percentile of the weights is at least 0.1 (that is, the proportion of observations with weights  $< 0.1$  is less than 10%) (Langfelder and Horvath, 2012) and (2) for each individual genotype, 40th percentile of the weights is at least 0.9 (that is, at least 40% of the observation have a high coefficient of at least 0.9). If both conditions are met,  $\text{MMAD} = \text{MAD}$ . If either condition is not met, MMAD equals the lowest value for which both conditions are met. The rationale is to exclude outliers but ensure that the

number of outliers is not too large either overall or in each genotype group. This approach has previously been used in (Lee et al., 2018).

DE analysis was carried out using DESeq2 (Love et al., 2014) version 1.22.2 with default arguments except for disabling outlier replacement (since we use weights to downweigh potential outliers) and independent filtering (since we have pre-filtered genes based on expression levels).

### **Enrichment analysis**

Enrichment calculations were carried out using R package anRichment (<https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/GeneAnnotation/>) that implements standard Fisher exact test and a multiple-testing correction across all query and reference gene sets.

### **Antigen Retrieval and Immunohistochemical Staining**

Immunostaining was carried out using published methods (Gu et al., 2009). S830 (gift from Dr. G. Bates; 1:10,000 dilution) antibody was used for mHTT aggregate detection. GFAP antibody (Sigma, 1:20,000 dilution) was used for detection of reactive astrocytes; NeuN antibody (Millipore, 1:1,000) was used to identify neurons. Ubiquitin antibody (Dako, 1:1,000 dilution) was used for detection of ubiquitin. Actn2 antibody (1:500) was used to quantification of its expression in striatum. For double immunofluorescent staining (e.g. S830/NeuN and S830/Ubiquitin) we added the S830 antibody and incubated at 4°C for one day, then added NeuN or Ubi antibody and incubated at 4°C overnight.

Images were acquired using a Dragonfly High-Speed Confocal Microscope 200 and the Fusion 2.0 software package (Andor, Oxford Instrument). Laser and detector settings were maintained

constant for the acquisition of each immunostaining. For all analyses, at least three images were taken per brain region and slide using  $\times 60$  oil objective lens, at  $2,048 \times 2,048$  pixel resolution, with z-step size of  $1 \mu\text{m}$  at  $10 \mu\text{m}$  thickness.

### **Immunofluorescent Staining.**

12mo mice from the *Fan1*-KD x Q140 cross were used for neuropathology studies (N = 6 per genotype). 6m GFP-*Fan1* mice were used for the *Fan1* cell-type expression study. All mice were transcardially perfused using 0.1M PBS and post-fixed with 4% PFA . Fixed brains were incubated in 30% sucrose in 0.1M PSB for 3 days at 4oC. Brains were then flash frozen and kept in -80oC freezer. Brains were sectioned using a cryostat cut  $40\mu\text{m}$  thick brain sections. Immunofluorescent staining was performed using the following antibodies. Anti-Actn2 (Epitomics EP2529Y, 1:800), Anti-VGlu1 (EMD Millipore MAB5502, 1:500), Anti-Em48 (Molecular Biology Reagents MAB5374; 1:350), anti-NeuN (Millipore MAB377, 1:1,000), and anti-GFP (Novus NB600-308, 1:2000). These were followed by the compatible Alexa Fluor conjugated secondary antibodies (Life technologies goat anti-rabbit and goat anti-mouse Alexa Fluor-594 A11037 and A11005, goat anti-rabbit Alexa Fluor-488 A11008, 1:300). Sections were imaged using a Zeiss LSM800 confocal laser scanning microscope and immunofluorescent intensity measured using ImageJ (NIH). All images for each study were taken under identical conditions and quantitated blind to genotype.

## References

- Agus, F., Crespo, D., Myers, R.H., and Labadorf, A. (2019). The caudate nucleus undergoes dramatic and unique transcriptional changes in human prodromal Huntington's disease brain. *BMC Med Genomics* 12, 137.
- Airik, R., Schueler, M., Airik, M., Cho, J., Porath, J.D., Mukherjee, E., Sims-Lucas, S., and Hildebrandt, F. (2016). A FANCD2/FANCI-Associated Nuclease 1-Knockout Model Develops Karyomegalic Interstitial Nephritis. *J Am Soc Nephrol* 27, 3552-3559.
- Aizman, O., Brismar, H., Uhlen, P., Zettergren, E., Levey, A.I., Forsberg, H., Greengard, P., and Aperia, A. (2000). Anatomical and physiological evidence for D1 and D2 dopamine receptor colocalization in neostriatal neurons. *Nat Neurosci* 3, 226-230.
- Andresen, J.M., Gayan, J., Djousse, L., Roberts, S., Brocklebank, D., Cherny, S.S., Group, U.S.-V.C.R., Group, H.M.C.R., Cardon, L.R., Gusella, J.F., *et al.* (2007). The relationship between CAG repeat length and age of onset differs for Huntington's disease patients with juvenile onset or adult onset. *Ann Hum Genet* 71, 295-301.
- Arrasate, M., and Finkbeiner, S. (2012). Protein aggregates in Huntington's disease. *Exp Neurol* 238, 1-11.
- Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., and Finkbeiner, S. (2004). Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* 431, 805-810.
- Aziz, N.A., Pijl, H., Frolich, M., van der Graaf, A.W., Roelfsema, F., and Roos, R.A. (2010). Leptin secretion rate increases with higher CAG repeat number in Huntington's disease patients. *Clin Endocrinol (Oxf)* 73, 206-211.
- Banez-Coronel, M., Ayhan, F., Tarabochia, A.D., Zu, T., Perez, B.A., Tusi, S.K., Pletnikova, O., Borchelt, D.R., Ross, C.A., Margolis, R.L., *et al.* (2015). RAN Translation in Huntington Disease. *Neuron* 88, 667-677.
- Banez-Coronel, M., Porta, S., Kagerbauer, B., Mateu-Huertas, E., Pantano, L., Ferrer, I., Guzman, M., Estivill, X., and Marti, E. (2012). A pathogenic mechanism in Huntington's disease involves small CAG-repeated RNAs with neurotoxic activity. *PLoS Genet* 8, e1002481.
- Bartzokis, G., Lu, P.H., Tishler, T.A., Fong, S.M., Oluwadara, B., Finn, J.P., Huang, D., Bordelon, Y., Mintz, J., and Perlman, S. (2007). Myelin breakdown and iron changes in Huntington's disease: pathogenesis and treatment implications. *Neurochem Res* 32, 1655-1664.
- Bates, G.P., Dorsey, R., Gusella, J.F., Hayden, M.R., Kay, C., Leavitt, B.R., Nance, M., Ross, C.A., Scahill, R.I., Wetzel, R., *et al.* (2015). Huntington disease. *Nat Rev Dis Primers* 1, 15005.
- Bayram-Weston, Z., Stone, T.C., Giles, P., Elliston, L., Janghra, N., Higgs, G.V., Holmans, P.A., Dunnett, S.B., Brooks, S.P., and Jones, L. (2015). Similar striatal gene expression profiles in the striatum of the YAC128 and HdhQ150 mouse models of Huntington's disease are not reflected in mutant huntingtin inclusion prevalence. *BMC Genomics* 16, 1079.

Ben-Shachar, S., Lanpher, B., German, J.R., Qasaymeh, M., Potocki, L., Nagamani, S.C., Franco, L.M., Malphrus, A., Bottenfield, G.W., Spence, J.E., *et al.* (2009). Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J Med Genet* 46, 382-388.

Benraiss, A., Wang, S., Herrlinger, S., Li, X., Chandler-Militello, D., Mauceri, J., Burm, H.B., Toner, M., Osipovitch, M., Jim Xu, Q., *et al.* (2016). Human glia can both induce and rescue aspects of disease phenotype in Huntington disease. *Nat Commun* 7, 11758.

Bezzi, P., Gundersen, V., Galbete, J.L., Seifert, G., Steinhauser, C., Pilati, E., and Volterra, A. (2004). Astrocytes contain a vesicular compartment that is competent for regulated exocytosis of glutamate. *Nat Neurosci* 7, 613-620.

Bronner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoe, M.K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., and *et al.* (1994). Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* 368, 258-261.

Brooks, S., Higgs, G., Jones, L., and Dunnett, S.B. (2012a). Longitudinal analysis of the behavioural phenotype in Hdh(CAG)150 Huntington's disease knock-in mice. *Brain Res Bull* 88, 182-188.

Brooks, S., Higgs, G., Jones, L., and Dunnett, S.B. (2012b). Longitudinal analysis of the behavioural phenotype in HdhQ92 Huntington's disease knock-in mice. *Brain Res Bull* 88, 148-155.

Brooks, S.P., Betteridge, H., Trueman, R.C., Jones, L., and Dunnett, S.B. (2006). Selective extra-dimensional set shifting deficit in a knock-in mouse model of Huntington's disease. *Brain Res Bull* 69, 452-457.

Brown, A.M., and Ransom, B.R. (2007). Astrocyte glycogen and brain energy metabolism. *Glia* 55, 1263-1271.

Browne, S.E., and Beal, M.F. (2006). Oxidative damage in Huntington's disease pathogenesis. *Antioxid Redox Signal* 8, 2061-2073.

Burrus, C.J., McKinstry, S.U., Kim, N., Ozlu, M.I., Santoki, A.V., Fang, F.Y., Ma, A., Karadeniz, Y.B., Worthington, A.K., Dragatsis, I., *et al.* (2020). Striatal Projection Neurons Require Huntingtin for Synaptic Connectivity and Survival. *Cell Rep* 30, 642-657 e646.

Carroll, J.B., Warby, S.C., Southwell, A.L., Doty, C.N., Greenlee, S., Skotte, N., Hung, G., Bennett, C.F., Freier, S.M., and Hayden, M.R. (2011). Potent and selective antisense oligonucleotides targeting single-nucleotide polymorphisms in the Huntington disease gene / allele-specific silencing of mutant huntingtin. *Mol Ther* 19, 2178-2185.

Carson, M.J., Thrash, J.C., and Walter, B. (2006). The cellular response in neuroinflammation: The role of leukocytes, microglia and astrocytes in neuronal death and survival. *Clin Neurosci Res* 6, 237-245.

Carty, N., Berson, N., Tillack, K., Thiede, C., Scholz, D., Kottig, K., Sedaghat, Y., Gabrysiak, C., Yohrling, G., von der Kammer, H., *et al.* (2015). Characterization of HTT inclusion size, location, and timing in the zQ175 mouse model of Huntington's disease: an in vivo high-content imaging study. *PLoS One* 10, e0123527.

Cattaneo, E., Zuccato, C., and Tartari, M. (2005). Normal huntingtin function: an alternative approach to Huntington's disease. *Nat Rev Neurosci* 6, 919-930.

Cepeda, C., Hurst, R.S., Calvert, C.R., Hernandez-Echeagaray, E., Nguyen, O.K., Jocoy, E., Christian, L.J., Ariano, M.A., and Levine, M.S. (2003). Transient and progressive electrophysiological alterations in the corticostriatal pathway in a mouse model of Huntington's disease. *J Neurosci* 23, 961-969.

Cepeda, C., Wu, N., Andre, V.M., Cummings, D.M., and Levine, M.S. (2007). The corticostriatal pathway in Huntington's disease. *Prog Neurobiol* 81, 253-271.

Chung, D.W., Rudnicki, D.D., Yu, L., and Margolis, R.L. (2011). A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. *Hum Mol Genet* 20, 3467-3477.

Conlon, E.G., Lu, L., Sharma, A., Yamazaki, T., Tang, T., Shneider, N.A., and Manley, J.L. (2016). The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *Elife* 5.

Cooke, M.S., Evans, M.D., Dizdaroglu, M., and Lunec, J. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J* 17, 1195-1214.

Crook, Z.R., and Housman, D. (2011). Huntington's disease: can mice lead the way to treatment? *Neuron* 69, 423-435.

Crotti, A., Benner, C., Kerman, B.E., Gosselin, D., Lagier-Tourenne, C., Zuccato, C., Cattaneo, E., Gage, F.H., Cleveland, D.W., and Glass, C.K. (2014). Mutant Huntingtin promotes autonomous microglia activation via myeloid lineage-determining factors. *Nat Neurosci* 17, 513-521.

Cudkowicz, M., and Kowall, N.W. (1990). Degeneration of pyramidal projection neurons in Huntington's disease cortex. *Ann Neurol* 27, 200-204.

Cummings, D.M., Cepeda, C., and Levine, M.S. (2010). Alterations in striatal synaptic transmission are consistent across genetic mouse models of Huntington's disease. *ASN Neuro* 2, e00036.

Dandelot, E., and Gourdon, G. (2018). The flash-small-pool PCR: how to transform blotting and numerous hybridization steps into a simple denatured PCR. *Biotechniques* 64, 262-265.

de la Monte, S.M., Vonsattel, J.P., and Richardson, E.P., Jr. (1988). Morphometric demonstration of atrophic changes in the cerebral cortex, white matter, and neostriatum in Huntington's disease. *J Neuropathol Exp Neurol* 47, 516-525.

De Souza, R.A., and Leavitt, B.R. (2015). Neurobiology of Huntington's Disease. *Curr Top Behav Neurosci* 22, 81-100.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., *et al.* (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245-256.

Deng, Y.P., Wong, T., Bricker-Anthony, C., Deng, B., and Reiner, A. (2013). Loss of corticostriatal and thalamostriatal synaptic terminals precedes striatal projection neuron pathology in heterozygous Q140 Huntington's disease mice. *Neurobiol Dis* 60, 89-107.

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

Djousse, L., Knowlton, B., Hayden, M., Almqvist, E.W., Brinkman, R., Ross, C., Margolis, R., Rosenblatt, A., Durr, A., Dode, C., *et al.* (2003). Interaction of normal and expanded CAG repeat sizes influences age at onset of Huntington disease. *Am J Med Genet A* 119A, 279-282.

Dragatsis, I., Levine, M.S., and Zeitlin, S. (2000). Inactivation of Hdh in the brain and testis results in progressive neurodegeneration and sterility in mice. *Nat Genet* 26, 300-306.

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

Drummond, J.T., Li, G.M., Longley, M.J., and Modrich, P. (1995). Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* 268, 1909-1912.

Dumas, E.M., van den Bogaard, S.J., Ruber, M.E., Reilman, R.R., Stout, J.C., Craufurd, D., Hicks, S.L., Kennard, C., Tabrizi, S.J., van Buchem, M.A., *et al.* (2012). Early changes in white matter pathways of the sensorimotor cortex in premanifest Huntington's disease. *Hum Brain Mapp* 33, 203-212.

Duyao, M.P., Auerbach, A.B., Ryan, A., Persichetti, F., Barnes, G.T., McNeil, S.M., Ge, P., Vonsattel, J.P., Gusella, J.F., Joyner, A.L., and *et al.* (1995). Inactivation of the mouse Huntington's disease gene homolog Hdh. *Science* 269, 407-410.

Evans, S.J., Douglas, I., Rawlins, M.D., Wexler, N.S., Tabrizi, S.J., and Smeeth, L. (2013). Prevalence of adult Huntington's disease in the UK based on diagnoses recorded in general practice records. *J Neurol Neurosurg Psychiatry* 84, 1156-1160.

Farrer, L.A., Cupples, L.A., Wiaters, P., Conneally, P.M., Gusella, J.F., and Myers, R.H. (1993). The normal Huntington disease (HD) allele, or a closely linked gene, influences age at onset of HD. *Am J Hum Genet* 53, 125-130.

Ferrari Bardile, C., Garcia-Miralles, M., Caron, N.S., Rayan, N.A., Langley, S.R., Harmston, N., Rondelli, A.M., Teo, R.T.Y., Watti, S., Anderson, L.M., *et al.* (2019). Intrinsic mutant HTT-

mediated defects in oligodendroglia cause myelination deficits and behavioral abnormalities in Huntington disease. *Proc Natl Acad Sci U S A* 116, 9622-9627.

Fiacco, T.A., and McCarthy, K.D. (2004). Intracellular astrocyte calcium waves in situ increase the frequency of spontaneous AMPA receptor currents in CA1 pyramidal neurons. *J Neurosci* 24, 722-732.

Finkbeiner, S. (2011). Huntington's Disease. *Cold Spring Harb Perspect Biol* 3.

Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., and Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* 75, 1027-1038.

Fisher, E.R., and Hayden, M.R. (2014). Multisource ascertainment of Huntington disease in Canada: prevalence and population at risk. *Mov Disord* 29, 105-114.

Franich, N.R., Basso, M., Andre, E.A., Ochaba, J., Kumar, A., Thein, S., Fote, G., Kachemov, M., Lau, A.L., Yeung, S.Y., *et al.* (2018). Striatal Mutant Huntingtin Protein Levels Decline with Age in Homozygous Huntington's Disease Knock-In Mouse Models. *J Huntingtons Dis* 7, 137-150.

Frost, J.L., and Schafer, D.P. (2016). Microglia: Architects of the Developing Nervous System. *Trends Cell Biol* 26, 587-597.

GeM-HD (2015). Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell* 162, 516-526.

GeM-HD (2019). CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell* 178, 887-900 e814.

Glass, C.K., Saijo, K., Winner, B., Marchetto, M.C., and Gage, F.H. (2010). Mechanisms underlying inflammation in neurodegeneration. *Cell* 140, 918-934.

Godin, J.D., Colombo, K., Molina-Calavita, M., Keryer, G., Zala, D., Charrin, B.C., Dietrich, P., Volvert, M.L., Guillemot, F., Dragatsis, I., *et al.* (2010). Huntingtin is required for mitotic spindle orientation and mammalian neurogenesis. *Neuron* 67, 392-406.

Gokce, O., Bonhoeffer, T., and Scheuss, V. (2016a). Clusters of synaptic inputs on dendrites of layer 5 pyramidal cells in mouse visual cortex. *Elife* 5.

Gokce, O., Stanley, G.M., Treutlein, B., Neff, N.F., Camp, J.G., Malenka, R.C., Rothwell, P.E., Fuccillo, M.V., Sudhof, T.C., and Quake, S.R. (2016b). Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. *Cell Rep* 16, 1126-1137.

Gomes-Pereira, M., Fortune, M.T., Ingram, L., McAbney, J.P., and Monckton, D.G. (2004). Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum Mol Genet* 13, 1815-1825.

Goold, R., Flower, M., Moss, D.H., Medway, C., Wood-Kaczmar, A., Andre, R., Farshim, P., Bates, G.P., Holmans, P., Jones, L., and Tabrizi, S.J. (2019). FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. *Hum Mol Genet* 28, 650-661.

Goula, A.V., Berquist, B.R., Wilson, D.M., 3rd, Wheeler, V.C., Trottier, Y., and Merienne, K. (2009). Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. *PLoS Genet* 5, e1000749.

Goula, A.V., Stys, A., Chan, J.P., Trottier, Y., Festenstein, R., and Merienne, K. (2012). Transcription elongation and tissue-specific somatic CAG instability. *PLoS Genet* 8, e1003051.

Gray, M., Shirasaki, D.I., Cepeda, C., Andre, V.M., Wilburn, B., Lu, X.H., Tao, J., Yamazaki, I., Li, S.H., Sun, Y.E., *et al.* (2008). Full-length human mutant huntingtin with a stable polyglutamine repeat can elicit progressive and selective neuropathogenesis in BACHD mice. *J Neurosci* 28, 6182-6195.

Group, T.H.s.D.C.R. (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, 971-983.

Gutkunst, C.A., Li, S.H., Yi, H., Mulroy, J.S., Kuemmerle, S., Jones, R., Rye, D., Ferrante, R.J., Hersch, S.M., and Li, X.J. (1999). Nuclear and neuropil aggregates in Huntington's disease: relationship to neuropathology. *J Neurosci* 19, 2522-2534.

Gwon, G.H., Kim, Y., Liu, Y., Watson, A.T., Jo, A., Etheridge, T.J., Yuan, F., Zhang, Y., Kim, Y., Carr, A.M., and Cho, Y. (2014). Crystal structure of a Fanconi anemia-associated nuclease homolog bound to 5' flap DNA: basis of interstrand cross-link repair by FAN1. *Genes Dev* 28, 2276-2290.

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A., Shaw, P.G., Kim, M.S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., *et al.* (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507, 195-200.

Hammond, T.R., Dufort, C., Dissing-Olesen, L., Giera, S., Young, A., Wysoker, A., Walker, A.J., Gergits, F., Segel, M., Nemesh, J., *et al.* (2019). Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. *Immunity* 50, 253-271 e256.

Hedreen, J.C., Peyser, C.E., Folstein, S.E., and Ross, C.A. (1991). Neuronal loss in layers V and VI of cerebral cortex in Huntington's disease. *Neurosci Lett* 133, 257-261.

Heikkinen, T., Lehtimäki, K., Vartiainen, N., Puolivali, J., Hendricks, S.J., Glaser, J.R., Bradaia, A., Wadel, K., Touller, C., Kontkanen, O., *et al.* (2012). Characterization of neurophysiological and behavioral changes, MRI brain volumetry and 1H MRS in zQ175 knock-in mouse model of Huntington's disease. *PLoS One* 7, e50717.

Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suarez-Farinas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J., *et al.* (2008). A translational profiling approach for the molecular characterization of CNS cell types. *Cell* 135, 738-748.

- Heng, M.Y., Duong, D.K., Albin, R.L., Tallaksen-Greene, S.J., Hunter, J.M., Lesort, M.J., Osmand, A., Paulson, H.L., and Detloff, P.J. (2010). Early autophagic response in a novel knock-in model of Huntington disease. *Hum Mol Genet* 19, 3702-3720.
- Hickey, M.A., Kosmalska, A., Enayati, J., Cohen, R., Zeitlin, S., Levine, M.S., and Chesselet, M.F. (2008). Extensive early motor and non-motor behavioral deficits are followed by striatal neuronal loss in knock-in Huntington's disease mice. *Neuroscience* 157, 280-295.
- Hickey, M.A., Zhu, C., Medvedeva, V., Lerner, R.P., Patassini, S., Franich, N.R., Maiti, P., Frautschy, S.A., Zeitlin, S., Levine, M.S., and Chesselet, M.F. (2012). Improvement of neuropathology and transcriptional deficits in CAG 140 knock-in mice supports a beneficial effect of dietary curcumin in Huntington's disease. *Mol Neurodegener* 7, 12.
- Hodges, A., Strand, A.D., Aragaki, A.K., Kuhn, A., Sengstag, T., Hughes, G., Elliston, L.A., Hartog, C., Goldstein, D.R., Thu, D., *et al.* (2006). Regional and cellular gene expression changes in human Huntington's disease brain. *Hum Mol Genet* 15, 965-977.
- Hodgson, J.G., Agopyan, N., Gutekunst, C.A., Leavitt, B.R., LePiane, F., Singaraja, R., Smith, D.J., Bissada, N., McCutcheon, K., Nasir, J., *et al.* (1999). A YAC mouse model for Huntington's disease with full-length mutant huntingtin, cytoplasmic toxicity, and selective striatal neurodegeneration. *Neuron* 23, 181-192.
- Huang, B., Wei, W., Wang, G., Gaertig, M.A., Feng, Y., Wang, W., Li, X.J., and Li, S. (2015). Mutant huntingtin downregulates myelin regulatory factor-mediated myelin gene expression and affects mature oligodendrocytes. *Neuron* 85, 1212-1226.
- Jain, A., and Vale, R.D. (2017). RNA phase transitions in repeat expansion disorders. *Nature* 546, 243-247.
- Jeste, D.V., Barban, L., and Parisi, J. (1984). Reduced Purkinje cell density in Huntington's disease. *Exp Neurol* 85, 78-86.
- Kadyrov, F.A., Dzantiev, L., Constantin, N., and Modrich, P. (2006). Endonucleolytic function of MutLalpha in human mismatch repair. *Cell* 126, 297-308.
- Kadyrova, L.Y., Gujar, V., Burdett, V., Modrich, P.L., and Kadyrov, F.A. (2020). Human MutLgamma, the MLH1-MLH3 heterodimer, is an endonuclease that promotes DNA expansion. *Proc Natl Acad Sci U S A* 117, 3535-3542.
- Kay, C., Collins, J.A., Miedzybrodzka, Z., Madore, S.J., Gordon, E.S., Gerry, N., Davidson, M., Slama, R.A., and Hayden, M.R. (2016). Huntington disease reduced penetrance alleles occur at high frequency in the general population. *Neurology* 87, 282-288.
- Kennedy, L., Evans, E., Chen, C.M., Craven, L., Detloff, P.J., Ennis, M., and Shelbourne, P.F. (2003). Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum Mol Genet* 12, 3359-3367.

Keogh, N., Chan, K.Y., Li, G.M., and Lahue, R.S. (2017). MutSbeta abundance and Msh3 ATP hydrolysis activity are important drivers of CTG\*CAG repeat expansions. *Nucleic Acids Res* 45, 10068-10078.

Kim, M., Lee, H.S., LaForet, G., McIntyre, C., Martin, E.J., Chang, P., Kim, T.W., Williams, M., Reddy, P.H., Tagle, D., *et al.* (1999). Mutant huntingtin expression in clonal striatal cells: dissociation of inclusion formation and neuronal survival by caspase inhibition. *J Neurosci* 19, 964-973.

King, I.F., Yandava, C.N., Mabb, A.M., Hsiao, J.S., Huang, H.S., Pearson, B.L., Calabrese, J.M., Starmer, J., Parker, J.S., Magnuson, T., *et al.* (2013). Topoisomerases facilitate transcription of long genes linked to autism. *Nature* 501, 58-62.

Koenning, M., Jackson, S., Hay, C.M., Faux, C., Kilpatrick, T.J., Willingham, M., and Emery, B. (2012). Myelin gene regulatory factor is required for maintenance of myelin and mature oligodendrocyte identity in the adult CNS. *J Neurosci* 32, 12528-12542.

Kordasiewicz, H.B., Stanek, L.M., Wancewicz, E.V., Mazur, C., McAlonis, M.M., Pytel, K.A., Artates, J.W., Weiss, A., Cheng, S.H., Shihabuddin, L.S., *et al.* (2012). Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. *Neuron* 74, 1031-1044.

Kovalenko, M., Dragileva, E., St Claire, J., Gillis, T., Guide, J.R., New, J., Dong, H., Kucherlapati, R., Kucherlapati, M.H., Ehrlich, M.E., *et al.* (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, e44273.

Kratz, K., Schopf, B., Kaden, S., Sendoel, A., Eberhard, R., Lademann, C., Cannavo, E., Sartori, A.A., Hengartner, M.O., and Jiricny, J. (2010). Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. *Cell* 142, 77-88.

Kremer, B., Almqvist, E., Theilmann, J., Spence, N., Telenius, H., Goldberg, Y.P., and Hayden, M.R. (1995). Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. *Am J Hum Genet* 57, 343-350.

Kremer, H.P., Roos, R.A., Dingjan, G.M., Bots, G.T., Bruyn, G.W., and Hofman, M.A. (1991). The hypothalamic lateral tuberal nucleus and the characteristics of neuronal loss in Huntington's disease. *Neurosci Lett* 132, 101-104.

Kudo, T., Schroeder, A., Loh, D.H., Kuljis, D., Jordan, M.C., Roos, K.P., and Colwell, C.S. (2011). Dysfunctions in circadian behavior and physiology in mouse models of Huntington's disease. *Exp Neurol* 228, 80-90.

Kuhn, A., Goldstein, D.R., Hodges, A., Strand, A.D., Sengstag, T., Kooperberg, C., Becanovic, K., Pouladi, M.A., Sathasivam, K., Cha, J.H., *et al.* (2007). Mutant huntingtin's effects on striatal gene expression in mice recapitulate changes observed in human Huntington's disease brain and do not differ with mutant huntingtin length or wild-type huntingtin dosage. *Hum Mol Genet* 16, 1845-1861.

Labadorf, A., Hoss, A.G., Lagomarsino, V., Latourelle, J.C., Hadzi, T.C., Bregu, J., MacDonald, M.E., Gusella, J.F., Chen, J.F., Akbarian, S., *et al.* (2015). RNA Sequence Analysis of Human Huntington Disease Brain Reveals an Extensive Increase in Inflammatory and Developmental Gene Expression. *PLoS One* 10, e0143563.

Labbadia, J., and Morimoto, R.I. (2013). Huntington's disease: underlying molecular mechanisms and emerging concepts. *Trends Biochem Sci* 38, 378-385.

Lachaud, C., Slean, M., Marchesi, F., Lock, C., Odell, E., Castor, D., Toth, R., and Rouse, J. (2016). Karyomegalic interstitial nephritis and DNA damage-induced polyploidy in Fan1 nuclease-defective knock-in mice. *Genes Dev* 30, 639-644.

Landles, C., and Bates, G.P. (2004). Huntingtin and the molecular pathogenesis of Huntington's disease. Fourth in molecular medicine review series. *EMBO Rep* 5, 958-963.

Langfelder, P., Cantle, J.P., Chatzopoulou, D., Wang, N., Gao, F., Al-Ramahi, I., Lu, X.H., Ramos, E.M., El-Zein, K., Zhao, Y., *et al.* (2016). Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. *Nat Neurosci* 19, 623-633.

Larson, E., Fyfe, I., Morton, A.J., and Monckton, D.G. (2015). Age-, tissue- and length-dependent bidirectional somatic CAG\*CTG repeat instability in an allelic series of R6/2 Huntington disease mice. *Neurobiol Dis* 76, 98-111.

Law, S., Gillmore, J., Gilbertson, J.A., Bass, P., and Salama, A.D. (2020). Karyomegalic interstitial nephritis with a novel FAN1 gene mutation and concurrent ALECT2 amyloidosis. *BMC Nephrol* 21, 74.

Leach, F.S., Nicolaides, N.C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L.A., Nystrom-Lahti, M., and *et al.* (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* 75, 1215-1225.

Leavitt, B.R., and Tabrizi, S.J. (2020). Antisense oligonucleotides for neurodegeneration. *Science* 367, 1428-1429.

Lee, C.Y., Cantle, J.P., and Yang, X.W. (2013). Genetic manipulations of mutant huntingtin in mice: new insights into Huntington's disease pathogenesis. *FEBS J* 280, 4382-4394.

Lee, J.M., Pinto, R.M., Gillis, T., St Claire, J.C., and Wheeler, V.C. (2011). Quantification of age-dependent somatic CAG repeat instability in Hdh CAG knock-in mice reveals different expansion dynamics in striatum and liver. *PLoS One* 6, e23647.

Lee, J.M., Ramos, E.M., Lee, J.H., Gillis, T., Mysore, J.S., Hayden, M.R., Warby, S.C., Morrison, P., Nance, M., Ross, C.A., *et al.* (2012). CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology* 78, 690-695.

Lee, J.M., Zhang, J., Su, A.I., Walker, J.R., Wiltshire, T., Kang, K., Dragileva, E., Gillis, T., Lopez, E.T., Boily, M.J., *et al.* (2010). A novel approach to investigate tissue-specific trinucleotide repeat instability. *BMC Syst Biol* 4, 29.

Lei, W., Jiao, Y., Del Mar, N., and Reiner, A. (2004). Evidence for differential cortical input to direct pathway versus indirect pathway striatal projection neurons in rats. *J Neurosci* 24, 8289-8299.

Levine, M.S., Klapstein, G.J., Koppel, A., Gruen, E., Cepeda, C., Vargas, M.E., Jokel, E.S., Carpenter, E.M., Zanjani, H., Hurst, R.S., *et al.* (1999). Enhanced sensitivity to N-methyl-D-aspartate receptor activation in transgenic and knockin mouse models of Huntington's disease. *J Neurosci Res* 58, 515-532.

Li, H., Li, S.H., Cheng, A.L., Mangiarini, L., Bates, G.P., and Li, X.J. (1999). Ultrastructural localization and progressive formation of neuropil aggregates in Huntington's disease transgenic mice. *Hum Mol Genet* 8, 1227-1236.

Li, S.H., Li, H., Torre, E.R., and Li, X.J. (2000). Expression of huntingtin-associated protein-1 in neuronal cells implicates a role in neuritic growth. *Mol Cell Neurosci* 16, 168-183.

Lin, C.H., Tallaksen-Greene, S., Chien, W.M., Cearley, J.A., Jackson, W.S., Crouse, A.B., Ren, S., Li, X.J., Albin, R.L., and Detloff, P.J. (2001). Neurological abnormalities in a knock-in mouse model of Huntington's disease. *Hum Mol Genet* 10, 137-144.

Lin, Y., Dent, S.Y., Wilson, J.H., Wells, R.D., and Napierala, M. (2010). R loops stimulate genetic instability of CTG.CAG repeats. *Proc Natl Acad Sci U S A* 107, 692-697.

Lin, Y., Hubert, L., Jr., and Wilson, J.H. (2009). Transcription destabilizes triplet repeats. *Mol Carcinog* 48, 350-361.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* 362, 709-715.

Liu, X., Wang, C.E., Hong, Y., Zhao, T., Wang, G., Gaertig, M.A., Sun, M., Li, S., and Li, X.J. (2016). N-terminal Huntingtin Knock-In Mice: Implications of Removing the N-terminal Region of Huntingtin for Therapy. *PLoS Genet* 12, e1006083.

Lobo, M.K., Karsten, S.L., Gray, M., Geschwind, D.H., and Yang, X.W. (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nat Neurosci* 9, 443-452.

Loh, D.H., Kudo, T., Truong, D., Wu, Y., and Colwell, C.S. (2013). The Q175 mouse model of Huntington's disease shows gene dosage- and age-related decline in circadian rhythms of activity and sleep. *PLoS One* 8, e69993.

Loo, L., Simon, J.M., Xing, L., McCoy, E.S., Niehaus, J.K., Guo, J., Anton, E.S., and Zylka, M.J. (2019). Single-cell transcriptomic analysis of mouse neocortical development. *Nat Commun* 10, 134.

Macdonald, D., Tessari, M.A., Boogaard, I., Smith, M., Pulli, K., Szynol, A., Albertus, F., Lamers, M.B., Dijkstra, S., Kordt, D., *et al.* (2014). Quantification assays for total and polyglutamine-expanded huntingtin proteins. *PLoS One* 9, e96854.

MacDonald, V., and Halliday, G.M. (2002). Selective loss of pyramidal neurons in the pre-supplementary motor cortex in Parkinson's disease. *Mov Disord* 17, 1166-1173.

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trotter, Y., Lehrach, H., Davies, S.W., and Bates, G.P. (1996). Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* 87, 493-506.

Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M., and Thornton, C.A. (2000). Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* 289, 1769-1773.

Manley, K., Shirley, T.L., Flaherty, L., and Messer, A. (1999). Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet* 23, 471-473.

Mann, D.M., Oliver, R., and Snowden, J.S. (1993). The topographic distribution of brain atrophy in Huntington's disease and progressive supranuclear palsy. *Acta Neuropathol* 85, 553-559.

Massey, T.H., and Jones, L. (2018). The central role of DNA damage and repair in CAG repeat diseases. *Dis Model Mech* 11.

Mattsson, B., Gottfries, C.G., Roos, B.E., and Winblad, B. (1974). Huntington's chorea: pathology and brain amines. *Acta Psychiatr Scand Suppl* 255, 269-277.

McCarthy, S., Das, S., Kretschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., *et al.* (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 48, 1279-1283.

McKinstry, S.U., Karadeniz, Y.B., Worthington, A.K., Hayrapetyan, V.Y., Ozlu, M.I., Serafin-Molina, K., Risher, W.C., Ustunkaya, T., Dragatsis, I., Zeitlin, S., *et al.* (2014). Huntingtin is required for normal excitatory synapse development in cortical and striatal circuits. *J Neurosci* 34, 9455-9472.

McNeil, S.M., Novelletto, A., Srinidhi, J., Barnes, G., Kornbluth, I., Altherr, M.R., Wasmuth, J.J., Gusella, J.F., MacDonald, M.E., and Myers, R.H. (1997). Reduced penetrance of the Huntington's disease mutation. *Hum Mol Genet* 6, 775-779.

Menalled, L., El-Khodori, B.F., Patry, M., Suarez-Farinas, M., Orenstein, S.J., Zahasky, B., Leahy, C., Wheeler, V., Yang, X.W., MacDonald, M., *et al.* (2009). Systematic behavioral evaluation of Huntington's disease transgenic and knock-in mouse models. *Neurobiol Dis* 35, 319-336.

Menalled, L.B. (2005). Knock-in mouse models of Huntington's disease. *NeuroRx* 2, 465-470.

Menalled, L.B., Kudwa, A.E., Miller, S., Fitzpatrick, J., Watson-Johnson, J., Keating, N., Ruiz, M., Mushlin, R., Alosio, W., McConnell, K., *et al.* (2012). Comprehensive behavioral and molecular characterization of a new knock-in mouse model of Huntington's disease: zQ175. *PLoS One* 7, e49838.

- Menalled, L.B., Sison, J.D., Dragatsis, I., Zeitlin, S., and Chesselet, M.F. (2003). Time course of early motor and neuropathological anomalies in a knock-in mouse model of Huntington's disease with 140 CAG repeats. *J Comp Neurol* 465, 11-26.
- Mirkin, S.M. (2007). Expandable DNA repeats and human disease. *Nature* 447, 932-940.
- Modrich, P., and Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu Rev Biochem* 65, 101-133.
- Moffitt, H., McPhail, G.D., Woodman, B., Hobbs, C., and Bates, G.P. (2009). Formation of polyglutamine inclusions in a wide range of non-CNS tissues in the HdhQ150 knock-in mouse model of Huntington's disease. *PLoS One* 4, e8025.
- Monteys, A.M., Ebanks, S.A., Keiser, M.S., and Davidson, B.L. (2017). CRISPR/Cas9 Editing of the Mutant Huntingtin Allele In Vitro and In Vivo. *Mol Ther* 25, 12-23.
- Morrison, P.J., Harding-Lester, S., and Bradley, A. (2011). Uptake of Huntington disease predictive testing in a complete population. *Clin Genet* 80, 281-286.
- Morton, A.J. (2013). Circadian and sleep disorder in Huntington's disease. *Exp Neurol* 243, 34-44.
- Morton, A.J., Wood, N.I., Hastings, M.H., Hurelbrink, C., Barker, R.A., and Maywood, E.S. (2005). Disintegration of the sleep-wake cycle and circadian timing in Huntington's disease. *J Neurosci* 25, 157-163.
- Nakamori, M., Pearson, C.E., and Thornton, C.A. (2011). Bidirectional transcription stimulates expansion and contraction of expanded (CTG)<sup>n</sup>(CAG) repeats. *Hum Mol Genet* 20, 580-588.
- Nave, K.A. (2010). Myelination and support of axonal integrity by glia. *Nature* 468, 244-252.
- Neil, A.J., Kim, J.C., and Mirkin, S.M. (2017). Precarious maintenance of simple DNA repeats in eukaryotes. *Bioessays* 39.
- Nestler, E.J., and Hyman, S.E. (2010). Animal models of neuropsychiatric disorders. *Nat Neurosci* 13, 1161-1169.
- Neueder, A., Landles, C., Ghosh, R., Howland, D., Myers, R.H., Faull, R.L.M., Tabrizi, S.J., and Bates, G.P. (2017). The pathogenic exon 1 HTT protein is produced by incomplete splicing in Huntington's disease patients. *Sci Rep* 7, 1307.
- Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31, 3812-3814.
- Nicolaides, N.C., Papadopoulos, N., Liu, B., Wei, Y.F., Carter, K.C., Ruben, S.M., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., and et al. (1994). Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* 371, 75-80.
- Orr, H.T., and Zoghbi, H.Y. (2007). Trinucleotide repeat disorders. *Annu Rev Neurosci* 30, 575-621.

- Osipovitch, M., Asenjo Martinez, A., Mariani, J.N., Cornwell, A., Dhaliwal, S., Zou, L., Chandler-Militello, D., Wang, S., Li, X., Benraiss, S.J., *et al.* (2019). Human ESC-Derived Chimeric Mouse Models of Huntington's Disease Reveal Cell-Intrinsic Defects in Glial Progenitor Cell Differentiation. *Cell Stem Cell* 24, 107-122 e107.
- Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., Truong, O., Hsuan, J.J., and Jiricny, J. (1995). GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. *Science* 268, 1912-1914.
- Palpagama, T.H., Waldvogel, H.J., Faull, R.L.M., and Kwakowsky, A. (2019). The Role of Microglia and Astrocytes in Huntington's Disease. *Front Mol Neurosci* 12, 258.
- Panigrahi, G.B., Lau, R., Montgomery, S.E., Leonard, M.R., and Pearson, C.E. (2005). Slipped (CTG)<sup>n</sup>(CAG)<sup>m</sup> repeats can be correctly repaired, escape repair or undergo error-prone repair. *Nat Struct Mol Biol* 12, 654-662.
- Panigrahi, G.B., Slean, M.M., Simard, J.P., Gileadi, O., and Pearson, C.E. (2010). Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired. *Proc Natl Acad Sci U S A* 107, 12593-12598.
- Panigrahi, G.B., Slean, M.M., Simard, J.P., and Pearson, C.E. (2012). Human mismatch repair protein hMutLalpha is required to repair short slipped-DNAs of trinucleotide repeats. *J Biol Chem* 287, 41844-41850.
- Papadopoulos, N., Nicolaidis, N.C., Wei, Y.F., Ruben, S.M., Carter, K.C., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., Adams, M.D., and *et al.* (1994). Mutation of a mutL homolog in hereditary colon cancer. *Science* 263, 1625-1629.
- Paulsen, J.S., Langbehn, D.R., Stout, J.C., Aylward, E., Ross, C.A., Nance, M., Guttman, M., Johnson, S., MacDonald, M., Beglinger, L.J., *et al.* (2008). Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *J Neurol Neurosurg Psychiatry* 79, 874-880.
- Paulsen, J.S., Nopoulos, P.C., Aylward, E., Ross, C.A., Johnson, H., Magnotta, V.A., Juhl, A., Pierson, R.K., Mills, J., Langbehn, D., *et al.* (2010). Striatal and white matter predictors of estimated diagnosis for Huntington disease. *Brain Res Bull* 82, 201-207.
- Pavese, N., Gerhard, A., Tai, Y.F., Ho, A.K., Turkheimer, F., Barker, R.A., Brooks, D.J., and Piccini, P. (2006). Microglial activation correlates with severity in Huntington disease: a clinical and PET study. *Neurology* 66, 1638-1643.
- Pearl, L.H., Schierz, A.C., Ward, S.E., Al-Lazikani, B., and Pearl, F.M. (2015). Therapeutic opportunities within the DNA damage response. *Nat Rev Cancer* 15, 166-180.
- Peng, Q., Wu, B., Jiang, M., Jin, J., Hou, Z., Zheng, J., Zhang, J., and Duan, W. (2016). Characterization of Behavioral, Neuropathological, Brain Metabolic and Key Molecular Changes in zQ175 Knock-In Mouse Model of Huntington's Disease. *PLoS One* 11, e0148839.
- Pennell, S., Declais, A.C., Li, J., Haire, L.F., Berg, W., Saldanha, J.W., Taylor, I.A., Rouse, J., Lilley, D.M., and Smerdon, S.J. (2014). FAN1 activity on asymmetric repair intermediates is

mediated by an atypical monomeric virus-type replication-repair nuclease domain. *Cell Rep* 8, 84-93.

Petkau, T.L., Hill, A., Connolly, C., Lu, G., Wagner, P., Kosior, N., Blanco, J., and Leavitt, B.R. (2019). Mutant huntingtin expression in microglia is neither required nor sufficient to cause the Huntington's disease-like phenotype in BACHD mice. *Hum Mol Genet* 28, 1661-1670.

Pfriege, F.W., and Ungerer, N. (2011). Cholesterol metabolism in neurons and astrocytes. *Prog Lipid Res* 50, 357-371.

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

Plotkin, J.L., Day, M., Peterson, J.D., Xie, Z., Kress, G.J., Rafalovich, I., Kondapalli, J., Gertler, T.S., Flajolet, M., Greengard, P., *et al.* (2014). Impaired TrkB receptor signaling underlies corticostriatal dysfunction in Huntington's disease. *Neuron* 83, 178-188.

Pluciennik, A., Burdett, V., Baitinger, C., Iyer, R.R., Shi, K., and Modrich, P. (2013). Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLalpha endonuclease activation. *Proc Natl Acad Sci U S A* 110, 12277-12282.

Pouladi, M.A., Stanek, L.M., Xie, Y., Franciosi, S., Southwell, A.L., Deng, Y., Butland, S., Zhang, W., Cheng, S.H., Shihabuddin, L.S., and Hayden, M.R. (2012). Marked differences in neurochemistry and aggregates despite similar behavioural and neuropathological features of Huntington disease in the full-length BACHD and YAC128 mice. *Hum Mol Genet* 21, 2219-2232.

Pouladi, M.A., Xie, Y., Skotte, N.H., Ehrnhoefer, D.E., Graham, R.K., Kim, J.E., Bissada, N., Yang, X.W., Paganetti, P., Friedlander, R.M., *et al.* (2010). Full-length huntingtin levels modulate body weight by influencing insulin-like growth factor 1 expression. *Hum Mol Genet* 19, 1528-1538.

Ranen, N.G., Stine, O.C., Abbott, M.H., Sherr, M., Codori, A.M., Franz, M.L., Chao, N.I., Chung, A.S., Pleasant, N., Callahan, C., and *et al.* (1995). Anticipation and instability of IT-15 (CAG)<sub>n</sub> repeats in parent-offspring pairs with Huntington disease. *Am J Hum Genet* 57, 593-602.

Ratovitski, T., Gucek, M., Jiang, H., Chighladze, E., Waldron, E., D'Ambola, J., Hou, Z., Liang, Y., Poirier, M.A., Hirschhorn, R.R., *et al.* (2009). Mutant huntingtin N-terminal fragments of specific size mediate aggregation and toxicity in neuronal cells. *J Biol Chem* 284, 10855-10867.

Raymond, L.A., Andre, V.M., Cepeda, C., Gladding, C.M., Milnerwood, A.J., and Levine, M.S. (2011). Pathophysiology of Huntington's disease: time-dependent alterations in synaptic and receptor function. *Neuroscience* 198, 252-273.

Reddy, K., Schmidt, M.H., Geist, J.M., Thakkar, N.P., Panigrahi, G.B., Wang, Y.H., and Pearson, C.E. (2014). Processing of double-R-loops in (CAG)<sub>n</sub>(CTG)<sub>n</sub> and C9orf72 (GGGGCC)<sub>n</sub>(GGCCCC)<sub>n</sub> repeats causes instability. *Nucleic Acids Res* 42, 10473-10487.

Reenan, R.A., and Kolodner, R.D. (1992). Isolation and characterization of two *Saccharomyces cerevisiae* genes encoding homologs of the bacterial HexA and MutS mismatch repair proteins. *Genetics* 132, 963-973.

Reilmann, R. (2019). Parkinsonism in Huntington's disease. *Int Rev Neurobiol* 149, 299-306.

Reindl, W., Baldo, B., Schulz, J., Janack, I., Lindner, I., Kleinschmidt, M., Sedaghat, Y., Thiede, C., Tillack, K., Schmidt, C., *et al.* (2019). Meso scale discovery-based assays for the detection of aggregated huntingtin. *PLoS One* 14, e0213521.

Richfield, E.K., Maguire-Zeiss, K.A., Vonkeman, H.E., and Voorn, P. (1995). Preferential loss of preproenkephalin versus preprotachykinin neurons from the striatum of Huntington's disease patients. *Ann Neurol* 38, 852-861.

Rikitake, M., Fujikane, R., Obayashi, Y., Oka, K., Ozaki, M., and Hidaka, M. (2020). MLH1-mediated recruitment of FAN1 to chromatin for the induction of apoptosis triggered by O(6) - methylguanine. *Genes Cells* 25, 175-186.

Rodriguez, G.P., Romanova, N.V., Bao, G., Rouf, N.C., Kow, Y.W., and Crouse, G.F. (2012). Mismatch repair-dependent mutagenesis in nondividing cells. *Proc Natl Acad Sci U S A* 109, 6153-6158.

Rosas, H.D., Wilkens, P., Salat, D.H., Mercaldo, N.D., Vangel, M., Yendiki, A.Y., and Hersch, S.M. (2018). Complex spatial and temporally defined myelin and axonal degeneration in Huntington disease. *Neuroimage Clin* 20, 236-242.

Ross, C.A., and Tabrizi, S.J. (2011). Huntington's disease: from molecular pathogenesis to clinical treatment. *Lancet Neurol* 10, 83-98.

Sancar, A., and Hearst, J.E. (1993). Molecular matchmakers. *Science* 259, 1415-1420.

Sapp, E., Ge, P., Aizawa, H., Bird, E., Penney, J., Young, A.B., Vonsattel, J.P., and DiFiglia, M. (1995). Evidence for a preferential loss of enkephalin immunoreactivity in the external globus pallidus in low grade Huntington's disease using high resolution image analysis. *Neuroscience* 64, 397-404.

Sathasivam, K., Neueder, A., Gipson, T.A., Landles, C., Benjamin, A.C., Bondulich, M.K., Smith, D.L., Faull, R.L., Roos, R.A., Howland, D., *et al.* (2013). Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc Natl Acad Sci U S A* 110, 2366-2370.

Sathasivam, K., Woodman, B., Mahal, A., Bertaux, F., Wanker, E.E., Shima, D.T., and Bates, G.P. (2001). Centrosome disorganization in fibroblast cultures derived from R6/2 Huntington's disease (HD) transgenic mice and HD patients. *Hum Mol Genet* 10, 2425-2435.

Saudou, F., and Humbert, S. (2016). The Biology of Huntingtin. *Neuron* 89, 910-926.

Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., *et al.* (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* 174, 1015-1030 e1016.

- Schmidt, M.H.M., and Pearson, C.E. (2016). Disease-associated repeat instability and mismatch repair. *DNA Repair (Amst)* 38, 117-126.
- Schmitz, C., Rub, U., Korr, H., and Heinsen, H. (1999). Nerve cell loss in the thalamic mediodorsal nucleus in Huntington's disease. II. Optimization of a stereological estimation procedure. *Acta Neuropathol* 97, 623-628.
- Semaka, A., Kay, C., Doty, C., Collins, J.A., Bijlsma, E.K., Richards, F., Goldberg, Y.P., and Hayden, M.R. (2013a). CAG size-specific risk estimates for intermediate allele repeat instability in Huntington disease. *J Med Genet* 50, 696-703.
- Semaka, A., Kay, C., Doty, C.N., Collins, J.A., Tam, N., and Hayden, M.R. (2013b). High frequency of intermediate alleles on Huntington disease-associated haplotypes in British Columbia's general population. *Am J Med Genet B Neuropsychiatr Genet* 162B, 864-871.
- Shelbourne, P.F., Killeen, N., Hevner, R.F., Johnston, H.M., Tecott, L., Lewandoski, M., Ennis, M., Ramirez, L., Li, Z., Iannicola, C., *et al.* (1999). A Huntington's disease CAG expansion at the murine Hdh locus is unstable and associated with behavioural abnormalities in mice. *Hum Mol Genet* 8, 763-774.
- Skotte, N.H., Southwell, A.L., Ostergaard, M.E., Carroll, J.B., Warby, S.C., Doty, C.N., Petoukhov, E., Vaid, K., Kordasiewicz, H., Watt, A.T., *et al.* (2014). Allele-specific suppression of mutant huntingtin using antisense oligonucleotides: providing a therapeutic option for all Huntington disease patients. *PLoS One* 9, e107434.
- Slow, E.J., van Raamsdonk, J., Rogers, D., Coleman, S.H., Graham, R.K., Deng, Y., Oh, R., Bissada, N., Hossain, S.M., Yang, Y.Z., *et al.* (2003). Selective striatal neuronal loss in a YAC128 mouse model of Huntington disease. *Hum Mol Genet* 12, 1555-1567.
- Smith, K.J., Kapoor, R., and Felts, P.A. (1999). Demyelination: the role of reactive oxygen and nitrogen species. *Brain Pathol* 9, 69-92.
- Smogorzewska, A., Desetty, R., Saito, T.T., Schlabach, M., Lach, F.P., Sowa, M.E., Clark, A.B., Kunkel, T.A., Harper, J.W., Colaiacovo, M.P., and Elledge, S.J. (2010). A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol Cell* 39, 36-47.
- Sotrel, A., Paskevich, P.A., Kiely, D.K., Bird, E.D., Williams, R.S., and Myers, R.H. (1991). Morphometric analysis of the prefrontal cortex in Huntington's disease. *Neurology* 41, 1117-1123.
- Southwell, A.L., Kordasiewicz, H.B., Langbehn, D., Skotte, N.H., Parsons, M.P., Villanueva, E.B., Caron, N.S., Ostergaard, M.E., Anderson, L.M., Xie, Y., *et al.* (2018). Huntingtin suppression restores cognitive function in a mouse model of Huntington's disease. *Sci Transl Med* 10.
- Squitieri, F., Andrew, S.E., Goldberg, Y.P., Kremer, B., Spence, N., Zeisler, J., Nichol, K., Theilmann, J., Greenberg, J., Goto, J., and *et al.* (1994). DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. *Hum Mol Genet* 3, 2103-2114.

Stout, J.C., Glikmann-Johnston, Y., and Andrews, S.C. (2016). Cognitive assessment strategies in Huntington's disease research. *J Neurosci Methods* 265, 19-24.

Stout, J.C., Paulsen, J.S., Queller, S., Solomon, A.C., Whitlock, K.B., Campbell, J.C., Carlozzi, N., Duff, K., Beglinger, L.J., Langbehn, D.R., *et al.* (2011). Neurocognitive signs in prodromal Huntington disease. *Neuropsychology* 25, 1-14.

Surmeier, D.J., Song, W.J., and Yan, Z. (1996). Coordinated expression of dopamine receptors in neostriatal medium spiny neurons. *J Neurosci* 16, 6579-6591.

Swami, M., Hendricks, A.E., Gillis, T., Massood, T., Mysore, J., Myers, R.H., and Wheeler, V.C. (2009). Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum Mol Genet* 18, 3039-3047.

Tabrizi, S.J., Ghosh, R., and Leavitt, B.R. (2019a). Huntingtin Lowering Strategies for Disease Modification in Huntington's Disease. *Neuron* 101, 801-819.

Tabrizi, S.J., Langbehn, D.R., Leavitt, B.R., Roos, R.A., Durr, A., Craufurd, D., Kennard, C., Hicks, S.L., Fox, N.C., Scahill, R.I., *et al.* (2009). Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurol* 8, 791-801.

Tabrizi, S.J., Leavitt, B.R., Landwehrmeyer, G.B., Wild, E.J., Saft, C., Barker, R.A., Blair, N.F., Craufurd, D., Priller, J., Rickards, H., *et al.* (2019b). Targeting Huntingtin Expression in Patients with Huntington's Disease. *N Engl J Med* 380, 2307-2316.

Tai, Y.F., Pavese, N., Gerhard, A., Tabrizi, S.J., Barker, R.A., Brooks, D.J., and Piccini, P. (2007). Microglial activation in presymptomatic Huntington's disease gene carriers. *Brain* 130, 1759-1766.

Taneja, K.L., McCurrach, M., Schalling, M., Housman, D., and Singer, R.H. (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J Cell Biol* 128, 995-1002.

Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., *et al.* (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563, 72-78.

Teo, R.T., Hong, X., Yu-Taeger, L., Huang, Y., Tan, L.J., Xie, Y., To, X.V., Guo, L., Rajendran, R., Novati, A., *et al.* (2016). Structural and molecular myelination deficits occur prior to neuronal loss in the YAC128 and BACHD models of Huntington disease. *Hum Mol Genet* 25, 2621-2632.

Thompson, J.C., Harris, J., Sollom, A.C., Stopford, C.L., Howard, E., Snowden, J.S., and Craufurd, D. (2012). Longitudinal evaluation of neuropsychiatric symptoms in Huntington's disease. *J Neuropsychiatry Clin Neurosci* 24, 53-60.

Thongthip, S., Bellani, M., Gregg, S.Q., Sridhar, S., Conti, B.A., Chen, Y., Seidman, M.M., and Smogorzewska, A. (2016). Fan1 deficiency results in DNA interstrand cross-link repair defects, enhanced tissue karyomegaly, and organ dysfunction. *Genes Dev* 30, 645-659.

- Timchenko, N.A., Iakova, P., Cai, Z.J., Smith, J.R., and Timchenko, L.T. (2001). Molecular basis for impaired muscle differentiation in myotonic dystrophy. *Mol Cell Biol* 21, 6927-6938.
- Todd, P.K., and Paulson, H.L. (2010). RNA-mediated neurodegeneration in repeat expansion disorders. *Ann Neurol* 67, 291-300.
- Tome, S., Holt, I., Edelmann, W., Morris, G.E., Munnich, A., Pearson, C.E., and Gourdon, G. (2009). MSH2 ATPase domain mutation affects CTG\*CAG repeat instability in transgenic mice. *PLoS Genet* 5, e1000482.
- Van Raamsdonk, J.M., Metzler, M., Slow, E., Pearson, J., Schwab, C., Carroll, J., Graham, R.K., Leavitt, B.R., and Hayden, M.R. (2007). Phenotypic abnormalities in the YAC128 mouse model of Huntington disease are penetrant on multiple genetic backgrounds and modulated by strain. *Neurobiol Dis* 26, 189-200.
- Van Raamsdonk, J.M., Pearson, J., Rogers, D.A., Bissada, N., Vogl, A.W., Hayden, M.R., and Leavitt, B.R. (2005). Loss of wild-type huntingtin influences motor dysfunction and survival in the YAC128 mouse model of Huntington disease. *Hum Mol Genet* 14, 1379-1392.
- Veldman, M.B., and Yang, X.W. (2018). Molecular insights into cortico-striatal miscommunications in Huntington's disease. *Curr Opin Neurobiol* 48, 79-89.
- Vonsattel, J.P., and DiFiglia, M. (1998). Huntington disease. *J Neuropathol Exp Neurol* 57, 369-384.
- Vonsattel, J.P., Myers, R.H., Stevens, T.J., Ferrante, R.J., Bird, E.D., and Richardson, E.P., Jr. (1985). Neuropathological classification of Huntington's disease. *J Neuropathol Exp Neurol* 44, 559-577.
- Wall, N.R., De La Parra, M., Callaway, E.M., and Kreitzer, A.C. (2013). Differential innervation of direct- and indirect-pathway striatal projection neurons. *Neuron* 79, 347-360.
- Wang, N., Gray, M., Lu, X.H., Cattle, J.P., Holley, S.M., Greiner, E., Gu, X., Shirasaki, D., Cepeda, C., Li, Y., *et al.* (2014a). Neuronal targets for reducing mutant huntingtin expression to ameliorate disease in a mouse model of Huntington's disease. *Nat Med* 20, 536-541.
- Wang, R., Persky, N.S., Yoo, B., Ouerfelli, O., Smogorzewska, A., Elledge, S.J., and Pavletich, N.P. (2014b). DNA repair. Mechanism of DNA interstrand cross-link processing by repair nuclease FAN1. *Science* 346, 1127-1130.
- Warby, S.C., Montpetit, A., Hayden, A.R., Carroll, J.B., Butland, S.L., Visscher, H., Collins, J.A., Semaka, A., Hudson, T.J., and Hayden, M.R. (2009). CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. *Am J Hum Genet* 84, 351-366.
- Wexler, N.S., Young, A.B., Tanzi, R.E., Travers, H., Starosta-Rubinstein, S., Penney, J.B., Snodgrass, S.R., Shoulson, I., Gomez, F., Ramos Arroyo, M.A., and *et al.* (1987). Homozygotes for Huntington's disease. *Nature* 326, 194-197.

- Wheeler, V.C., Auerbach, W., White, J.K., Srinidhi, J., Auerbach, A., Ryan, A., Duyao, M.P., Vrbanac, V., Weaver, M., Gusella, J.F., *et al.* (1999). Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Hum Mol Genet* 8, 115-122.
- Wild, E.J., and Tabrizi, S.J. (2017). Therapies targeting DNA and RNA in Huntington's disease. *Lancet Neurol* 16, 837-847.
- William Yang, X., and Gray, M. (2011). Mouse Models for Validating Preclinical Candidates for Huntington's Disease. In *Neurobiology of Huntington's Disease: Applications to Drug Discovery*, D.C. Lo, and R.E. Hughes, eds. (Boca Raton (FL)).
- Wojciechowska, M., and Krzyzosiak, W.J. (2011). Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum Mol Genet* 20, 3811-3821.
- Wood, T.E., Barry, J., Yang, Z., Cepeda, C., Levine, M.S., and Gray, M. (2019). Mutant huntingtin reduction in astrocytes slows disease progression in the BACHD conditional Huntington's disease mouse model. *Hum Mol Genet* 28, 487-500.
- Xiang, Z., Valenza, M., Cui, L., Leoni, V., Jeong, H.K., Brilli, E., Zhang, J., Peng, Q., Duan, W., Reeves, S.A., *et al.* (2011). Peroxisome-proliferator-activated receptor gamma coactivator 1 alpha contributes to dysmyelination in experimental models of Huntington's disease. *J Neurosci* 31, 9544-9553.
- Yang, X.W., and Gong, S. (2005). An overview on the generation of BAC transgenic mice for neuroscience research. *Curr Protoc Neurosci* Chapter 5, Unit 5 20.
- Zeitler, B., Froelich, S., Marlen, K., Shivak, D.A., Yu, Q., Li, D., Pearl, J.R., Miller, J.C., Zhang, L., Paschon, D.E., *et al.* (2019). Allele-selective transcriptional repression of mutant HTT for the treatment of Huntington's disease. *Nat Med* 25, 1131-1142.
- Zeitlin, S., Liu, J.P., Chapman, D.L., Papaioannou, V.E., and Efstratiadis, A. (1995). Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nat Genet* 11, 155-163.
- Zhang, J., and Walter, J.C. (2014). Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA Repair (Amst)* 19, 135-142.
- Zhang, T., Huang, J., Gu, L., and Li, G.M. (2012). In vitro repair of DNA hairpins containing various numbers of CAG/CTG trinucleotide repeats. *DNA Repair (Amst)* 11, 201-209.
- Zhao, Q., Xue, X., Longerich, S., Sung, P., and Xiong, Y. (2014). Structural insights into 5' flap DNA unwinding and incision by the human FAN1 dimer. *Nat Commun* 5, 5726.
- Zhao, X.N., and Usdin, K. (2018). FAN1 protects against repeat expansions in a Fragile X mouse model. *DNA Repair (Amst)* 69, 1-5.
- Zhou, W., Otto, E.A., Cluckey, A., Airik, R., Hurd, T.W., Chaki, M., Diaz, K., Lach, F.P., Bennett, G.R., Gee, H.Y., *et al.* (2012). FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. *Nat Genet* 44, 910-915.

Zuccato, C., and Cattaneo, E. (2014). Huntington's disease. *Handb Exp Pharmacol* 220, 357-409.

Zuccato, C., Ciammola, A., Rigamonti, D., Leavitt, B.R., Goffredo, D., Conti, L., MacDonald, M.E., Friedlander, R.M., Silani, V., Hayden, M.R., *et al.* (2001). Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. *Science* 293, 493-498.

Zuccato, C., Liber, D., Ramos, C., Tarditi, A., Rigamonti, D., Tartari, M., Valenza, M., and Cattaneo, E. (2005). Progressive loss of BDNF in a mouse model of Huntington's disease and rescue by BDNF delivery. *Pharmacol Res* 52, 133-139.

Zylka, M.J., Simon, J.M., and Philpot, B.D. (2015). Gene length matters in neurons. *Neuron* 86, 353-355.