
Defining the Htt-exon 1 aggregation state induced transcriptomic changes in Huntington's disease

Nagaraja Sundara Moily

ORCID: 0000-0002-3045-9563

Submitted in total fulfilment of the requirements of the degree
of Doctor of Philosophy

January 2021

Supervised by - Prof. Danny Hatters,
Prof. Alicia Oshlack & Prof. Anthony Hannan

Department of Biochemistry and Molecular Biology
Bio21 Molecular Science and Biotechnology Institute
The University of Melbourne

Abstract

Huntington's disease (HD) is a trinucleotide repeat expansion disorder that is caused due to a defect in the exon 1 of the Huntingtin gene. The defect is in the number of polyglutamine coding CAG repeats. If the number of glutamine repeats in the huntingtin protein (Htt) exceeds 35 the soluble Htt protein becomes sticky, aggregate prone and clumps together to form micro-meter sized inclusions, which is the key feature in HD.

There is mounting evidence that the soluble form of mutant Htt is highly proteotoxic to cells and increases the risk of cell death prior to aggregation, whereas aggregates impart distinct changes to cellular health, potentially adapting the cell to stressors and negating toxicity, as soluble Htt is sequestered into inclusions. Nonetheless, the molecular mechanisms underpinning these transitions remain unclear. This has been primarily due to a lack of methods to separate cells with and without aggregates.

In this thesis we have elucidated the transcriptional signatures that are specifically attributed to the different aggregation states seen in HD, and to the discrete steps in the aggregation process involved. Soluble mutant Htt states produced the largest transcriptional dysregulation with dampened CREB signalling specifically invoked by the soluble mutant Httex1 being the most striking change. We also demonstrated CREB signalling is critical to survival in a cell culture model of HD with cell toxicity ameliorated by activation of CREB signalling cascade. The various biological processes linked to the different aggregation states include transcription regulation by higher chromatin modifiers, altered NF- κ B signalling, autophagy, NAD⁺ biosynthesis, SUMOylation, altered HIF-1 signalling and ribosome biogenesis.

We have further explored the relevant biology of Httex1 aggregation states and splicing defects associated with different aggregation states. Many genes showed altered splice forms which could be relevant to disease. Further using proteomics informed through the transcriptomics, we have identified splicing defects in HD. We have identified spliceosome changes in HD and have also found changes in small nuclear RNA biogenesis which could lead to alterations in spliceosome function in human isogenic Huntingtin iPSC cell lines. These findings could pave the way for therapeutic strategies aimed at key molecular changes before and after pathological aggregation of proteins.

Declaration

This is to certify that

1. The thesis comprises only my original work towards the PhD except as indicated in the Preface,
2. Due acknowledgment has been made in the text to all other material used,
3. The thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Dr. Nagaraj S Moily., M.B.B.S

Dept. of Biochemistry and Molecular Biology
Faculty of Medicine, Health and Dentistry
The University of Melbourne

Awards, Presentations and Publications

Awards:

Recipient of the Bill Sawyer Medal

December 2017

Awarded for outstanding research achievements in PhD candidature

Recipient of the Department of Biochemistry Travel Award

April 2017

To attend CSHL conference on systems biology in USA

Recipient of Student Poster Award

February 2016

Awarded at 41st Lorne Protein Conference

Recipient of MIRS and MIFRS for Graduate Research 2013-2017

Oral Presentations:

Huntington Disease Research Day, August 2016 at Monash University, Clayton, Victoria, Australia.

Abstract selected for an oral presentation

Stem Cell symposium at the Melbourne Brain Center, Florey institute of mental health and neurosciences, Melbourne, September 2016

Invited Speaker at Florey institute of mental health and neurosciences, Melbourne

2nd Annual EMBL Australia PhD Symposium- Completing the Pipeline: From Biology to Bioinformatics and Back Again, November 2015; Bio21 Institute, Melbourne

Abstract selected for an oral presentation

Proteostasis and disease symposium at Wollongong, November 2014

Abstract selected for an oral presentation

Publications from PhD:

1. **Nagaraj S. Moily**, Angelique R. Ormsby, Aleksandar Stojilovic, Yasmin M. Ramdzan, Jeannine Diesch, Ross D. Hannan, Michelle S. Zajac, Anthony J. Hannan, Alicia Oshlack and Danny M. Hatters. Transcriptional profiles for distinct aggregation states of mutant Huntingtin exon 1 protein unmasks new Huntington's disease pathways. *Molecular and Cellular Neuroscience*, Volume 83, 2017, Pages 103-112, ISSN 1044-7431,
2. Moore Z. Chen, **Nagaraj S Moily**, Rebecca J. Wood, Jess Bridgford, Trevor A. Smith, Zhegang Song, Ben Zhong Tang, Leann Tilley, Xiaohong Xu, Mahmoud A. Pouladi, Yuning Hong and Danny M. Hatters. A thiol probe for measuring unfolded protein load and proteostasis in cells. *Nature Communications-* 8, Article number: 474 (2017) doi:10.1038/s41467-017-00203-5.
3. Yasmin M. Ramdzan, Mikhail M. Trubetskov, Estella A. Newcombe, Angelique R. Ormsby, Xiaojing Sui, Marie Bongiovanni, Sally L. Gras, Grant Dewson, Jason M. L. Miller, Steven Finkbeiner, **Nagaraj S Moily**, Jonathan Niclis, Clare L. Parish, Anthony W. Purcell, Michael J. Baker, Jackie A. Wilce, Saboora Waris, Diana Stojanowski , Till Böcking , Gavin Reid, Danny M. Hatters. Huntingtin inclusions trigger cellular quiescence, deactivate apoptosis and drive slow progression to necrosis. Volume 19, Issue 5, 2 May 2017, Pages 919–927. <https://doi.org/10.1016/j.celrep.2017.04.029>. (2017).
4. **Nagaraj S. Moily**, Mona Radwan, Angelique R. Ormsby, Mahmoud Paloudi, Gavin Reid, Alicia Oshlack, Danny M. Hatters. Httex1 aggregation states induce spliceosome defects and aberrant splicing in Huntington's disease (Manuscript in preparation).

Poster Presentations:

Lorne Conference on Protein structure and function –

2014, 2016, 2017

Cold Spring Harbor Laboratories conference - Systems Biology: Global Regulation of Gene Expression meeting on February 2017, NYC, USA

Preface

The following work was carried out in collaboration:

- Aleksandar Stojilovic (Honors student- Hatters Lab) contributed to the data presented in the Figure 2, Figure 4 in Chapter 2.
- Mona Radwan assisted in proteomics experiment and analysis done in chapter 3.

Nagaraj Moily performed proteomics experiments and proteomic data analysis and contributed to the data presented in the Figure 4e and Figure 6 in Chapter 4.

Acknowledgements

First and foremost, I would like to thank my supervisors Prof. Danny Hatters, Prof. Alicia Oshlack & Prof. Anthony Hannan for providing me supervision, guidance, and support during my PhD.

I am especially indebted to Danny for his supervision and for his valuable critics throughout my PhD work. I am also grateful for his continued support and guidance in letting me explore new avenues of research in the lab. They have helped me immensely to develop as a scientist.

I would like to express my gratitude to Alicia for her assistance with the project's analysis. Her ability to break down difficult ideas and analyze them has always inspired me. I thank Tony for his support and guidance and advice. He has always been a wonderful mentor, and I will be eternally grateful for his assistance through my PhD and my time in Melbourne.

A very special thanks to all my committee members Assoc. Prof. Paul Gooley, Assoc. Prof. Kathryn Holt, Dr Jonathon Niclis, Prof. Andrew Hill, Prof. Ross Hannan for their guidance and helpful suggestions.

To all my fellow PhD students at Bio21 Institute and past members of Hatters Lab, Yasmin Ramdzan, Angelique Ormsby, Mona Radwan, Aleks Stojilovic, Mikhail Trubetskov, Estella Newcombe, Zhe Chen, Autumn Bricker joy, Rebecca Wood, Xiaojing Sui, Andrew Ngo, Candi Raeburn. Thanks for providing a friendly and supportive environment to complete my PhD. I wish you all the very best in your future endeavors.

I would like to thank the Mass spectrometry and proteomics facility, the Biological Optical Microscopy Platform at Bio21 Institute for providing top class facilities for research. I would also like to thank Anna Michalska and Kathryn at Stem cell Australia, Florey Institute of mental Health and Neurosciences for providing me Human embryonic stem cells and Assoc. Prof Mahmoud Pouladi, at Agency for Science, Technology and Research, Singapore for induced pluripotent patient stem cells for my experiments.

I am immensely grateful to my wife Divya, she has always been a source of strength to me. We have been through lots of ups and downs in our life, but you have always remained a pillar of love,

devotion and inspiration. I am eternally grateful to my parents, my dad Prof. Sundara Moily for inculcating a scientific spirit and curiosity in me and my mom Shakunthala for being steadfast in her love and care. They have made me what I am today, and I wouldn't be here without their support. Love you mum, dad. Love to my sister – Dr. Nagarathna Baadkar, my brother Premraj Moily. I would also like to thank my in-laws Korambath Narayanan and Indira for their love and support. My love to Deepesh Korambath, Nivedita Kunder, Chetan Baadkar, Medha and Jay.

Very special thanks to all my friends in Melbourne Rahul Rane, Amol Ghodke, Nitin Patil, Neerja, Varsha and back home in India who has always been an immense source of support and motivation that helped me to keep going in my PhD.

I dedicate this work to my parents who always believed in the richness of learning and my patients who my shown me what life is and how it is to be lived!

Thanks to all mighty for all the blessings and love I have received so far in my life.

Abbreviations

Abbreviations	Full description
%	Percentage
°C	Degree Celsius
A3SS	Alternative 3' splice sites
A5SS	Alternative 5' splice sites
ACN	Acetonitrile
Adcyap1r1	Adenylate Cyclase Activating Polypeptide 1 (Pituitary) Receptor Type I
ADORA2A	Adenosine A2a Receptor
ALS	Amyotrophic lateral sclerosis
AS	Alternative splicing
ASE	Alternative splicing events
ATF6	Activating transcriptional factor 6
ATP	Adenosine triphosphate
B2	5-[4-(4-chlorobenzoyl)-1-piperazinyl]-8-nitroquinoline
BAM	Binary alignment map
BCA	Bicinchoninic acid
BDNF	Brain-derived neurotrophic factor
bp	Base pairs
BRD4	Bromodomain-containing protein 4
BSA	Bovine serum albumin
CAG	Cytosine-Adenosine-Guanine
cAMP	Cyclic adenosine monophosphate
CBP	cAMP response element binding protein
CCG	Cytosine-Cytosine-Guanine
CDK1	Cyclin-dependent kinase 1
Cftr	CF transmembrane conductance regulator
CHOP	C/EBP homologous protein
CLASP	Cytoplasmic linker associated protein
Clta	Clathrin light polypeptide
c-MYC	Myc proto-oncogene protein
CPM	Counts per million

CRCTs	CREB regulated transcription co-activators
CREB	cAMP-response–element-binding protein
CRISPR	Clustered regularly interspaced short palindromic repeat
CTFA	Trifluoroacetic acid
Cys	Cysteine
DAVID	Database for Annotation, Visualization and Integrated Discovery
DFFB	DNA Fragmentation Factor Subunit Beta
DHFR	Dihydrofolate reductase
Dhx9	Dexh-Box Helicase 9
Dlg4	Discs large homolog 4
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DnaJa2	DnaJ Heat Shock Protein Family (Hsp40) Member A
DnaI4	Dynein axonemal light chain 4
Ehmt2	Euchromatic histone lysine methyltransferase 2
ELAVL1	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1(Hu antigen R)
EMBOSS	European Molecular Biology Open Software Suite
EP300	E1A binding protein p300
Eps15	Epidermal growth factor receptor substrate 15
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated degradation
nESI	Nano electrospray ionization
FACS	Fluorescence-activated cell sorted
FADD	Fas-associated protein with death domain
FCS	Fluorescence correlation microscopy
FDR	False discovery rate
FRET	Förster resonance energy transfer
FSC	Forward scatter
GABA	Gamma aminobutyric acid
GABARAP	Gamma-aminobutyric acid receptor-associated protein
GABARAPL	GABA Type A Receptor Associated Protein Like
GFP	Green fluorescent protein
GO	Gene ontology

GRC38	Genome Reference Consortium Build 38
Gria2	Glutamate receptor ionotropic AMPA2
GRIK2	Glutamate Ionotropic Receptor Kainate Type Subunit 2)
Grin1	Glutamate receptor ionotropic, NMDA1
GRIN2A	Glutamate Ionotropic Receptor NMDA Type Subunit 2A
GRIN2B	Glutamate Ionotropic Receptor NMDA Type Subunit 2B
GRP78	Glucose-regulated protein of 78 kda
HAP1	Huntingtin Associated Protein 1
HAP40	Huntingtin-Associated Protein 40
HD	Huntington's disease
HDAC	Histone deacetylase
HEAT	Huntingtin, elongation factor 3, protein phosphatase 2A, and TOR1
HH	HEAT repeat domains
HIF-1	Hypoxia-inducible factor 1
HIP1	Huntingtin-interacting protein 1
HIP1R	Huntingtin interacting protein 1
HN4FA	Hepatocyte nuclear factor 4 alpha
HPLC	High Performance Liquid Chromatography
HSP40	Heat shock protein family 40
HSP70	Heat shock protein family 70
HSP90	Heat shock protein family 90
Htt	Huntingtin
Httex1	Htt exon 1
i	Cells with Inclusions
IAA	Iodoacetamide
IB	Inclusion bodies
IKK	IκB kinase
IPSC	Induced pluripotent stem cells
IRE1α	Inositol-requiring enzyme 1 α
ISA1α	Inositol-requiring enzyme 1 α
IT15	Interesting transcript 15
JMJD6	Jumonji domain-containing 6
JNK	Jun N-terminal kinase
KEGG	Kyoto Encyclopedia of Genes and Genomes

KIAA0101	PCNA-Clamp-Associated Factor
KIF5C	Kinesin Family Member 5C
LC3	Microtubule-associated protein 1A/1B-light chain 3
LC-MS/MS	Liquid chromatography
MS/MS	Tandem mass spectrometry
LIF	LIF Interleukin 6 Family Cytokine
MAP2K6	Mitogen-Activated Protein Kinase 6
MAP3K5	Mitogen-Activated Protein Kinase 5
MAPK	Mitogen-activated protein kinase
MBNL	Muscleblind like splicing regulator
MDM2	MDM2 Proto-Oncogene
mHtt	Mutant huntingtin
MLH1	MutL Homolog 1
Morf4l1	Mortality factor 4 like 1
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MSH3	MutS Homolog 3
MTRNR2L2	MT-RNR2 Like 2
MXE	Mutually exclusive exons
Mycbp2	MYC Binding Protein 2
N2B27	N2B27 neurobasal media
NaCNBH3	Sodium cyanoborohydride
NAD	Nicotinamide adenine dinucleotide
NAMPT	Nicotinamide Phosphoribosyl transferase
Ndufa12	NADH dehydrogenase 1 alpha subcomplex 12
NES	Nuclear export signal
NFkB1	Nuclear factor kappa b subunit 1
ni	Cells with diffuse Httex1
NMDA	N-methyl-D-aspartate
NMR	Nuclear magnetic resonance
nNOS	Neuronal nitric oxide synthase
NP40	Nonyl phenoxypolyethoxylethanol
Nrbf2	Nuclear receptor binding factor 2
OMIM	Online Mendelian Inheritance in Man

PACSN1	Protein Kinase C And Casein Kinase Substrate In Neurons 1
PAXIP1	PAX Interacting Protein 1
PBS	Phosphate-buffered- saline
PCR	Polymerase chain reaction
PERK	Protein kinase R-like endoplasmic reticulum kinase
PET	Positron emission tomography
Pik3r2	Phosphoinositide-3-kinase regulatory subunit 2
PolyQ	Polyglutamine
Pparg	Peroxisome proliferator activated receptor gamma
PQC	Protein Quality Control
PRD	Polyproline-rich domain
PSI	Percent spliced in
PulSA	Pulse shape analysis
QC	Quality control
qPCR	Quantitative PCR
Racgap1	Rac gtpase activating Protein 1
Raf1	Raf-1 proto-oncogene, serine/threonine kinase
RAN	Repeat-associated non-AUG
Rap1a	Ras-related protein rap-1a
Rapgef1	Rap guanine nucleotide exchange factor 1
RELA	RELA Proto-Oncogene, NF-KB Subunit
REST	Repressor element-1 transcription factor/neuron-restrictive silencer factor (REST/NRSF)
RFU	Relative fluorescence units
RI	Retained intron
RIPK1	Receptor-interacting serine/threonine-protein kinase 1
rMATs	Replicate Multivariate Analysis of Transcript Splicing
RN	Ras-related protein rap-1a (rap1a)
ROI	Region of Interest
Rpl40	Ribosomal protein L40
RPS/RPL	Ribosomal associated protein
RRM	RNA recognition motif
Rrs1	Regulator of ribosome synthesis
SC-35	Serine/Arginine-rich splicing factor (SRSF2)

SE	Skipped exon
Sema4c	Semaphorin 4C
SMA	Spinal muscular atrophy
Smarca4	SWI/SNF-Related Matrix-Associated Actin-Dependent Regulator of Chromatin Subfamily A Member 4
snRNA	Small nuclear RNA
SQSTM1	Sequestosome 1
SR	Serine and Arginine
SREK1	Splicing Regulatory Glutamic Acid and Lysine Rich Protein 1
SRSF2	Serine and Arginine Rich Splicing Factor 2
SRSF4	Serine and Arginine Rich Splicing Factor 4
SRSF6	Serine and Arginine Rich Splicing Factor 6
SSC	Side Scatter
STAR	Spliced Transcripts Alignment to a Reference
STAT3	Signal Transducer and Activator Of Transcription 3
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SUMO2	Small ubiquitin-related modifier 2
TAFII130	TATA-box-binding protein-associated factor II
TBP	TATA-binding protein
TCEP	Tris (2-carboxyethyl) phosphine
TCERG1	Transcription Elongation Regulator 1
TEAB	Tetraethylammonium tetra hydroborate
TNFR	Tumor necrosis factor receptor
TOR1	Serine/threonine-protein kinase TOR1
TP53	Tumor protein p53
TPE-MI	Tetraphenylethene maleimide
TRADD	Tumor necrosis factor receptor type 1-associated DEATH domain protein
TRAF2	TNF Receptor Associated Factor 2
tRNA	Transfer RNA
TRNT1	Transfer nucleotidyl transferase 1
U2AF	U2 Small Nuclear RNA Auxiliary Factor
Uba52	Ubiquitin A-52 Residue Ribosomal Protein Fusion Product 1
UBC	Ubiquitin C
UPR	Unfolded protein response

UPS	Ubiquitin-proteasome system
U-snRNA	Uridyl-rich small nuclear RNA
U-snRNP	Uridyl-rich small nuclear ribonucleoproteins
VHL	Von Hippel-Lindau Tumor Suppressor
WT	Wild-type
μg	Microgram
μL	Microlitre

List of Contents

Abstract	i
Declaration	ii
Awards, Presentations and Publications	iii
Preface	v
Acknowledgements	vi
Abbreviations	viii
Table of contents	xv
List of figures	xvii
List of tables	xviii

Table of Contents

<i>Chapter 1:Introduction</i>	1
1.1. Huntington's disease	2
1.2. History of Huntington's disease	2
1.3. Epidemiology of Huntington's disease.....	4
1.4. Genetics of the Huntington's disease	5
1.5. Clinical aspects of Huntington's disease.....	7
1.6. Neuropathological hallmarks of Huntington's disease	9
1.7. Huntingtin protein structure and function	10
1.8. Huntington disease models	14
1.9. Htt aggregation states	16
1.10. Huntington disease gene expression changes.....	18
1.11. Unfolded protein response in Huntington's disease	21
1.12. Thesis outline and aims.....	24
<i>Chapter 2:Htt exon 1 aggregation state induced transcriptomic changes in Huntington's disease</i>	25
2.1. Abstract.....	26
2.2. Transcriptional profiles for distinct aggregation states of mutant Huntingtin exon 1 protein unmask new Huntington's disease pathways.....	27
2.3. Conclusion and future directions.....	37
2.4. Supplementary material	38
<i>Chapter 3:Httex1 aggregation states induce spliceosome defects and aberrant splicing in Huntington's disease</i>	47
3.1. Abstract.....	48
3.2. Introduction	49
3.3. Materials and methods.....	53

3.3.1. DNA constructs	53
3.3.2. Cell culture	53
3.3.3 RNA sequence processing and alternative splicing analysis	54
3.3.4. Enrichment of GO Category and KEGG Pathway.....	55
3.3.5. Network Analyst - Minimal interaction networks	56
3.3.6. Collection of cells by Pulse Shape Analysis	56
3.3.7. Whole cell proteomics.....	56
3.3.8. Immunofluorescence.....	61
3.3.9. Quantitative real-time PCR	61
3.4 Results and Discussion.....	64
3.4.1. Soluble mutant Httex1 affects spliceosome pathway and related process.....	64
3.4.2. Spliceosome integrity is defective in HD.....	70
3.4.3. Aberrant splicing is seen in different aggregation of Httex1	72
3.5 Conclusions	92
<i>Chapter 4: A novel proteomic approach for measuring unfolded protein load and proteostasis...</i>	<i>95</i>
4.1. Abstract.....	96
4.2. A thiol probe for measuring unfolded protein load and proteostasis in cells.	97
4.3 Conclusion and future directions.....	109
4.4. Supplementary material	110
<i>Chapter 5: Perspectives and Conclusion.....</i>	<i>111</i>
<i>Bibliography.....</i>	<i>116</i>

List of figures

Figure 1.1: Roadmap of Huntington disease discovery	4
Figure 1.2: Global prevalence of HD.....	5
Figure 1.3: Clinical progression of Huntington disease	8
Figure 1.4: The different brain regions affected by HD.....	10
Figure 1.5: Schematic representation of the full-length structure of Huntingtin	12
Figure 3.1: Diagram depicting spliceosome assembly on nascent mRNA	51
Figure 3.2: Spliceosome function is altered in mHtt-72Q-ni compared to 25Q-ni	65
Figure 3.3: SRSF2 activity is increased cells in Neur2a cells 72Q aggregation states	67
Figure 3.4: SRSF2 activity is increased cells in HD - hiPSCs	69
Figure 3.6: Different classes of alternate splicing events across aggregation states	73
Figure 3.7: Global exon skipping events seen in different aggregation states of Httex1.....	77
Figure 3.8: Heatmap of biological processes altered by splicing across Httex1 aggregation states	80
Figure 3.9: Httex1 aggregation state related splicing signatures	86
Figure 3.10: Volcano plots of protein expression changes mapped to aggregation states of Httex1	89
Figure 3.11: Schematic of aggregation state related splice variants identified in transcriptome data verified as protein expression changes	90

List of tables

Table 3.1: List of human iPSC HD neuronal lines and their corresponding CAG sizes.....	53
Table 3.2: Real-time PCR reactions custom master mix preparation	62
Table 3.3: List of primers used for the study	63

Chapter 1

Introduction

1.1. Huntington's disease

Huntington's disease (HD, OMIM # 143100) is a devastating, autosomal dominant, highly penetrant eventually fatal neurodegenerative disorder. The onset of HD is typically in the 3rd to 4th decade of adult life and is gradually progressive. The disease is characterised by a triad of involuntary movements (hence also called Huntington's Chorea), cognitive decline, emotional and behavioural problems. The defect that results in HD is the unstable number of CAG repeats present in the Huntingtin gene. In mutation carriers, the abnormally long CAG repeats (of greater than or equal to CAG of 36) gets translated into an abnormally long polyglutamine(polyQ) sequence in the Huntingtin protein (Htt)[1]. This CAG mutation results in either a toxic gain in function or a loss of normal Htt protein function and more recently, RNA mediated toxic gain of function has been implicated in the pathogenesis as well [2, 3]. Although the mutant Htt (mHtt) protein is initially soluble, it ultimately leads to the formation of insoluble aggregates through various intermediates. HD is characterised by a progressive loss of striatal and cortical neurons in the brain. There is no known cure for HD and death occurs 15 - 20 years after onset. Treatment options are limited and are mostly symptomatic[4].

1.2. History of Huntington's disease

HD is also called Huntington's chorea. Paracelsus, (1493-1541) first coined the term "chorea", which is recognised as an early symptom of HD[5]. The term "chorea" is derived from Greek word "choreia" meaning dance. Since HD involves a loss of voluntary control of body movements, patients with HD were thought to be possessed and were treated as such. HD was named after George Huntington, who in 1872 first accurately described the symptoms of the disease and emphasised its hereditary nature in a landmark paper in the Philadelphia journal, The Medical and Surgical Reporter[6]. In 1906 Punnet described HD as autosomal dominant[7]. The US-Venezuela Huntington's Disease Collaborative Research Project started in 1979 laid the foundation for the discovery of the Huntington gene. Based on the work started by Americo Negrette, Nancy Wexler researched and gathered pedigrees of families with a history of HD from the region around Lake Maracaibo, Venezuela, where HD was highly prevalent[8, 9].

A pedigree containing more than 3,000 individuals was compiled by 1983, and by detailed genetic linkage analysis James Guesella and colleagues, at the Massachusetts

General Hospital, Boston determined the location of the disease causing gene on human chromosome 4[10]. Finally, in 1993, the Huntington disease study group[1] identified the precise location of the causal gene at 4p16.3, the terminal band on the short arm of chromosome 4, making it the first autosomal locus to be detected using genetic linkage analysis. The gene was alluded to as "interesting transcript 15 - *IT15*" and is now identified as *HTT* – *Huntingtin* gene and the protein named "Huntingtin (Htt)"[1]. Further fine mapping of microsatellite markers revealed a defect in the exon 1 of mutant Huntingtin (mHtt) gene. The defect related to the instability in the number of CAG repeats[11]. The first animal models for HD were developed by 1996, and subsequently in 1997, micro aggregates or inclusion bodies were described as the characteristic feature in both patient and animal models[12, 13]. One of the first gene expression studies performed in transgenic mice expressing exon 1 of the *Htt* gene showed altered neurotransmitter receptors in HD brain and transcriptional dysregulation was attributed as possible mechanism of pathogenesis[14] (Figure 1.1). The first phase III randomised controlled clinical trials of coenzyme Q₁₀ and remacimide in HD were conducted by the Huntington study group in 2001[15]. In 2015, the first patients were treated with tominersen, with late 2017 marking the completion of phase 1/2a clinical trials for this gene silencing therapy for HD[16, 17]. Tominersen is a potential antisense mRNA therapy for HD. It binds to the faulty mHtt mRNA targeting it for degradation and reducing mHtt protein produced by cells. Other antisense therapies from Wave Life Sciences are in advanced trials[18]. However, the potential risk of the drug targeting wild-type Htt mRNA is a complicating factor and the requirement for intrathecal injections makes the treatment difficult.

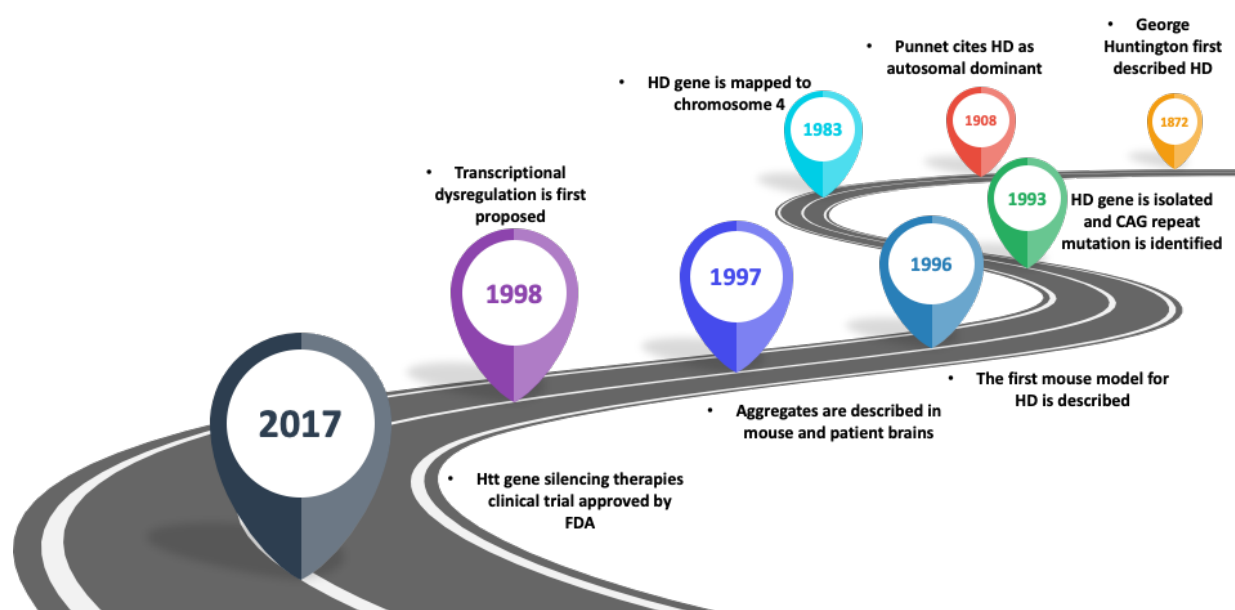


Figure 1.1: Roadmap of Huntington disease discovery. Key milestones in Huntington disease research from the first description of HD by George Huntington, to the identification of the genetic loci responsible for disease to the other advances in more than 25 years of HD research.

1.3. Epidemiology of Huntington's disease

Though HD is a rare disorder, it is one of the most common in the family of neurodegenerative disorders that is caused by polyglutamine expansions. While the disease occurs in all racial groups, it is more common in people of Western European origin. It affects both male and female equally. Although Venezuela has the highest per capita HD population in the world, North America and Europe have the highest HD prevalence, with an incidence of 4-8 cases for every 100,000 controls[19]. In Australia, it is about 5-7/100,000, with a higher prevalence of 12-13/100,000 in Tasmania[20]. The disease may also be frequent in parts of central Asia and in the Indian subcontinent (1.75 per 100,000) because of Caucasian admixture in the native population. The Chinese (0.3/100,000), Japanese (0.65/100,000) and African populations show a significantly lower prevalence of HD[21, 22] as seen in the Figure 1.2.

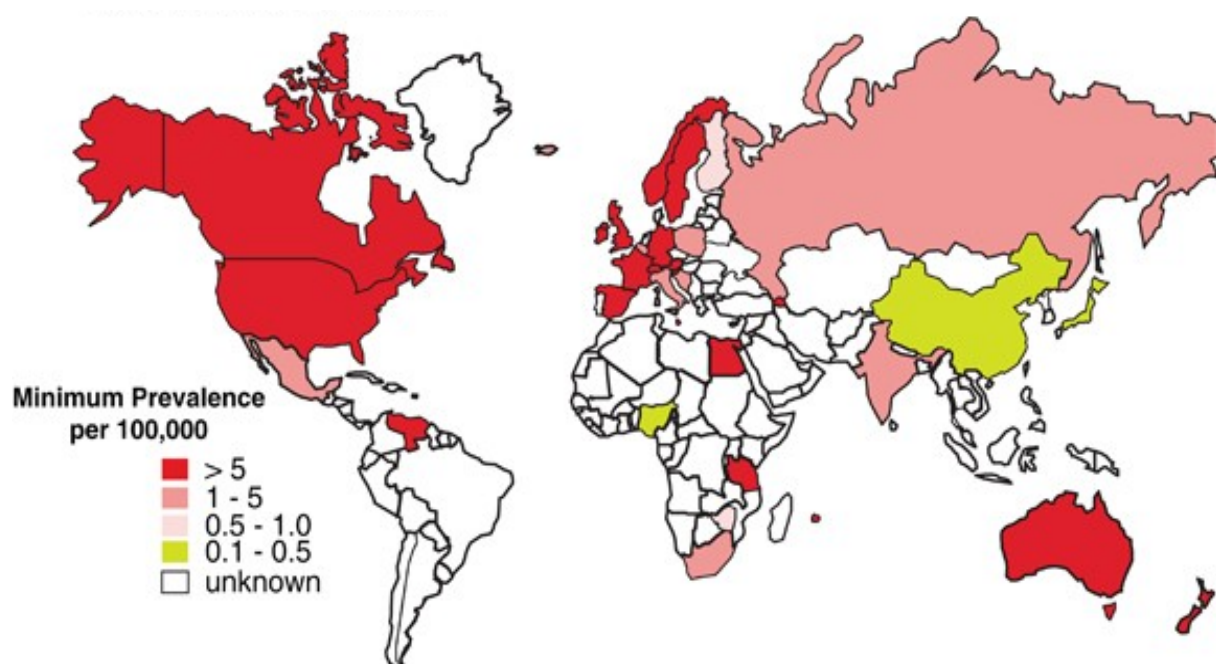


Figure 1.2: Global prevalence of HD. HD prevalence is substantially higher on a global scale in European populations, than in Asian populations. Venezuela has the highest recorded prevalence of HD worldwide (700 per 100,000). The origins of HD in Venezuela can be traced back to European settlers who introduced the disease into the indigenous populations, as in most cases in Asia. There is limited data on HD prevalence in Africa, compounded by limited sample sizes, and lack of reliable molecular testing verifying the HD diagnosis, making the statistics unreliable in its estimation (Adapted from Warby et al., 2015[22]).

1.4. Genetics of Huntington's disease

HD was the first autosomal dominant disease whose cause was identified by genetic linkage analysis[1]. The genetic defect was proven to be from an expanded series of CAG repeats in the first exon of *HTT* gene (*Httex1*) which resulted in an abnormally long polyglutamine (polyQ) tract near the N-terminus of the Huntingtin protein[23]. The length of CAG repeat in exon-1 is highly heterogenous with healthy individuals having 8-36 CAG repeats with the mean CAG repeat of approximately 23[24]. If the number of CAG repeats exceeds 40 or more, the expansion is pathologic, fully penetrant and causes disease. If the number of CAG repeats is between 36-40, there will be reduced penetrance, and affected individuals will experience late-onset HD after 60 years[25, 26]. CAG repeat lengths greater than sixty produce a more aggressive phenotype with onset before the second decade of life called juvenile-onset HD [27]. Although HD has a highly variable age of onset (mean age of onset ~45 years), there is a strong inverse correlation between the age of onset and the number of CAG repeats, with higher expansions resulting in a faster onset of disease[28]. The maximum length of the polymorphic CAG repeat in the *Htt*-gene mutant allele is the most significant determinant of age of onset of HD and accounts for most of the variability(~56%) seen in the age of onset[29]. In addition, the size of normal CAG allele also influences the severity of HD, but to a lesser extent[30]. The presence of normal allele with large normal CAG repeats is associated with later onset[31], reduced clinical severity, progression of the disease[32]. Further, the normal CAG allele and its interplay with the expanded CAG allele is also known to affect basal ganglia volumes in premanifest HD mutation carriers[30]. A competitive length-dependent interaction of the polyQ domains of normal *Htt* and mutant *Htt* with their binding partners might underpin these findings. Since mHtt sequesters normal *Htt* and promotes fibrillogenesis at a rate that is proportional to the length of the polyQ stretch in either normal or mutant protein[33, 34], increasing length of polyQ in normal *Htt* could lead

to a stronger association with mHtt fragments, thereby promoting their sequestration and hence preventing them from aberrantly interfering with other essential proteins[35].

The CAG repeat in Htt-exon 1 is also subject to meiotic instability, with an increase in CAG repeat length in successive generations leading to an earlier onset of disease, a phenomenon known as anticipation. Anticipation can occur through the maternal and paternal line though is greater through paternal transmission during spermatogenesis[36]. Furthermore, somatic instability of the CAG repeats in post-mitotic neurons have also been reported and might alter the age of onset [37]. In addition to CAG repeat length in Httex1, other genetic and environmental factors can contribute to disease[38]. To understand the contribution of genetic factors other than CAG repeats in the *HTT* gene, genome-wide association studies have been conducted in large multi-clinic cohorts of HD patients. The Genetic Modifiers of Huntington's Disease Consortium, 2015 identified two loci in chromosome 15 and a single locus in chromosome 8 that altered the neurological onset of disease. Further pathway analysis revealed genes related to DNA repair and maintenance (*MLH1* located on chromosome 3) as a potential modifier of disease onset[39]. Another genome-wide study to find genetic variants of disease progression identified three genes (*MSH3*, *DHFR*, and *MTRNR2L2*) in chromosome 5 to modify progression of disease. Both *Mlh1* and *Msh3* associated with DNA mismatch repair are critical modifiers of somatic instability in mouse models of HD[40, 41]. Indeed, DNA repair pathways are a common genetic mechanism affecting multiple polyQ diseases[42]. Though studies have shown the disease-modifying effect of DNA damage response genes on both the onset[43] and progression of HD[44], a greater understanding of the underlying molecular mechanisms is an essential precursor in developing disease-modifying therapies targeting DNA repair associated genes in HD. Other candidate genes encoding including Htt-interacting proteins (*HAPI*, *TCERG1*), neurotransmitter receptors (*ADORA2A*, *GRIN2A*, *GRIN2B*, *GRIK2*), synaptic processes, energy metabolism (*PPARGC1A*, *Co1*, *NRF-1*, *TFAM*), stress response and apoptosis (*MAP2K6*, *MAP3K5*, *TP53*, *DFFB*), and sequence variations in *HTT* gene (CCG repeat polymorphism and $\Delta 2642$ polymorphism)[38, 45] have also been implicated as modifiers of HD but these findings are yet to be robustly replicated as disease modifying targets.

While HD has a major genetic component in its pathogenesis, environmental factors and lifestyle have also been tied to the onset and progression of HD. Wexler et al., 2004 first determined that major variability in the age of onset could be ascribed to unidentified shared and non-shared environmental factors[46]. Clinical studies and animal models have shown support for this variability, with a lack of physical activity and chronic stress exacerbating

symptoms of HD development and progression[47]. Environmental interventions via cognitive stimulation, physical activity, stress reduction and diet, combined with future pharmacological therapies could help ameliorate the burden of HD.

1.5. Clinical aspects of Huntington's disease

From a clinical standpoint a diagnosis of HD is made when a patient exhibits an unexplained characteristic movement disorder, with a positive family history of similar illness. The existence of a defined mutation in CAG repeat numbers in the *Htt* gene allows individuals to undergo predictive and diagnostic genetic testing[28]. The CAG repeat expansion being the key diagnostic criterion for positive confirmation. The CAG repeats can also help predict phenoconversion in the clinic. CAG length between 40 and 60, phenoconversion is in the third and fourth decade of life. In cases where there is a large expansion of the greater than 60 CAG repeats, the disease phenoconversion occurs much early and is called Juvenile Huntington's disease with onset before 20 years[27]. Individuals carrying CAG repeats between 36-39 develop late-onset HD post 60 years as described earlier[25, 48]. Hence genotyping is essential for diagnostics, support and planning services for the patient and family. Even though development of motor symptoms is indicative of phenoconversion or onset of disease, psychiatric illness and subtle motor deficits can be present up to 10-15 years before the onset of clear movement disorder and must be managed appropriately[28].

Clinically HD has a long pre-manifest phase (asymptomatic) and prodromal phase (subtle signs and symptoms), and once symptoms develop also called the manifest phase, has a long disease course over 10-20 years[28] as shown in Figure 1.3. HD usually presents as a triad of motor, psychiatric and cognitive symptoms. The characteristic motor symptoms include incoordination, involuntary movements, difficulties with speech and swallowing, difficulties in balance, and walking and is usually the first symptom to present. Cognitive symptoms include inability to cope, lack of concentration, cognitive decline and eventually dementia. Psychiatric symptoms include irritability, depression, anxiety, mood swings, social withdrawal, and suicidal ideation. In addition, there is an increased suicide rate among patients of HD[4]. Even though motor symptoms are indicative of phenoconversion to HD, the cognitive and psychiatric aspects of HD may develop many years before this. Clinical characterisation of HD progression is carried out using the Unified Huntington's Disease Rating Scale[49] and Shoulson-Fahn Total Functional Capacity rating scales[50]. The CAG repeat length is correlated with the clinical progression of motor and cognitive deficits seen

in HD[4, 51, 52]. A measure called the standardized CAG age product score (CAP score) has been developed and used in longitudinal studies of HD pathology in large cohorts of patients with of different age groups and CAG repeat sizes. The standardized CAP score used to predict HD progression is a function of CAG repeat expansions and duration of exposure to the effects of expansion. The measure based on large scale HD datasets is derived using $100 \times \text{Age} \times [(\text{CAG} - L) \div S]$, where Age is the current age, CAG is the repeat length, and L (scaling constant), S (normalizing constant) are constants[28]. Though efforts have been made to develop detailed predictive models and characterize HD progression[25, 53], we do not have a complete understanding of the factors determining clinical features of HD. Currently, there are no approved effective treatments for HD, although tetrabenazine and deutetrabenazine have been approved for treatment of the HD related chorea[54]. Psychiatric symptoms are managed conventionally, with antidepressives and antipsychotics. Newer therapeutics targeting mHtt production by targeting *Htt* mRNA have shown promising results and are currently in trial[55].

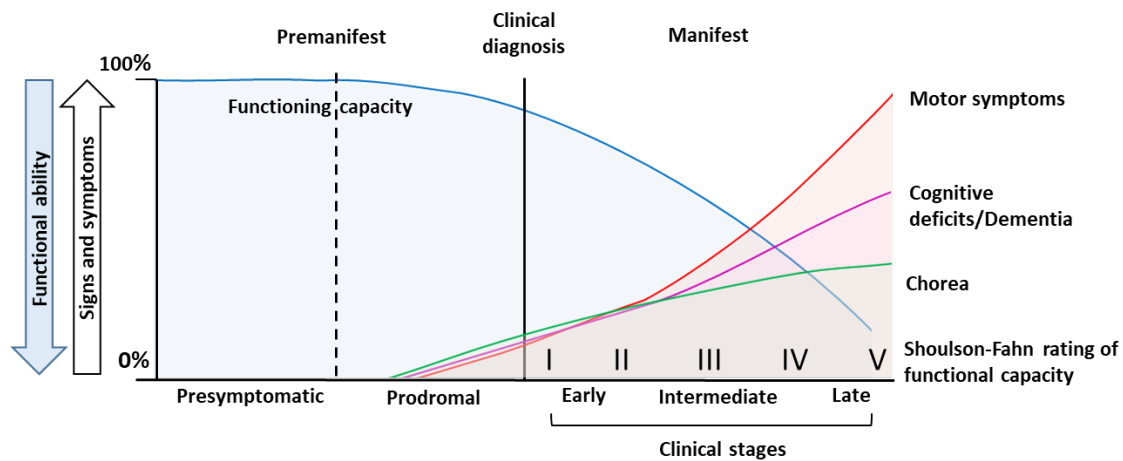


Figure 1.3: Clinical progression of Huntington disease. The clinical progression can be classified into premanifest and manifest phases. The initial premanifest phase occurs before there is clear diagnosable symptoms of HD. In the premanifest phase HD can go through stages of asymptomatic (symptoms have not been displayed) and prodromal HD (symptoms are very subtle). The Manifest HD phase is characterized by gradually progressive impairments of motor and cognitive abilities and dementia. Chorea develops in patients early on but plateaus with disease progression. The manifest phase can be divided into early, moderate and late stage of disease, classified according to the Shoulson-Fahn Functional Capacity rating. Typically, disease onset is seen at ~45 years of age which correlates with the normalized CAP score of ~100. Thus, CAP score facilitates the mapping of progression of the disease in individuals with different CAG expansion lengths onto the same chart. Adapted from[28].

1.6. Neuropathological hallmarks of Huntington's disease

The neuropathology of HD is marked by gradually worsening degeneration in the striatum and basal ganglia, i.e., caudate and putamen. Moreover, recent pathoanatomical and neuroimaging studies[56-58] of the HD brain have shown that it is a multisystem degenerative disorder that affects various parts of the brain, including the cerebral cortex, thalamus, brainstem, as well as the cerebellum. HD affects all the cerebral lobes and is further characterised by degradation and thinning of their cortical mantle accompanied by atrophy and shrinkage in various areas of the cerebral cortex[59]. There is layer specific neuronal death in cerebral cortex especially in layers III, V and VI[60]. Cognitive decline, psychiatric symptoms, and neurophysiological defects are now thought to be associated with this loss of cortical neurons. The choreiform movements are primarily attributed to striatal neurodegeneration and impairment of corticostriatal projection neurons and their afferent inputs[59, 60]. The Vonsattel grading system classifies the extent of neuropathological involvement in the striatum into five grades ranging from zero to four with zero showing undiscernible neuropathology to 90-95 per cent neuronal loss in grade four[61]. The loss of both cortical and striatal white matter accompanied by loss of basal ganglia grey matter have been confirmed as an early feature of HD in vivo by neuroimaging studies[62] (Figure 1.4A). Due to the widespread neuropathological changes seen in HD, it is now widely believed that HD exhibits similarities with other spinocerebellar ataxias caused by polyglutamine expansions than previously thought.

The susceptibility of striatal neuronal loss correlates with the length of the CAG repeat expansion[63, 64], with the enkephalin-containing striatal neurons most vulnerable to degeneration. In presymptomatic HD patients both striatal neurons projecting to the substantia nigra and cortical pyramidal neurons that project into the striatum demonstrate early degeneration[65, 66]. Cortical neurons exhibit reduced staining for phosphorylated neurofilaments, tubulin, and key proteins such as microtubule-associated protein 2, complexin 2, dynamin and neuro specific phosphoprotein PACSIN 1[67, 68]. These proteins are essential in neurons for their cytoskeletal integrity, axonal transport and synaptic function indicating that cortical dysfunction is important in the pathogenesis of the disorder. Evidence also indicates a major glial component in HD pathogenesis. Activated microglia that cause neuronal death are seen in the early stages of HD pathology in both the cortex and striatum and persists through the late stages of the disease[69]. This is also accompanied by an increase in the activated astrocytes and oligodendroglia in the striatum[61, 70].

One of the key neuropathological hallmarks of HD is the appearance of inclusions enriched with polyQ (Figure 1.4B)[12, 59]. The presence of excess glutamines due to their polar nature causes links to form both within and between proteins making mHtt stickier. Hence rather than folding into functional proteins, mHtt develops into tangled, rigid groupings resulting in the formation of insoluble inclusions in affected neurons. The aggregation related changes in molecular pathology are described in detail in section 1.9.

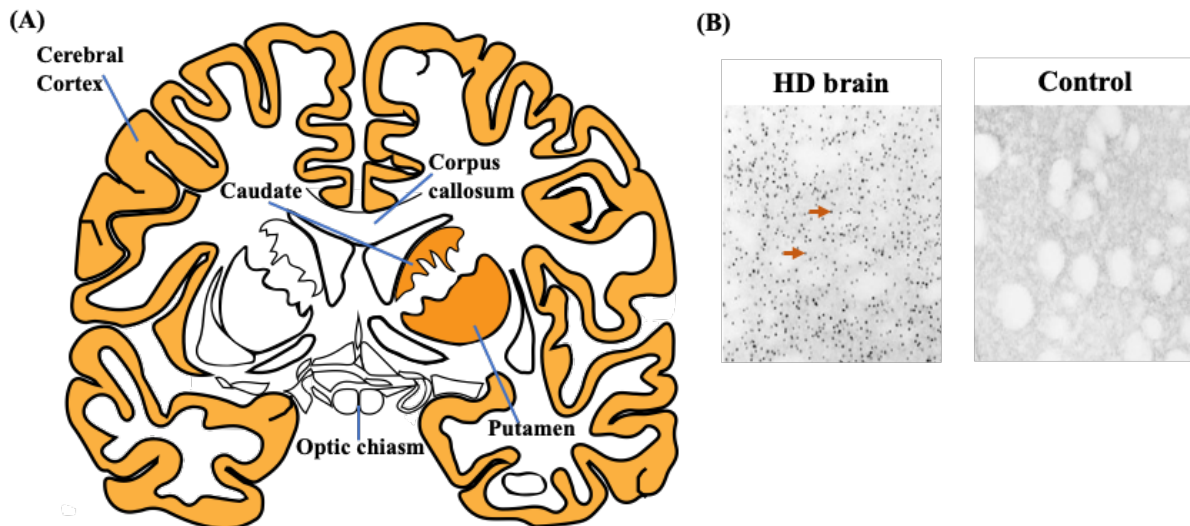


Figure 1.4: The different brain regions affected by HD. (A) Current pathoanatomical changes depicted in HD include severe involvement of the basal ganglia including the caudate nucleus and putamen as well as cerebral cortex. Affected areas are shaded in orange. Adapted from Rub et al., 2016[59]. (B) Photomicrographs of Htt immunoreactivity in mouse striatum. The images show Htt immunoreactive nuclear inclusion as indicated by arrow in transgenic mouse brain carrying pathogenic repeats. Normal control brain with 18 glutamine repeats no such inclusions can be seen. Adapted from Davies et. al., 1997[12].

1.7. Huntingtin protein structure and function

Huntingtin is a large protein of 348 kDa that contains 3144 amino acids(NCBI Reference Sequence: NP_002102.4) (Huntington's Disease Collaborative Research Group 1993)[71]. There are 67 exons in the *Htt* gene with the gene. The *Htt* gene is ~180 kilo bases in length and has two mRNA transcripts of 10,366 and 13,711 base pairs(bp) in length. The second transcript differed in the 3' untranslated region of the mRNA by a sequence of 3,360 bp and showed tissue specific enrichment in the brain. These transcripts of varying length could be produced due to differential polyadenylation[72]. Two studies on alternative splicing in the

Htt gene also showed the presence 5 additional transcripts resulting in the generation of polyQ containing variants of Htt protein missing exons 10, 12, 29, and 46 or instead having an additional exon 41b or keeping a 57-bp portion of intron 28. However, none of the novel isoforms generated a truncated Htt protein. Instead, 4 of the 5 novel isoforms identified specifically delete domains and alterations to generate smaller Htt proteins[73, 74]. Loss of these regions may alter the protein-protein interaction and change the susceptibility to cleavage as compared to canonical Htt which can alter the function of these isoforms.

The first exon of *Htt* gene (*Httex1*) encodes around 100 amino acid long N -terminal region of the Htt protein. It encompasses the amino terminal domain (1-17 amino acids), the polyglutamine tract (polyQ) which begins immediately after the amino terminal region and a 51 amino acid long polyproline-rich domain (PRD). The polyQ tract and PRD domain which flanks the poly glutamine region are variable in humans [2]. The amino terminal domain plays a very important role in HD pathogenesis[75]. It adopts an amphipathic α -helical structure[76] and is believed to mediate interactions with cellular membrane encased organelles such as the mitochondria, Golgi apparatus and endoplasmic reticulum[76-79]. In neurons there is disruption of organellar endomembranes and membrane dynamics by amyloid-like fibrils of polyQ inclusions[80, 81]. These aberrant interactions are mediated by the amino terminal of Htt and contribute to the damaging cellular effects of protein aggregations[82]. This domain also serves as the nuclear export signal (NES) and was identified as a critical region playing a role in the localisation of Htt during stress, aggregation, and associated toxicity[76]. Further the amino terminal is subjected to post translational modifications like sumoylation, acetylation, ubiquitination at lysine's 6, 9, 15 and phosphorylation at serine's 13, 16 that affect the subcellular localization and clearance of Htt[2]. During periods of stress, phosphorylation of serine residues 13 and 16 induce a conformational change in the α -helical N17 region, and subsequent targeting of Htt to subregions within the nucleus. mHtt is found to be hypo phosphorylated at these residues, and as such fails to display this translocators response to stress[79].

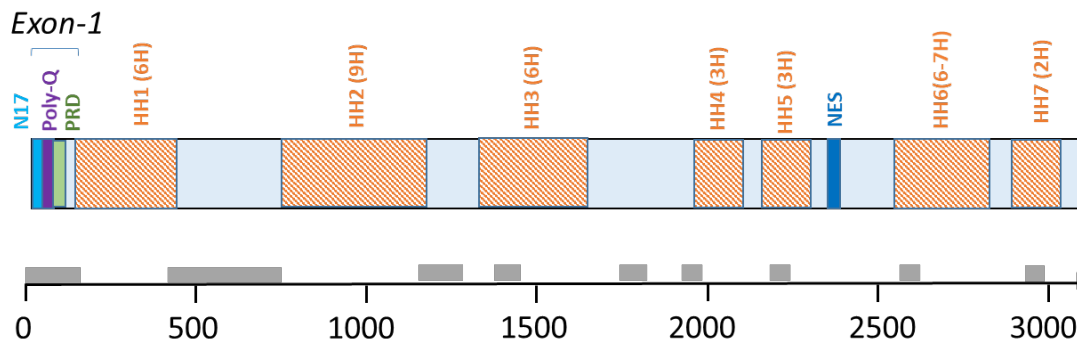


Figure 1.5: Schematic representation of the full-length structure of Huntingtin. Top: the Htt domain sequence. N17 – 17 amino acid sequence of the N-terminal domain of exon 1; polyQ – polyglutamine tract; PRD – proline-rich domain. HH1-7 corresponds to the functionally predicted HEAT domains. HH1 and 3 contain 6 repeats; HH2 has 9 repeats; HH4 and 5, 3 repeats; HH6, 6-7 repeats; HH7, 2 repeats. NES – signal for nuclear exports. Bottom: position of amino acids. Gray bars depict highly disordered regions. Figure adapted from Saudou et al., 2016[2].

The polyQ region of the Htt protein influences the structural and functional properties of the protein. Recent electron microscopy of Htt protein co-expressed and co-purified with the huntingtin-associated protein HAP40 has revealed that Httex1, with its polyQ domain, is disordered and flexible in the native wildtype protein, as no electron density was observed for this part of the protein[83]. This flexibility in wildtype Htt contrasts with the rigid non-native state of the polyQ domain in misfolded Httex1 aggregates, previously seen in various studies using electron microscopy[82], electron paramagnetic resonance spectroscopy[84], and solid-state nuclear magnetic resonance[85]. Httex1 by itself has been previously shown to cause the pathology in HD[86] and has been more studied than the full-length huntingtin due to its role in disease. Moreover, due to its conformational heterogeneity, size, and versatility, the structural characterization of full length Htt protein has been difficult. The full length Htt protein after Httex1 contain series of intrinsically disordered regions interspersed with several (up to 36) HEAT-repeats, clustered into 7 large alpha rod domains[2, 87] as shown in Figure 1.5. The HEAT repeat domains are sequences of around fifty amino acids separated by a non-helical region in anti-parallel alpha-helices and are found in Huntingtin, elongation factor 3, protein phosphatase 2A, and TOR1[2, 88]. These repeats are hypothesized to serve as solenoid-like structure scaffolding numerous protein complexes and mediating protein-protein interactions[89].

Furthermore, multiple protease cleavage sites are also present in the disordered regions in Htt and undergo extensive post-translational alterations, including SUMOylation, acetylation, phosphorylation, ubiquitination, palmitoylation and myristoylation[90-95] which might be involved in the mechanism regulating Htt activity and affect characteristics including protein half-life, aggregation and can even modulate toxicity by increasing Htt clearance through the autophagy related mechanisms[91, 92] as well as the ubiquitin-proteasome system (UPS)[96, 97].

Both *Htt* mRNA and protein are ubiquitously expressed and are widely distributed in the brain and other tissues[98, 99]. Htt production in neurons is higher than in other tissues, beginning early in development and continuing through adulthood. Htt is evolutionarily well conserved and interacts with many protein partners. It is considered essential for cell function as deletion of Htt results in embryonic lethality in mice[100]. Though the exact function of Htt is not yet understood, it is known to play an important role in multiple processes in the cell including cellular transport, endocytosis, vesicle trafficking and recycling, cell division coordination, autophagy[2] as well as transcriptional regulation[101].

Htt is designed to serve as a molecular scaffold due to its large size and multiplicity in reported molecular partners. Htt serves as a hub, linking multiple partners to a complex and coordinating different cellular processes. Htt scaffolds dynein/dynactin molecular motors in the cell[102], to regulate many cellular processes including anterograde and retrograde transport of organelles and vesicles through neuronal processes[103, 104].

The cellular molecular motors interact with Htt to regulate the cellular movement of essential cargoes, including presynaptic vesicles, autophagosomes, endosomes and lysosomes. This is achieved through direct interaction of Htt with dynein or indirectly with the p150 subunit of dynactin[103] and the kinesin 1 member KIF5C[105] via Huntingtin-associated protein 1 (HAP1) to Htt. In addition to the interactions with these molecular motors, Htt and HAP40 also interact via optineurin/Rab8 with actin-based motors that connect myosin VI to the Golgi complex[106] [107] to assist the trans-Golgi endosomal trafficking and via Rab5 for endosomal cargo transfer between microtubules and actin cytoskeleton[108]. Htt also interacts with several endocytosis related proteins, such as huntingtin-interacting protein 1 (HIP1) and huntingtin-interacting protein 1 related (HIP1R), which are involved in clathrin-mediated endocytosis by facilitating membrane invagination and clathrin coating assembly.[109]. As a result, Htt is crucial for endosomal trafficking, vesicular transport, and disruption of this function can lead to neuronal dysfunction and death.

Htt is also thought to be anti-apoptotic. It inhibits the main inducers of apoptosis, the caspase 9[110] and caspase 3[111]. HTT has been shown to increase brain-derived neurotrophic factor levels (BDNF) via multiple mechanisms including stimulation of transcription with Htt mediated inhibition of the silencing action of repressor element-1 transcription factor/neuron-restrictive silencer factor (REST/NRSF)[112] on BDNF promoter II. Downstream NRSF-regulated genes are important for neuronal growth and maintenance. Gauthier et al. also reported that huntingtin is required for vesicular transport of BDNF[113]. This is significant, because the striatum does not generate BDNF and relies almost entirely on BDNF, which is delivered by endocytosis and retrograde transportation through cortico-striatal afferences to striatal cell bodies[114, 115]. Thus *BDNF* transcription, transport, subsequent release from cortical neurons to endocytosis and trafficking into striatal neurons could be disrupted by mHtt which results in loss of trophic support and cell death[116]. Htt is also active in cell division processes. It is distributed in both differentiated neurons and dividing cells and is found in dividing cells at the spindle poles, mitotic spindles and microtubule networks[117]. During cell division particularly the transition from prophase to anaphase Htt via the dynein/dynactin/NUMA/LGN complex is directed to the spindle poles and is essential for the assembly and orientation of mitosis spindles[118]. Htt also plays pivotal role as a regulator of autophagy as well and as a master regulator of transcription[2, 101]. The transcriptional dysregulation by mHtt is described in section 1.10.

1.8. Huntington disease models

Over the last two decades, our understanding of HD disease pathology has improved considerably, but effective treatment to prevent or control the progression of the disease is still unavailable. Thus, animal and cell models will continue to play a key role in understanding disease development, discovering new therapeutic agents and screening potential drug targets before testing in clinical trials.

Numerous cell models have been developed to study HD. These range from neuronal cell lines with CAG repeat knock-ins to induced pluripotent stem cells (iPSCs)[119] and iPSC-derived brain organoid models[120]. However, it is difficult to understand disease pathogenesis by studying individual cell lines or their derivatives in isolation. Animal models facilitate a better understanding of disease pathogenesis. One of the earliest model systems used to understand HD pathogenesis were acute toxin models for HD[121]. Since striatal cell loss is the primary neuropathological hallmark of HD excitotoxin kainic acid, 3-

nitropropionic acid and quinolinic acid has been injected into animals. The model animals replicated the neuropathology seen in HD, including stratum lesions similar to those seen in HD[122-124]. Although toxin models have been useful, toxin models do not adequately simulate the gradual neuronal degeneration and molecular effects associated with disease progression.

Various transgenic models from worms to primates have been developed for HD. The HD worm (*Caenorhabditis elegans*) was generated by expression of the N - terminal 171 amino acid fragment of human huntingtin protein containing the extended polyglutamine (150Qs) in neurons. Htt171 form aggregates and are toxic to neurons in a polyQ length-dependent manner[125]. This model represents one of the relatively simple genetic animal models of PolyQ neurotoxicity. Other nematode models in which 513 amino acid fragments of human Htt expressed in muscle walls have been developed to study Htt proteotoxicity. These HD worms also show motor impairment and a shortened lifespan[126]. Similarly animal models were established in species ranging from fruit flies[127], mice and rats to large animals like pigs, sheep and monkeys offering important insights and recapitulation of pathological features of polyQ dependent formation of nuclear inclusions and progressive neuronal degeneration[128]. Main distinguishing factors among models of HD are the transgene constructs used (*Htt* mutation) and approach used to produce them, which in turn specifies the pathology and symptomatology of the animal model.

Rodent models have proved to be very useful providing significant insights into HD pathogenesis related processes as well providing a platform for testing therapeutics, with more than 20 different models produced to date[128]. The first transgenic mouse models produced for studying Huntington's disease were the R6/1 and R6/2 transgenic mice expressing exon 1 of *Htt* with 115 and 150 CAG repeats respectively. A genomic fragment of 1.9 kb containing a human huntingtin promoter drives the transgene expression. The resulting levels of transgene expression in R6/1 and R6/2 mice were approximately 31 per cent and 75 per cent, respectively, compared to the endogenous huntingtin. This resulted in progressive behavioural and neuropathological symptoms[86]. In nearly all neurons ubiquitinated mutant huntingtin-containing intranuclear inclusions appeared in both R6/1 and R6/2 mice by eight to ten weeks of age, albeit at a slower rate in R6/1 mice with a smaller number of CAG repeats[12]. Both transgenic mice with full-length huntingtin[129, 130] and knock-in mouse models with extra CAG repeats being incorporated into the endogenous sites[131, 132] have been developed. Although full-length models have the greater advantage than mice expressing truncated version of Htt in more accurately simulating the genetic mutation of

human HD, these mice show slower onset of abnormal behaviour and neuropathological defects compared to mice expressing truncated huntingtin transgene.

Observations that the formation of inclusions correlate with the development of neurochemical and neurological changes[133, 134] and similar inclusion formation in HD patients[135, 136] increased the likelihood that inclusions may be the primary means by which the mutant protein causes neuronal injury in HD. Nevertheless, many studies have raised concerns about the direct contribution of inclusions to neuronal damage[137-139], with studies indicating neuronal inclusions reducing risk of neuronal death, as further elaborated in section 1.9.

1.9. Htt aggregation states

Abnormal protein folding and aggregation is thought to be involved in a number of human pathologies[140, 141]. These disorders are characterised by the eventual accumulation of large protein deposits with disease progression. The fundamental cause of these changes is the mutation induced destabilisation of native protein conformations, which can result in abnormal protein folding into non - native conformations. These changes further trigger assembly of proteins into β -sheet rich amyloid fibrils and eventually contributes to creation of the protein aggregates with disease[142].

HD is a prototypical protein aggregation disease which is characterized by the presence of inclusion bodies in striatal and cortical neurons[143]. PolyQ expansion mutations in mHtt protein cause incorrect protein folding and result in the formation of inclusion bodies. Inclusion bodies are punctate accumulations containing mHtt protein and N terminal fragments of mHtt encompassing approximate lengths of exon 1[135]. Purified Htt polyQ peptides also form fibrils with classical β -sheet amyloid properties[141, 144], with the rate of fibril formation directly proportional to increasing polyQ lengths, once the polyQ length exceeds the disease threshold[145].

Htt undergoes multiple configurational changes[2, 146]. Many studies have been conducted to ascertain the kinetics and develop models of the Htt aggregation leading to formation of insoluble fibrillar aggregates[147-149]. Studies have shown that the Htt-exon 1 fragments aggregate to generate a heterogenous mixtures with specific forms of aggregate in a concentration and length-dependent manner[148]. Numerous X-ray crystallography, fluorescence correlation microscopy (FCS), electron microscopy, atomic force microscopy, and solid-state nuclear magnetic resonance studies have also been instrumental in

characterizing the oligomers of the expanded mHtt as well as studying the oligomerization mechanism both *in vitro*[150], *in vivo*[151-154] and clarifying the morphology, organization and assembly of fibrils[148, 149, 155, 156]. Even though the mechanism(s) and conformational basis by which mHtt leads to toxicity is unclear, the composition, abundance, and persistence of toxic conformers that mediate neuronal damage in HD also must depend on the length of the polyQ[155]. Following broad classes were generally identified: monomers, soluble oligomers, amorphous aggregates, protofibrils, amyloid fibrils and other significant variants like spherical and annular oligomeric structures[150, 155, 157-160].

Htt containing 36 or less polyQ are soluble, however in mHtt, as this number increases it becomes highly aggregation-prone and insoluble[146]. This association between polyQ lengths and aggregation rates was also demonstrated in HD cell culture models[161-163]. Further a four - phase aggregation cascade was proposed to explain the aggregation of mHtt protein based on an HD model of cell culture[164]. Aggregation starts with the phase of accumulation during which only soluble, abnormally folded monomers are present at low concentrations. Small oligomers are formed in the second phase and remain in equilibrium with monomers. Nucleation is activated at one or more sites during the third phase which ultimately leads to the formation of larger inclusions, the fourth phase of aggregation[164, 165]. There is evidence that cells dynamically convert mHtt proteins to aggregates progressing from soluble monomer to oligomeric forms and finally resulting in the formation of insoluble inclusion bodies[19, 146, 166, 167]. However, the role of inclusion bodies/aggregates in HD is a contentious[168, 169]. Initially, aggregates were suggested to cause toxicity by sequestering vital cellular components[170-175]. However, other evidence suggests that aggregates may not be deleterious but are products of sophisticated quality control systems that sequester the soluble and more toxic mHtt forms in the cell[138, 161, 176]. Neurons transfected with Httex1 that had formed aggregates were more likely to survive longer in culture than neurons retaining dispersed Htt throughout the cytosol[138]. Additionally, pharmacological treatment with compounds like B2 (5-[4-(4-chlorobenzoyl)-1-piperazinyl]-8-nitroquinoline) increased aggregate formation and lessened Htt-mediated proteasome dysfunction and cellular pathology[177]. However, more recent studies have shown that both early aggregates as well as late mature aggregates to be toxic and lead to cell death by different mechanisms. The early aggregates were found to trigger apoptosis, while the mature late aggregate containing cells had delayed necrosis[178].

Various studies have also indicated that soluble forms of mHtt to be toxic. A novel *in vitro* study combining NMR and hydrogen-exchange techniques, compared the structures of monomeric non-pathogenic and pathogenic Httex1. This study revealed tadpole-like structures of soluble mHttex1 causing gain of toxic interactions[179]. Another study using fluorescence resonance energy transfer (FRET) confocal microscopy of poly-Q self-interaction reporters that distinguished oligomers from monomers and aggregates, indicated that aggregate formation resulted in rapid oligomer depletion with very little cytosolic monomers left[154]. Further longitudinal analysis of these FRET reporters indicated a higher susceptibility of death for cells with higher diffuse FRETs, in keeping with the accumulation of oligomers which causes increased proteotoxicity[154]. A “numbers and brightness” analysis also showed that when a certain critical threshold concentration of Httex1 is reached, monomers get assembled into oligomers in a two-step process. The study further confirmed the previous finding of oligomeric depletion with aggregate formation[164].

Increasing evidence appears to indicate that mHtt induced conformational changes prior to aggregation are central to the molecular pathogenesis of HD[180]. Moreover, mHtt transitions freely and rapidly between different aggregation states, and this dynamic remodeling makes it almost impossible to isolate the effects of any specific aggregation state. Challenges remain in defining how different aggregation states influences cellular activities at the molecular level, as well as in deciphering the link between Htt aggregation process and cellular responses to these changes. A better understanding of these cellular activities and their corresponding responses will help us understand the factors that are pathogenic or modify polyQ aggregation or has potential therapeutic roles. These changes have been well characterised in the study published in Chapter 2.

1.10. Huntington disease gene expression changes

A large body of studies show that transcriptional dysregulation is a key early step in HD pathogenesis[181-189]. Gene expression profiling of animal models and HD patient brains indicate a large impact of mHtt on gene expression with a significantly disrupted transcriptome from a minimally symptomatic stage of disease, with worsening seen with disease progression [190, 191]. Widespread transcriptional change occurs extensively and early, even in less affected Vonsattel grades 0–2 HD brains[192]. It has been shown that different functional regions of the brain have significantly different levels of dysregulated gene expression, with motor areas having greater effects than cognitive areas[192]. Motor

areas including the caudate nucleus of the post mortem HD brains, a third of all genes had altered expression[184]. Early studies revealed GABA levels in HD caudate was considerably reduced[193]. The emergence of choreiform movements and dystonia is thought to be correlated with degeneration of neurons containing enkephalin followed by P-containing neurons[66, 193].

Striatum, the most severely impaired brain area in HD[194],[195] showed significant gene expression variations in HD patients[192, 196-198]. One of the biggest questions in the field is how the ubiquitously expressed mHtt protein causes severe neuropathological changes, primarily in the striatum. Recent transcriptomic analysis of human prefrontal cortex and striatum of prodromal HD carriers further confirmed striatum showing expression representing extensive inflammatory processes before any appearance of symptoms and significant neurogenesis as compared to the cortex[199]. Although selective neuronal death is a prominent feature in HD, there is now a plethora of research suggesting that major neuronal dysfunction precedes the death of neurons[116]. Neuronal impairment and neuronal cell death seen in HD is linked to widespread neuronal transcriptional disruption induced by mHtt as indicated by aberrant gene expression changes seen in HD brain[200, 201].

Initial in situ hybridization studies showed decreased mRNA levels of signalling neuropeptides enkephalin, somatostatin, substance P, dopamine receptors (D1 and D2) and neuronal nitric oxide synthase (*nNOS*) in the HD brain[202-204]. This downregulation was further pronounced on disease progression with the findings supported by various studies of levels of neurotransmitters, the protein levels of their receptors and their specific metabolites[205]. Analysis of positron emission tomography (PET) scans of HD asymptomatic individuals with confirmed genetic diagnosis showed decreased potential for dopamine neurotransmitter receptor D2 binding supporting the idea that functional brain changes precede symptoms in HD[206]. However, changes in cell population, such as the loss of neurons and the gain of astrocytes and microglia, could confound some of these early differential expression changes[192, 196-198].

Single gene expression studies performed in transgenic mice overall exhibited similar pattern of gene alterations as in human HD. Reduced mRNA levels of dopamine D1 and D2 receptors, metabotropic glutamate receptors (*mGluR1*, *mGluR2* and *mGluR3*) have been confirmed in R6/1 and R6/2 mice and corresponding reduced protein levels were confirmed[14, 86, 207] implicating transcriptional dysregulation in HD. These findings suggest that as disease progress, HD brain undergoes multiple changes. The types of genes to

be defined as downregulated in single gene studies fall into many broad classes, including neurotransmitter receptors, intracellular signalling and cytoskeletal proteins. Additionally, second messengers and synaptic proteins in transgenic R6 mice were found to be changed[189].

Subsequently, when DNA microarrays were developed in the early 2000s, HD was one of the first diseases in which DNA microarrays[197] were used to produce a detailed examination of gene changes in R6/2 mice striatum[208]. As in single gene studies there has been remarkable concordance in extensive gene expression array studies across HD post-mortem brain tissues and animal models[209-211]. Cell culture and biochemical studies also indicate that mHtt can interfere with gene transcription[173, 212, 213]. The altered expression of genes in HD results from both disease-related cellular consequences and adaptive mechanisms to compensate for these changes. In general, there is downregulation in genes associated with neurotransmitter receptors (dopamine – D1 and D2, serotonin, cAMP-coupled receptors- adenosine A2a, cannabinoid CB1 receptors, GABA and glutamate receptors); components of cellular second messenger signalling pathways (regulators of calcium signalling, protein kinases-A, protein kinases-C related proteins and growth factor signaling); neuropeptides (somatostatin and proenkephalin); synaptic transmission (anchoring and vesicle components) as well as genes related to DNA metabolism and repair mechanisms[210, 214, 215]. Upregulation is generally seen in genes related to protein folding /clearance and cellular stress responses[189, 216].

The presence of a larger polyQ domain in the Htt protein could lead to transcriptional disruption through a variety of mechanisms. Gerber et al., first demonstrated in their study the activation of transcription using simple homopolymeric repeat stretches of glutamine *in vitro*. Numerous transcription factors including cAMP response element-binding protein (CREB) binding protein (CREBBP), TATA-box binding protein (TBP) and specificity protein 1 (Sp1) are known to contain glutamine-rich activation domains[200]. mHtt containing expanded polyQ can function directly as a transcriptional regulator by mimicking their behaviour - by binding to their recognition elements or can affect the ordinary function of these glutamine activation domain containing transcription factors. Additionally Htt exon 1 is also known to directly bind to DNA in the absence of other proteins, which could alter the DNA conformation and this aberrant DNA binding may also play a part in HD pathogenesis[101]. One of the theories behind transcriptional dysregulation in HD is sequestration of the components of transcriptional machinery by PolyQ aggregates[217]. Numerous transcriptional regulators like CREB, Sp1, TATA-box-binding protein-associated

factor II - (TAFII130), co-activators - CBP and TBP get sequestered along with mHtt aggregates[216, 218]. mHtt has higher affinity for these transcriptional regulators compared to the wild-type Htt suppressing their function. But interestingly, transcription factors lacking polyQ domains, like Sin3A[219], have been found in aggregates, suggesting that there are other protein – protein interactions besides polyQ interactions play a role. Moreover, numerous interactions between mHtt and core transcriptional machinery components exist[220]. Both *in vitro* and *in vivo* studies have also highlighted the direct interactions between mHtt, TFIID and TFIIF (RAP30 subunit) of the pre-initiation transcription complex[221]. Further the effect of mHtt on the basal transcriptional machinery may include loss of RNA polymerase activity. The large subunit RNA pol II can be sequestered in inclusions in cell lines expressing mutant Htt[191]. These observations suggest mHtt is capable of disrupting the core transcription machinery. mHtt can bind to the histone acetyltransferase domains interfering with histone acetylation and deacetylation further impacting gene transcription[222]. In addition, there is aggregation-induced inhibition of enzymes involved in chromatin remodeling[186, 218] and further alterations in non-coding RNAs culminating in alterations of transcription[186, 216]. These mechanisms are not mutually exclusive as mHtt can disrupt transcription via multiple mechanisms.

Finally, transcriptome profiling in HD shows an image of neurons with significant changes in gene expression. In the context of mHtt aggregation, it may be presumed that these shifts in gene expression are due to disease-related cellular effects or to compensated adaptive homeostasis of protein misfolding. mHtt aggregation based adaptive transcriptional modifications, although necessary under normal conditions is unable to respond to increased polyQ load resulting in neuronal dysfunction, neuronal toxicity, and symptoms of HD. In fact, recent studies indicate that the expected loss or gain of mHtt functions leads in systematic modification of both coding and non-coding genes in HD, thus undermining transcription pathways in cells[223, 224]. The association of transcriptomic changes to the different aggregation states of mHtt will give us a deeper understanding of the changed biological processes in response to protein aggregation.

1.11. Unfolded protein response in Huntington's disease

The capacity of a cell to recognise and respond to different stresses is crucial to cell viability. One of the vital organelles required for a broad range of proteins to be properly folded and processed is Endoplasmic reticulum (ER). In order to deal with misfolded protein

accumulation as seen in HD, the cells have evolved complex quality control (QC) mechanisms that help to correct protein folding in the ER. The unfolded protein response (UPR) protects against ER stress caused by a build-up of unfolded proteins in the cell[225]. Several measures including temporary translational attenuation, transcription induction via transcriptional regulators like activating transcriptional factor 6 (ATF6) and C/EBP homologous protein (CHOP)[226], activation of protein degradation via the proteasome and synthesis of ER chaperone proteins such as binding immunoglobulin protein (BiP)[227], are initiated to safeguard the integrity and fidelity of protein folding in the ER in response the build-up of unfolded or misfolded proteins. UPR pathway is shown to be constitutively activated in neurodegenerative diseases[228-230], with HD post-mortem brains expressing ER stress markers such as BiP, CHOP and RRS1 (regulator of ribosome synthesis). Interestingly, in HD pre-symptomatic mouse and cell models, ER stress markers were shown to be triggered indicating early ER dysfunction in HD[231, 232]. UPR signalling in the ER is mediated by three stress-sensing transmembrane proteins, the inositol-requiring enzyme 1 α (ISA1A), the protein kinase R (PKR)-like endoplasmic reticulum kinase (PERK), and ATF6, which together provide a coordinated response that helps to resolve disrupted proteostasis[233]. Although UPR helps alleviate ER stress, cell death can be caused by persistent ER stress resulting in an activated UPR. A significant defect in endoplasmic reticulum-associated degradation (ERAD)[234] of misfolded proteins has also been documented to cause neuronal toxicity due to the sequestration of critical ERAD proteins UFD1, P97 and NPI4 by polyQ mHtt fragments resulting in ERAD failure. Recent analysis have indicated extensive induction of UPR in various models of HD[235]. Further a strong association between polyQ length and UPR pathway gene expression has been established[235].

The protein folding in the cells is guided by a substantial network of chaperones, quality control factors and cellular trafficking pathway jointly called the proteostasis network. Stress responses, like UPR, mediate the complex control of the proteostasis network. The specialised ER stress sensing UPR pathway and others in the cell detects and integrates developing protein-folding issues and resolves them by directing translational attenuation to diminish protein load and reshapes the proteostatic network composition[236]. Modulation of the UPR in HD and other neurodegeneration is therefore a promising target for future therapeutic treatments.

Huntington's disorder is a severe neurodegenerative illness with signs of motor failure, neurological disability, and psychological problems gradually escalating over a span of 10-20 years, eventually leading to death. HD associated morbidity and mortality put immense social and financial pressures on impacted people, their caregivers and communities. Medical options for the treatment of HD are limited and provide symptomatic relief, but do not affect the pathology of the disease itself. After the discovery of the Htt gene, HD research has progressed in leaps and bounds. While more than two decades of work in molecular genetics and molecular pathology of HD have taken place, there has been no effective therapy to reduce the pathological effects of mHtt. This thesis examines how aggregation state of PolyQ impacts transcription and other cellular functions. A wide range of PolyQ expansion effects on transcriptome could be drawn from these experiments. This work contributes knowledge to the field with the potential to unmask new pathways to dysfunctional mechanisms. Changes in gene expression (which occur prior to onset and during disease progression) have become a key focus of neurodegenerative disease research. Identification and validation of patterns of gene expression that occur in response to specific aggregation states could provide guidance on mechanisms involved in neurodegenerative molecular pathogenesis and may facilitate the discovery of new clinically relevant therapeutic targets. Since HD results from conformational changes in Htt protein, the identification of conformation-related toxic and adaptive changes will provide therapeutically amenable targets.

1.12. Thesis outline and aims

The broad aim of the thesis is to understand the HD biology better through the angle of aggregation states. The thesis tries to underpin the toxic mechanisms as well as the adaptive changes undertaken by the cell to better cope with mHtt aggregation states as compared to the wildtype Htt.

Chapter 2 of the thesis aims to obtain a better picture of aggregation states related gene expression changes. The main aims of the study were to

1. How do different Huntingtin (Htt) protein aggregation states alter gene expression?
2. Identify specific gene signatures altered across Htt protein aggregation states
3. Functional validation of key targets altered by Htt aggregation states

These have been addressed and published as in Molecular and Cellular Neurosciences.

Chapter 3 further builds on chapter 2 and elaborates on splicing related alterations seen in HD with respect to aggregation states. Chapter 3 is a manuscript in preparation for submission and is written as such. The manuscript describes

1. The changes in small RNA alterations and spliceosomal defects seen in HD
2. Defines aggregation related splicing changes
3. Proteomics of aggregation states and proteogenomic confirmation of spliced isoforms.

The chapter 4 showcases the development of a novel thiol probe to quantify unfolded protein response in the cell. I highlight the use of novel proteomic approach to identify and quantify unfolded peptides in the cell to stressors.

Chapter 2:

Htt exon 1 aggregation state induced transcriptomic changes in Huntington's disease

2.1. Abstract

Transcriptional profiles for distinct aggregation states of mutant Huntingtin exon 1 protein unmask new Huntington's disease pathways

Huntington's disease is caused by polyglutamine (polyQ)-expansion mutations in the CAG tandem repeat of the *Huntingtin* gene. The central feature of Huntington's disease pathology is the aggregation of mutant Huntingtin (Htt) protein into micrometer-sized inclusion bodies. Soluble mutant Htt states are most proteotoxic and trigger an enhanced risk of death whereas inclusions confer different changes to cellular health and may even provide adaptive responses to stress. Yet the molecular mechanisms underpinning these changes remain unclear. Using the flow cytometry method of pulse-shape analysis (PulSA) to sort neuroblastoma (Neuro2a) cells enriched with mutant or wild-type Htt into different aggregation states, we clarified which transcriptional signatures were specifically attributable to cells before versus after inclusion assembly. Dampened CREB signalling was the most striking change overall and invoked specifically by soluble mutant Httex1 states. Toxicity could be rescued by stimulation of CREB signalling. Other biological processes mapped to different changes before and after aggregation included NF-kB signalling, autophagy, SUMOylation, transcription regulation by histone deacetylases and BRD4, NAD⁺ biosynthesis, ribosome biogenesis and altered HIF-1 signalling. These findings open the path for therapeutic strategies targeting key molecular changes invoked prior to, and subsequently to, Httex1 aggregation.

2.2. Transcriptional profiles for distinct aggregation states of mutant Huntingtin exon 1 protein unmask new Huntington's disease pathways

Molecular and Cellular Neuroscience 83 (2017) 103–112



Contents lists available at ScienceDirect

Molecular and Cellular Neuroscience

journal homepage: www.elsevier.com/locate/ymcne



Transcriptional profiles for distinct aggregation states of mutant Huntingtin exon 1 protein unmask new Huntington's disease pathways



Nagaraj S. Moily^a, Angelique R. Ormsby^a, Aleksandar Stojilovic^a, Yasmin M. Ramdzan^a, Jeannine Diesch^{b,c}, Ross D. Hannan^{b,d}, Michelle S. Zajac^{e,f}, Anthony J. Hannan^{e,f}, Alicia Oshlack^{g,*}, Danny M. Hatters^{a,*}

^a Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, VIC 3010, Australia

^b Research Division, Peter MacCallum Cancer Centre, 305 Grattan Street, Melbourne, VIC 3000, Australia

^c Josep Carreras Leukaemia Research Institute, ICO Hospital Germans Trias i Pujol, Badalona, Spain

^d The John Curtin School of Medical Research, Australian National University, Acton, ACT, Australia

^e Florey Institute of Neuroscience and Mental Health, Melbourne Brain Centre, University of Melbourne, Parkville, VIC, Australia

^f Department of Anatomy and Neuroscience, University of Melbourne, Parkville, VIC, Australia

^g Murdoch Childrens Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia

ARTICLE INFO

Keywords:

Protein misfolding
Amyloid
Neurodegenerative disease
Huntington's disease

ABSTRACT

Huntington's disease is caused by polyglutamine (polyQ)-expansion mutations in the CAG tandem repeat of the *Huntingtin* gene. The central feature of Huntington's disease pathology is the aggregation of mutant Huntingtin (Htt) protein into micrometer-sized inclusion bodies. Soluble mutant Htt states are most proteotoxic and trigger an enhanced risk of death whereas inclusions confer different changes to cellular health, and may even provide adaptive responses to stress. Yet the molecular mechanisms underpinning these changes remain unclear. Using the flow cytometry method of pulse-shape analysis (PulSA) to sort neuroblastoma (Neuro2a) cells enriched with mutant or wild-type Htt into different aggregation states, we clarified which transcriptional signatures were specifically attributable to cells before versus after inclusion assembly. Dampened CREB signalling was the most striking change overall and invoked specifically by soluble mutant Httex1 states. Toxicity could be rescued by stimulation of CREB signalling. Other biological processes mapped to different changes before and after aggregation included NF- κ B signalling, autophagy, SUMOylation, transcription regulation by histone deacetylases and BRD4, NAD⁺ biosynthesis, ribosome biogenesis and altered HIF-1 signalling. These findings open the path for therapeutic strategies targeting key molecular changes invoked prior to, and subsequently to, Httex1 aggregation.

1. Introduction

Protein misfolding and aggregation into amyloids underlies many of the major human neurodegenerative diseases, including Alzheimer's, Parkinson's, Huntington's and motor neuron disease (Sipe et al., 2016). In these diseases, the appearance of macroscopic intracellular protein aggregates known as inclusions form as a hallmark of disease development. Huntington's disease (HD) is one of the classic examples of the amyloid diseases involving intracellular protein aggregating into inclusions. HD involves an expansion of a CAG trinucleotide repeat sequence, encoding polyglutamine (polyQ), within exon 1 of huntingtin (Htt) beyond the disease threshold of 35 repeats (MacDonald et al.,

1993). The polyQ expansion leads to a greatly enhanced capacity of the Htt exon 1 (Httex1) domain to aggregate into amyloid fibrils in vitro (Scherzinger et al., 1999) and for N-terminal fragments similar in length to Httex1 to form microscopic inclusion bodies in pathology (DiFiglia et al., 1997; Kazantsev et al., 1999). The link to disease from polyQ-expanded Httex1 has been well documented. Namely, the transgenic expression of polyQ-expanded Httex1 is sufficient to produce a HD-like pathology in rodent and primate models (Davies et al., 1997; von Horsten et al., 2003; Yang et al., 2008) and confers toxicity in cell culture models as well (Arrasate et al., 2004). As such Httex1 has been used extensively to model HD biology.

An outstanding question remains as to the role of aggregation in

Abbreviations: PolyQ, Polyglutamine; Htt, Huntingtin; Httex1, Htt Exon 1; HD, Huntington's disease; PulSA, Pulse Shape Analysis; i, Cells With Inclusions; ni, Cells With Diffuse Httex1; RFU, Relative Fluorescence Units; IPA, Ingenuity Pathway Analysis; FDR, False Discovery Rate; WT, Wild-Type; GFP, Green Fluorescent Protein; ROI, Region Of Interest

* Corresponding authors.

E-mail addresses: alicia.oshlack@mcri.edu.au (A. Oshlack), dhatters@unimelb.edu (D.M. Hatters).

<http://dx.doi.org/10.1016/j.mcn.2017.07.004>

Received 12 April 2017; Received in revised form 23 June 2017; Accepted 21 July 2017

Available online 23 July 2017

1044-7431/ © 2017 Elsevier Inc. All rights reserved.

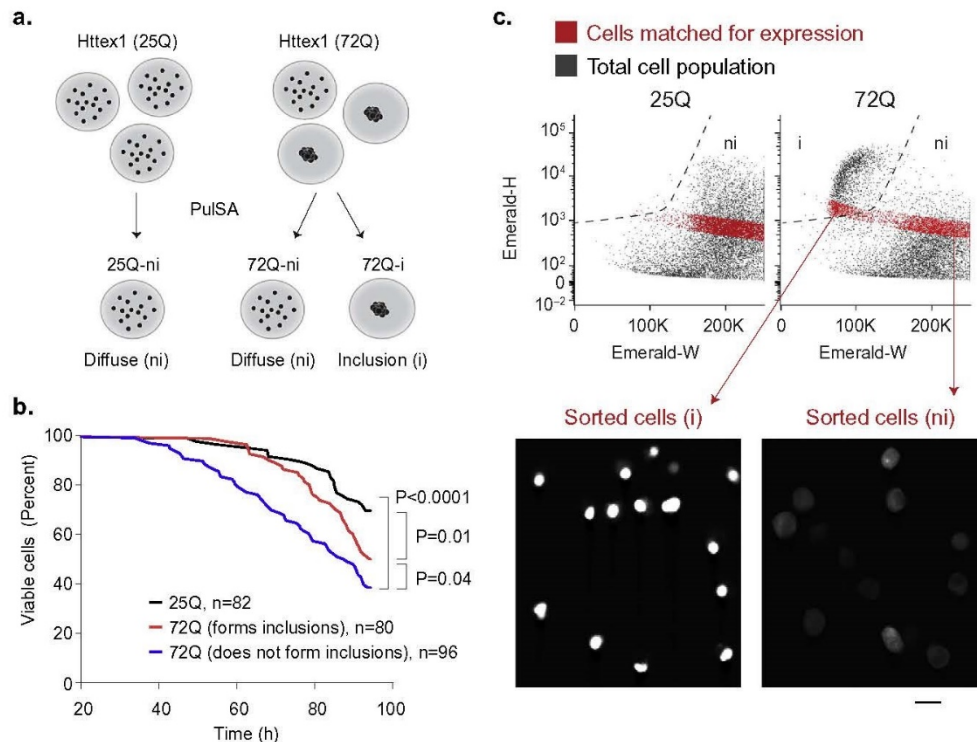


Fig. 1. Strategy to separate cells into populations enriched with polyQ-expanded Httex1 in distinct aggregation states. **a)** Pulse Shape Analysis (PulSA) by flow cytometry can separate cell populations heterogeneous in aggregation state. **b)** Neuro2a cells that form inclusions live longer than those that do not. Shown are Kaplan-Meier survival curves tracked by longitudinal imaging of Httex1-Emerald fusions from 20 h after transfection. **c)** Flow cytograms showing the PulSA gating strategy to sort the mixed cell populations for analysis by whole transcriptome analysis. Shown are gates for i and ni and for matched expression (red; Further details in Fig. S1). The images show representative cells collected by sorting flow cytometry and imaged for GFP fluorescence by confocal microscopy. Scale bar, 20 μ m. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

pathogenesis. Evidence increasingly suggests that cellular malfunctioning originates prior to aggregation of Htt into the inclusion body – indeed formation of inclusions seems to aid survival (reviewed in Arrasate and Finkbeiner, 2012; Hatters, 2012). Therefore, soluble polyQ-expanded Httex1 states (including monomers and oligomers) may directly confer cellular toxicity and that toxicity could become nullified as the soluble states are sequestered into the visible aggregates. The mechanisms mediating the origins of cellular toxicity and inclusion assembly remain to be properly elucidated.

Until recently, it has not been straightforward to decipher the cellular changes arising from polyQ-expanded Httex1 as it shifts from soluble states into the inclusion body because cell populations contain mixtures of cells enriched with Httex1 in different aggregation states. This heterogeneity prevents a clean delineation of effects attributable to mutation from aggregation state.

We previously developed a flow cytometry based method (Pulse Shape Analysis: PulSA) to separate cells expressing polyQ-expanded Httex1 into two major groups: those enriched with Httex1 in inclusions (i) and those enriched with soluble Httex1 states (predominantly monomer) (ni) (Ramdhan et al., 2012) (Fig. 1a). This methodology provides a unique opportunity to explicitly understand how aggregation state of Httex1 (or indeed any mutant protein that forms an inclusion) relates to molecular dysfunction. Therefore, in this study, we sought to comprehensively determine how the transcriptome is altered

by Httex1 aggregation in a cell culture model that is both sensitive to mutant Httex1 toxicity and which can be robustly controlled for covariates such as time and expression level. We show that Httex1 provides the most profound changes to the transcriptome prior to aggregation and that the toxicity can be accounted for by a sustained shutdown of baseline CREB signalling from soluble Httex1.

2. Materials and methods

2.1. DNA vectors and constructs

pT-REX vectors (Life Technologies) encoding Httex1 with 25Q or 72Q polyQ sequence lengths were prepared as described previously (Ramdhan et al., 2012; Ramdhan et al., 2010). pGW1-based Httex1-mCherry constructs were prepared as described previously (Tsvetkov et al., 2010).

2.2. Cell culture

Neuro2a cells, obtained originally from the American Type Culture Collection, were maintained in Optimem medium (Life Technologies) supplemented with 10% v/v FCS, 1 mM glutamine, 200 U/ml penicillin and 200 μ g/ml streptomycin in a humidified incubator with 5% v/v atmospheric CO₂. Neuro2a cells were transiently transfected with the

vectors using Lipofectamine 2000 reagent (Life Technologies) using the manufacturer's recommendations.

2.3. Collection of cells by Pulse Shape Analysis, RNA extraction and sequencing

48 h after transfection, cells were fluorescence-activated cell sorted using a BD FACS ARIA III cell sorter. In essence, the cells were gated for live cell population by SSC and FSC gates and then the 'ni' and 'i' gates by PuLSA as described (Ramdhan et al., 2012). To match for expression, cells were further gated to the same median GFP intensity of 2200 RFU by varying the window of expression (Fig. S1). The 25Q and 72Q Httex1 constructs were transfected and sorted in parallel across four days and performed as four matched replicates. 50,000 cells were collected for each sample (25Q-ni, 72Q-ni and 72Q-i) and sorted directly into TRI-reagent (Zymo Research, Irvine, CA). For imaging, a parallel sample set was collected and imaged as described previously by confocal microscopy (Ramdhan et al., 2012). Total RNA was purified with the Direct Zol RNA extraction kit (Zymo Research, Irvine, CA) and verified for quality and yield ($> 1 \mu\text{g}$ and RNA Integrity number of > 9) using the Bioanalyzer 2100 and RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA). rRNA was depleted and the library prepared per the manufacturer's protocols (TruSeq, Illumina) for whole transcriptome sequencing using Illumina Hi-Seq 2000 (Illumina, San Diego, CA, USA). FastQC report generated from the RNA Sequencing analysis showed good per base quality scores.

2.4. RNA sequence processing

50 base pair paired-end reads with an average library size of 15 million tags with at least one reported alignment ($\sim 90\%$ mapping rate) was obtained. The reads were mapped to the reference mouse genome (GRCm38/mm10) using Tophat2 (2.0.13) (Kim et al., 2013) aligner. The gene level read counts were obtained (Ensembl transcripts release 78) using Rsubread - featureCounts (v1.4.6-p1) (Liao et al., 2013) using RStudio (Version 0.99.1246)/R (Version 3.3.1).

2.5. Differential expression analysis of RNA sequencing data

Differential expression analysis of RNA sequencing data was conducted using voom transformation and limma differential expression analysis pipeline in R-bioconductor (Version 3.3). Genes were first filtered if they did not have at least 1 counts per million mapped fragments, in more than three libraries. Counts were converted to $\log_2$ counts per million, normalized using the trimmed mean of M-values normalization method (Robinson and Oshlack, 2010) and precision weighted with the 'voom' function of the limma package (v3.24.15) (Law et al., 2014). A linear model was fitted to each gene with effects for the sample types (25Q-ni, 72Q-ni and 72Q-i) as well as a further fixed effect for the four replicate batches that the samples were PuLSA sorted across four days. Empirical Bayes moderated t-statistics were used to assess differences in gene expression. Statistical tests for differences in expression levels were evaluated by comparing between the sample groups: i.e. the 72Q-ni vs. 25Q-ni, 72Q-ni vs 72Q-i, 72Q-i vs 25Q-ni and total 72Q (72Q-i + 72Q-ni) vs.25Q. Genes were adjusted for multiple testing and significant genes were designated as having a false discovery rate (FDR) < 0.05 using decideTest function in limma. The upregulated genes were labelled 1; downregulated genes -1 ; and not differentially expressed genes 0 (Table S1c). KEGG pathway analysis was performed with DAVID v6.8 using a minimum gene count of 10 and a modified Fisher Exact P-Value of < 0.01 with FDR correction (Huang et al., 2009a, 2009b).

2.6. STRING analysis

The differentially expressed genes between the 72Q-ni and 25Q-ni

with FDR corrected P-values < 0.05 and fold change of > 2 were used to create a protein: protein interaction network with the STRING v.10 algorithm (Szklarczyk et al., 2015). Data were analyzed with parameters including only experimentally obtained data and a medium confidence interaction score of 0.40. 1st shell interactors of > 5 proteins was used.

2.7. Upstream regulator analysis

Upstream regulator analysis was performed on the genes differentially expressed between the 72Q-ni and 25Q-ni with Ingenuity Pathway Analysis (IPA) (Qiagen). IPA compares gene expression in a dataset with lists of genes that are known to be regulated by specific upstream transcriptional regulators, to infer whether a regulator is likely to be activated or inhibited. Since there is a considerable overlap in the lists of gene regulated by different transcriptional regulators, multiple points along a pathway can be suggested as key regulators, which can lead to a significant redundancy in the analysis. In IPA, this effect is quantified by the activation z-score statistics, which increases or decreases depending on the inferred activation/inhibition state of an upstream regulator. Significance in IPA was attributed to upstream regulators that have an overlap P-value of < 0.01 and an activation z-score of > 2 or less than -2 .

2.8. Heuristic post hoc analysis of gene expression data

Gene expression data were classified into different transcriptional signatures using significant results (FDR < 0.05) obtained by decideTest function in limma on different comparisons. The transcription signature groups included genes with the criteria as described below (in both positive and negative directions (i.e. -1 or 1) for each signature. Progressive fold change to determine the direction of expression change across aggregation states was calculated using the formula (fold change 72Q-ni vs 25Q-ni) * [Absolute(fold change of 72Q-i vs 25Q-ni - fold change of 72Q-ni vs 25Q-ni)].

Gene expression signature	Decide test 72Q-ni vs. 25Q-ni	Decide test 72Q-i vs. 25Q-ni	Decide test 72Q-ni vs. 72Q-i	Progressive fold change
Transcription signature group 1	1	1	0	–
Transcription signature group 2	0	1	1	–
Transcription signature group 3	1	1	1	> 0
Transcription signature group 4	1	0	–	< 0

For each transcription signature group 1 the genes were subdivided into upregulated and downregulated gene lists with regard to direction of expression in 72Q-ni vs 25Q-ni comparison. The genes which did not fall into the categories mentioned above were excluded from the analysis.

2.9. Network Analyst – minimal interaction networks

The gene expression signatures were mapped to known protein: protein interaction networks using network analyst toolkit (Xia et al., 2014; Xia et al., 2015). In essence, genes in a transcriptome signature group were mapped to a validated network of all known binary protein:

protein interactions. The network was trimmed to contain only proteins that connect the input genes. The most connected genes remaining in the network that had a degree centrality of > 30 formed the hubs. The most locally well-connected subnetworks with a statistical threshold of Wilcoxon rank-sum P -values < 0.05 were extracted as modules.

2.10. R6/1 mouse tissue processing, RNA extraction and microarray analysis

R6/1 hemizygote males (Mangiarini et al., 1996) were obtained from the Jackson Laboratory (Bar Harbor, ME, USA) and bred with CBB6 (CBAXC57/B6) F1 females to establish an R6/1 colony. Genotypes were determined by PCR (Mangiarini et al., 1996) with genomic DNA obtained from toe clips. Wild-type (WT) and R6/1 HD male mice were collected after weaning at 3.5 weeks of age and transferred into standard mouse cage boxes, 10 cm $\times$ 16 cm $\times$ 31 cm in size, in groups of 3–6.

WT and HD mice were killed at 12 weeks of age by cervical dislocation between 9 and 11:30 am. Different experimental groups were randomly allocated into balanced batches to eliminate batch effects. Brains were removed, rinsed in PBS and placed on ice for ≤ 5 –10 min prior to dissection. Brains were bisected and striatum was dissected out. Samples were immediately frozen on dry ice and transferred to a -80°C freezer for storage until required.

Total RNA from the striatum was extracted using RNeasy Mini kits (Qiagen). On-column DNase clean-up was performed on all samples using RNase-Free DNase set (Qiagen). The concentration and integrity of the extracted total RNA present in the final eluates were determined using an Agilent Bioanalyser 2100. All the samples had a RIN of > 9.3 . GeneChip Mouse Exon 1.0ST arrays (Affymetrix, Santa Clara, CA, USA) were used, with hybridization and scanning conducted by the Australian Genomics Research Facility (Parkville, Australia) per standard Affymetrix protocols, with appropriate quality control (QC).

Microarray data was analyzed using the limma package (v3.24.15) in R-bioconductor (Version 3.3). The array data was preprocessed and quality control was done using the oligo package. The normalization was done at the probe level using Robust Multi-Array Average method in oligo (Carvalho and Irizarry, 2010). The probe set annotation was done using the Affymetrix moGene10 annotation data package (version 8.4.0) (MacDonald, 2016). The statistical testing of changes in gene expression of transgenic R6/1 HD mice was compared to wild-type littermates using a linear model in limma (Supplementary Table S2a). The pathways were obtained using IPA core analysis with gene changes attributed specifically to neuronal tissue in IPA knowledge base.

2.11. Human caudate gene analysis

Previously published raw microarray data (Hodges et al., 2006) was retrieved from the EBI Array Express database (Kolesnikov et al., 2015) (<http://www.ebi.ac.uk/arrayexpress>) under accession number E-AFMX-6. The raw data was processed and normalized using the bioconductor affy package (v1.52.0) (Gautier et al., 2004) with the Robust Multi-Array Average method (Bolstad et al., 2003) and differential expression was tested in limma (v3.24.15) in R-bioconductor (Version 3.3). Gene-set enrichment analyses were conducted using the caudate genes as the reference set and our cell-based RNA-seq data as the gene set with the Roast function in limma (Wu et al., 2010).

2.12. Longitudinal live cell imaging

The cells were plated at a density of 1×10^5 cells/well in a 24 well plate. After 24 h, cells were co-transfected with 72Q-Httex1-mCherry, 25Q-Httex1-mCherry or mCherry alone and GFP using Lipofectamine 3000 reagent (Life Technologies). Media was refreshed 6 h after transfection. The GFP was used as a cellular tracer of death and proxy for Httex1 expression level. After 24 h cells were imaged using a JuLI

stage live cell imaging system with fluorescent images acquired at 15 min intervals for 96 h (Nanoentek, Seoul, South Korea). Cells were tracked for visible aggregate formation and the time point of death (by loss of GFP fluorescence) with image processing tool Fiji (Version 2.0.0-rc-44) (Schindelin et al., 2012). Survival curves were analyzed using Prism software. Cells that drifted from focus were censored in the analysis. Expression level was calculated by measuring mean GFP levels in an ROI within the cytosol. The cells included for analysis had a mean fluorescence ± 2 standard deviation of the sample with the lowest variance. P -values were determined using log-rank (Mantel Cox) test. For drug treatments, forskolin (Cayman Chemical) was added after transfection to a concentration of 0.5 μM , which corresponds to the EC₅₀ dosage. CREB inhibitors SGC-CBP30 (20 nM) and I-CBP112 (150 nM) were added after transfection (Sigma-Aldrich). Both drugs were added at 6 h post transfection during media refreshment.

3. Results & discussion

3.1. Cell populations expressing matched Httex1 levels divided into those with inclusions versus those without

Our model system was a mouse neuroblastoma cell line (Neuro2a), which is a widely used model for studying Htt-induced pathogenic mechanisms. These cells were selectively vulnerable to toxicity of polyQ-expanded Httex1 (72Q) over the wild-type polyQ length (25Q) (both as fusions to mEmerald fluorescent protein (Olshina et al., 2010) (Fig. 1b). In accord with prior observations, the formation of inclusions repressed the rate of death relative to cells retaining exclusively diffuse mutant Httex1 (Arrasate et al., 2004).

In transfected cell populations the levels of Httex1 and aggregation rates are heterogeneous (Ramdhan et al., 2012). To remove this heterogeneity and enrich for cells with aggregates versus those without we used PulSA (Ramdhan et al., 2012). PulSA enabled the separation of cells with inclusions (i) from those without (ni) in non-expanded, wild-type (25Q) and expanded, mutant (72Q) forms of Httex1, using a gating strategy that matched expression levels across our three groups (25Q-ni, 72Q-ni and 72Q-i) (Fig. 1c; Fig. S1). This strategy captured a snapshot of cells representing the extreme ends of the spectrum of aggregation states, and which enriches cell groups with > 90 –95% fidelity (Polling et al., 2014; Polling et al., 2015).

3.2. The largest changes in the transcriptome are stimulated by soluble mutant Httex1

Next-generation RNA sequencing (RNAseq) yielded 12,876 high abundance genes on four biological matched replicates. To first test whether our dataset yielded results in accord with previous HD datasets, we tested for differential expression between the disease like 72Q samples (72Q-ni + 72Q-i) and 25Q-ni as control. Comparing these two groups with the previously described human caudate microarray data from HD brain (versus non-HD controls) (Hodges et al., 2006) we observed 1109 differentially expressed genes in common to both. The upregulated genes detected in mouse were found to be significantly enriched for upregulated expression in the human samples. That is, we took the significantly upregulated genes detected in our mouse study ($n = 428$ genes) and used this as a gene set for testing with Roast (Wu et al., 2010) in the human study (FDR = 0.01). These results are indicative of a general consistency in upregulated genes between our cell model and human caudate samples.

As an additional test, we found that our 72Q samples treated as one group (72Q-ni + 72Q-i) versus the 25Q-ni group yielded several HD-relevant KEGG pathways (Supplementary Table S1a). This included DNA repair (Genetic Modifiers of Huntington's Disease, Consortium, 2015), p53 signalling (Bae et al., 2005) and RNA transport (Gasset-Rosa et al., 2017).

To understand how aggregation state affected transcription, we

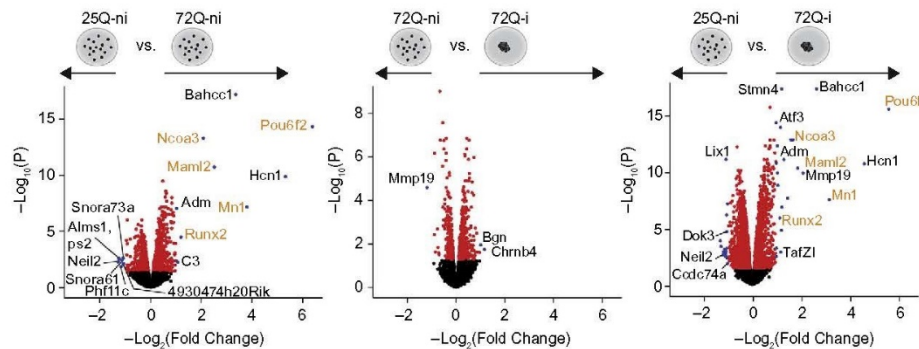


Fig. 2. Volcano plots of gene expression changes attributable to aggregation states of Httex1. Shown are comparisons between the three treatment groups: cells lacking visible aggregates but differing in polyQ length (25Q-ni and 72Q-ni cells) and cells with expanded polyQ length that differ in aggregation state (72Q-i). Genes with a FDR less than $P = 0.05$ are shown in red. Genes with > 2 -fold change are shown in blue. Genes encoding proteins with an intrinsic native polyQ sequence have orange labels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

measured the differential expression patterns for all pair-wise comparisons between the three sample Httex1 groups: 25Q-ni, 72Q-ni and 72Q-i. The Httex1 72Q-ni versus the 25Q-ni comparison yielded changes to how soluble mutant Httex1 affected expression while the 72Q-i versus the 72Q-ni comparison assessed more specifically how aggregation of mutant Httex1 into inclusions affected expression (Fig. 2). These analyses indicated that the largest and most significant changes occurred from mutant Httex1 when it was soluble prior to inclusion assembly. Furthermore, most of the changes were sustained in cells with inclusions (Fig. 2 and Table S1c).

3.3. Inactivated CREB-CREM signalling is the most profound impact arising from soluble Httex1

Next, we probed the dataset for clues as to how the transcriptional changes triggered by the soluble mutant Httex1 relate to mechanism of toxicity. First, we performed a STRING network analysis (Szklarczyk et al., 2015) (on human protein counterparts) of the genes most altered in expression to see if they correlated with validated protein-protein interactions. Of the 9 genes upregulated > 2 -fold in the 72Q-ni condition (versus 25Q-ni), 3 formed part of the CREB signalling network (NCOA3, MDM4 and RUNX2) (Fig. 3a). CREB is a key transcription factor that has been previously implicated as defective in HD (Luthi-Carter et al., 2000; Shimohata et al., 2000; Wyttenbach et al., 2001; Mantamadiotis et al., 2002; Steffan et al., 2000, 2001; Dunah et al., 2002). Furthermore, CREB signalling has been suggested to be mediated by deactivation of CREB binding protein (CREBBP) activity through abnormal interactions with soluble Httex1 (Steffan et al., 2001; Jiang et al., 2006; Schaffar et al., 2004; Cong et al., 2005). There was no significant change in CREBBP or CREB transcript levels.

We next sought to measure the extent to which the changes in transcriptome can be explained by alteration of CREB activity. To do this we performed an unbiased Upstream Regulator Analysis using Ingenuity Pathway Analysis (IPA). The principle behind this analysis is to search for changes in transcription factor activity that accounts for the changes in downstream gene expression as represented by our dataset. 16 transcription factors were identified that are known to regulate the genes in our dataset. Critically, only 2 of these transcription factors, CREB and the closely functionally related CREM, could explain the altered expression patterns of our data with high confidence by their inactivation (Fig. 3b, Table S1b). These data collectively suggest that CREB signalling defects are among the most pervasive of the changes in the transcriptome (including before and after aggregation) and that these effects are attributable to mutant Httex1 before

aggregation.

To test for the in vivo relevance of these findings, we also examined the mRNA of R6/1 Httex1 transgenic mouse (Mangiarini et al., 1996) striatum tissue at a time point prior to motor onset (12 weeks of age) by microarray analysis. Analysis of the gene expression differences in the R6/1 genotype ($n = 3$) versus non-transgenic littermates ($n = 4$) indicated CREB signalling (and related cAMP signalling) were in the top 5 pathways negatively impacted in an IPA canonical pathway analysis (Table S2a,b).

3.4. Dampened CREB signalling explains the molecular basis of toxicity

To further examine the importance of CREB inactivation in toxicity we performed survival curve analyses focusing on manipulating CREB signalling (Fig. 3c). Prior studies have found that deletion of CREB-CREM expression in postnatal forebrain results in extensive neurodegeneration reminiscent of HD pathology (Mantamadiotis et al., 2002). Inhibition of CREB signalling with inhibitors SGC-CBP30 and I-CBP112, which target CREBBP and EP300 bromodomains with high specificity (Hammitzsch et al., 2015; Picard et al., 2015), in our Neuro2a culture system led to a high death rate indicating that baseline CREB signalling is critical for cell survival (Fig. 3d). Cells expressing mutant Httex1 were negatively impacted by inhibition of CREB signalling, as anticipated (Fig. 3c). However, activation of CREB signalling with forskolin, which triggers downstream phosphorylation and activation of CREB (Insel and Ostrom, 2003), significantly rescued the toxicity caused by mutant Httex1 and had otherwise no effect on baseline survival rates in cells with the wild-type Httex1. Collectively, these results suggested that toxicity of soluble mutant Httex1 manifests predominately through impairment of baseline CREB signalling and that this be remedied by triggering the CREB signalling.

3.5. Additional novel pathways unearthed as transcriptional signatures for Httex1 aggregation

We next focused on extracting other patterns in the dataset using a transcriptional signature approach as shown in Fig. 4a using a post hoc heuristic analysis. These signatures can be used to make predictions about how cells respond to Httex1 aggregation state in both pathogenic and adaptive manners. This analysis yielded hundreds of genes in each transcriptional signature group. Features were extracted from this data using the Network analyst software (Xia et al., 2014), which links together genes based on their known protein-protein interactions. The most connected genes in the networks (set with a minimum threshold of

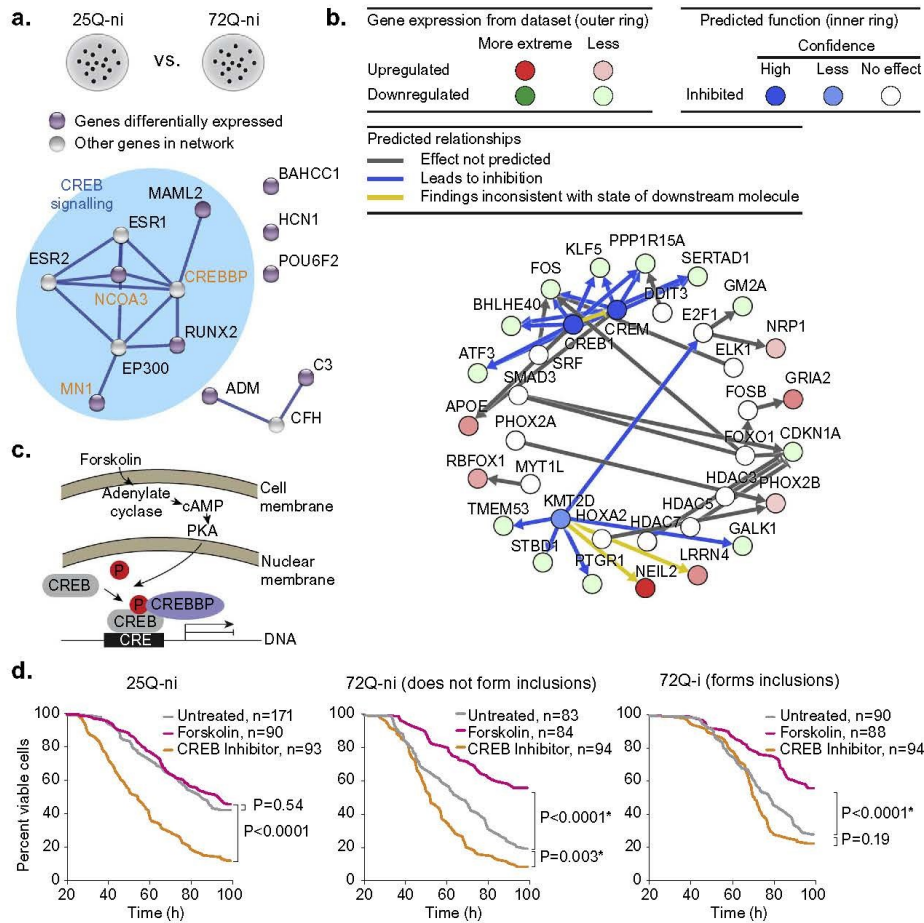


Fig. 3. CREB signalling is dampened by soluble mutant Httex1 and explains the major source of toxicity to Neuro2a cells. a) A protein-protein interacting network of the genes enriched > 2-fold in cells enriched with soluble mutant Httex1. Shown are STRING (v10) validated protein-protein interactions of genes corresponding to those enriched in the dataset. The network was built to include first shell interactors (i.e. interacting proteins corresponding to genes not enriched in the dataset). b) Upregulator analysis (Ingenuity) to explain how changes in differential expression patterns of our dataset relate to changed protein activity of transcription factors. The inner circle shows transcription factors that control genes seen differentially expressed in the dataset. The outer ring shows the key genes in our dataset that are targets of the transcription factors and their patterns of expression. c) Schematic of CREB signalling activity and role of activator forskolin. d) Effect of upregulating (with forskolin) or suppressing (with inhibitors CREB signalling SGC-CBP30 and I-CBP112) on survival curves. P-values corresponds to log-rank (Mantel Cox) tests.

30 interactions) were identified as hub genes of interest (Table S3). Functional subnetworks (modules) that relate more discretely to biological functions were also identified using a Walktrap algorithm in Network analyst (Fig. 4b). The final list (as discussed below) of modules and hub genes has an extensive overlap with established HD mechanisms. There was also a strong overlap with CREB signalling, which suggests these modules are affected directly by the dampened CREB signalling response or are adaptive responses to counteract CREB signalling defects.

We explain the key details of these modules below and provide a discussion of how they may connect to CREB signalling and HD biology (and expanded on in Supplementary Note 1). This discussion outlines what we see are the key features of the data set and point to connections with established links to Huntington's disease biology and other

features of potential interest.

3.5.1. Transcriptional signature 1

This signature describes the changes invoked by soluble mutant Httex1 and which have no further change when inclusions form. One explanation for this signature is to describe changes occurring as a result of toxic responses. This may apply for the HDAC5 and BRD4 modules, for which HDAC5 and BRD4 genes are chromatin modifiers that have been shown to influence CREB signalling, which are consistent with the broader changes seen in CREB signalling (Ha et al., 2010; Tropberger et al., 2013; Devaiah et al., 2016). HDAC proteins have also been previously linked to underlying HD pathogenesis (see Supplementary Note 1 for more discussion).

Other modules are more likely to reflect responses to counter the

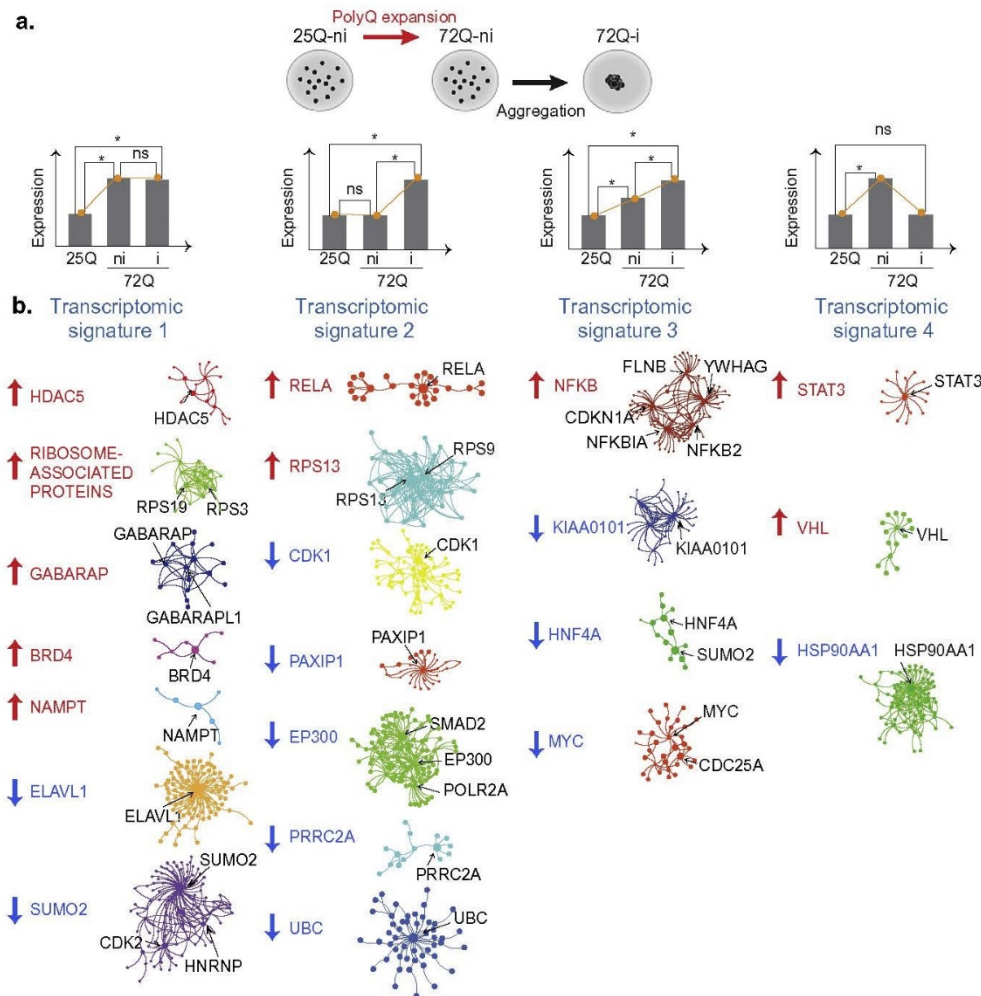


Fig. 4. Transcriptional signatures for Httex1 aggregation state. a) Shown is our model for how transcriptome data can be mapped to the changes attributable to mutation, and then aggregation into inclusions. Data are filtered into transcriptional signatures as represented by the bar graphs, based on changes with FDR < 0.05 ($n = 4$). b) Shown are functional modules (with Wilcoxon rank-sum P -values < 0.05), and their designated names, extracted from protein-protein interaction networks corresponding to all the genes in each transcriptional signature using Network Analyst Walkmap software. These modules can be upregulated or downregulated in the transcriptional signature (and the direction is shown). Each module shows the genes (points) and their connections. The most connected genes in the modules are labelled.

toxic effects of Httex1. Such upregulated modules (named after the key genes in them) included HDAC5, BRD4, NAMPT, ribosome-associated proteins and GABARAP. GABARAP, ELAVL1 and SQSTM1 genes have roles in regulating macroautophagy and proteasomal-mediated degradation and have been shown to be involved in HD pathology (see Supplementary Note 1 for more discussion). Their upregulation thus may indicate attempted increased clearance mechanisms to cope with the toxic Httex1 response. However, for GABARAP the changes may also reflect a block in autophagy (discussed in more detail in Supplementary Note 1). NAMPT gene encodes the rate-limiting enzyme in NAD production, which is an essential cellular metabolite involved in energy production, redox homeostasis, Ca^{2+} signalling and post-translational modifications. Prior work has shown that misfolded prion

protein can reduce oxidised NAD (NAD^+) levels (Zhou et al., 2015), hence the upregulation of NAD^+ may reflect a compensation response. Increased SUMOylation of Htt has been suggested to facilitate disaggregation of Htt, thereby releasing the soluble mutant Htt monomers and increasing neurotoxicity (Steffan et al., 2004). Hence reduction of SUMO pathways may indicate an adaptive response to aid sequestration of the soluble Httex1 states.

3.5.2. Transcriptional signature 2

This signature describes the transcriptional changes that are only altered when inclusions form. One possibility to explain this signature are the genes that change expression upon adaptation responses as inclusions form on the basis of prior work suggesting inclusions are

actively formed to sequester toxic soluble protein states (Arrasate et al., 2004; Kopito, 2000; Kaganovich et al., 2008). Up-regulated modules include RELA, which is a subnetwork within the NF- κ B signalling pathway that has been implicated in HD (Trager et al., 2014) and ribosomal associated protein module, which has functions in regulating ribosome biogenesis (Warner and McIntosh, 2009). As such their up-regulation may be part of mechanisms to globally regulate protein translation. Of note to the broader connection to CREB signalling alterations, CREB signalling regulates ribosomal protein production and aberrant production has been linked to ribosomal stress and apoptosis (Deisenroth and Zhang, 2011). Five modules were downregulated in this group. Within these modules, the PAXIP1 gene is involved in DNA repair and the CDK1 gene is involved in cell cycle re-entry. Both have been previously linked to HD pathology and in toxicity (explained further in Supplementary Note 1). A notable downregulated module was EP300. The EP300 gene encodes a co-activator of CREB, which suggests this response is influenced by the broader deregulation of CREB signalling. Also downregulated is the UBC module. UBC encodes ubiquitin, which is involved in ubiquitin-proteasome mediated protein degradation. This response is likely part of broader adaptation changes to protein folding stress (more details are explained in Supplementary Note 1).

3.5.3. Transcriptional signature 3

This signature describes a change of transcript levels triggered by soluble Httex1 and extended further when inclusions form. This signature may reflect a sustained or delayed toxicity response; or alternatively a delayed or amplifying adaptive response as aggregation proceeds. Upregulated modules included NF- κ B pathway (more generally than the RELA subset). This may reflect a compensation response since increased NF- κ B signalling aids cell survival (Yu et al., 2000). Down-regulated modules included KIAA0101, HNF4A and MYC. KIAA0101 gene is involved in oxidative stress mediated DNA repair (Klein and Ackerman, 2003). HNF4A gene is involved in the metabolic regulation of glucose and lipids has been shown to be impaired in HD (Olah et al., 2008; Block et al., 2010) and MYC gene is a master regulator of protein transcription (and indirectly translation) and is influenced by CREB signalling (Riggelen et al., 2010). Thus, the changes in MYC likely reflect an amplifying adaptive response to bypass defects in CREB signalling as well as aiding cell survival by reducing translation under stressed conditions.

3.5.4. Transcriptional signature 4

This signature describes transcript changes invoked by soluble 72Q Httex1 that then return to control levels when inclusions form. This signature could reflect genes that are specifically rebalanced by aggregation in adaptive responses from inclusion formation. The up-regulated modules included STAT3, of which the STAT3 gene works in concert with NF- κ B to regulate cell survival (Grivnikov and Karin, 2010) and VHL, which directs Hypoxia-inducible factor-1 (HIF-1) for degradation via the ubiquitin-proteasome system and suppresses apoptosis. The sole down-regulated module included HSP90AA. HSP90 and related chaperones link closely to systems regulating protein folding, degradation and potentially disaggregation (Luo et al., 2010).

4. Conclusions

In conclusion, our study has enabled the first comprehensive view of how aggregation state of Httex1 influences the transcriptome in HD. Most noteworthy was the finding that soluble Httex1 propagates the most substantial changes in the transcriptome by blocking CREB signalling, which supports an extensive body of work indicating CREB signalling defects as central to HD pathomechanisms. Importantly, our study is the first to our knowledge that has been able to identify this defect as specifically attributed to the soluble mutant Httex1 and in context of the system-wide transcriptional changes. Indeed, the

blockage of CREB signalling seemingly accounted for the majority of the toxicity caused by Httex1. As such, our data reaffirms CREB signalling as a highly attractive therapeutic target in Huntington's disease. This conclusion is supported findings that an enhancement of CREB levels via Sirt1 overexpression in the R6/2 HD mouse model can improve survival time (Cohen et al., 2012). In contrast, the formation of inclusions offers (by comparison) trivial changes in the magnitude of transcriptional changes.

This finding is also consistent with a novel mechanism we recently discovered showing that inclusion formation establishes a state of cellular quiescence and switches off a fast death mechanism triggered by soluble Httex1 states (Ramdhan et al., 2017). That is, a state of quiescence would impair further changes initiated from the soluble Httex1 states and the evidence herein is that there are limited adaptation responses centered around protein quality control and clearance and compensation to a loss of CREB signalling. The signatures point to those involved in adapting to defects in CREB signalling such as the rewiring of CREB-related signally networks to compensate for CREB deficiency.

Another feature of our dataset was the pronounced enrichment of genes encoding polyQ. Five of the 9 genes upregulated > 2-fold by the soluble polyQ-expanded Httex1 contained endogenous polyQ sequences suggesting that polyQ-expanded Htt may more broadly deactivate by soluble co-aggregation these proteins leading to their compensatory upregulation (Fig. 2). Three of the upregulated polyQ-encoding genes are involved in CREB signalling, which may explain why CREB signalling appears to be the most profoundly affected pathway (Fig. 3a).

Our new roadmaps of the transcriptional changes associated with Htt aggregation states offer a powerful resource to understanding how Httex1, notably the soluble forms, propagates toxicity as well as in yielding CREB signalling as a therapeutic target for further investigation.

Author contributions

NSM, JD, RDH, AJH, AO, and DMH designed research; NSM, ARO, AS, YMR, and MSZ performed research; NSM, ARO, AS, AO and DMH analyzed data; and NSM, A O, and DMH co-wrote the paper.

Acknowledgements

This work was supported by grants to DMH from the Australian Research Council (grant number FT120100039); grants/fellowships from the National Health and Medical Research Council Project to DMH (grant numbers APP1049458, APP1049459 and APP1102059), and a grant from the Hereditary Disease Foundation (USA). AJH is an NHMRC Principal Research Fellow.

The authors declare that they have no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.mcn.2017.07.004>.

References

- Arrasate, M., Finkbeiner, S., 2012. Protein aggregates in Huntington's disease. *Exp. Neurol.* 238, 1–11.
- Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., Finkbeiner, S., 2004. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* 431, 805–810.
- Bae, B.I., Xu, H., Igarashi, S., Fujimuro, M., Agrawal, N., Taya, Y., Hayward, S.D., Moran, T.H., Montell, C., Ross, C.A., Snyder, S.H., Saw, A., 2005. p53 mediates cellular dysfunction and behavioral abnormalities in Huntington's disease. *Neuron* 47, 29–41.
- Block, R.C., Dorsey, E.R., Beck, C.A., Brenna, J.T., Shoulson, I., 2010. Altered cholesterol and fatty acid metabolism in Huntington disease. *J. Clin. Lipidol.* 4, 17–23.
- Bolstad, B.M., Irizarry, R.A., Astrand, M., Speed, T.P., 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193.
- Carvalho, B.S., Irizarry, R.A., 2010. A framework for oligonucleotide microarray

- preprocessing. *Bioinformatics* 26, 2363–2367.
- Cohen, D.E., Cui, L., Supinski, A., Savas, J.N., Mazzulli, J.R., Yates, J.R., Bordone, L., Guarente, L.P., Krainc, D., 2012. Sirt1 mediates neuroprotection from mutant huntingtin by activation of TORC1 and CREB transcriptional pathway. *Nat. Med.* 18, 159–165.
- Cong, S.Y., Peppers, B.A., Evert, B.O., Rubinstein, D.C., Roos, R.A., van Ommen, G.J., Dorsman, J.C., 2005. Mutant huntingtin represses CBP, but not p300, by binding and protein degradation. *Mol. Cell. Neurosci.* 30, 560–571.
- Davies, S.W., Turmaine, M., Cozens, B.A., DiFiglia, M., Sharp, A.H., Ross, C.A., Scherzinger, E., Wanker, E.E., Mangiarini, L., Bates, G.P., 1997. Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* 90, 537–548.
- Deisenroth, C., Zhang, Y., 2011. The ribosomal protein-Mdm2-p53 pathway and energy metabolism: bridging the gap between feast and famine. *Genes Cancer* 2, 392–403.
- Devalia, B.N., Case-Borden, C., Geggion, A., Hsu, C.H., Chen, Q., Meerzaman, D., Dey, A., Ozato, K., Singer, D.S., 2016. BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin. *Nat. Struct. Mol. Biol.* 23, 540–548.
- DiFiglia, M., Sapp, E., Chase, K.O., Davies, S.W., Bates, G.P., Vonsattel, J.P., Aronin, N., 1997. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* 277, 1990–1993.
- Dunah, A.W., Jeong, H., Griffin, A., Kim, Y.M., Standaert, D.G., Hersch, S.M., Mouradian, M.M., Young, A.B., Tanese, N., Krainc, D., 2002. Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science* 296, 2238–2243.
- Gasset-Rosa, F., Chillon-Marinas, C., Gogginashvili, A., Atwal, R.S., Artates, J.W., Tabet, R., Wheeler, V.C., Bang, A.G., Cleveland, D.W., Lagier-Tourenne, C., 2017. Polyglutamine-expanded huntingtin exacerbates age-related disruption of nuclear integrity and nucleocytoplasmic transport. *Neuron* 94 (48–57), e44.
- Gautier, L., Cope, L., Bolstad, B.M., Irizarry, R.A., 2004. Affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20, 307–315.
- Genetic Modifiers of Huntington's Disease, Consortium, 2015. Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell* 162, 516–526.
- Grivennikov, S.I., Karin, M., 2010. Dangerous liaisons: STAT3 and NF-kappaB collaboration and crosstalk in cancer. *Cytokine Growth Factor Rev.* 21, 11–19.
- Ha, C.H., Kim, J.Y., Zhao, J., Wang, W., Jhun, B.S., Wong, C., Jin, Z.G., 2010. PKA phosphorylates histone deacetylase 5 and prevents its nuclear export, leading to the inhibition of gene transcription and cardiomyocyte hypertrophy. *Proc. Natl. Acad. Sci. U. S. A.* 107, 15467–15472.
- Hammitzsch, A., Tallant, C., Fedorov, O., O'Mahony, A., Brennan, P.E., Hay, D.A., Martinez, F.O., Al-Mossawi, M.H., de Wit, J., Vecellio, M., Wells, C., Wordsworth, P., Müller, S., Knapp, S., Bowness, P., 2015. CBP30, a selective CBP/p300 bromodomain inhibitor, suppresses human Th17 responses. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10768–10773.
- Hatters, D.M., 2012. Putting huntingtin "aggregation" in view with windows into the cellular milieu. *Curr. Top. Med. Chem.* 12, 2611–2622.
- Hodges, A., Strand, A.D., Aragaki, A.K., Kuhn, A., Sengstad, T., Hughes, G., Elliston, L.A., Hartog, C., Goldstein, D.R., Thu, D., Hollingsworth, Z.R., Collin, F., Synek, B., Holmans, P.A., Young, A.B., Weder, N.S., Delorenzi, M., Kooperberg, C., Augood, S.J., Faulstich, R.L., Olson, J.M., Jones, L., Luthi-Carter, R., 2006. Regional and cellular gene expression changes in human Huntington's disease brain. *Hum. Mol. Genet.* 15, 965–977.
- von Horsten, S., Schmitt, I., Nguyen, H.P., Holzmann, C., Schmidt, T., Walther, T., Bader, M., Pabst, R., Köbbe, P., Krotova, J., Stiller, D., Kask, A., Vaarmann, A., Rathke-Hartlieb, S., Schulz, J.B., Grasshoff, U., Bauer, I., Vieira-Saecker, A.M., Paul, M., Jones, L., Lindenberg, K.S., Landwehrmeyer, B., Bauer, A., Li, X.J., Riess, O., 2003. Transgenic rat model of Huntington's disease. *Hum. Mol. Genet.* 12, 617–624.
- Huang da, W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13.
- Huang da, W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.
- Insel, P.A., Ostrom, R.S., 2003. Forskolin as a tool for examining adenylyl cyclase expression, regulation, and G protein signaling. *Cell. Mol. Neurobiol.* 23, 305–314.
- Jiang, H., Poirier, M.A., Liang, Y., Pei, Z., Weiskittel, C.E., Smith, W.W., DeFranco, D.B., Ross, C.A., 2006. Depletion of CBP is directly linked with cellular toxicity caused by mutant huntingtin. *Neurobiol. Dis.* 23, 543–551.
- Kaganovich, D., Kopito, R., Frydman, J., 2008. Misfolded proteins partition between two distinct quality control compartments. *Nature* 454, 1088–1095.
- Kazantsev, A., Preisinger, E., Dranovsky, A., Goldhaber, D., Housman, D., 1999. Insoluble detergent-resistant aggregates form between pathological and nonpathological lengths of polyglutamine in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 96, 11404–11409.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.
- Klein, J.A., Ackerman, S.L., 2003. Oxidative stress, cell cycle, and neurodegeneration. *J. Clin. Invest.* 111, 785–793.
- Kolenskikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandt, M., Burdett, T., Megy, K., Pilicheva, E., Rustici, G., Tikhonov, A., Parkinson, H., Petryszak, R., Sarkans, U., Brazma, A., 2015. ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.* 43, D1113–D1116.
- Kopito, R.R., 2000. Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* 10, 524.
- Law, C.W., Chen, Y., Shi, W., Smyth, G.K., 2014. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15, R29.
- Liao, Y., Smyth, G.K., Shi, W., 2013. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* 41, e108.
- Luo, W., Sun, W., Taldone, T., Rodina, A., Chiosis, G., 2010. Heat shock protein 90 in neurodegenerative diseases. *Mol. Neurodegener.* 5, 24.
- Luthi-Carter, R., Strand, A., Peters, N.L., Solano, S.M., Hollingsworth, Z.R., Menon, A.S., Frey, A.S., Spector, B.S., Penney, E.B., Schilling, G., Ross, C.A., Borchelt, D.R., Tapscott, S.J., Young, A.B., Cha, J.H., Olson, J.M., 2000. Decreased expression of striatal signaling genes in a mouse model of Huntington's disease. *Hum. Mol. Genet.* 9, 1259–1271.
- MacDonald, J.W., 2016. mogene10stprobeset.db.
- MacDonald, M.E., Ambrose, C.M., Duyao, M.P., Myers, R.H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S.A., James, M., Groot, N., MacFarlane, H., Jenkins, B., Anderson, M.A., Wexler, N.S., Gusella, J.F., Bates, G.P., Baxendale, S., Hummerich, H., Kirby, S., North, M., Youngman, S., Mott, R., Zehetner, G., Sedlacek, Z., Poustka, A., Frischauf, A.-M., Lehrach, H., Buckler, A.J., Church, D., Doucette-Stamm, L., O'Donovan, M.C., Riba-Ramirez, L., Shah, M., Stanton, V.P., Strobel, S.A., Draths, K.M., Wales, J.L., Dervan, P., Housman, D.E., Altherr, M., Shlang, R., Thompson, L., Fielder, T., Wasmuth, J.J., Tagle, D., Valdes, J., Elmer, L., Allard, M., Castilla, L., Swaroop, M., Blanchard, K., Collins, F.S., Snell, R., Holloway, T., Gillespie, K., Datson, N., Shaw, D., Harper, P.S., 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971–983.
- Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trotter, Y., Leach, H., Davies, S.W., 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.
- Mantamadiotis, T., Lemberger, T., Bleckmann, S.C., Kern, H., Kretz, O., Martin Villalba, A., Tronche, F., Kellendonk, C., Gau, D., Kapfhammer, J., Otto, C., Schmid, W., Schutz, G., 2002. Disruption of CREB function in brain leads to neurodegeneration. *Nat. Genet.* 31, 47–54.
- Olah, J., Klivenyi, P., Gardian, G., Vecsei, L., Orosz, F., Kovacs, G.G., Westerhoff, H.V., Ovadi, J., 2008. Increased glucose metabolism and ATP level in brain tissue of Huntington's disease transgenic mice. *FEBS J.* 275, 4740–4755.
- Olshina, M.A., Angley, L.M., Ramdzan, Y.M., Tang, J., Bailey, M.F., Hill, A.F., Hatters, D.M., 2010. Tracking mutant huntingtin aggregation kinetics in cells reveals three major populations that include an invariant oligomer pool. *J. Biol. Chem.* 285, 21807–21816.
- Picaud, S., Fedorov, O., Thanassopoulou, A., Leonards, K., Jones, K., Meier, J., Olzscha, H., Monteiro, O., Martin, S., Philpott, M., Tumber, A., Filippakopoulos, P., Yapp, C., Wells, C., Che, K.H., Bannister, A., Robson, S., Kumar, U., Parr, N., Lee, K., Lugo, D., Jeffrey, P., Taylor, S., Vecellio, M.L., Bountra, C., Brennan, P.E., O'Mahony, A., Velichko, S., Müller, S., Hay, D., Daniels, D.L., Uhr, M., La Thangue, N.B., Kouzarides, T., Prinjha, R., Schwallier, J., Knapp, S., 2015. Generation of a selective small molecule inhibitor of the CBP/p300 bromodomain for leukemia therapy. *Cancer Res.* 75, 5106–5119.
- Pollings, S., Mok, Y.F., Ramdzan, Y.M., Turner, B.J., Yerbury, J.J., Hill, A.F., Hatters, D.M., 2014. Misfolded polyglutamine, polyaniline, and superoxide dismutase 1 aggregate via distinct pathways in the cell. *J. Biol. Chem.* 289, 6669–6680.
- Pollings, S., Ormsby, A.R., Wood, R.J., Lee, K., Shoubridge, C., Hughes, J.N., Thomas, P.Q., Griffin, M.D., Hill, A.F., Bowden, Q., Bocking, T., Hatters, D.M., 2015. Polyaniline expansions drive a shift into alpha-helical clusters without amyloid-fibril formation. *Nat. Struct. Mol. Biol.* 22, 1008–1015.
- Ramdzan, Y.M., Nisbet, R.M., Miller, J., Finkbeiner, S., Hill, A.F., Hatters, D.M., 2010. Conformation sensors that distinguish monomeric proteins from oligomers in live cells. *Chem. Biol.* 17, 371–379.
- Ramdzan, Y.M., Pollings, S., Chia, C.P., Ng, I.H., Ormsby, A.R., Croft, N.P., Purcell, A.W., Bogoyevitch, M.A., Ng, D.C., Gleeson, P.A., Hatters, D.M., 2012. Tracking protein aggregation and mislocalization in cells with flow cytometry. *Nat. Methods* 9, 467–470.
- Ramdzan, Y.M., Trubetskoy, M.M., Ormsby, A.R., Newcombe, E.A., Sui, X., Tobin, M.J., Bongiovanni, M.N., Gras, S.L., Dawson, G., Miller, J.M.L., Finkbeiner, S., Moily, N.S., Niclis, J., Parish, C.L., Purcell, A.W., Baker, M.J., Wilce, J.A., Waris, S., Stojanowski, D., Bocking, T., Ang, C.S., Ascher, D.B., Reid, G.E., Hatters, D.M., 2017. Huntingtin Inclusions Trigger Cellular Quiescence, Deactivate Apoptosis, and Lead to Delayed Necrosis. *Cell Rep.* 19, 919–927.
- Riggelsen, J.V., Yetil, A., Felsner, D.W., 2010. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat. Rev. Cancer* 10, 301–309.
- Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11, R25.
- Schaffar, G., Breuer, P., Boteva, R., Behrends, C., Tsvetkov, N., Strippel, N., Sakahira, H., Siegers, K., Hayer-Hartl, M., Hartl, F.U., 2004. Cellular toxicity of polyglutamine expansion proteins: mechanism of transcription factor deactivation. *Mol. Cell* 15, 95–105.
- Scherzinger, E., Sittler, A., Schweiger, K., Heiser, V., Lurz, R., Hasenbank, R., Bates, G.P., Leach, H., Wanker, E.E., 1999. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4604–4609.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Elieiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological image analysis. *Nat. Methods* 9, 676–682.
- Shimohata, T., Nakajima, T., Yamada, M., Uchida, C., Onodera, O., Naruse, S., Kimura, T., Koide, R., Nozaki, K., Sano, Y., Ishiguro, H., Sakoe, K., Ooshima, T., Sato, A., Ikeuchi, T., Oyake, M., Sato, T., Aoyagi, Y., Hozumi, I., Nagatsu, T., Takiyama, Y., Nishizawa, M., Goto, J., Kanazawa, I., Davidson, I., Tanese, N., Takahashi, H., Tsuji, S., 2000. Expanded polyglutamine stretches interact with TAFII130, interfering with CREB-dependent transcription. *Nat. Genet.* 26, 29–36.
- Sipe, J.D., Benson, M.D., Buxbaum, J.N., Ikeda, S.I., Merilini, G., Saraiya, M.J.,

- Westmark, P., 2016. Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines. *Amyloid* 23, 209–213.
- Steffan, J.S., Kazantsev, A., Spasic-Boskovic, O., Greenwald, M., Zhu, Y.Z., Gohler, H., Wanker, E.E., Bates, G.P., Housman, D.E., Thompson, L.M., 2000. The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6763–6768.
- Steffan, J.S., Bodai, L., Pallos, J., Poelman, M., McCampbell, A., Apostol, B.L., Kazantsev, A., Schmidt, E., Zhu, Y.Z., Greenwald, M., Kurokawa, R., Housman, D.E., Jackson, G.R., Marsh, J.L., Thompson, L.M., 2001. Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in *Drosophila*. *Nature* 413, 739–743.
- Steffan, J.S., Agrawal, N., Pallos, J., Rockabrand, E., Trotman, L.C., Slepko, N., Iles, K., Lukacovich, T., Zhu, Y.Z., Cattaneo, E., Pandolfi, P.P., Thompson, L.M., Marsh, J.L., 2004. SUMO modification of Huntingtin and Huntington's disease pathology. *Science* 304, 100–104.
- Szkilarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., von Mering, C., 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–452.
- Trager, U., Andre, R., Lahiri, N., Magnusson-Lind, A., Weiss, A., Grueninger, S., McKinnon, C., Srinathsinghji, E., Kahlon, S., Pfister, E.L., Moser, R., Hummerich, H., Antoniou, M., Bates, G.P., Luthi-Carter, R., Lowdell, M.W., Bjorkqvist, M., Ostroff, G.R., Aronin, N., Tabrizi, S.J., 2014. HTT-lowering reverses Huntington's disease immune dysfunction caused by NF-kappaB pathway dysregulation. *Brain* 137, 819–833.
- Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., Buhler, M., Margueron, R., Schneider, R., 2013. Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. *Cell* 152, 859–872.
- Tsvetkov, A.S., Miller, J., Arrasate, M., Wong, J.S., Pleiss, M.A., Finkbeiner, S., 2010. A small-molecule scaffold induces autophagy in primary neurons and protects against toxicity in a Huntington disease model. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16982–16987.
- Warner, J.R., McIntosh, K.B., 2009. How common are extra-ribosomal functions of ribosomal proteins? *Mol. Cell* 34, 3–11.
- Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M.L., Visvader, J.E., Smyth, G.K., 2010. ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* 26, 2176–2182.
- Wyttenbach, A., Swartz, J., Kita, H., Thykjaer, T., Carmichael, J., Bradley, J., Brown, R., Maxwell, M., Schapira, A., Orntoft, T.F., Kato, K., Rubinsztein, D.C., 2001. Polyglutamine expansions cause decreased CRE-mediated transcription and early gene expression changes prior to cell death in an inducible cell model of Huntington's disease. *Hum. Mol. Genet.* 10, 1829–1845.
- Xia, J., Benner, M.J., Hancock, R.E., 2014. NetworkAnalyst—integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res.* 42, W167–174.
- Xia, J., Gill, E.E., Hancock, R.E., 2015. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat. Protoc.* 10, 823–844.
- Yang, S.-H., Cheng, P.-H., Banta, H., Piotrowska-Nitsche, K., Yang, J.-J., Cheng, E.C.H., Snyder, B., Larkin, K., Liu, J., Orkin, J., Fang, Z.-H., Smith, Y., Bachevalier, J., Zola, S.M., Li, S.-H., Li, X.-J., Chan, A.W.S., 2008. Towards a transgenic model of Huntington's disease in a non-human primate. *Nature* 453, 921–924.
- Yu, Z., Zhou, D., Cheng, G., Mattson, M.P., 2000. Neuroprotective role for the p50 subunit of NF-kappaB in an experimental model of Huntington's disease. *J. Mol. Neurosci.* 15, 31–44.
- Zhou, M., Ottenberg, G., Sferazza, G.F., Hubbs, C., Fallahl, M., Rumbaugh, G., Brantley, A.F., Lasmezas, C.I., 2015. Neuronal death induced by misfolded prion protein is due to NAD⁺ depletion and can be relieved in vitro and in vivo by NAD⁺ replenishment. *Brain* 138, 992–1008.

2.3. Conclusion and future directions

In conclusion, our study has enabled the first comprehensive view of how aggregation state of Httex1 influences the transcriptome in HD. Most noteworthy was the finding that soluble Httex1 propagates the most substantial changes in the transcriptome by blocking CREB signalling, which supports an extensive body of work indicating CREB signalling defects as central to HD pathomechanisms. Importantly, our study is the first to our knowledge that has been able to identify defects as specifically attributed to the soluble mutant Httex1 in context of the system-wide transcriptional changes. This was enabled using a unique flowcytometry based toolkit -PulSA that helped enrich cells with soluble mutant Httex1. Indeed, the blockage of CREB signalling seemingly accounted for majority of the toxicity caused by Httex1. As such, our data reaffirms CREB signalling as a highly attractive therapeutic target in Huntington's disease. This conclusion is supported findings that an enhancement of CREB levels via Sirt1 overexpression in the R6/2 HD mouse model can improve survival time [237]. In contrast, the formation of inclusions offers (by comparison) trivial changes in the magnitude of transcriptional changes. Moreover, of the changes observed, the signatures point to those involved in adapting to defects in CREB signalling such as the rewiring of CREB-related signally networks to compensate for CREB deficiency. In addition, other changes associated with protein homeostasis responses were also documented in these signatures.

Another feature of our dataset was the pronounced enrichment of genes encoding polyQ. Five of the 9 genes upregulated more than 2-fold by the soluble polyQ-expanded Httex1 contained endogenous polyQ sequences suggesting that polyQ-expanded Htt may more broadly deactivate by soluble co-aggregation these proteins leading to their compensatory upregulation (Fig 2). Three of the upregulated polyQ-encoding genes are involved in CREB signalling, which may explain why CREB signalling appears to be the most profoundly affected pathway (Fig 3a). While our study was able to elucidate several Htt aggregation related mechanisms, the use of mouse neuroblastoma cell line is a limitation of the study as it does not ideally replicate *in vivo* changes in the brain. Though efficient in elucidating mechanisms at work, the use of immortal neural cell lines and overexpression of mutant Httex1 fragments does not accurately reflect the intricate interplay of diverse cell types and their effect on target gene expression as seen in the brain. The cross validation using R6/2 mouse brain data alleviates some of these concerns. An additional point of note is that R6/2

mice, depending on the length of their CAG repeats, develop nuclear (90 CAG) or cytoplasmic aggregates (200 CAG) [238]. The nuclear aggregates form early, are punctate and have distinct characteristics to cytoplasmic aggregates, which polymerise into larger inclusion bodies in the HD mouse brain. While our cell line Httex1 overexpression model system produced primarily cytosolic aggregates, as demonstrated in the R6/2-CAG 200 mouse brain, it may not represent the complexity of Htt aggregation species and their subcellular site-specific consequences.

Our new roadmaps of the transcriptional changes associated with Htt aggregation states offer a powerful resource to understanding how Httex1, notably the soluble forms, propagates toxicity as well as in yielding CREB signalling as a therapeutic target for further investigation. By utilisation of newer functional tools like CRISPR inducible cell lines (both activation and knock-out) we can further understand the biology of toxic and adaptive changes in HD. The targets identified in our study has broad relevance in terms of other aggregation related disorders. This study can be used as a template to understand the biology of other neurodegenerative conditions and develop better targeted therapeutics.

2.4. Supplementary material

Supplementary materials for:

Transcriptional profiles for distinct aggregation states of mutant Huntingtin exon 1 protein unmask new Huntington's disease pathways

Nagaraj S. Moily, Angelique R. Ormsby, Aleksandar Stojilovic, Yasmin M. Ramdzan, Jeannine Diesch, Ross D. Hannan, Michelle S. Zajac, Anthony J. Hannan, Alicia Oshlack and Danny M. Hatters

Link for peer reviewed publication: <https://doi.org/10.1016/j.mcn.2017.07.004>

Link for Table S1, Table S2, Table S3: <https://ars.els-cdn.com/content/image/1-s2.0-S1044743117301306-mmc2.xlsx>

- **Table S1: a)** Key pathways altered on comparison of grouped 72Q datasets (72Q-i + 72Q-ni) vs. 25Q-ni. **b)** Ingenuity- upstream regulator analysis was done on differentially expressed genes between the soluble mutant Httex1 (72Q-ni) and soluble wild-type Httex1 (25Q-ni). **c)** The differential gene expression data for all

comparisons done in the study, including total 72Q (72Q-i + 72Q-ni) vs.25Q; 72Q-ni vs. 25Q-ni; 72Q-ni vs. 72Q-i and 72Q-i vs. 25Q-ni.

- **Table S2:** a) Differential striatal gene expression gene list of R6/1 Huntington disease mouse compared to the wild type litter mates. b) Ingenuity- canonical pathway analysis was performed to determine key pathways altered in the striatum of HD mice.
- **Table S3:** Gene hubs identified in the hub analysis of different transcriptomic signatures.

Supplementary figures and Supplemental Note 1 are below.

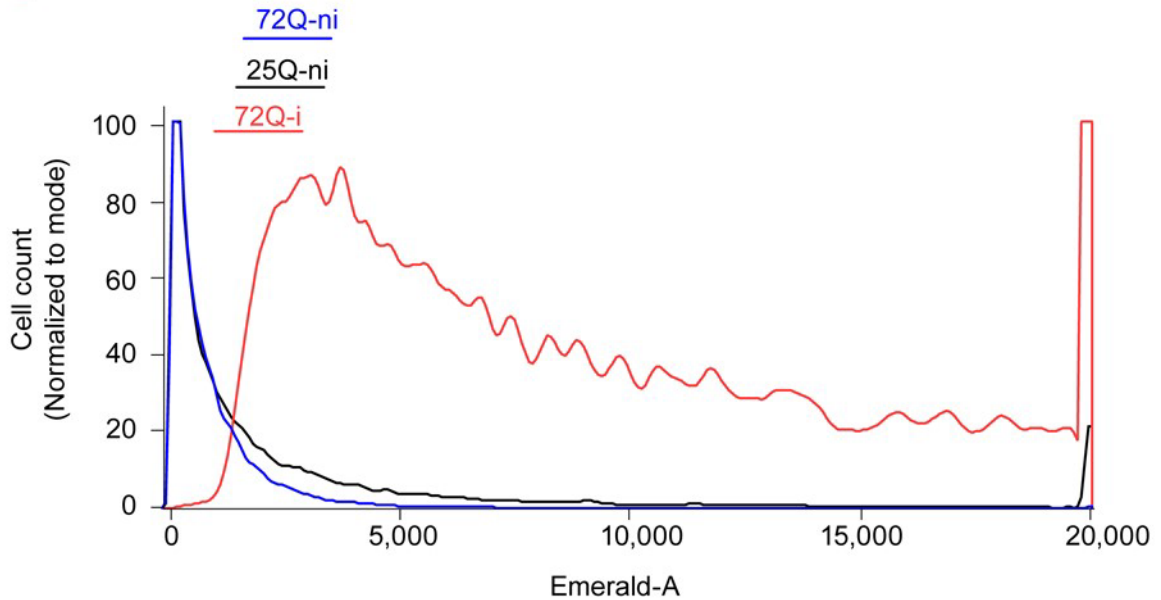


Fig S1: Flow cytometry gates used to match the expression level of cells collected for RNA sequencing (equivalent to the red data points shown in Fig 1c). Shown are histograms of expression levels (based on Emerald fluorescence) of cells transfected with Httex1-Emerald at 48 h post transfection. Bandwidths for the gates are shown above the histograms by the bars. Each gate describes a matched median fluorescence (at 2,200 fluorescence units) and the same bandwidths of 2,000 fluorescence units.

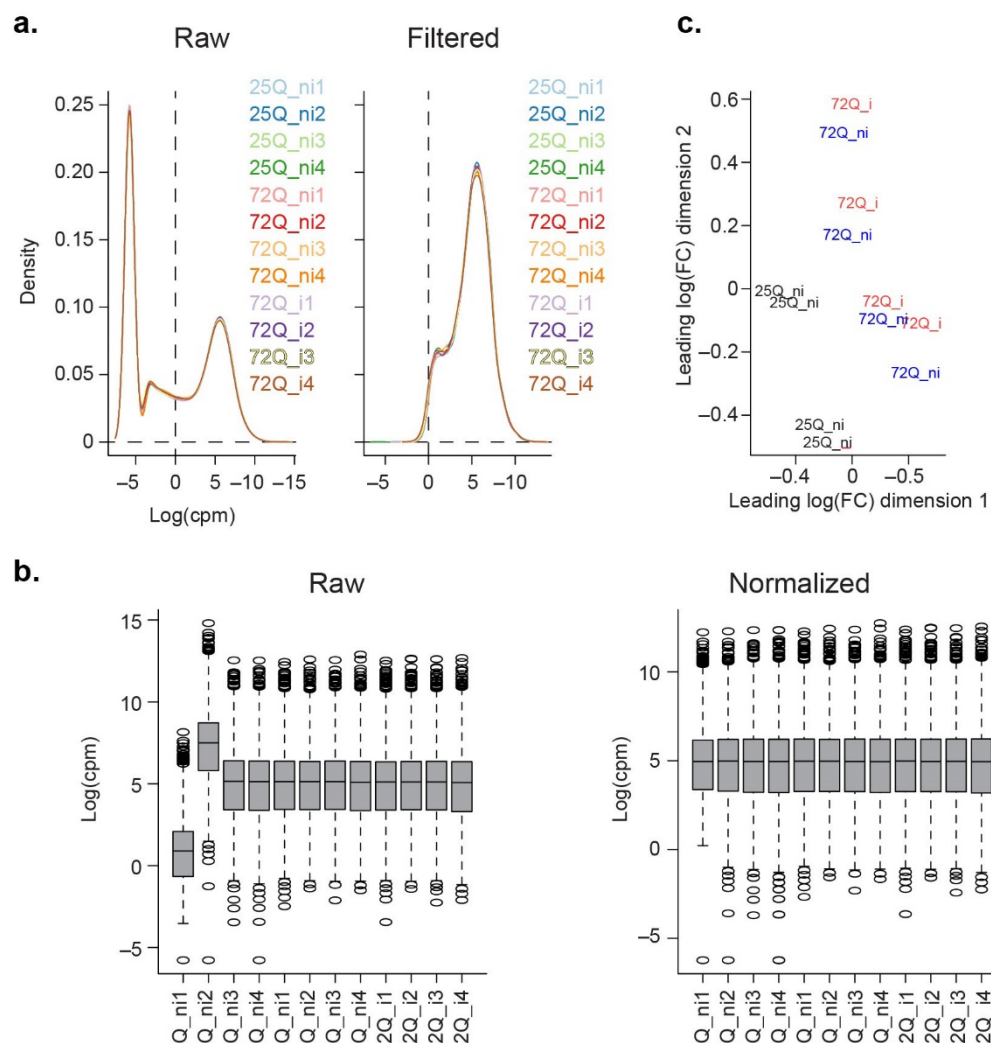


Fig S2: Quality control assessments and data corrections of raw sequencing data (corresponds to data in Fig 2). **a)** The density of log counts per million (log(CPM)) values for raw pre-filtered data and post-filtered data are shown for each sample. Dotted vertical lines mark the log counts per million of zero threshold (equivalent to a count per million value of one used in the filtering step.) **b)** Boxplots of log(CPM) values showing expression distributions for raw data and normalised data for each sample in the modified dataset where the counts in samples 1 and 2 have been scaled from their original values respectively. **c)** Multidimensional scaling plots of log-CPM values over dimensions 1 and 2 with samples coloured and labelled by sample groups (left graph). Distances on the plot correspond to the leading fold-change, calculated based on the 500 genes most variable genes in the dataset.

SUPPLEMENTAL NOTE 1

Histone deacetylase 5 (HDAC5). Gene transcription is governed in part by the acetylation and deacetylation of histones, the latter of which is mediated by histone deacetylases (HDACs). Whereas the role of HDAC1, HDAC3 has well been documented in HD[239, 240], the role of HDAC5 in HD is not well characterized. However, CREB1-CBP signalling is switched off by displacement of CREB-CBP from the promoter by CREB2 and HDAC5 suggesting a potential connection between the CREB signalling pathway maladaptation and HDAC5 function. HDAC5 also is known to deacetylate histone H4[241] and H4 deacetylation has been observed as a feature in HD pathology[212].

Bromodomain-containing protein 4 (BRD4). BRD4 is a member of the BET (bromodomain and extra-terminal domain) family, which function as epigenetic regulators. They acetylate histones (H3 and H4) and facilitate the transcription of target genes[242, 243]. BRD4 acetylates the same target histones as CREBBP, suggesting a functional connection at this level[244, 245].

Nicotinamide phosphoribosyl transferase (NAMPT). NAMPT is the rate-limiting enzyme in the nicotinamide adenine dinucleotide (NAD) biosynthetic pathway. Oxidised and reduced NAD (NAD⁺ and NADH respectively) are essential cellular metabolites involved in energy production, redox homeostasis, Ca²⁺ signalling and post-translational modifications[246, 247]. It was recently shown that NAD⁺ levels were depleted in cells expressing a soluble misfolded monomeric form of prion protein via excessive cellular protein ADP-ribosylation and resulted in metabolic failure[248]. Furthermore, administration of NAD⁺ specifically protected against soluble monomeric prion toxicity. Hence upregulated NAMPT likely reflects a coping mechanism against soluble toxic Httex1 forms.

Gamma-aminobutyric acid receptor-associated protein (GABARAP). The GABARAP family members (GABARAP, GABARAPL1 and GABARAPL2) are involved in the intracellular trafficking of receptors and the macroautophagy pathway. Autophagy offers protection to cells by removing toxic misfolded protein in response to proteotoxic stress[249, 250]. GABARAP is involved in autophagosome formation and maturation for autophagy involving specific cytosolic cargo[249, 250]. The hub gene sequestosome 1 (SQSTM1), is a ubiquitin-binding scaffold protein that colocalizes with ubiquitinated protein aggregates in HD [251]. SQSTM1 binds directly to GABARAP family proteins and may serve to link Lys-63-linked ubiquitylated proteins to the autophagosome[252]. Critically, SQSTM1 and

GABARAP accumulate when autophagy is inhibited and vice versa, these are depleted when autophagy is stimulated. Hence the upregulation of SQSTM1 and GABARAP hubs in cells with soluble mutant Htt suggests a block in autophagy. This may point to a specific defect in autophagy by soluble Httex1 forms.

ELAV-like protein 1 (ELAVL1). ELAVL1 or HuR (human antigen R) is an RNA-binding protein involved in stabilizing mRNA. HuR is recruited to stress granules, which are important structures implicated as central to the dysfunction underpinning Amyotrophic Lateral Sclerosis [253, 254]. ELAVL1 is also thought to be involved in autophagy by binding to and post-transcriptionally regulating SQSTM1 mRNA in cells with proteasome inhibition[255]. Furthermore, the expression of ELAVL1 is thought to be regulated NF- κ B and CDK1 signalling[256, 257]. Collectively these results suggest ELAVL1 is tightly linked to the changes in the GABARAP, NF- κ B and CDK1 modules.

Small ubiquitin-related modifier 2 (SUMO2). The SUMO protein family participate in post-translational protein modifications (SUMOylations) that enable cells to respond to extracellular stimuli with various roles in cellular function and pathology[258]. Mutant Htt undergoes SUMO modification in the amino-terminal region by the stress-inducible SUMO-2[259]. SUMO-2 modification regulates the accumulation of insoluble HTT and increased SUMOylation results in disaggregation of the insoluble Htt aggregates, thereby releasing the soluble mutant Htt and increasing neurotoxicity[94]. Thus, a downregulation of SUMO2 would help in reducing SUMOylation of soluble mutant Htt thereby reducing soluble mutant Htt toxicity.

RELA proto-oncogene, NF- κ B subunit (RELA). RELA, also known as TF65/p65, is involved in NF- κ B heterodimer formation and its nuclear translocation and activation. RELA links tightly with the BRD4 and CREB modules. BRD4 acetylates RELA as part of its function to activate NF- κ B. The transactivation domains within the carboxy-terminal region of RELA interact with p300 and CREBBP for transcriptional coactivation[260].

Ribosome associated proteins (RPS/RPL). Ribosomal proteins play an essential role in ribosome assembly and protein translation however, they have key ribosome-independent functions as well[261]. Disruption of ribosome biogenesis triggers a stress response that activates p53 signalling and apoptosis[261, 262]. This stress response involves protein MDM2 associating with the ribosome proteins to promote p53 activation[263]. Hence abnormal activation of MDM2 and ribosome proteins in cells enriched with mutant Httex1

may be a trigger for this apoptotic stimulus or a response to mitigate stress. A further link is that CREB mediates RP gene transcription[264].

Cyclin-dependent kinase-1 (CDK-1). CDK1 a serine/threonine kinase, belongs to the family of cyclin-dependent kinases which plays a key role in cell cycle regulation[265]. CDK1 is central to the activation of p53 under stress, leading to apoptosis[266, 267].

PAX-interacting protein 1 (PAXIP1). This gene is a member of the paired box (PAX) gene family mediating signal transduction events in the DNA damage response pathway[268, 269]. Mutant Htt-induced DNA damage and repair defects have been previously connected with HD [270], with the activation of the DNA damage response preceding the formation of protein aggregates[271]. Upon DNA damage, PAXIP1 localizes to DNA damage nuclear foci and interacts with proteins involved in sensing and repairing DNA damage[272]. Additionally, PAXIP1 contains a long polyQ stretch that may interact with mutant Htt through its naturally occurring polyQ domain. Further in another polyQ disease (spinal and bulbar muscular atrophy (SBMA), also known as Kennedy's disease), the polyQ-expanded mutant protein (androgen receptor) sequesters PAXIP1 attenuating DNA repair and increasing genomic instability[273]. Thus, Htt may co-aggregate with PAXIP1 and directly interfere with its function.

E1A binding protein p300 (EP300). EP300 (p300) is a histone acetyltransferase that regulates transcription[274]. It mediates cAMP-mediated gene regulation by binding specifically to phosphorylated CREB protein[275]. This gene has also been identified as a co-activator of HIF-1 α (which is part of the VHL module described below), and, thus, plays a role in the stimulation of hypoxia-induced genes[276].

Polyubiquitin-C (UBC). Polyubiquitin-C encodes ubiquitin protein. UBC gene transcription is induced during stress and results in the formation of polyubiquitin precursor proteins of around 3–4 or 9–11 monomers[277]. This provides an extra pool of ubiquitin for clearance of excess misfolded proteins. Ubiquitin is an important signalling protein for proteasome-mediated degradation and has been suggested to be dysregulated in HD brain and mouse models[278].

Nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B). NF- κ B signalling regulates expression of key pro-survival apoptotic genes. Several lines of evidence have linked Epsmutant Htt to NF- κ B. For example, Htt has been reported to interact with IKK to influence NF- κ B –signalling in immune cells[279]. In addition, the p65 subunit of

NF- κ B has been shown to sequester in the Htt aggregates along with CBP/p300[280] and the mice lacking the p50 subunit of NF- κ B suffered increased damage to striatal neurons after administration 3-nitropropionic acid[281], thereby indicating the key role NF- κ B plays in as mutant Htt-induced cell death and aggregation.

KIAA0101. This gene encodes for a protein called PCNA-associated factor, which is an essential protein in both DNA replication and DNA damage repair[282]. Defects in DNA mismatch repair mainly related to persistent oxidative stress[283] have been implicated in HD and other neurodegenerative disorders[284]. This target could be a novel regulator of oxidative stress-related cell health and could indicative of progressive cellular DNA damage.

Hepatocyte nuclear factor 4 alpha (HNF4A). HNF4A belongs to a superfamily of nuclear hormone receptor (NHR) proteins that control transcription in diverse metabolic pathways. It plays an important role in glucose metabolism and insulin signalling and is also involved in cholesterol and fatty acid metabolism. HNF-4 α is a nuclear transcription factor and has been linked to CREB signalling, where CBP functions as a key transcriptional coactivator for HNF4A[285], [286]. Impaired fatty acid metabolism and cholesterol dyshomeostasis has been observed in HD and HNF4A could function as impaired lipid regulation targets[287]. Further increased glucose metabolism and ATP level in brain tissue of Huntington's disease transgenic mice[288] shown might indicate CBP-HNF4A related control of cellular energy dynamics in HD.

c-MYC (MYC). MYC has been described as a global regulator of cellular processes, including transcription, regulation of chromatin structure, translation and ribosome biogenesis[289]. CREBBP has been reported to bind and acetylate c-MYC to positively regulate c-MYC activity [289, 290]. Moreover, MYC inactivation irreversibly suppresses ribosome biogenesis and protein translation with marked changes in the total rate of protein synthesis[289].

Signal transducer and activator of transcription 3 (STAT3). STAT3 is an important transcription factor which is expressed in response to cellular stress and controls the expression of several anti-apoptotic genes. Once activated STAT3 works in concert with NF- κ B to regulate cell survival[291, 292]. CREBBP can activate STAT3 by acetylating the C-terminal transcriptional activation domain leading to transactivation of downstream pro-survival genes such as Bcl-2 and Bcl-X[293].

Von Hippel–Lindau tumor suppressor (VHL). VHL is an E3 ligase that targets Hypoxia-inducible factor-1 (HIF-1) for degradation by the proteasome[294]. HIF-1 has a key role in cellular responses to hypoxia, including the regulation of genes involved in energy metabolism and apoptosis [281]. HIF-1 pathway is established in mechanisms of neurodegeneration[295]. HIF-1 activity requires co-transactivation factors such as CREBBP[296].

Heat shock protein 90 (HSP90AA). Hsp90 is a molecular chaperone with important roles in maintaining protein stability. Hsp90 also regulates the activity of the HIF-1, independently to VHL function[297]. Furthermore, inhibition of Hsp90 activates heat shock factor-1 (HSF-1), which stimulates Hsp70 and Hsp40 expression that can promote disaggregation and protein degradation[298].

Chapter 3:

Httex1 aggregation states induce spliceosome defects and aberrant splicing in Huntington's disease.

3.1. Abstract

HD is a genetic disease caused by an expanded CAG tract in the Htt gene that allows translation into an elongated polyQ-stretch in the mHtt protein. The primary characteristic molecular pathology in HD is the accumulation of mHtt into micrometre-sized aggregates. The soluble form of mHtt has been shown to be most neurotoxic, whereas aggregates lower the risk of cellular death. We have previously described transcriptional changes related to different aggregation states of mHtt. Altered splicing has been implicated in HD causing exacerbation of pathogenesis. In this study we identified defects in the uridine rich small nuclear RNAs as well in the splicing factor SRSF2 which is required for formation of a functional spliceosomal complex in cells with mutant Htt. By utilising flow cytometry method of PulSA we further analysed splicing signature in cells enriched with various aggregation states of mHtt. Further using mass spectrometry, we also confirmed modifications to transcript-specific peptides resulting from alternative splicing before versus after inclusion assembly. These findings indicate defects in splicing machinery and detail aggregation state specific splicing changes in Huntington disease as well as avenues for future research.

3.2. Introduction

Aggregation of proteins is a critical feature of neurodegenerative disorders including HD[299]. HD results from an expansion of CAG trinucleotide repeats within Httex1 beyond the pathogenic threshold of 35 repeats[1], translating into an elongated polyQ stretch in the mHtt protein. This pathogenic polyQ expansion increases the propensity of soluble mutant Huntingtin (Htt) to form amyloid fibrils and the N-terminal polyQ fragments to form intracellular protein aggregates called inclusions[13][1]. The soluble form of Httex1 has been shown to cause toxicity much before aggregation occurs. Though the exact mechanism of Httex1 induced pathogenesis is not clear, we have detailed (in Chapter 2) the consequences of soluble mutant Httex1 in a cell culture model of HD that transcriptional dysregulation mediated through CREB signalling is a major feature of HD pathogenesis[300]. Similarly, other studies have also implicated transcriptional dysregulation via CREB signalling alterations as key to transcriptional alterations in various animal models of HD[172, 301]. The transcriptional changes seen in HD are widespread and occur in various areas of the HD brain long before symptoms appear[68, 210]. Similarly, post-transcriptional alteration in mRNA splicing have been documented in HD[302-305] and aberrant splicing changes have also been confirmed in HD brains[303, 306]. However, the post-transcriptional splicing changes associated with different aggregation states and their regulatory mechanisms have not been characterised.

Pre-mRNA splicing is a critical step necessary for gene expression. It is performed by a spliceosome that recognizes splicing signals and catalyses the elimination of non-coding intron sequences and assembles mature mRNAs from protein-coding exons before their export and translation in the cytoplasm[307, 308]. More than 95% of genes in the human genome are recognised to generate transcripts that undergo alternative splicing, enhancing the proteomic diversity in the cell[309, 310]. The regulation of gene splicing can modify the availability of protein isoforms by altering when, where and how mRNA transcripts are expressed, alterations in cellular splicing patterns can influence many cellular functions and cellular response to stress[311-313]. Cellular splicing is undertaken in the nucleus by uridyl-rich small nuclear RNAs (U-snRNAs). These constitutively expressed small RNAs through ribonucleoproteins (U-snRNPs) dictate precisely where splicing occurs and hence are crucial to mRNA splicing[314-317]. The core U-snRNP complexes together with 200–300 proteins constitute the spliceosome. About ninety-nine per cent of introns are handled by the major U2-type spliceosome which consists of five snRNAs, *U1*, *U2*, *U4*, *U5* and *U6* and other

related proteins[318, 319]. The minor U12-type spliceosome is similar to the major U2 spliceosome complex, although it comprises four distinct snRNA, *U11*, *U12*, *U4atac*, *U6atac* and splices a minor class of introns (<800)[320]. Both the classes of spliceosomes share the *U5* snRNA

Of the hundreds of spliceosome-associated proteins majority are believed to be common between the major and minor classes of spliceosomes, and a few related to U11 and U12 snRNPs are associated with only the minor spliceosome[321, 322]. Spliceosome formed by the U-snRNAs and associated proteins assembles via U-snRNP intermediate complexes as detailed in Figure 3.1.

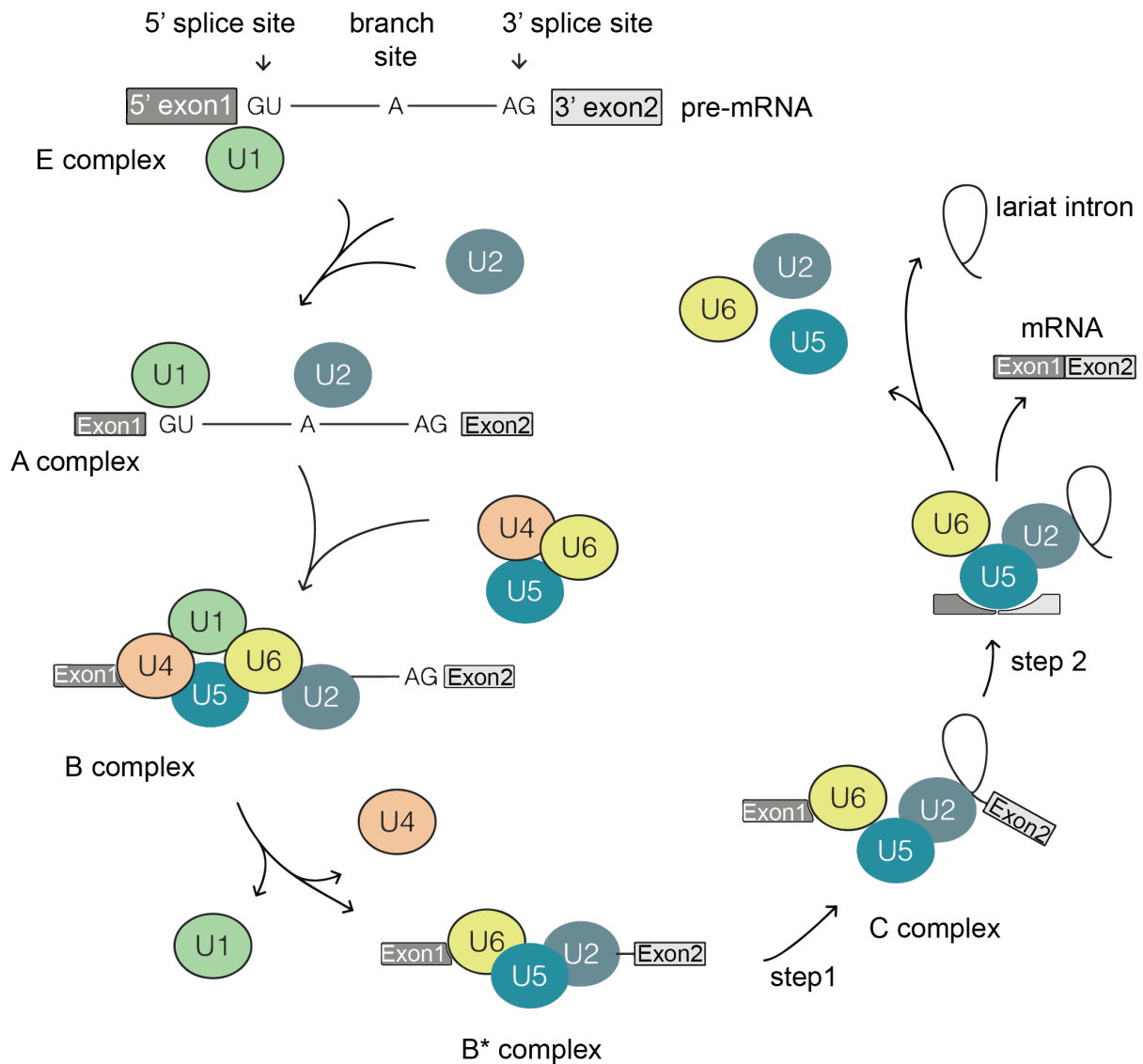


Figure 3.1: Diagram depicting spliceosome assembly on nascent mRNA. The spliceosome is assembled from the U-snRNP components in stepwise fashion on the nascent mRNA. The various U-snRNPs involved in spliceosome assembly and splicing cycle vary in their sequence of appearance and differ in their composition. There is an ordered interaction of spliceosomal snRNPs in the splicing cycle.

The schematic for stepwise spliceosome assembly and catalysis of splicing is explained in Figure 3.1. Spliceosome assembly is initiated by the U1 snRNP attachment to the 5' guanine-uracil splice site of an exon to form the early complex E. At the same time, the U2 snRNPs assemble at the branch site in an ATP-dependent reaction to form complex A. The precatalytic spliceosome B complex then forms by incorporation of the U5 and U4–U6 snRNPs. The catalytic activation of complex B requires several additional rearrangements, including the separation of U1 snRNP from the 5' splice site and unravelling of the base-pair interaction of the *U4/U6* snRNA duplex resulting in *U4* snRNA loss. The catalytically active complex B* is then produced by a new base-pairing arrangement between *U6* and *U2* snRNAs[314]. This reconfiguration makes the 2' hydroxyl group of the branch point adenosine to form a phosphodiester bond at 5' splice site resulting in the cleaving of the 5' spliced site (step 1), generating exon 1 and a looped structure called lariat intron[323]. The spliceosome then undergoes further remodelling by aligning the exons with the conserved *U5* snRNA loop for the second trans-esterification reaction (step2) by which the two exons are covalently linked[324]. Subsequently the spliced mature mRNA is released, the lariat intron is degraded and the spliceosome disassembled and its components recycled [325]. The dynamic nature spliceosome assembly, low levels of the spliceosome and associated machinery have hindered structural studies of snRNPs or intact spliceosomes. At least four spliceosomal complexes are observed in vitro: E, A, B and C[326, 327]. These spliceosomal complexes, as denoted in Figure 3.1, vary in their composition as well as sequential order of appearance. In general spliceosomal remodelling events by the intermediate complexes are characterized by ATP dependent stable interactions of the snRNAs with specific proteins which results in the formation of intermediate spliceosomal ribonucleoprotein complexes (snRNPs)[328]. These intermediate complexes guided by U-snRNAs recognize the splice sites present in pre-mRNA sequences committing the pre-mRNA to the splicing cycle along with numerous other splicing factors[318, 329, 330]. Of the various splicing factors, the SR family of splicing factors play a key role in regulating spliceosome function[331].

SR proteins comprises several phylogenetically conserved proteins are distinguished by the presence of a RNA recognition motif (RRM) and an arginine and serine rich RS domain which enable these splicing factors to concurrently interact with RNA and other protein components [332]. In addition to playing significant regulatory role in mRNA splicing, they also participate in post-splicing operations, including nuclear export of mature mRNA, nonsense-mediated decay to remove defective mRNA and mRNA translation[333]. The SR proteins are typically located in nuclear compartments called speckles, which are sites of active genes transcription and co-transcriptional splicing as well as regions of pre-mRNA splicing factors storage and reassembly[334, 335]. The SR proteins also help in “exon definition” by recruiting the splicing machinery to exon splicing enhancer sites, which function as a buffer against exon skipping during gene splicing. As a result, the linear order of exons from 5' to 3' in spliced mRNAs is guaranteed[336, 337]. The splicing factor SRSF2 (also called SC-35) is required for mammalian spliceosome assembly[338]. SRSF2 is an essential protein involved in forming the first ATP-dependent splicing complex. SRSF2 binds to the 5' and 3'-splice sites on the nascent mRNA and interacts during the spliceosome assembly with spliceosomal components at the 5'- and 3'-splice sites. It mediates the ATP-dependent pre-mRNA interactions with U1 and U2 snRNPs [339] and mediates complex interactions of splicing elements bound to the 5' and 3' splice sites via the Arginine/Serine-rich domains to bridging the 5' and 3' splice site binding components, U1 snRNP and U2 auxiliary factor (U2AF)[331].

Studies have shown that the expression of the mutant Htt induces mis-splicing in HD patient brains[303, 306, 340] and these changes have been recapitulated in the HD cellular and animal models[305, 340, 341]. The associated toxicity is primarily attributed to the trapping of splicing factors like serine/arginine-rich splicing factor 6(SRSF6)[304], Muscleblind Like Splicing Regulator (MBNL)[342] and U2AF2[343] by the insoluble aggregates. Further RNA binding proteins like MBNL are also involved in RAN translation and RAN related pathology[344, 345] and accumulation of RAN translation products[346] have been documented in HD aggregates.

Reports also indicate that mHtt and wild-type Htt differ in their interaction with many splicing factors, indicating that mHtt through aberrant protein-protein interactions can interfere with the splicing regulatory networks by associating with other proteins[347]. Further aberrant mis-splicing leading to the generation of small N-terminal fragments of mHtt exon 1 has also been demonstrated to cause HD pathogenesis[348]. These studies collectively show that splicing alterations may, at least in part, underlie HD pathogenesis.

One emergent hypothesis is that splicing defects in HD are mainly caused due to sequestration of splicing factors associated with aggregated state of mutant Htt although aberrant protein-protein interaction of soluble mHtt disrupting splicing networks cannot be ruled out. In this chapter, we look at the post-transcriptional mRNA splicing changes associated with Httex1 aggregation states and ascertain how altered transcripts of individual genes differ in their expression across aggregation states. We were further able to identify transcript isoforms at the peptide level via whole-cell proteomics in cells with different Httex1 aggregation states. Using neuronal cells with and without Httex1 aggregates as well as HD iPSC patient and gene corrected neuronal precursors, we elaborate the dysregulation in spliceosomal associated non-coding snRNAs and associated splicing machinery in HD.

3.3. Materials and methods

3.3.1. DNA constructs

Httex1 of both 25Q or 72Q sequence lengths were cloned into pT-REx vectors (Life Technologies) as described previously[349, 350].

3.3.2. Cell culture

Mouse neuronal Neuro2a cells, originally obtained from the American Type Culture Collection, was used in the study. These cells were grown in OptiMEM medium (Life Technologies) supplemented with 10% v/v fetal calf serum, 1 mM glutamine, 200 U/ml penicillin and 200 µg/ml streptomycin in a 5% v/v atmospheric CO₂ humidified incubator at 37°C. Using the Lipofectamine 2000 reagent (Life Technologies), Neuro2a cells were transiently transfected with DNA constructs as per manufacturer's recommendations.

Table 3.1: List of human iPSC HD neuronal lines and their corresponding CAG sizes

hiPSCs	CAG size
NDD00081	27, 18
NDD00083	30, 19
NDD00086	45, 19
NDD00088	45, 19
NDD00091	65, 19
NDD00092	65, 19
NDD00096	81, 27
NDD00097	81, 19

Human isogenic induced pluripotent neuronal stem cells (hiPSCs) cell lines were a gift from the Pouladi lab and generated as described in Xiaohong Xu et al., 2017[351]. Human iPSC HD pathogenic and gene corrected neuronal precursor cell lines and their CAG repeat sizes are included in Table 3.1. The hiPSC lines obtained were induced into a neuronal lineage and characterised by immunofluorescence (Pax6, Nestin) and qPCR (Pax6, Nestin, Sox1) to ascertain pluripotency and for normal karyotype at the Pouladi lab.

Human iPSC was induced into precursor neural stem cells by a previously established and published protocols[352]. In brief, human iPSCs(hiPSCs) were maintained in N2B27 media made up of DMEM-F12/Neural Basal medium in 1:1 ratio with addition of N2 supplement one per cent, B27 supplement – two per cent, 10 ng/ml human LIF, and 5 µg/ml BSA containing 3 µM CHIR99021 (Tocris), 2 µM SB431542 (Sigma). The hiPSCs at about 40-50% confluence was split into 1:2 ratio using Accutase (Sigma) and plated onto Matrigel-coated plates. RHO/ROCK pathway inhibitor -Y27632 (Stem Cell Technologies) was used during plating of cells to prevent apoptotic cell death. hiPSCs were maintained at 37°C with 5% v/v atmospheric CO₂ at adequate humidity. Cells at passage 5 were used for experiments.

3.3.3 RNA sequence processing and alternative splicing analysis

Whole transcriptomic data generated in Chapter 2 with four biological replicates of cells enriched with soluble wild-type Htt (25Q-ni) or with soluble mHtt (72Q-ni) and aggregated mHtt (72Q-i) was reanalysed in this study. Paired end 50bp paired were assigned and mapped to the mouse reference genome(GRCm38/mm10) using STAR (2.5.3) aligner[353]. The average library size of the mapped samples was 15 million reads with at least one reported alignment (~90% mapping rate). Differential splicing analysis of the sequencing data was conducted using replicate multivariate analysis of transcript splicing (rMATS) (Version 4.0.1)[354] software to individually quantify and analyse differences in alternative splicing events (ASE) in Neuro2a cells expressing either the 72Q-ni- Httex1, 72Q-i-Httex1 or 25Q-ni-Httex1 wild-type control proteins. rMATS takes sorted- Binary Alignment Map (BAM) files and defines splicing events that are differentially represented between two conditions. The sorted-BAM files of cells enriched with different aggregation states 72Q-ni, 72Q-i, 25Q-ni wild-type controls were used as sample inputs. Additionally, an Ensemble gene transfer format (GRC38 version 89)[355] was provided for mouse transcript annotation to be used for rMATS analysis. The ASE events in different aggregation states was identified using default thresholds in the rMATS analysis pipeline. The rMATS ASE analysis classified splicing

events into five classes by defaults as skipped exon (SE), retained intron (RI), alternative 5' splice sites (A5SS), alternative 3' splice sites (A3SS), and mutually exclusive exons (MXE) as described in Figure 3.4B[354]. A standard parameter to define the degree or magnitude of alternative splicing is the exon-inclusion ratio, often referred to as percent spliced in (PSI)[356]. The PSI value is calculated for the exon and its two flanking exon neighbours and is an indicator of isoform abundance. PSI is the relative abundance ratio of all isoforms containing a particular exon to the relative abundance of all the gene isoforms containing the said exon. It is calculated by the formula

$$PSI = \frac{S/LS}{\left(\frac{S}{LS}\right) + \left(\frac{R}{LR}\right)}$$

$$\Delta PSI = PSI(Test) - PSI(Control)$$

Where S indicates the number of reads corresponding to the spliced transcript, R indicates the number of reads corresponding to the reference transcript. LS and LR are the lengths of splicing transcript and reference transcript respectively. In order to return the probability of differential splicing and the degree of alternative splicing for each ASE, individual values are then compared between groups of samples to obtain ΔPSI . The value of ΔPSI is the difference in the mean percent-spliced-in between test and control group. Significant splicing change defined as $FDR < 0.05$ were used for functional enrichment analysis. Skipped exons were mainly considered for downstream analysis unless mentioned otherwise. Genomic visualisation of global splicing events was done using MASER[357] in RStudio (Version 1.1.447) /R (Version 3.6.1).

3.3.4. Enrichment of GO Category and KEGG Pathway

The Gene ontology (GO) and pathway enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID)[358] version 6.8 (<https://david.ncifcrf.gov/>) tool. The genes with significant differential ASE were used in the analysis. False discovery rate (FDR) correction was implemented to account for and obtain adjusted significance values. Using differential splicing profiles, we identified and sorted ASEs out into genes that were enriched in GO categories and KEGG pathways.

3.3.5. Network Analyst - Minimal interaction networks

The post-transcription gene splicing signatures were mapped using the network analyst toolkit[359, 360] to known protein: protein interaction networks. Essentially all spliced genes in the splicing signature were connected to an experimentally verified protein: protein interaction network with STRING v.10 algorithm[361].

Subsequently the network was pruned so that only proteins that connect the input spliced genes were retained. The most centrally linked genes in the network play a central role in regulating the network, creating a hub that connected the rest of the network. The hubs were formed by the most linked genes remaining in the network that had a degree of centrality of >5.

3.3.6. Collection of cells by Pulse Shape Analysis

Using a BD FACS ARIA III cell sorter mouse neuronal Neuro2a cells were fluorescence-activated cell sorted 48 hours post transfection and used for whole cell proteomics. The cells were basically gated by SSC and FSC gates for the live cell population and then using PulSA- 'ni' and 'i' gates, as described[349]. Cells were further gated to the same median GFP intensity of 4,000 RFU, to normalise and match expression by adjusting the width of the expression window across replicates. For Quantitative PCR experiments smaller median GFP window of 2,500 RFU was used to collect 50,000 cells in 4 replicate sets. The Httex1 constructs of wild-type 25Q and pathogenic 72Q were transfected and sorted in parallel across five days and used as five matched replicates. For each replicate sample of 25Q-ni, 72Q-ni and 72Q-i, a minimum of 200,000 cells were sorted and collected for whole cell proteomics. Samples were sorted into PBS for proteomics and TRI-reagent (Zymo Research, Irvine, CA) for RNA extraction. Cells were kept on ice for all stages of planning and handling of the sorting process. Cells were pelleted at 120 g for 6 minutes, and the resultant pellet was snap frozen using liquid nitrogen to be stored at - 80 °C. An identical sample collection was performed for imaging and analysed with confocal microscopy[349].

3.3.7. Whole cell proteomics

Preparing whole cell protein extracts

The stored snap frozen cell pellets were thawed then lysed using a cell lysis reagent - RIPA lysis buffer (25mM Tris-HCl, pH 7.4, 0.1% SDS, 1% NP40, 1% Sodium deoxycholate, 150mM NaCl, 1x complete mini-protease inhibitor cocktail; Roche). The pellets were treated with around 100µL of RIPA lysis buffer and incubated for 30 minutes. The protein concentration was subsequently determined by using the Pierce™ BCA Protein Assay (Thermo Fisher Scientific, USA), as directed by the manufacturer in each sample. To remove any residual detergent that can interfere with subsequent MS analysis, equal measured quantities of protein were combined with six-fold volumes of acetone (at -20°C) and the precipitation was allowed to proceed overnight. Next day, the samples were spun at 13,000 rpm at 4°C for 10 minutes and acetone removed without the protein pellet being disturbed. The pellets were again rinsed with pre-chilled acetone, and were allowed to dry fully for ten minutes. The pellet was dissolved in 100 µL 0.1M TEAB, vortexed and sonicated thrice for 30 seconds to aid solubilization.

Protein digestion

The protein cysteines were reduced using TCEP reducing agent (tris(2-carboxyethyl) phosphine) (10 mM, pH 8.0), at 60°C for 45 minutes accompanied by iodoacetamide (IAA) (10 mM) alkylation in the dark for 15 minutes at room temperature. The mixture was then treated with (25mM TEAB, pH 8) following which sequencing-grade trypsin (Promega) with 1:50 ratio of enzyme to substrate was added. The reaction was taken to completion during overnight incubation at 37 °C, and was terminated by the addition of trifluoroacetic acid (TFA) to one per cent.

Desalting of the acidified tryptic peptides was performed by means of Oasis HLB 1 cc Vac SPE Cartridge, (Waters Corp, USA) prewashed with a mixture of acetonitrile (ACN) with trifluoroacetic acid (TFA). 80 per cent v/v of ACN with 0.1 per cent v/v TFA was used as both prewash and elution solution. Subsequent washes were performed with 1.2 ml of TFA (0.1% v/v) thrice utilising a vacuum manifold. The samples were then loaded onto the SPE cartridge and washed with 1.5 ml of TFA (0.1% v/v). The samples were then eluted with 0.8 ml of ACN and TFA elution solution. The eluent was collected in 1.5 ml microcentrifuge tubes for lyophilization through freeze drying (Virtis, SP Scientific) and reconstituted using 100 µl distilled water. Quantification of peptides was done in triplicates using a micro-BCA assay (Thermo Fischer Scientific, USA) with bovine serum albumin as standard before labelling.

Peptide labelling

For dimethyl labelling, 25 µg of peptides was reconstituted to 100 µL with TEAB (100 mM), before dimethyl labelling. Peptides were differentially labelled by adding 4 µL of 4 per cent v/v formaldehyde-CH₂O (light label) or CD₂O (Medium label) (Sigma, USA) with 4 µL of sodium cyanoborohydride (600 mM) (NaCNBH₃) (Sigma, USA). On the other hand, 4 µL of 4 per cent v/v D₂-formaldehyde-¹³CD₂O (reconstituted from 20 per cent solution, 98 per cent D₂; Cambridge Isotope Laboratories, USA) and 4 µL of sodium cyanoborodeuteride NaCNBD₃ (600 mM) in 1 M NaOH (98.7 per cent D₂; CDN Isotopes, Canada) was used to heavy-label peptides. The peptide mixtures were incubated at room temperature for 1 hour while mixing to confirm that > 99% of the peptides were labelled. Quenching was performed by adding 16 µL of one per cent ammonium hydroxide and by the addition of 8 µL formic acid. The light dimethyl label, medium dimethyl label and heavy dimethyl label were then mixed up in equal volumes. Five biological replicates for each sample which were labelled as follows: light label for 72Qni, Medium label for 72Qi and heavy label for 25Qni. Peptide mixtures were then combined and were run as a single experiment on LC-MS/MS spectrometer.

Nano ESI-LC-MS/MS analysis

Analysis of the peptide mixtures were performed using liquid chromatography-nano electrospray ionization-tandem mass spectrometry (LC-nESI-MS/MS) using the Orbitrap Lumos mass spectrometer (Thermo Scientific, USA) equipped with nanoflow reversed-phase-HPLC (Ultimate 3000 RSLC, Dionex). The following columns - Acclaim Pepmap nano-trap column (Dionex – C18, 100 Å, 75 µm × 2 cm) and Acclaim Pepmap RSLC analytical column (Dionex – C18, 100 Å, 75 µm × 50 cm) were used in the nano-HPLC system.

1 µg of the whole proteome or 0.135 µg of peptides L from the peptide mix were loaded into the enrichment (trap) column for every LC-MS/MS experiment. A constant isocratic flow of 5 µL/minute of three per cent CH₃CN with 0.1 per cent formic acid for six minutes was maintained before the switching the enrichment column in-line with the analytical column. The elution in the LC conducted was done using five per cent DMSO with 0.1 per cent v/v formic acid (solvent A) and 100 per cent CH₃CN with five per cent DMSO and 0.1 per cent v/v formic acid. In the first 95 minutes, the gradients switched from three per cent B to 20 per cent B followed by 20 per cent to 40 per cent B, the next ten minutes. The gradient increased

to 80 per cent B subsequently in five minutes and maintained the same in the last five minutes before the switch to three per cent B gradient for ten minutes for equilibration before the next LC run. Data-dependant acquisition was undertaken in a positive ionization mode with MS spectra scanned from 400-1500 m/z, with a resolution set at 120,000 with an automatic gain control target value of 5e5. The spectrometer was operated with the spray voltage set at 1.9 kV and the source temperature set at 275 °C. DMSO lockmass of 401.92272 was used. The instrument was run in “top speed” acquisition method mode with a cycle time of 3 seconds, considering only the most intense charge state of the precursor whereby peptide ions with charge states ≥ 2 -5 were isolated with a precursor isolation window of 1.6 m/z and fragmented with high energy collision mode with 30 ± 5 per cent stepped collision energy. The fragment ion spectra were acquired in Orbitrap at mass resolving power of 15000 resolution. Dynamic exclusion had a duration of 30 seconds.

Data analysis

The data analysis was performed using the Proteome Discoverer (version 2.1.1.21; Thermo Fisher Science, USA) software using the search engine Mascot (Matrix Science version 2.4.1). In combination with common contaminant proteins, the GFP sequence (UniProt ID: P42212), the SwissProt Mus Musculus database (version 2016-07; 16794 proteins) and a custom database of peptides generated from RNA transcripts in the sequencing data sequences were added to create a comprehensive database of different peptide isoforms against which the MS data was filtered. The custom database of peptides was generated from RNA transcripts in the sequencing data by in-silico translation of nucleic acid sequences (the 3 forward reading frames) to their corresponding protein sequences using the Transeq method in the EMBOSS software suite.

For protein identification, the search was conducted with the precursor MS mode tolerance of 10 ppm, fragment MS/MS mode tolerance of 0.6 Da, with enzyme specificity determined to include those fully cleaved by trypsin. The minimum peptide length required was six with the maximum missed cleavage sites permitted set to two. The variable modifications allowed were: Oxidation (M), Dimethylation (K), Dimethylation (N-Term), Dimethylation: 2H(4) (K), Dimethylation 2H(4) (N-term), 2H(6)13C(2) Dimethylation (K), 2H(6)13C(2) Dimethylation (N-term) (Variable), N-term Acetylation (Protein N-term) and Carbamidomethyl (C) was set as a fixed modification. The FDR maximum of estimation was set to one per cent at the protein identification level and 0.5 per cent at the peptide identification level, while actual was 0.1 per cent for each replicate at both the protein and

peptide levels. At least one unique peptide was required for protein report in all three replicates with the common contaminant, Keratin, excluded from the dataset.

Proteome Discoverer v.2.1.0.81 was used for peptide quantification using the precursor ions quantifier node. Pairs of dimethyl labelled peptide among two comparisons of light, medium or heavy, were developed with a mass precision of two parts per million and a signal to noise threshold of three. For peptide identification we applied 0.6 minutes of retention time tolerance for isotope pattern multiplets and allowed spectra with two single peaks or missing channels to be quantified. For each replicate protein abundances were calculated by determining the mean abundance of the top three unique peptide abundances used for quantitation. Peptide group abundances and protein abundances were normalized using Peptide and Protein Quantifier node in Proteome discoverer (v 2.1) and imported into Perseus software (v 1.6.5.0).

Protein log₂ abundances were computed and samples replicates were grouped. Statistical analysis was performed using GraphPad Prism software (version 8.0.2.263) (Graphpad Software Inc, USA). Pairwise comparisons of proteomes were done using paired two-sample t-test to determine significant differences in protein abundances on the assumption that data obtained is normally distributed. Protein log₂ abundances were computed and samples replicates were grouped. Statistical analysis was performed using GraphPad Prism software (version 8.0.2.263) (Graphpad Software Inc, USA). Pairwise comparisons of proteomes were done using paired two-sample t-test to determine significant differences in protein abundances on the assumption that data obtained is normally distributed. Though it is not ideal to automatically assume that the data are normally distributed, non-parametric tests do not have enough sensitivity for comparisons with few replicates and hence we applied parametric paired t-tests.

Proteogenomic analysis for identification of spliced peptide isoforms was performed using Splicify[362]. The differentially expressed splice isoform analysis with rMATS identified splicing isoforms for further validation. These splice variants were translated into a database of protein isoform (FASTA) sequences by 3-frame translation. This custom database combined with mouse Uniprot protein (GRCm38)[363] database was further used with MaxQuant[364], to define MS/MS spectra of PulSA sorted proteomic samples. The extraction and quantification of isoform-specific peptides was enabled by utilising the splice isoform generated from whole transcriptomic analysis. Differentially expressed splice protein isoforms was calculated after. The code for the analysis is accessible at

<https://github.com/NKI-TGO/SPLICIFY>. The analysis was performed using the default parameters.

3.3.8. Immunofluorescence

In an eight well Ibidi chamber 1×10^5 cells/ well were plated 24 hours before transfection. Cells were subsequently transfected with 72Q-Httex1-Emerald, 25Q-Httex1-Emerald or Emerald only using Lipofectamine 2000 reagent (Life Technologies). The transfection media was replaced 6 hours post transfection. After 48 hours, the cells were fixed at room temperature by four percent w/v of paraformaldehyde for ten minutes and permeabilized with 0.1% TritonX100 for 30 minutes and blocked in 5% BSA for 1 hour. Primary antibody anti-SC35 (1:750 dilution) was incubated at room temperature for 1 hour. Post wash with PBS, secondary antibody (1:1000 anti-mouse Cy5) was added for 30 minutes followed by nuclei counterstaining with Hoechst (1:250) for 25 minutes. Post fixation all the steps were conducted at room temperature. The cells were imaged using SP5 confocal microscope. The mean nuclear fluorescence intensity was measured with cell profiler software (v 2.2.1)[365]. The hiPSC HD cells were fixed, stained as mentioned above. The human isogenic neuronal cells were imaged using the SP8 confocal microscope and mean nuclear fluorescence intensity was measured with image processing tool Fiji (Version 1.53)[366]. The results were analysed using Prism software.

3.3.9. Quantitative real-time PCR

For analysis of small nuclear RNA by quantitative PCR, the 72Q and 25Q plasmid DNA were transfected into Neuro2a cells for 48 hours, before sorting by flowcytometry- PulSA into TRI-reagent (Zymo Research, USA) and stored in -80°C before proceeding for RNA extraction. hiPSC cell lines of different CAG sizes as in Table 3.1 of passage five were used to validate the snRNA levels. Total RNA was extracted using the Direct Zol RNA extraction kit (Zymo Research, USA) and RNA concentration measured using Qubit fluorometer (Qubit RNA BR Assay Kit). 300-500ng of RNA was used for cDNA conversion. SuperScript III first- strand synthesis system (Invitrogen) was used for cDNA conversion with oligo(dT) primers as per to the protocols provided. RNaseH treatment was performed to remove any residual RNA.

Table 3.2: Real-time PCR reactions custom master mix preparation

Master mix Reagent	X 1000 sample
Water	3276 μ l
MgCl ₂ (50mM)	400 μ l
ThermoPol reaction buffer (10X with 20mM MgCl ₂)	1000 μ l
High Resolution Melting Master	250 μ l
dNTPs (25 mM)	64 μ l
Immolase DNA polymerase (20 Units/ μ l)	10 μ l
Total:	5000 μl
Reagent	X 1000 sample
Master mix reagent	5000 μ l
F +R primer combined (50mM)	80 μ l
Water	2920 μ l
Total:	8000 μl

Plus 2 ml of diluted (1: 10) cDNA = total volume per well in 384 well plate = 8 + 2 = 10 ml. Samples were prepared in a total volume of 10 μ l per reaction with a final concentration of 10 μ M of each primer (Table 3.3). The reaction was run in Lightcycler 480 System Real-Time PCR system (Roche). The $\Delta\Delta C_t$ values were calculated to determine fold changes in gene expression.

Table 3.3: List of primers used for the study

Primer	Primer Sequence
U1 snRNA Forward	GATACCATGATCACGAAGGTGGTT
U1 snRNA Reverse	CACAAATTATGCAGTCGAGTTTCC
U2 snRNA Forward	TTTGGCTAAGATCAAGTGTAGTATCTGTTC
U2 snRNA Reverse	AATCCATTTAATATATTGTCCTCGGATAGA
U4 snRNA Forward	GCGCGATTATTGCTAATTGAAA
U4 snRNA Reverse	AAAAATTGCCAATGCCGACTA
U5 snRNA Forward	GGTTTCTCTTCAGATCGCATAAATC
U5 snRNA Reverse	CTCAAAAAATTGGGTAAAGACTCAGA
U6 snRNA Forward	GCTTCGGCAGCACATATACTAAAAT
U6 snRNA Reverse	ACGAATTTGCGTGTCATCCTT
U11 snRNA Forward	GTGCGGAATCGACATCAAGAG
U11 snRNA Reverse	CGCCGGGACCAACGAT
U12 snRNA Forward	AACTTATGAGTAAGGAAAATAACGATTCTG
U12 snRNA Reverse	CGACCTTTACCCGCTCAAAA
U4atac snRNA Forward	GCGCATAGTGAGGGCAGTACT
U4atac snRNA Reverse	GCACCAAAATAAAGCAAAAGCTCTA
U6atac snRNA Forward	AGGTTAGCACTCCCCTTGACAA
U6atac snRNA Reverse	TGGCAATGCCTTAACCGTATG
5.8S rRNA Forward	CGGCTCGTGCGTCGAT
5.8S rRNA Reverse	CCGCAAGTGCGTTCGAA
5S rRNA Forward	CGGCCATACCACCCTGAAC
5S rRNA Reverse	GCGGTCTCCCATCCAAGTAC

3.4 Results and Discussion

3.4.1. Soluble mutant Httex1 affects spliceosome pathway and related process

We have previously shown in chapter 2 that soluble mutant Httex1 induces toxicity in cells. Briefly, expression levels of the pathogenic soluble 72Qni, aggregated 72Qi and the wild type 25Qni cell populations were matched using flowcytometry using the PuLSA toolkit. This expression matched enriched cell populations were subject to RNA sequencing and cell survival of individual populations was conducted by live cell imaging. The cells containing soluble mutant Httex1 showed significantly lower survival compared to aggregated and wildtype Httex1. Interestingly, one of the key pathways altered in 72Q-ni as compared to 25Q-ni was the spliceosome pathway (FDR = 0.0145, Fold enrichment = 2.27). The gene ontology analysis of 72Q-ni as compared to 25Q-ni also revealed alterations in mRNA processing and RNA splicing (FDR <0.0001)[300]. The expression of spliceosome pathway genes was upregulated in mHtt 72Q-ni as compared to wild type. We performed network analysis of the spliceosomal pathway genes identified in the study which showed that the SR family of splicing factors are enriched in 72Q-ni group compared to 25Q-ni (FDR <0.0001, Fold enrichment = 4.13) (Figure 3.2).

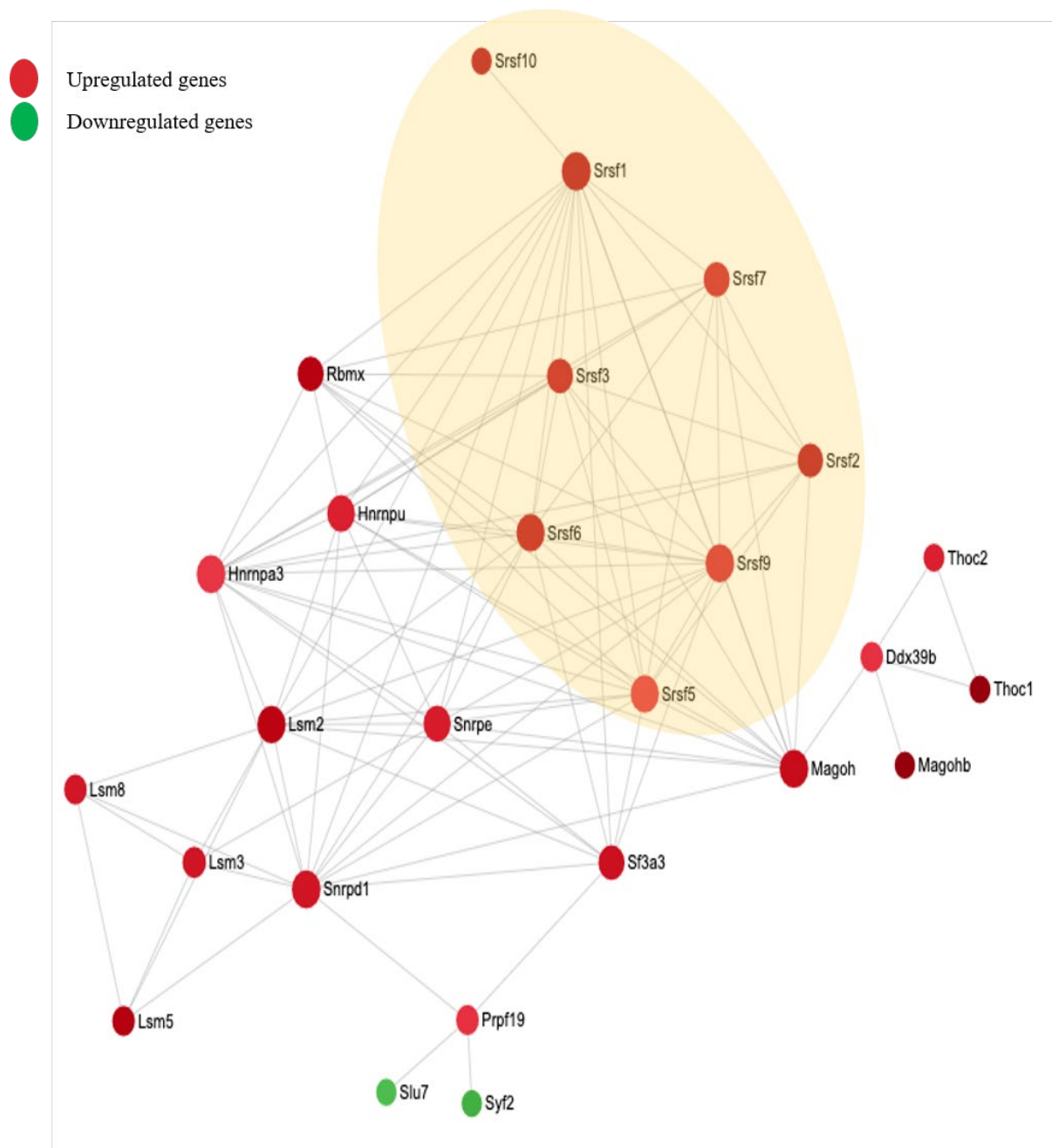


Figure 3.2: Spliceosome function is altered in mHtt-72Q-ni compared to 25Q-ni. Protein-protein interaction zero-order network of the spliceosome pathway genes plotted according to their gene expression fold change in cells enriched with soluble mutant 72Q-ni as compared to 25Q-ni. Validated protein-protein interactions of genes matching those enriched in the dataset as shown by STRING algorithm (v10). The network was created using Network analyst to include only the genes which show significant differential expression. The enriched SR family of splicing factors are highlighted in yellow.

Pre-mRNA splicing is spatially and temporally regulated by splicing factors such as SR proteins that are concentrated in nuclear speckles[367]. Since SR splicing factors play important role in initiation of splicing and assembly of the spliceosome complex[331, 338], we quantified SRSF2 levels in cells of different Htt aggregation states. Neuro2a cells from a single experimental culture were transfected with Httex1 72Q, 25Q and emerald only control for 48 hours. Subsequently the cells were fixed and immunostained against the key activated splicing factor SRSF2 and analysis by confocal microscopy was performed. 72Q Httex1 formed aggregates as seen in Figure 3.3A. The analysis of images was done using cell profiler software (v 2.2.1). The mean fluorescence nuclear intensity of SRSF2 immunostained nuclear speckles was quantified (mean $\pm$ SD), and $n \geq 90$ cells were analysed in each group. A higher SRSF2 activity is seen in cells with diffuse 72Q-ni and aggregated 72Q-i as compared to 25Q-ni. This result indicates the spliceosome related activity is higher in expanded 72Q as compared to the wildtype 25Q. Further, there was no difference in mean fluorescence intensity in cells with and without aggregates (Figure 3.3B).

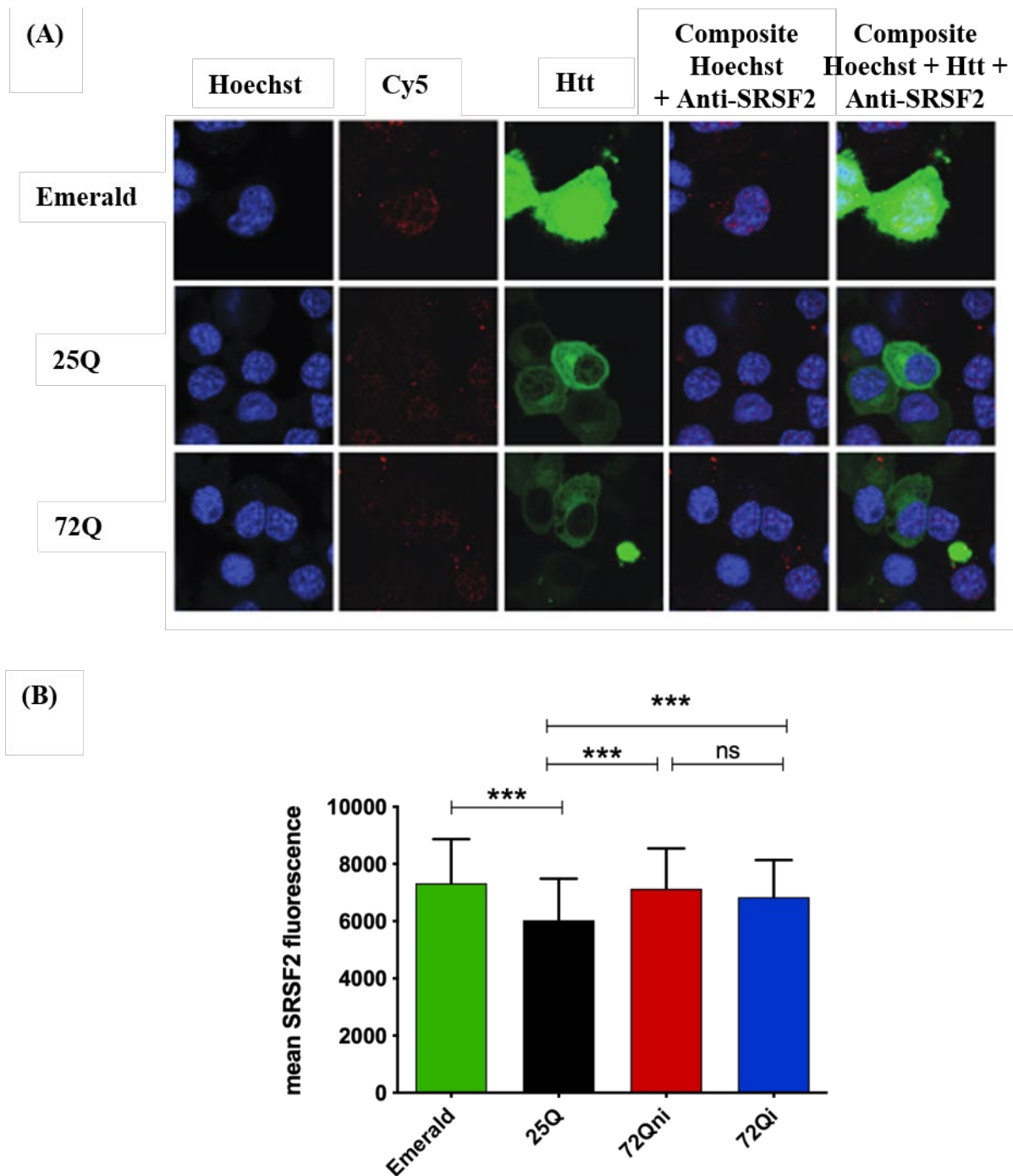


Figure 3.3: SRSF2 activity is increased cells in Neur2a cells 72Q aggregation states. (A) Neuro2a cells from one experimental culture were transfected with Httex1 72Q, 25Q and emerald only control for 48 hours and fixed and immuno-stained with anti-SRSF2 antibody with Cy5 secondary antibody and Hoechst as nuclear marker and viewed on confocal microscope. The analysis of images was done using cell profiler software (v 2.2.1). **(B)** Confocal images from (A) were used to determine the mean fluorescent intensity of SRSF2 in the nucleus (mean $\pm$ SD), $n \geq 90$ cells were analyzed in each group.

A similar observation was seen in isogenic HD hiPSCs cell lines of different polyQ lengths (65 and 81 CAG repeats) as compared to gene corrected hiPSCs with 27 CAG repeats (Figure 3.4A). There was significant difference in mean fluorescence intensity in hiPSCs with pathological expansions compared to gene corrected hiPSCs with normal polyQ length (Figure 3.4B). This further validated our results of altered spliceosome activity in HD. As hiPSCs do not form aggregates in culture[351], we can specifically attribute altered function of key spliceosome protein SRSF2 to the soluble mutant Htt protein in these human patient lines. Although our data did not reveal a CAG length dependent increase in SRSF2 activity additional assays with different CAG lengths of 45, 65 and 81 isogenic lines would probably help to tease out if there is CAG length dependent increase in SRSF2 activity in the nucleus.

It has been previously shown SRSF2 is a crucial factor required for mammalian spliceosomes and is involved in early E complex, as well as complex A and B formation required for splicing (Figure 3.1)[368], [339]. SRSF2 in the nucleus is localized to discrete regions and the loss of splicing activity is correlated with a reduction in the SRSF2 levels[338]. Hence increased SRSF2 activity seen in our data might correlate with aberrant splicing activity initiated in cells containing soluble 72Q-ni and in cells which later form aggregates 72Q-i. Increased SRSF2 levels have also been associated with nonsense-mediated mRNA decay (NMD), a quality control mechanism that degrades mRNAs with premature termination codons.

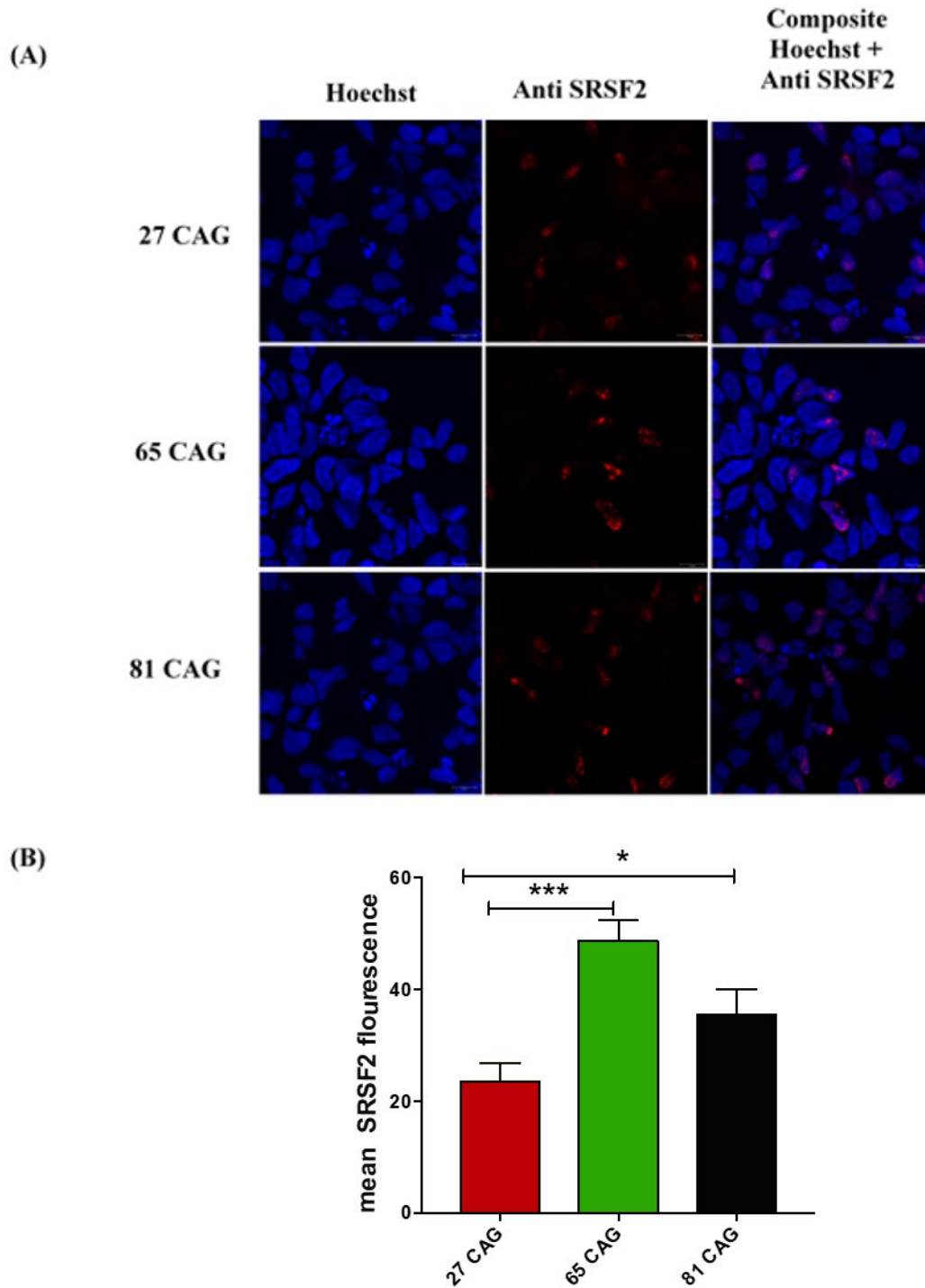


Figure 3.4: SRSF2 activity is increased cells in HD - hiPSCs. (A) Two clones from three different isogenic HD human iPSC cell lines were fixed and immuno-stained with anti-SRSF2 antibody. Cy5 was used as a secondary antibody and Hoechst as nuclear marker and images collected by confocal microscopy. Image analysis was performed using Fiji software (v 2.2.1). (B) Confocal images from (A) were used to determine the mean fluorescent intensity of SRSF2 in the nucleus (mean $\pm$ SE), $n \geq 40$ cells were analyzed in each group. The experiment was repeated twice.

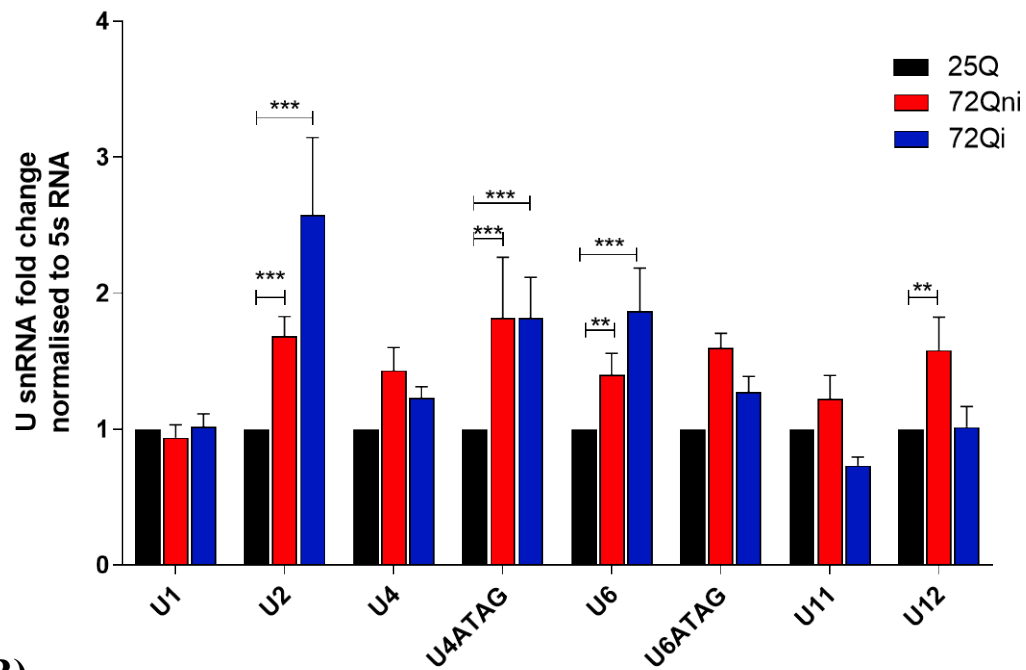
3.4.2. Spliceosome integrity is defective in HD

As spliceosome is assembled and regulated by basal snRNA and associated proteins, we were interested in investigating the non-coding RNAs involved in spliceosome formation. The snRNAs remain in steady stoichiometry inside the spliceosome and are an essential requirement for mRNA splicing. The 5' splice site and branch point upstream of the 3' splice site are initially recognised by *U1* and *U2* snRNAs, and later by *U4/U6* snRNAs. The *U5* tri-snRNP subsequently joins the spliceosome before its reconfigurations which eventually results in *U6* snRNA catalysing the final splicing reaction[369, 370]. Therefore, any perturbations in snRNA levels could lead to altered spliceosomal integrity[371]. We were intrigued to see whether snRNA levels were changed by Htt aggregation. We transfected 25Q, 72Q and Emerald only constructs in Neuro2a cells and using flowcytometry-PulSA sorted cells into soluble 72Q-ni, aggregated 72Q-i and wild-type 25Q-ni populations. Total RNA including small RNA was extracted from sorted cells and the amount of Uridine-rich small nuclear RNAs in the samples were measured. The snRNA levels of *U2*, *U4ATAG*, *U6* and *U12* were significantly upregulated in soluble mutant Httex1 as compared to 25Q-ni (Figure 3.5A). We did not see a significant upregulation of *U1*, *U4*, *U6ATAG* and *U11* snRNA when compared to 25Q-ni. A similar upregulation of *U2*, *U4ATAG* and *U6* snRNAs was seen in aggregated Httex1. Since 72Q-ni is associated with increased toxicity, the altered snRNA levels seen in 72Q-ni indicate splicing machinery defects as a potential mechanism of related toxicity. These aberrant non-coding snRNA defects are further carried over into the 72Q-i aggregated state (Figure 3.5A).

To further confirm our results, we measured levels of small nuclear RNAs in isogenic human Htt iPSC neuronal cell lines. These isogenic lines were obtained from HD patients and were gene corrected for the number of CAG repeats. Our data show that *U1*, *U2*, *U4 ATAG*, *U5*, *U6* snRNAs were significantly upregulated in 45 CAG repeat isogenic lines. *U1*, *U2*, *U4*, *U4 ATAG*, *U5*, *U12* snRNAs were significantly upregulated in 65 CAG repeat isogenic lines while 81 CAG repeat isogenic lines showed upregulation in *U1*, *U2*, *U4*, *U4 ATAG*, *U5* snRNAs. The expanded CAG hiPSCs with 45, 65, and 81 CAG repeats (Figure 3.5B), the levels of *U1*, *U2*, *U4ATAG*, and *U5* snRNA is significantly higher than in control lines with 30 CAG repeats, while *U6* snRNA is significantly higher in 45 CAG lines and *U12* snRNA is significantly higher in 65 CAG lines as compared to control lines. Overall, these results show that the CAG expansion causes mis regulation of U-snRNAs biogenesis and can cause an alteration in spliceosomal integrity. These changes in splicing machinery can result in

abnormal splicing of specific genes that might be preferentially mis spliced in response to the aberrant snRNA levels in cells.

A)



B)

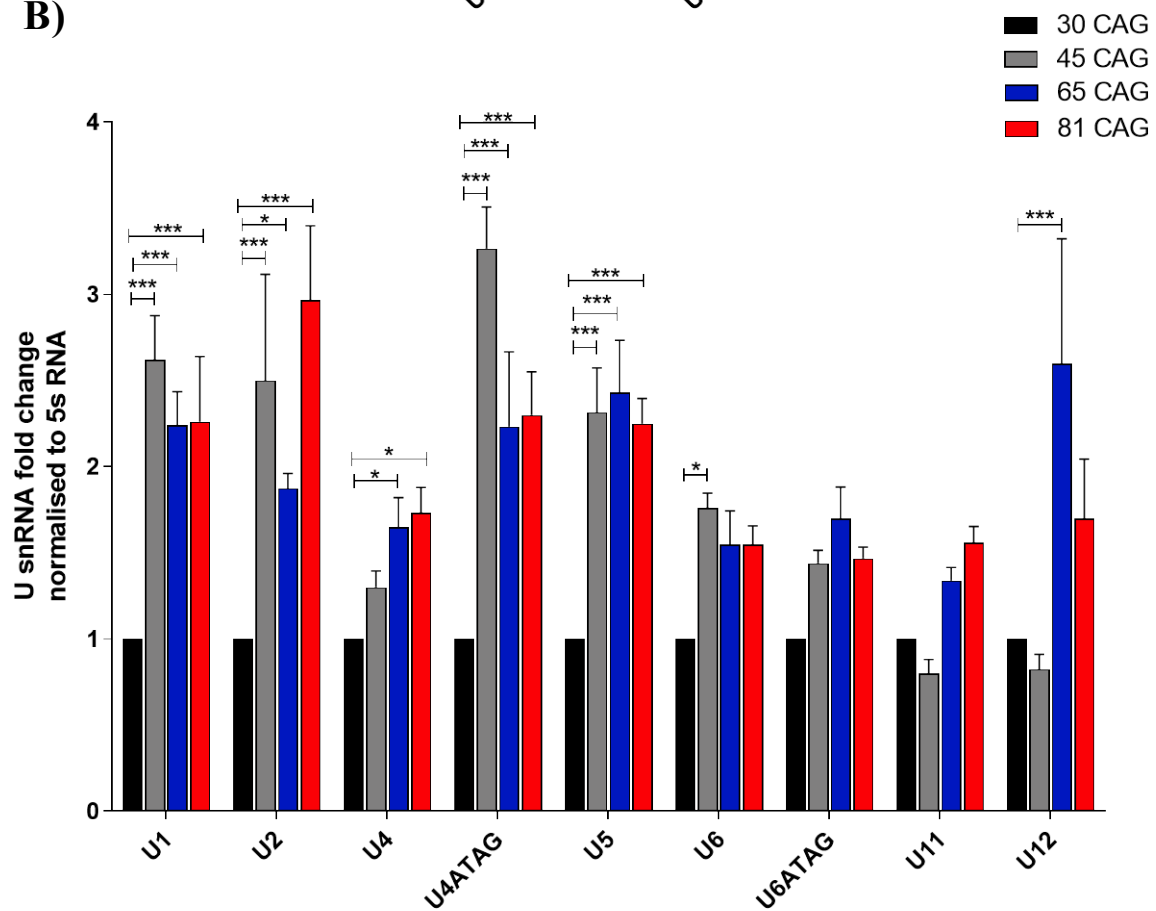


Figure 3.5: U snRNA levels are altered in soluble mHtt exon 1. (A) The RNAs were isolated from Neuro2a cells sorted by flowcytometry PulSA into 25Q-ni, 72Q-ni and 72Q-i aggregation state enriched populations and U-snRNA levels were assessed by quantitative real-time PCR. Average expression results were compiled from two independent assays with transfections performed in triplicates is plotted. (B) Mean U-snRNA expression of control and pathologic patient's hiPSC HD neuronal lines was determined from two independent experiments. Average snRNAs levels were normalised to the 5s RNA control. HD hiPSC patient lines present robust U-snRNA misregulation. (Bars: mean $\pm$ SE, * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, ANOVA with multiple testing).

3.4.3. Aberrant splicing is seen in different aggregation of Httex1

To understand the global splicing changes in relation to aggregation states we further explored spliced isoforms differences attributable to the different aggregation states (Figure 3.4A). Based on the whole transcriptome RNA sequencing data from chapter 2, we performed three comparisons for spliced isoform analysis. First, 72Q-ni against 25Q-ni, which examined how the soluble mutant Httex1 changed the isoforms of the gene. Second, 72Q-ni versus 72Q-i, which more precisely analysed how spliced isoforms were affected by the aggregation of mutant Httex1 into inclusions. Lastly, the expression of the 72Q-i versus 25Q-ni, which assessed how aggregates altered spliced isoforms compared to the wild-type baseline.

Using rMATS, we identified 115,214 splicing events across aggregation states. A total of 8,563 alternative splicing events were statistically significant in all comparisons between aggregation states in our dataset (Figure 3.6A) (FDR<0.05). Out of which 2,645 alternative splicing (AS) events were detected in the soluble 72Q-ni group and 2,855 alternative splicing events detected in aggregated 72Q-i group compared to wildtype soluble 25Q-ni. We also identified 3,063 AS events between soluble 72Q-ni compared to aggregated 72Q-i (Supplementary Table 3.1). Of the five major classes of alternative splicing events reported, the most prevalent events were skipped exon (SE), which accounted for 43.6% of total ASE detected, retained intron (RI) was next accounting for 18.6% of the total events. Exon skipping was the major class of alternative splicing seen in the dataset as compared to other AS events (Figure 3.6B). No significant difference among the different classes of splicing events was identified across aggregation states.

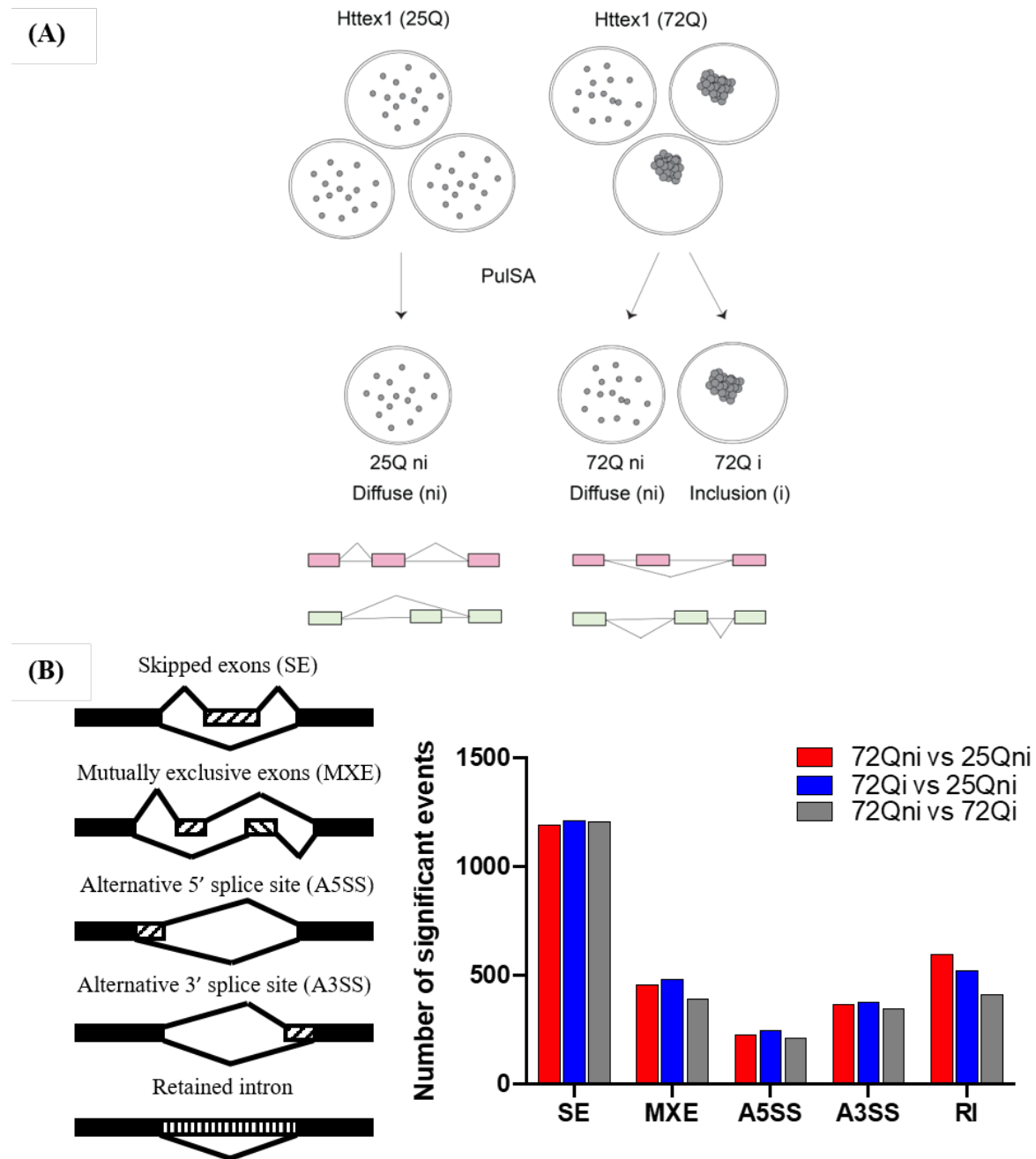


Figure 3.6: Different classes of alternate splicing events across aggregation states. (A) Cells with different aggregation states of mHtt were sorted into populations enriched with polyQ-expanded Httex1 with distinct aggregation wild-type 25Q-ni, 72Q-ni and 72Q-i states by flow cytometry- PulSA for differential isoform analysis. (B) The different types of alternative splicing events skipped exons (SE), mutually exclusive exons (MXE), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS) and retained intron (RI) detected by rMATS[354] across aggregation states 72Q-ni vs. 25Q-ni, 72Q-i vs. 25Q-ni and 72Q-ni vs. 72Q-I. Significant events had FDR <0.05.

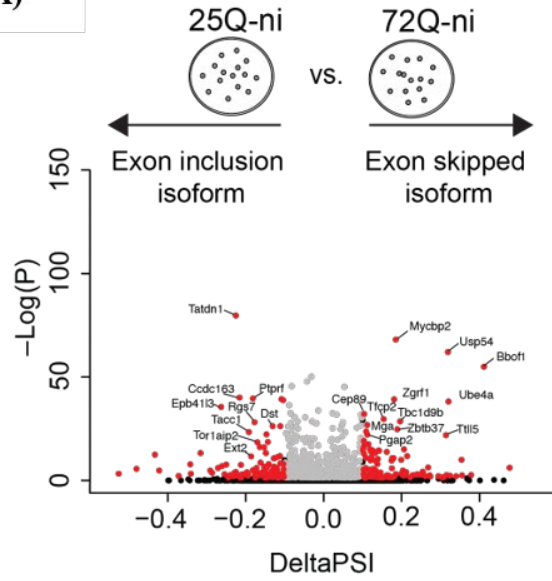
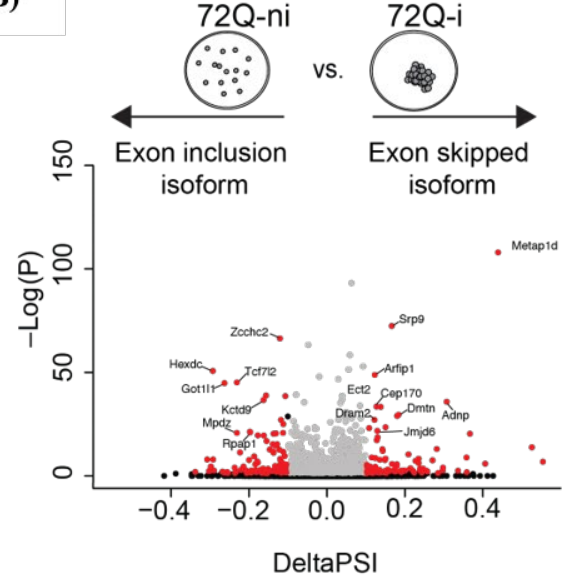
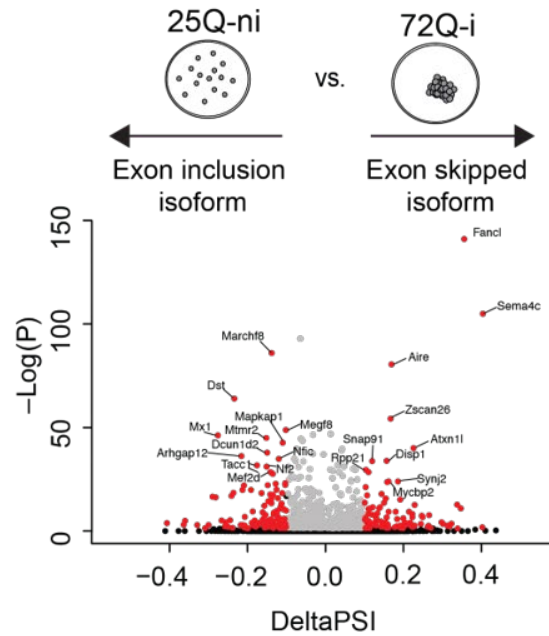
To obtain an overview of global skipped exon events between aggregation states, we plotted volcano plots highlighting significant skipped exons in each condition (Figure 3.7A-3.7C). The PSI values represent the frequency of occurrence of a particular exon in all known transcript isoform sequences of the gene. It is an indication of how well a particular exon is spliced into the transcript population. A PSI of 1 suggests constitutive exons that are used and never excluded in all transcripts, while values below 1 mean decreased inclusion of alternative exons and represent the percentage of exon-containing proteins relative to the total transcript population[372]. Δ PSI values detail the abundance of exon skipping and exon inclusion isoforms between the test and the control groups. Positive values indicate an excess of exon skipped isoforms in the test group and negative values abundant exon inclusion isoforms in the controls. Pairwise comparisons and correction for multiple testing was performed and false discovery rates <0.05 and Δ PSI $\geq \pm 0.1$ (10%)[354] was considered significant as denoted in Figure 3.7A-3.7C.

First, we checked if the genes spliced in our dataset showed any similarity with other human HD studies. The study from Lin et al.,2016[303] which examined splicing changes in HD patient motor cortex and control brains was used as reference. We saw a statistically significant overlap in genes that showed abnormal splicing ($n=51$, Fold change:2.3, FDR <0.0001) between human HD brains and our dataset (72Q vs 25Q) by hypergeometric gene enrichment testing[373] (Supplementary Table 3.2). While hypergeometric tests suffer drawbacks from assumptions that genes are independent and gene expression are not correlated, there is no easy alternative for enrichment analysis given the small sample size and difficulty of implementing linear model-based methods for alternative splicing events in our study. From Figure 3.7A we can see that global splicing changes occur in the soluble 72Q-ni as compared to 25Q-ni. There were 1,105 significant skipped exon transcripts in 72Q-ni as compared to 25Q-ni, 511 were exon skipped transcripts and 594 were exon inclusion transcripts. 1,109 significant skipped exon transcripts in 72Q-ni as compared to 72Q-i (Figure 3.7B) with 552 were exon skipped transcripts and 557 were exon inclusion transcripts. There were 1,108 significant skipped exon transcripts in 72Q-i as compared to 25Q-ni (Figure 3.7C) with 520 were exon skipped transcripts and 588 were exon inclusion transcripts. The respective events are denoted in supplementary table 3.3.

Interestingly one of the top spliced genes in 72Q-ni as compared to 25Q-ni, the E3 ubiquitin-protein ligase- Mycbp2 is known to regulate cAMP signalling pathway[374, 375]. We found other genes including Raf-1 Proto-Oncogene, Serine/Threonine Kinase (*Raf1*),

phosphoinositide-3-kinase regulatory subunit 2 (*Pik3r2*)[376], Mitogen-activated protein kinase 9 (*Mapk9*)[377], Ras-Related Protein Rap-1A (*Rap1a*) [378], Adenylate Cyclase Activating Polypeptide 1 (Pituitary) Receptor Type I (*Adcyap1r1*), CF Transmembrane Conductance Regulator (*Cftr*) of cAMP signalling pathway (KEGG: mmu04024)[379] in the dataset. In addition we also found genes - Dynein axonemal light chain 4(*Dnal4*), Nuclear receptor binding factor 2(*Nrbf2*), Glutamate receptor ionotropic AMPA2 (*Gria2*), NADH dehydrogenase 1 alpha subcomplex 12(*Ndufa12*), Peroxisome proliferator activated receptor gamma (*Pparg*), Glutamate receptor ionotropic, NMDA1 (*Grin1*), Discs large homolog 4 (*Dlg4*), Clathrin light polypeptide (*Clta*), Huntingtin interacting protein 1(*Hip1*), Mitogen-activated protein kinases – *Mapk9* and *Mapk10*, genes related to the HD pathway (KEGG:mmu05016)[380] to be alternatively spliced in 72Q-ni as compared to 25Q-ni. The MYCBP2 and cAMP-response element binding protein (CREB) is also significantly spliced in 72Q-i as compared to 25Q-ni. Other important genes alternatively spliced in 72Q-i as compared to 25Q-ni includes Semaphorin 4C (*Sema4c*) encodes the Semaphorin-4C protein that is needed to activate the p38 MAPK pathway which has been implicated in HD[381]. The histone demethylase Jumonji domain-containing protein 6 (*Jmjd6*) shows alternative splicing in 72Q-ni as compared to 72Q-i and the modulation of which has been shown to affect the survival of HD neurons[382].

To better understand the alternatively spliced genes between aggregation states, we looked at the skipped exons events in different aggregation states as depicted in the Venn diagram (Figure 3.5D). Interestingly, there seems to be genes which are preferentially spliced with different aggregation state. Around 66% of the SE were unique to individual aggregation states and 34% showing overlap in splicing. Similarly retained introns also showed a higher number of unique genes spliced (56%) between aggregation state. The enrichment analysis of the unique genes with skipped exon showed statistically significant enrichment of spliced genes related to regulation of actin cytoskeleton and the spliceosome in soluble mutant Httex1, whereas apoptotic genes were spliced in the aggregated Httex1 (Supplementary Table 3.4). Genes spliced for retained introns did not show any statistically significant enrichment.

(A)**(B)****(C)**

(D)

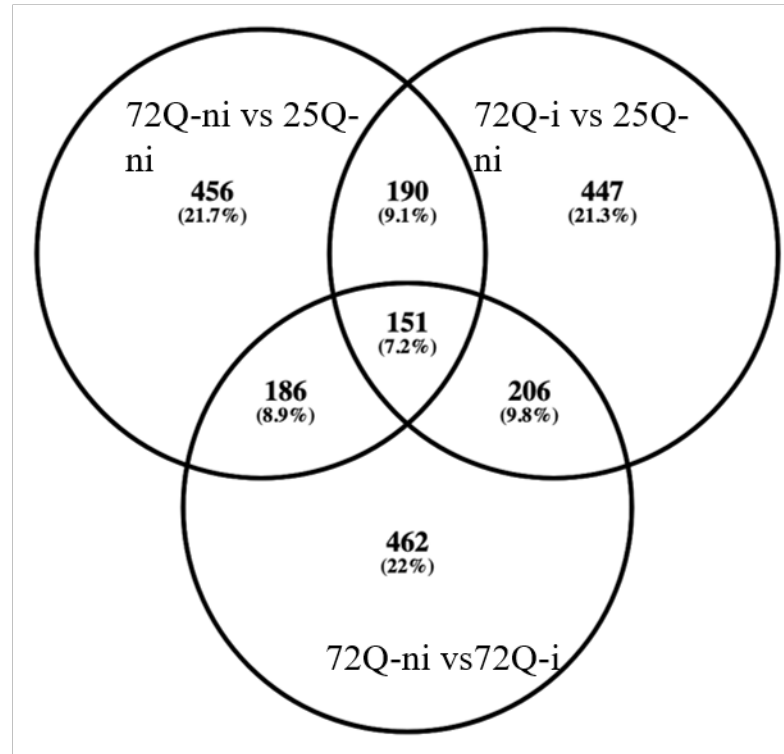


Figure 3.7: Global exon skipping events seen in different aggregation states of Httex1.

Volcano plots (A, B, C) showing the abundance of skipped exons events in different aggregation states. Percent spliced inclusion (PSI) values between conditions as Δ PSI values of all the ASE are represented on the X-axis and adjusted p-values (negative log) are plotted on Y-axis. Significant skipped exon events with significant adjusted p-values less than 0.05 with a minimum Δ PSI change equal or greater than 0.1 as investigated by rMATS are highlighted red. Grey dots show significant AS event and represent changes in PSI ratios of less than 0.1 and black dots denote non-significant events. D) Venn diagram showing skipped exons events in aggregation states.

To gain further insight into the changes in biological processes of aggregation specific spliced gene changes, we performed gene ontology and pathway analysis on the genes exhibiting alternative splicing (FDR <0.05; spliced skipped exon) of different aggregation states using DAVID-NCBI (version 6.8)[358]. Gene ontology terms with adjusted p-value of less than 0.05 with counts of 10 and greater were considered significant. DAVID analysis revealed that genes involved in transcription are significantly spliced in 72Q-ni as compared to the control 25Q-ni. We also observed splicing of genes involved in post transcriptional mechanisms like mRNA processing, RNA splicing, as well as chromatin modification which showed alternative splicing in soluble 72Q-ni. GO analysis of aggregated 72Q-i revealed similar alteration in splicing in genes involved in transcription, mRNA processing, RNA splicing, chromatin modification. Further, genes associated with cellular response to DNA damage stimulus and protein transport were also spliced. Comparison of soluble 72Qni and aggregated 72Q-i showed similar terms enriched. The heatmap of important GO biological processes analysis with the corresponding fold changes in different aggregation states is show in Figure 3.8.

Overall, based on the gene ontology and pathway analysis of alternate spliced genes in different aggregation states we can conclude that the genes mainly involved in transcription and post transcriptional processes are altered in different aggregation states of Httex1.



Figure 3.8: Heatmap of biological processes altered by splicing across Httex1 aggregation states. Heatmap of gene ontology terms enriched across different aggregation states compared – between the soluble forms of Htt - 72Q-ni vs 25Q-ni, between aggregated state and wild-type Htt - 72Q-i vs 25Q-ni and on aggregation 72Q-ni vs 72Q-i for alternatively spliced skipped exons. The heat map was generated using Morpheus tool from the Broad Institute using fold enrichment values of individual gene ontology terms (P. Value <0.05) across different aggregation states generated using DAVID (v.10). White to red colour scale (colour key, left) denotes minimum to maximum enrichment scores (fold score) and represents fraction of genes alternatively spliced for a given gene ontology gene set than expected by chance. The cells highlighted in green represent processes that are altered (FDR <0.05) in 72Q-ni that are worsened in 72Q-i. Gray represents unavailable values.

3.4.4. Altered splicing changes across Httex1 aggregation states.

In order to understand the variations in splicing patterns in genes with conversion of soluble Httex1 to aggregated state, we mapped the splicing signatures into 4 categories as described in the Figure 3.7A. Only skipped exon events in different aggregation states were considered. The FDR significance values across aggregation states were used to categorise the spliced gene into 4 groups as described in Figure 3.7A. The splicing signature 1 indicates the skipped exon events initiated by soluble Htt and continued in aggregated Htt. The splicing signature 2 indicates skipped exon events exclusively associated with the aggregated form of mHtt. Splicing signature 3 indicates events initiated by soluble form and exacerbated by aggregated state. Whereas signature 4 depicts skipped exons events initiated by soluble Htt and rescued with aggregation. While spliced signatures 1 and 3 indicate toxicity associated splicing changes, events in splicing signature 2 and 4 indicates rescue attempts. The different alternatively spliced genes categorised into four different splicing signatures are listed in supplementary table 3.5.

To further understand these splicing signatures and their effects we undertook a systems biology approach to pinpoint targets for aggregation related splicing defects. A protein-protein interaction network was created using the validated interactions from the STRING database(v10) using Network Analyst toolkit[383]. A minimum interaction network of spliced genes in each signature was created to identify key hub genes with high degree of connectivity (>5) with other genes in the network. We hypothesise that alterations in these highly networked genes would impact the functions of the whole splicing network of associated genes. The networks generated, and the hub genes identified are listed in Figure 3.5B and potential role in HD is described below.

These categories suggest only a working model of the mechanisms that will require further investigation. The below signatures only seek to provide elaboration on possible mechanisms of splicing changes that can be attributed to specific aggregation states but does not claim as definite mechanisms at play as these have been not verified in an orthologous system.

Splicing effect attributable to soluble Httex1 - Splicing signature 1: This signature would be consistent with the effects arising from soluble Httex1. The hub genes in splicing signature 1 include genes such as Epidermal Growth Factor Receptor Substrate 15 (*Eps15*), Mortality Factor 4 Like 1 (*Morf4l1*), DExH-Box Helicase 9 (*Dhx9*) and Euchromatic Histone Lysine Methyltransferase 2 (*Ehmt2*). EPS15 is a cofactor involved in clathrin mediated endocytosis [384]. EPS15 interacts with HIP1 which is an important partner of Htt and is important for clathrin mediated endocytosis in HD[385]. EPS15 is recruited into ubiquitin-rich cytoplasmic aggregates by Ubiquilin[386]. This presence in ubiquitin-rich cytoplasmic aggregates might demonstrate EPS15 function either in the membrane recruitment to aggregates or in autophagic fusion to lysosomes or endosomes[387]. MORF4L1 is involved in chromatin organisation, modification and regulates transcription of numerous genes primarily by the acetylation of histones H2A and H4 in the nucleosome[388]. MORF4L1 being part of NuA4 histone acetyl transferase complex might participate in effective repair of damaged DNA when recruited to the sites of DNA damage[389]. The *Dhx9* gene belongs to the DEAH-containing RNA helicase family. DHX9 functions as a transcriptional regulator. It is detected in cytoplasm and nucleus, where it catalyses and unwinds double-stranded RNA and DNA-RNA complexes in a ATP dependent reaction[390]. It is also part of the holo-RNA polymerase II (Pol II) complex and is regulates CREB mediated transcription by recruiting Pol II to the CREBBP protein. It has also been shown that DHX9 could regulate CREB-mediated transcription by other ATP-dependent processes[391]. The *Ehmt2* gene produces an histone methyltransferase enzyme that methylates lysine 9 residues of histone H3 resulting in repression of target genes expression[392]. This epigenetic regulator is crucial for transcriptional modulation and is one of the earliest of epigenetic-related transcription alterations in Huntington's disease mouse models[393]. Overall, the data suggests that the splicing of genes involved in crucial cellular processes are altered in soluble Httex1 aggregation state. The splicing of the genes related to endocytosis, chromatin organisation, DNA repair and transcription are altered by soluble mHtt.

Splicing effect attributable to aggregated Httex1- Splicing signature 2: This signature would be consistent with the effects arising from aggregated Httex1. Of the different hub genes identified, Nuclear Factor Kappa B Subunit 1 (*NFkB1*) (Figure 3.9B) is central in mHtt mediated neurotoxicity[394]. We have shown in chapter 2 that NFkB1 to be associated with transcriptomic changes that are triggered by soluble toxic Httex1 and extended by aggregated Httex1 state[300]. Similarly, we find that Nfkb1 hub seems to be involved with splicing changes associated with aggregation of Httex1. NFkB1 regulates pro-survival gene expression, remains essential for the functioning of neurotrophic factors, cytokines[395] and activates during NMDA-mediated excitotoxic cell death as seen in HD [396, 397]. *NFkB1* showed higher expression in human prefrontal cortex of HD patients compared to healthy control[185] as well as transgenic mice brain[394]. The increased NFkB activity is a protective response to ongoing oxidative stress and mitochondrial dysfunction seen in HD[394]. Further NFkB1 has significant cross talk with CREB via the co-activator CREB-binding protein[398]. These findings indicate that NFkB and its signalling cascade is central to HD pathogenesis and aberrant splicing of these modulators can cause perturbations in cell health.

Our lab has previously shown that inclusion formation supresses apoptotic cell death and later activates necrotic cell death pathways[178]. NFkB has been shown to be a key regulator of programmed cell death pathways, its signalling cascade either through canonical or non-canonical mechanisms with a complex interplay of downstream signalling molecules affecting the switch between anti-apoptotic, apoptotic and necrotic cell death. NFkB acts as a switch tipping the balance in the favour of cell survival or cell death following TNFR (Tumor necrosis factor receptor) activation which initiates three distinct signalling mechanisms that progress to different cell fates[399].

The presence and the subsequent co-operative recruitment of key signalling molecules directs the cells into these cascades. Recruitment of TNFR-associated death domain (TRADD), receptor-interacting protein kinase 1 (RIPK1), and TNFR-associated factor 2 (TRAF2) to TNFR promotes IKK-dependent activation of NFkB, which promotes the expression of antiapoptotic genes, leading to cell survival. In absence of RIPK1, TRADD binds to TNFR and effects FLICE-associated death domain (FADD) mediated mitochondrial release of cytochrome C and activation of effector caspases, resulting in apoptosis. RIPK1 when recruited to the TNFR signalling complex in absence of TRADD, activates the stress-activated Jun kinase (JNK) signalling cascade resulting in programmed cell death via

necrosis or apoptosis[399-402]. The metabolic state of the cell is important in determining whether the cell undergoes apoptosis or necrosis[403]. Cellular quiescence as represented by ATP depletion and other metabolic challenges can result in activation of NFkB[404], which further based on a severity or a threshold level of metabolic dysfunction can result in switch between necrosis or apoptosis[405]. Hence NFkB can potentially act as a controller of programmed cell death pathways. Alternative splicing of NFkB alters the function, location and activity of NFkB group of transcription factors[406, 407]. Hence aggregation associated aberrant splicing of NFkB could act as molecular switch that might suppresses apoptotic cell death with early aggregates and later activates necrotic cell death pathways with late aggregates and needs to be further elaborated.

SWI/SNF-Related Matrix-Associated Actin-Dependent Regulator of Chromatin Subfamily A Member 4 (Smarca4) is a member of the SWI/SNF family of proteins. Smarca4 forms part of ATP-dependent chromatin remodelling complex- SNF/SWI needed for transcriptional activation of chromatin-repressed genes[408]. SMARCA4 recruits topoisomerase 1 and can suppresses transcription-associated CAG instability seen in HD[409]. The HDAC 1,2,5 belong to the family of histone deacetylases that are important regulators of gene transcription. To mitigate the transcriptional dysregulation seen in HD, HDACs have recently emerged as important drug targets for treatment in HD[410-412]. HDAC 1, 2 have been well studied and have been shown to be neuroprotective in HD[413], [414] . HDAC6 also counteracts the effects of misfolded protein accumulation in ER by induction of HSP70 chaperones[415, 416].

Splicing effects worsened by aggregation - Splicing signature 3: This signature relates to the effects that are initiated by soluble Htt but worsened with Htt aggregation. The key hub gene identified in this signature is the Ubiquitin A-52 Residue Ribosomal Protein Fusion Product 1 (*Uba52*) gene. The Uba52 gene product consists of a fusion protein made up of N-terminal ubiquitin fused to ribosomal protein L40 (RPL40) at the C-terminus and serves to regulate the ribosomal protein complex and its function[417]. *Uba52* is an essential gene that regulates ubiquitination in cells[418] (Figure 3.7B). It is a major source of ubiquitin labelling of proteins that are used for different signalling outcomes or alternatively recycled by ubiquitin-proteasome system[419]. Increasing evidence now suggests that a HD and other neurodegenerative disorders show a compromised ubiquitin homeostasis[420]. Hence a defect UBA52 processing with Htt aggregation can have implications for cellular

ubiquitination compounding the proteostatic imbalance present in the cell. Additionally, a compromised UPS leading to proteostasis defects is also characteristic of many neurodegenerative disorders. UPS acts as a protein quality control(PQC) system by tagging peptides which fail PQC with ubiquitin which are then degraded by the proteosomes for clearance by the cell[421]. Any failure in this protein quality control and clearance is associated with increased protein aggregation of disease-specific proteins[422]. As evidenced by ubiquitin enrichment in mHtt inclusions observed in HD. As UBA52 is an important component of this protein quality control network, any loss of UBA52 function can contribute further to worsening protein aggregation[423]. Our data supports the role of ubiquitination systems in HD pathogenesis which progressively deteriorate with disease.

Splicing effects rescued by aggregation -Splicing signature 4: This signature is consistent with the hypothesis that inclusions are “protective” against toxicity. Rac GTPase Activating Protein 1 (RACGAP1) is a GTPase-activating protein the knockout of which results in embryonic lethality[424]. RACGAP1 binds to activated forms of Rho GTPases resulting in negative regulation of Rho-mediated signals. Rho GTPase signalling pathways have been implicated in HD as modifiers of mutant Htt toxicity[425]. Defective Rho GTPase signalling leads to development of toxic reactive oxygen species (ROS) and aggregation of mHtt[426]. The Rho GTPase signalling mediated generation of ROS occurs via the activation of NADPH oxidase, has been implicated in HD pathogenesis[427]. Similarly Rap Guanine Nucleotide Exchange Factor 1 (RAPGEF1) is a regulator of Rho GTPase activity[428]. RAPGEF1 also plays a critical roles in multiple signal transduction pathways regulating neuronal growth and survival[429]. Rho GTPase signalling pathways control a variety of cellular processes. They are crucial regulators of neuronal cytoskeleton, neuronal morphology, gene transcription and neuronal survival and death. Activation of Rac-mediated signalling facilitates neuronal survival, while activation of Rho-mediated signals also causes neuronal death[426, 430]. GTPase-activating proteins like Racgap1 by inactivating Rho mediated signalling might help protect against toxicity of soluble mHtt and can be a specific response induced by aggregation. Similarly, guanine nucleotide exchange factors like RAPGEF1 have been implicated in neurodegenerative disorders[431] and have been demonstrated to be pro survival in human neuronal cells[429] and might also indicate a protective response of aggregation. In addition CREB Cytoplasmic Linker Associated Protein 1 (CLASP1) is a nonmotor microtubule-associated proteins that promotes the stabilization of dynamic

microtubules[432]. This gene is a target of splicing factor Polypyrimidine Tract Binding Protein 1 and has been shown to be alternatively spliced in HD brain[303]. However further study into how these splicing defects relate to reduced toxicity is needed. Overall, our examination of this splicing signature revealed that the action of small GTPase mediated protective responses are likely mediated by aggregation to reduce the toxicity of soluble mHtt and provides potential avenues for further investigation.

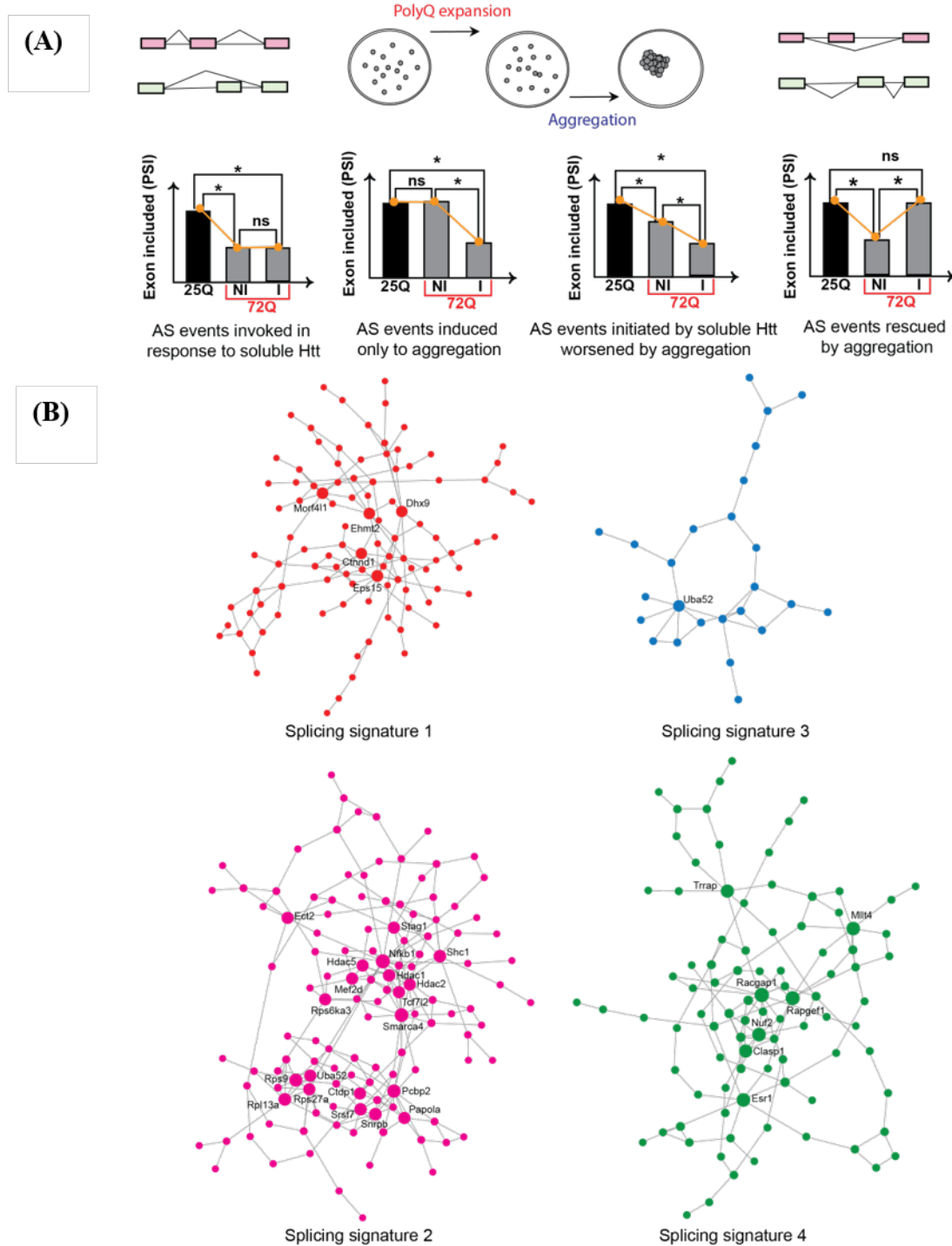


Figure 3.9: Httex1 aggregation state related splicing signatures. (A) Schematic representation depicting splicing patterns associated with various Httex1 aggregation states. ASE data was filtered with adjusted p-values <0.05 ($n=4$) and categorised based on splicing changes into different splicing signatures as depicted by the bar graphs. The exon inclusion levels can be changed in either direction in all splicing signatures (B) Shown are mapped protein: protein interaction networks of all spliced gene in a splicing signature generated using the Network Analyst Walk trap algorithm. The highly connected hub genes (with degree connectivity of > 5) are labelled.

3.4.6. Proteomic analysis of Httex1 aggregation states

In order to identify the global proteomic changes associated with different aggregation states, Neuro2a cells were transfected for 48 hours with 72Q and 25Q constructs with emerald only as control for the experiment. Flowcytometry by PulSA was performed over 5 days to get 5 independent replicates. The samples were dimethyl labelled and processed for whole cell proteomic LC-MS/MS analysis (Section 3.3.7). The proteomic data was analysed using proteome discoverer (v 2.1.1) and comparisons between different aggregation states of 72Q-ni, 72Q-i and 25Q-ni was performed. Statistically significant results are depicted in the Figure 3.8 and proteomics data is detailed in supplementary table 3.6. From Figure 3.8A we can see that tRNA nucleotidyl transferase 1(*Trnt1*) levels are increased in soluble 72Q-i. *Trnt1* is an essential gene essential to produce all mature tRNAs, required for protein synthesis[433]. The RNA polymerase enzyme produced by *Trnt1* is involved in tRNA quality control and stress response[434]. Mutations in *Trnt1* have been found to be associated in a range of neurodegenerative disorders[435, 436] with mutations causing altered protein homeostasis and autoinflammation[437]. TRNT1 has been shown to play a role in autophagy[438] and oxidative phosphorylation related cellular respiration[439]. Autophagy in cell is responsible for degradation of long-lived proteins and protein aggregates. *Trnt1* defects causes an increased lipidation of autophagic marker LC3-1 (microtubule-associated protein 1A/1B-light chain 3) with accumulation of LC3-II and decrease in lysosomal-associated membrane protein 1 expression required for autophagic clearance. In addition, *Trnt1* mutations in patients with Alzheimer's and HD can directly decrease the production of proteins like beclin-1, which are essential for autophagy[438]. We assume the increase in *Trnt1* levels seen in our data might increase autophagic clearance of toxic soluble mHtt, however, its exact role in HD is yet to be elucidated and needs further examination.

DnaJ Heat Shock Protein Family (Hsp40) Member A2 (*Dnaja2*) is a member of conserved DNAJ/HSP40 family of chaperones that maintain protein homeostasis in cells. Hsp40 proteins also serve as cofactors for Hsp70 chaperone activity[440]. Hsp40 family of proteins have been previously found to stably sequestered in the Httex1 inclusion core[441]. Our lab has previously shown DNAJA2 protein significantly reduced solubility as cells with mHtt of 97 polyQ converted from dispersed soluble unaggregated state to insoluble aggregated state[442]. In concurrence Radwan et al.,2020 also showed DNAJA2 enrichment in Httex1 inclusions[443]. Their sequestration can impair proteostasis[444] and the increase in this key

Hsp40 chaperone in our 72Q-i data (Figure 3.8C) can be a cellular strategy to cope with increased protein aggregation.

Serine and Arginine Rich Splicing Factor 4 (*Srsf4*) belongs to arginine/serine-rich SR splicing factor family. The SRSF4 protein is involved in RNA splicing and plays an important role in mRNA processing[445]. SRSF4 can shuttle between nucleus and cytoplasm mediating mRNA export, RNA stability, and mRNA translation through its shuttling activity[446]. The SR splicing factors are crucial splicing factors with respect to Httex1 aggregation related states as the expression of SR family of splicing factors is increased in both our gene expression (Figure 3.2) and proteomics data sets (Figure 3.8C). The increase in aggregates of Srsf4 could indicate possible sequestration of this SR family of splicing factor thereby reducing availability to the soluble form of mHtt.

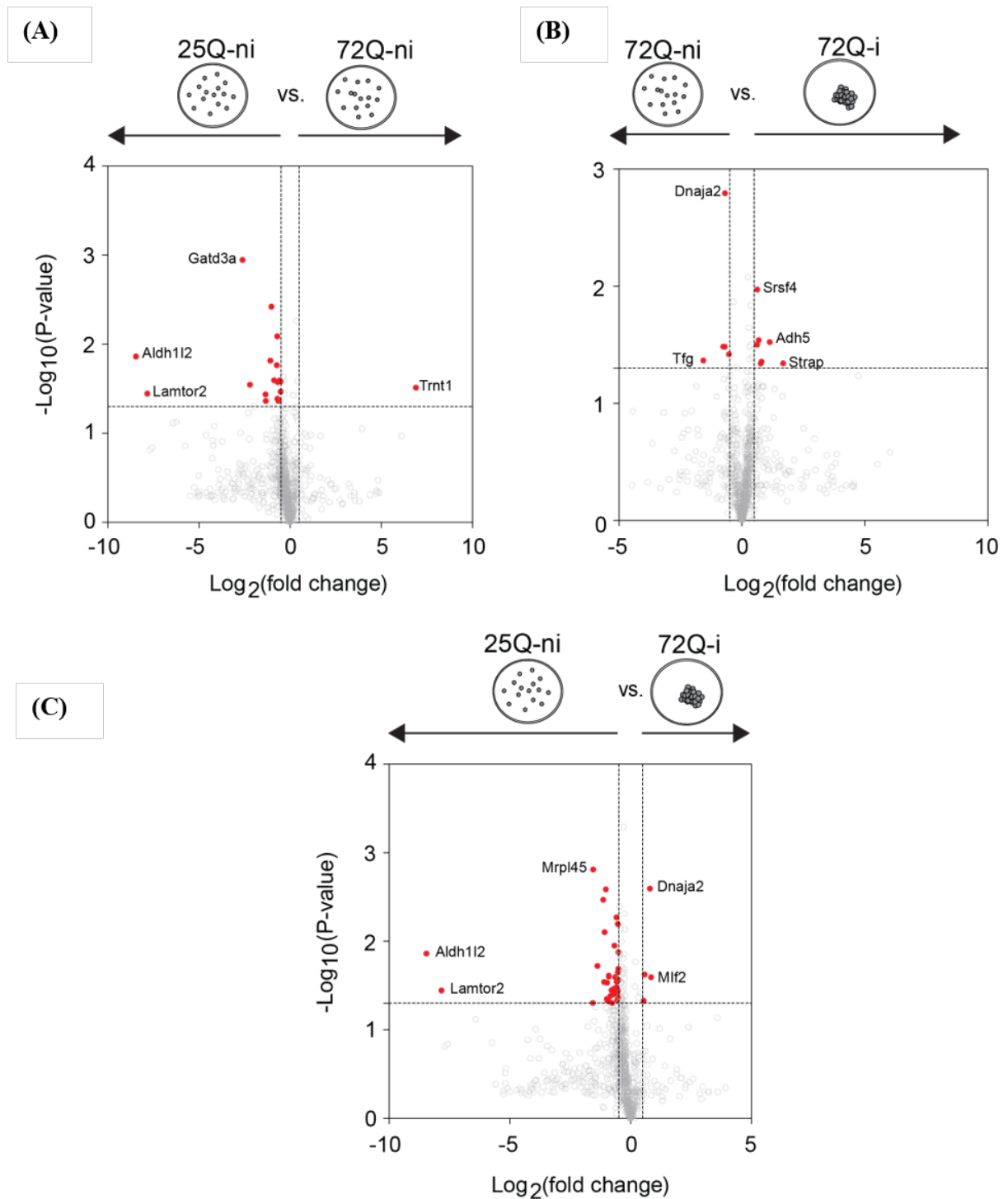


Figure 3.10: Volcano plots of protein expression changes mapped to aggregation states of Httex1. Depicted are protein expression changes in comparisons between (A) Cells that lack visible aggregates but differ in the length of polyQ (25Q-ni and 72Q-ni cells), (B) Cells with pathogenic polyQ that differ in aggregation profile (72Q-i) and (C) Cells with aggregates but varying in polyQ length (25Q-ni and 72Q-i cells). Genes with adjusted p-value < 0.05 with greater than 1.3-fold change are shown in red.

3.4.7. Identification and validation of alternative splicing isoforms

To validate the splicing events identified from our Httex1 aggregation state specific transcriptomic data, we used our Httex1 aggregation state matched proteomic data to find corresponding translated spliced peptide isoforms. This strategy of using flowcytometry PulSA matched transcriptomic and proteomic samples would enable us to verify multiple spliced isoforms at a proteome wide scale. Proteogenomic analysis was conducted using Splicify[362].

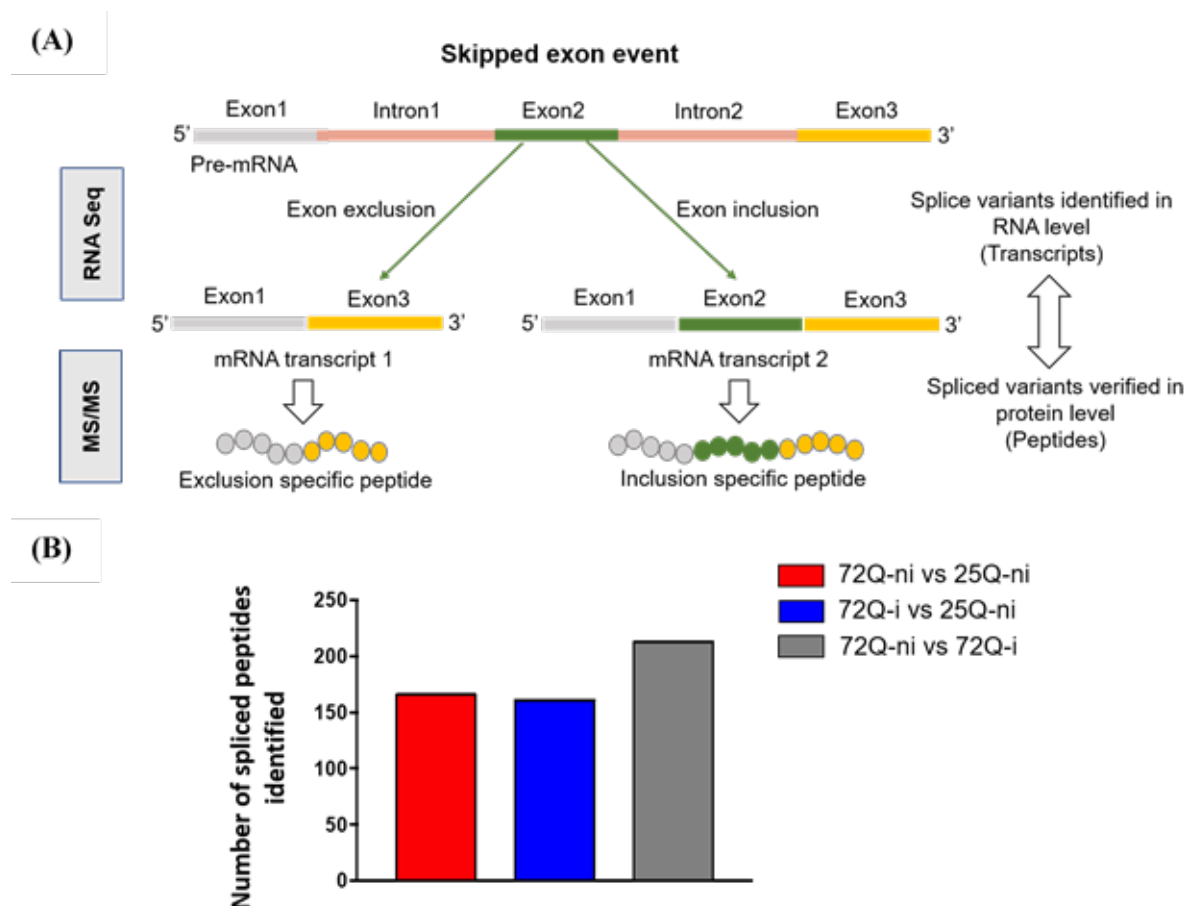


Figure 3.11: Schematic of aggregation state related splice variants identified in transcriptome data verified as protein expression changes. The RNA splicing variants were mapped onto the PulSA sorted 25Q-ni, 72Q-ni and 72Q-i MS/MS proteomics samples to identify MS/MS spectra arising from the matched set of aggregation state specific RNA-sequencing data using Splicify[362]. The spliced peptide isoforms levels were measured and differential splicing expression determined. (A) Representation of peptides fragments supporting translation of mRNA splicing events for skipped exons identified in MS/MS data. The exon inclusion or exon exclusion peptides can map to the target exon or span the exon-exon junction in skipped exon events. (B) Data showing number of aggregation state specific

spliced variants identified in both RNA-seq and MS/MS data across comparisons. No significant differential protein splicing expression was seen between comparisons.

We were able to identify 166 splice peptides in the cells with soluble 72Q-ni as compared to wild type 25Q-ni, 161 spliced peptides in soluble pathologic 72Q-ni group compared to aggregated 72Q-i and 213 spliced peptides were identified in aggregated 72Q-i as compared to 25Q-ni. We did not observe any significant differential peptide expression between the groups. However, through this analysis we validated the different splicing events attributable to the aggregation states of Httex1. The different types of splice events identified along with the splice isoforms are detailed in Supplementary Table 3.7.

3.5 Conclusions

Alternative splicing is a flexible and pervasive process for producing diverse but precisely regulated mRNAs from a single transcript. These ensuing mRNAs, in turn, produce functionally diverse protein. Accumulated evidence shows that aberrant RNA splicing is one of key features in a variety of neurodegenerative disorders including Huntington disease[303], [306]. The spliceosome machinery made up of snRNAs and associated splicing factors regulates this key process. In this study we explored the splicing mechanisms in different Httex1 aggregation states and dissected the associated splicing profiles. We found upregulation and enrichment of the genes involved in the splicing pathway in soluble 72Q-ni. Further analysis of these genes showed enrichment of SR family of splicing factors, including the key splicing initiator SRSF2. The quantification of this splicing factor revealed a higher activity in expanded polyglutamine 72Q cells as compared to wild type 25Q. A similar activation was also confirmed in isogenic neural precursor cells from HD patients. U-snRNAs are key regulators of splicing specificity and spliceosome assembly. They are present in constant stoichiometry inside the cell and any aberrant change in their expression affects splicing, hence we investigated these key molecules in neuronal cells with different aggregation states. Interestingly our study showed that there is aberrant production of snRNAs in cells with expanded polyglutamine 72Q with and without aggregates as compared to wild type 25Q. These changes were validated in gene corrected human iPSC cells of different polyQ lengths. This finding of defects in spliceosome integrity is in concurrence with other neurodegenerative disorders like ALS and SMA. In ALS, U snRNA levels are up-regulated with spliceosomal U snRNPs (made of U snRNA and other proteins) increasing in the motor neuron nuclei of patients with ALS. Whereas in SMA, spinal motor neurons show down-regulation of U snRNAs. Dysregulation of U snRNA's resulting in defects in the spliceosome function in both ALS and SMA[371]. In Alzheimer's the tau accumulation leads to spliceosome disruption and mRNA splicing errors. These findings have been correlated with neurofibrillary tangle burden in patient brains[447]. These findings together with our study suggest that a substantial loss of spliceosome integrity is a crucial mechanism common to neurodegeneration seen in HD, ALS, SMA and Alzheimer's disease.

The changes in the splicing machinery observed indicate that splicing defects are associated with early soluble mutant Htt and potentially could be a mechanism of toxicity. To correlate the aggregation state specific changes in splicing machinery, with changes in gene splicing we performed alternative splicing analysis between aggregation states. We detected

numerous splicing events between cells with soluble 72Q-ni as compared to wild type 25Q-ni, between cells with soluble 72Q-ni and aggregated 72Q-I and cells with aggregated 72Q-i as compared to wild type 25Q-ni. The genes spliced in our data were similar to the spliced genes identified in HD brain[303] validating our results. We did not find significant difference between different classes of AS events like skipped exons, retained introns and others in the different aggregation states. We also mapped gene splicing events across aggregation states to find AS events attributable to polyQ mutation, and subsequent conversion into inclusions. The splicing changes were classified into 4 different signatures and network analysis revealed key spliced hub genes that have been previously implicated in HD pathogenesis. We have previously identified CREB signalling defects as central to aggregation state related HD pathomechanisms[300]. Many of the genes spliced in expanded soluble Httex1 are cAMP signalling pathway genes or their targets. Splicing hubs like *Dhx9*, *NFkB1*, *Hdacs* identified are regulated by CREB related mechanisms. The CREB regulated transcription co-activators (CRCTs), are critical components of the cAMP signalling pathway and have been implicated in HD[448, 449]. They are bipartite regulators of gene expression i.e, they selectively and independently control pre-mRNA splicing or transcriptional activation of CRE-promoter containing genes[450]. This function unique to CRCTs influences splicing by modifying splice site selection of transcripts via proline rich domains that mediate CRCT-dependant alternative splicing[450]. The key splicing defects observed between different aggregation states of mHtt mainly relate to genes associated with CREB related transcription and co-transcriptional mRNA processing indicating a possible link back to CREB signalling. Though CRCTs (*Cret1*, *Crtc2*, *Crtc3*) were not significantly differentially expressed in our transcriptomic data, their role in mHtt mediated splicing defects cannot be ruled out and needs to be explored.

Though transcriptional dysregulation is well characterised in HD, splicing related changes are less studied. Our study explores post transcriptional mRNA changes in different aggregation states of Httex1. The identification of spliced peptides via Splicify analysis of the proteomic data validated the splicing defects seen in different aggregation states with splicing defects arising with soluble form of mHtt – aberrant splicing could be a mechanism of toxicity in HD. Overall, our data collectively suggests that splicing machinery is defective in HD and deregulated splicing is a prominent aggregation related change seen in HD. This study is the first to show altered snRNA biogenesis in HD and describes splicing defects and splicing patterns with key hub genes that might be specifically spliced in response to the Httex1 aggregation state providing further avenues for research.

3.6. Supplementary material

Supplementary materials for chapter 3 - “Httex1 aggregation states induce spliceosome defects and aberrant splicing in Huntington’s disease” is linked below

<https://www.dropbox.com/sh/qibir93gabtsib2/AAB6OhY3j5AGa2uBicCVn-iNa?dl=0>

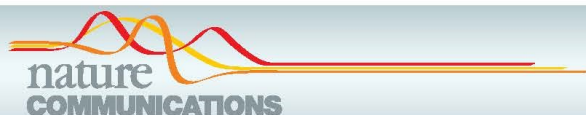
Chapter 4:

***A novel proteomic approach for
measuring unfolded protein load and
proteostasis.***

4.1. Abstract

Proteostasis refers to the cell's ability to keep its proteome folded. Proteostasis networks play a crucial role in maintaining a healthy proteome. When proteostasis balance is perturbed in the cell unfolded proteins will aggregate. A vast network of chaperones, QC factors, stress responses, trafficking and degradation pathways collectively termed the proteostasis network mediates the protein folding in the cells and are activated when there is an increased unfolded protein load. In this publication we note the dye, tetraphenylethene maleimide (TPE-MI) was used quantify unfolded protein loads in cells. TPE-MI fluorescence was triggered when cysteine (Cys) thiols usually embedded within the nucleus of globular proteins are subjected to unfolding and are subsequently labelled by TPE-MI. TPE-MI fluorescence is induced by cellular proteome under conditions that facilitate the aggregation of unfolded proteins including ER unfolded protein backlog caused by Tunicamycin, HSP90 inhibition with Novobiocin, and global oxidative stress and others. We further showed that TPE-MI reactivity can track unfolded proteins that under stress expose Cys residues using quantitative proteomics. This novel approach has utilities in further understanding protein unfolding kinetics as we can quantitatively track specific unfolded peptides in a high throughput manner in cells. TPE-MI can describe proteostasis defects in hiPSC and other cell models of HD and also detects drug induced protein disruption during treatment of malarial parasites. Therefore TPE-MI is a very promising as a method to check disease proteostasis mechanisms and unfolded protein load in the clinic.

4.2. A thiol probe for measuring unfolded protein load and proteostasis in cells.



ARTICLE

DOI: 10.1038/s41467-017-00203-5

OPEN

A thiol probe for measuring unfolded protein load and proteostasis in cells

Moore Z. Chen¹, Nagaraj S. Moily¹, Jessica L. Bridgford¹, Rebecca J. Wood¹, Mona Radwan¹, Trevor A. Smith², Zhegang Song³, Ben Zhong Tang³, Leann Tilley¹, Xiaohong Xu⁴, Gavin E. Reid^{1,2}, Mahmoud A. Pouladi^{4,5}, Yuning Hong^{1,2,6} & Danny M. Hatters¹

When proteostasis becomes unbalanced, unfolded proteins can accumulate and aggregate. Here we report that the dye, tetraphenylethene maleimide (TPE-MI) can be used to measure cellular unfolded protein load. TPE-MI fluorescence is activated upon labelling free cysteine thiols, normally buried in the core of globular proteins that are exposed upon unfolding. Crucially TPE-MI does not become fluorescent when conjugated to soluble glutathione. We find that TPE-MI fluorescence is enhanced upon reaction with cellular proteomes under conditions promoting accumulation of unfolded proteins. TPE-MI reactivity can be used to track which proteins expose more cysteine residues under stress through proteomic analysis. We show that TPE-MI can report imbalances in proteostasis in induced pluripotent stem cell models of Huntington disease, as well as cells transfected with mutant Huntington exon 1 before the formation of visible aggregates. TPE-MI also detects protein damage following dihydroartemisinin treatment of the malaria parasites *Plasmodium falciparum*. TPE-MI therefore holds promise as a tool to probe proteostasis mechanisms in disease.

¹Department of Biochemistry and Molecular Biology and Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, VIC 3010, Australia. ²School of Chemistry, The University of Melbourne, Parkville, VIC 3010, Australia. ³Department of Chemistry, The Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong, China. ⁴Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research, Singapore (A*STAR), 8A Biomedical Grove, Immunos, Level 5, Singapore 138648, Singapore. ⁵Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore. ⁶Department of Chemistry and Physics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC 3083, Australia. Correspondence and requests for materials should be addressed to Y.H. (email: Y.Hong@latrobe.edu.au) or to D.M.H. (email: dhatters@unimelb.edu.au)

The synthesis of new protein chains and their subsequent folding and targeting to specific cellular locations is one of the core activities of a cell. Because proteins typically fold inefficiently or display high inherent tendencies to misfold and aggregate, the folding process is strictly guided by quality control machinery¹. Small globular proteins may fold and unfold reversibly and spontaneously². However, complex proteins usually do not and this includes about two-thirds of the proteome³. In particular, proteins that fold in the endoplasmic reticulum (ER) need quality control assistance, including being directed through quality control checkpoints in the calnexin cycle^{3, 4}.

The quality control process and collective management of protein folding and turnover is defined as 'proteostasis', and involves a network of over 500 proteins in humans (as well as conserved systems in all life forms including bacteria and plants)⁵. When proteome foldedness is challenged through acute stresses, mechanisms are activated to improve baseline proteostasis capacity and to rectify imbalances. One example of this is the heat shock response whereby heat shock protein family chaperones are upregulated to help restore proteome foldedness⁶. Another example is the unfolded protein response which targets blocks in the folding pipeline of ER proteins⁷. Collectively these and other proteostasis mechanisms can buffer against short term strains on proteome folding.

In many diseases, especially neurodegeneration, proteostasis can become chronically imbalanced. Markers of this imbalance include expression of ER stress genes, upregulation of autophagy and inappropriate protein aggregation⁷⁻⁹. Indeed, chronic imbalance can be induced in animal models by ectopic expression of unstably folded proteins, which overwhelms the finite capacity of quality control reserves^{10, 11}.

One outstanding challenge in the field is determining the baseline efficiency with which cells maintain proteostasis. To date, the best approach to understand this is by expression of metastable proteins prone to aggregation, where the extent of aggregation is used as a proxy for proteostasis^{12, 13}. Here we describe the development of a strategy using a simple fluorogenic dye that can capture a snapshot of the balance of unfolded protein relative to folded states and does not require any ectopic expression of protein reporters. With purified proteins, and cellular models of proteome folding stress we validate this strategy and show three examples of how tetraphenylethene maleimide (TPE-MI) can measure proteostasis imbalance in disease contexts.

Results

TPE-MI is a turn-on fluorescent probe of unfolded proteins. We posited that under folding stresses cells will accumulate a backlog of unfolded proteins, consisting of proteins requiring

quality control to fold and/or already-folded proteins that are subjected to en masse unfolding. Hence, our strategy to monitor proteostasis efficiency was to label unfolded proteome. Non-disulphide bonded cysteine (i.e., with free thiols) is the least surface-exposed residue of all amino acids in proteins¹⁴. Thus, the abundance of accessible thiols should increase in cells as the proteome becomes more unfolded, enabling us to target these thiols as a proxy reporter for the level of unfolded proteins (Fig. 1a). Similar strategies have been used to study the protein folding state of purified globular proteins in vitro by measuring the loss of reactivity of free cysteines as they become buried in the core of proteins in the folded state, using the colorimetric Ellman's reagent (5,5'-dithio-bis-[2-nitrobenzoic acid])¹⁵. In cells, a key requirement for such a strategy is to use a dye that not only permeates the cell membrane but becomes fluorescent only once it has reacted with the target cysteines.

MI-functionalized TPE is a fluorescent dye with attributes that we predicted would make it suitable for this purpose (Fig. 1b). First, TPE-MI is inherently non-fluorescent until it is conjugated to a thiol via the maleimide¹⁶. This lack of fluorescence has been postulated to be due to the MI group quenching the TPE emission through the $n-\pi$ electronic conjugation of the carbonyl and olefinic groups, which are disrupted by thiol conjugation as evidenced by the blue-shift of the absorption maximum (Supplementary Fig. 1). TPE fluorescence is also contingent on its aggregation-induced emission (AIE) property whereby the fluorescence quantum yield is dependent on the motional restriction of the four phenyl rotamers of the TPE fluorophore that occurs in a rigid local molecular environment¹⁷.

To determine whether TPE-MI could be used to probe protein unfolding, we first investigated its reactivity to several model proteins: three that contain single free cysteines buried in the core of the folded state (bovine β -lactoglobulin, yeast enolase and human peroxiredoxin-3) and one negative control (for specificity of the reaction) that lacks cysteine (ubiquitin). Each of the proteins that contain a buried free cysteine thiol displayed far greater reactivity to TPE-MI when unfolded with guanidine hydrochloride, consistent with TPE-MI being selectively reactive with thiols that become exposed when globular proteins are unfolded (Fig. 2a-c). Ubiquitin showed no reactivity with TPE-MI, which demonstrated the specificity of the dye for thiols (Fig. 2a). Specificity for cysteine was further demonstrated by adding the competitive thiol reactant, *N*-methylmaleimide (NMM), in combination with TPE-MI to denatured β -lactoglobulin, which resulted in no increase in TPE fluorescence (Fig. 2d). TPE-MI alone, or protein alone also showed no changes in fluorescence signal (Supplementary Fig. 2).

Inside cells, glutathione (GSH) represents a major pool of non-protein thiols. It is present at concentrations of ~ 5 mM¹⁸,

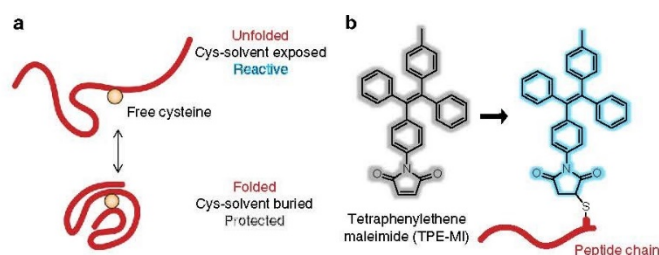


Fig. 1 Strategy for assaying protein foldedness via access to buried cysteine thiols. **a** Strategy for probing free cysteine thiols that become exposed to the solvent upon protein unfolding and permissive to maleimide reaction. **b** Structure of tetraphenylethene (TPE) conjugated to a maleimide (MI). Fluorescence is enabled upon conjugation to a protein and immobilisation of the phenyl rotamers.

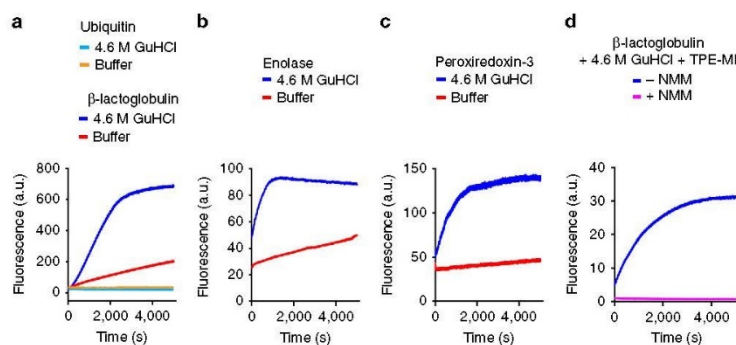


Fig. 2 TPE-MI preferentially reacts with buried cysteine thiols in unfolded proteins and switches on fluorescence. Shown are representative reaction kinetics for four proteins (note that the absolute fluorescence values between graphs here and in other figures cannot be compared to each other due to differences in instrument settings). **a** Bovine β -lactoglobulin, which contains five thiols: one buried free thiol and four disulphide-linked thiols; and bovine ubiquitin, which contains no thiols. **b** *Saccharomyces cerevisiae* enolase, which contains one buried free thiol and **c** human peroxiredoxin 3, which contains three thiols: one buried free thiol and two disulphide-linked surface-exposed thiols. Proteins were suspended in 100 mM sodium phosphate, pH 7.4, alone or supplemented with guanidine hydrochloride (GuHCl) to induce denaturation. At the start of the reaction 50 μ M TPE-MI was added. **d** Same design but with 50 μ M *N*-methylmaleimide (NMM) added before the addition of TPE-MI.

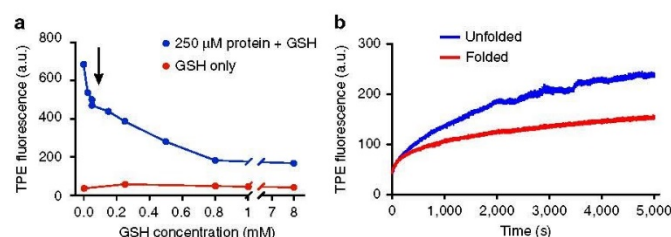


Fig. 3 TPE-MI fluorescence is not activated by glutathione reaction and can detect increases in unfolded protein load in cell lysates. **a** Reactivity of TPE-MI (50 μ M) with β -lactoglobulin (250 μ M) in the presence of the small thiol containing peptide glutathione (0.025 mM–8 mM GSH) and 4.6 M GuHCl. Fluorescence intensity of TPE-MI with different concentrations of GSH is shown for comparison (measured in 4.6 M GuHCl). Arrow indicates 7:3 protein thiol to glutathione thiol ratio found intracellularly. **b** Reaction kinetics of 50 μ M TPE-MI in the presence of intact and unfolded cell lysates from mouse Neuro2a neuroblastoma cells (0.5 mg ml⁻¹ total cellular protein). Shown are 'native' cell lysate in 100 mM sodium phosphate, pH 7.4 and 'unfolded' cell lysate containing 4.6 M GuHCl.

representing a GSH thiol to protein thiol molar ratio of approximately 3:7¹⁹. Hence GSH provides a major potential competitor that could interfere with the application of TPE-MI to probe the unfolded protein load. However, we discovered that GSH-conjugated TPE-MI is non-fluorescent (Fig. 3a and Supplementary Fig. 3a) (Supplementary Fig. 4a, b) confirmed the successful conjugation of GSH to TPE-MI. This contrasts with a previous report that TPE-MI conjugated to GSH is highly fluorescent¹⁶. However, in that case the GSH-TPE-MI conjugate was assessed in the solid phase¹⁶. Hence, it appears that in aqueous solution the phenyl rotamers of TPE-MI remain mobile when conjugated to the small GSH tripeptide. In contrast, it appears that proteins provide sufficient rigidity to enable AIE. This may arise from localised clustering of the phenyl rotamers with nearby hydrophobic amino acid sidechains that typically comprise the core of globular proteins. Support for this mechanism was revealed by time-resolved fluorescence measurements, which showed that TPE-MI emission decayed much faster when conjugated to GSH compared to the unfolded protein conjugate (Supplementary Fig. 5 and Supplementary Table 1). Collectively, these results suggested that in the cellular context GSH will have only a small dampening effect on the fluorescence

yields arising from TPE-protein thiol conjugates. Indeed the fluorescence yield of β -lactoglobulin remained high in the presence of glutathione at relevant cellular ratios of proteins and glutathione thiols (Fig. 3a). Further supporting this conclusion is the observation that denaturation of proteins in whole cell lysate with guanidine hydrochloride increased the reactivity with TPE-MI despite the presence of GSH (Fig. 3b).

TPE-MI detection of altered unfolded protein load in cells.

Next we assessed the behaviour of TPE-MI in live cells. TPE-MI exhibited a homogeneous cytoplasmic labelling pattern in live HeLa cells, with a lower level of labelling in the nucleus and apparent concentration in the region of the ER, which was anticipated as a major location for protein synthesis and folding (Fig. 4a). This indicated a good permeation of the plasma membrane and accessibility of the reagent to the intracellular proteome. Pre-treatment of the cells with NMM, which is known to permeate the cell membrane and block intracellular thiols^{20–22}, led to a reduction in labelling, suggesting that the TPE-MI fluorescence is associated with specific modification of thiol groups in the cell (Supplementary Fig. 6). TPE-MI was not

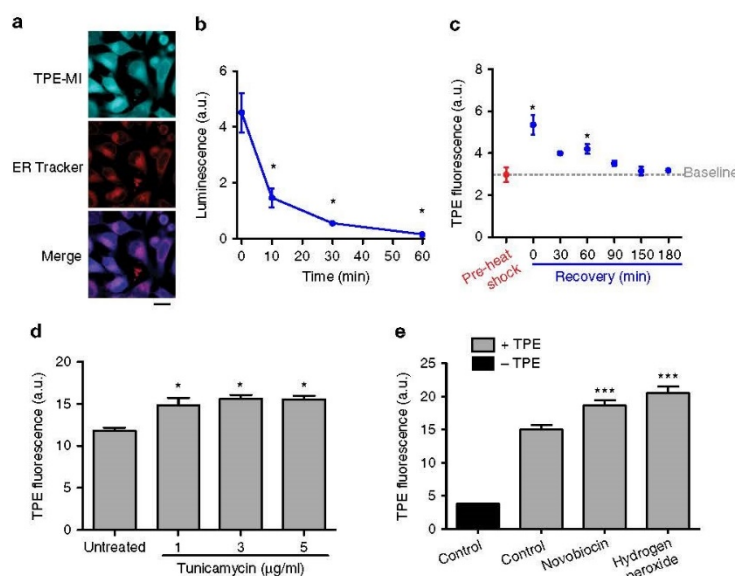


Fig. 4 TPE-MI fluorescence increases in accordance with diverse proteome stresses. **a** Confocal microscopy images for TPE-MI staining of HeLa cells with ER Tracker counterstain. Cells were fixed in 4% (v/v) paraformaldehyde for 15 min after staining with 50 μ M TPE-MI. Scale bar, 20 μ m. **b** Unfolding of the proteome by heat shock as assessed by denaturation of *Renilla* luciferase and luciferase activity. Data show luciferase luminescence of lysates taken after heat shock at 42 °C of HeLa cells transfected with *Renilla* luciferase for the times shown, $n = 5$; mean $\pm$ s.e.m. **c** Heat shock treatment of HeLa cells at 42 °C for 45 min and subsequent time course of recovery at 37 °C. Data show median TPE-MI fluorescence measured by flow cytometry of cells stained at the indicated time points, $n = 5$ technical replicates; mean $\pm$ s.e.m. Statistics for **b**, **c** were calculated between test groups and basal groups or zero time point. The details of the statistics are provided in Supplementary Data 1 where $P < 0.05$ (indicated by *). **d** Effect of overnight tunicamycin treatment on HeLa cells. Data show median fluorescence intensity of TPE-MI stained cells measured by flow cytometry, $n = 5$ biological replicates; mean $\pm$ s.e.m. **e** Effects of hsp90 inhibitor novobiocin (800 μ M for 6 h) and free radical generator, hydrogen peroxide (100 μ M for 1 h) on HEK293 cells. Data show means $\pm$ s.d. for five replicates. *** $P < 0.0001$. See Supplementary Data 1 for full statistics.

noticeably toxic to cells at concentrations and time points used in our assays (or at higher concentrations) as measured by death rates (Supplementary Fig. 7).

To test how effective TPE-MI was in monitoring unfolded protein load in live cells, we applied a variety of distinct stresses to proteome foldedness. First, we applied a 42 °C heat shock pulse to thermally unfold the proteome²³. As anticipated, the heat shock led to an increase in non-functional and unfolded proteins within ~30 min, as measured by a luciferase activity reporter assay (Fig. 4b; full details of statistical tests in Supplementary Data 1). TPE-MI treatment (on aliquots of cells) revealed an immediate increase in reactivity after heat shock. However, upon recovery at 37 °C, the levels of reactivity (on aliquots of cells) returned to baseline levels within about 150 min, which is in accord with the anticipated time for HeLa cells to recover from heat shock²⁴ (Fig. 4c). This result indicated that TPE-MI can measure net increases in the unfolded protein load and recovery from thermal denaturation. Other stresses also increased TPE fluorescence in accord with an anticipated accumulation of unfolded protein (Fig. 4d, e). These included tunicamycin, which is a glycosylation inhibitor and leads to a backlog of unfolded protein in the ER that cannot complete the calnexin cycle²⁵, inhibition of hub chaperone Hsp90 with novobiocin, which impairs Hsp90 activity without activating the heat shock response^{26, 27} and hydrogen peroxide, which confers damage by free radical reactions.

Next we tested whether TPE-MI could be used to track foldedness of individual proteins by proteomics approaches.

For these experiments we used tunicamycin treatment as the folding stress. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) profiles of cell lysates revealed tunicamycin to induce increased TPE reactivity of a broad range of proteins, including a prominent protein at ~50 kDa (Fig. 5a, b). As a control for specificity, we investigated a non-alkylating analogue of TPE-MI, which did not stain either purified proteins with a free cysteine thiol, or change fluorescence in cells treated with tunicamycin (Supplementary Fig. 8a–c). For deeper insight we next performed a quantitative proteomics analysis using a matched pair design with the method of stable isotope labelling with amino acids in cell culture (SILAC)²⁸. For this approach, we labelled control and tunicamycin-treated cells with TPE-MI (as the matched pairs) and then assessed the relative loss of peptides containing unmodified Cys residues from the mass spectrum. By performing nano-liquid chromatography-tandem mass spectrometry (LC-MS/MS), which was sufficient to sample a selection of the most abundant proteins of the proteome, we identified 154 Cys-containing proteins in all five replicates (Supplementary Data 2 for full proteomics summary). The median ratio of Cys-containing peptides in the tunicamycin/control conditions was significantly lower (by ~6%) than the ratio of non-Cys-containing peptides, which is consistent with tunicamycin treatment leading to a greater backlog of unfolded proteins (Fig. 6a).

Analysis of the individual Cys-containing peptides revealed a number of proteins that exhibited very large increases in

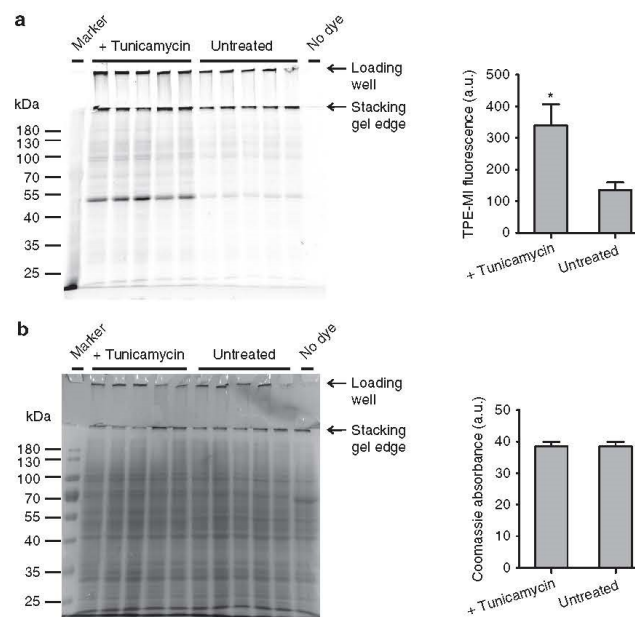


Fig. 5 Detection of proteome-wide Cys-reactivity upon tunicamycin treatment. **a** In-gel fluorescence of TPE-MI stained Neuro2a cell lysates by SDS-PAGE. Cells were treated with tunicamycin ($1\mu\text{g ml}^{-1}$, overnight) or vehicle control, then labelled with $50\mu\text{M}$ TPE-MI for 30 min before lysis. $40\mu\text{g}$ total cellular protein was loaded on the gel. Graphs show quantitation of fluorescence intensities in each lane ($n = 5$ biological replicates; mean $\pm$ s.e.m.; *, $P < 0.05$ —see Supplementary Data 1 for full statistics). **b** Corresponding Coomassie stained gel of **a**, and quantification. Graph shows quantitation of the Coomassie intensities ($n = 5$; mean $\pm$ s.e.m.; see Supplementary Data 1 for full statistics).

reactivity upon TPE-MI treatment following tunicamycin stress (Supplementary Fig. 9 and Fig. 6b). The six peptides with the greatest increase in reactivity came from THRB (prothrombin), G3P (glyceraldehyde-3-phosphate dehydrogenase (GAPDH)), ACTG (actin, cytoplasmic 2), FHL1 (four and a half LIM domains protein 1), FBRL (rRNA 2'-O-methyltransferase fibrillarin) and ENOB (β -enolase). These proteins have diverse and unrelated functions and none have known roles in thiol exchange mechanisms. The prominent protein observed by SDS-PAGE at $\sim 50\text{ kDa}$ may be ENOB, which has a molecular mass of 47 kDa and is highly abundant in cells (in the top 5% of protein abundance based on the PaxDb Protein Abundance database²⁹). Of note, ACTG contains four Cys-containing peptides that were detected in the data set, but only one of these (peptide 1–19) displayed a major increase in reactivity (Fig. 6c). Most noteworthy a very similar peptide (2–19), differing only by its lack of the amino terminal methionine, showed no change in TPE reactivity. Since N-terminal methionines are routinely removed from most proteins after synthesis³⁰, these data suggest that

TPE-MI selectively labels newly synthesised ACTG upon tunicamycin treatment, which presumably is incompetent to fold.

One cysteine residue in a peptide from ERF1 (residues 333–344) was notably protected from TPE reactivity under tunicamycin stress. ERF1 (eukaryotic release factor 1) is involved in releasing newly synthesised peptides from the ribosome during translation³¹. ERF1 was also fourfold less abundant (as measured from the non-Cys-containing peptide ratios) following tunicamycin treatment, and in light of lower levels of ERF1 reducing translation rates³², this would appear to be a protective response

to suppress protein synthesis. As part of its function, ERF1 forms a complex with cofactor ERF3 via a domain (residues 270–437) that encompasses the protected cysteine residue^{33, 34}. Hence, the reduction in TPE-MI reactivity may indicate a conformational change or binding event associated with a change in the function of ERF1 to reduce protein expression under stress.

Soluble mutant Huntingtin increases the unfolded proteome.

We examined whether TPE-MI could report on the impact of mutant protein expression on the global unfolded protein load in cells. Prior work has suggested that the expression of conformationally destabilised and/or aggregation-prone mutant proteins associated with neurodegenerative diseases can lead to a destabilization of proteostasis^{11, 12, 35}. For this experiment we examined the effect of Huntingtin, which when mutated to contain a polyglutamine sequence that is expanded beyond 36 contiguous residues, causes Huntington disease (HD)³⁶. We expressed the exon 1 fragment (Httex1), which is sufficient to invoke HD-like disease in mouse models³⁷, as a fusion to mCherry. At the high expression level, the mutant 97Q form of Httex1 was associated with an elevated TPE-MI fluorescence signal relative to a non-disease-causing 25Q form of Httex1 (Fig. 7a). TPE-MI did not directly label Httex1 (at least the aggregates), which does not contain any cysteine and hence the changes observed in TPE-MI reactivity are not due to direct labelling of the ectopic protein (Supplementary Fig. 10). This result suggests that mutant proteins associated with neurodegenerative disease can strain proteostasis capacity and lead to a backlog of unfolded proteins in the cell.

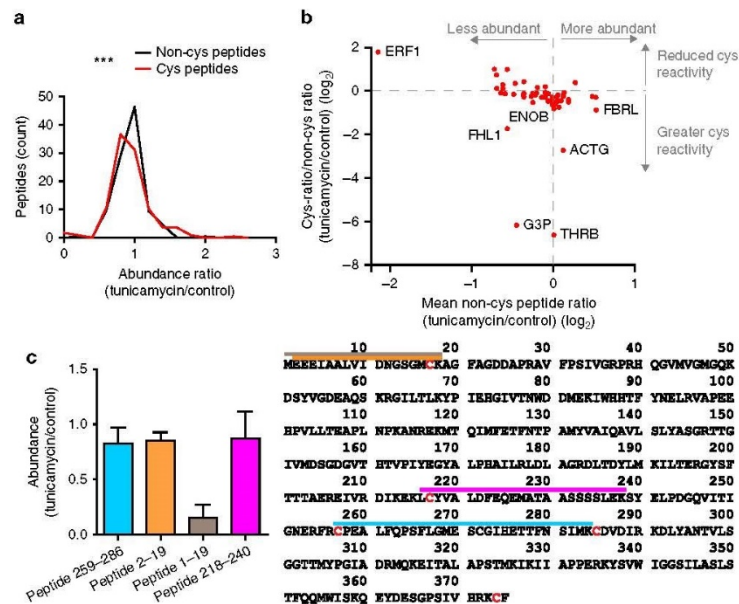


Fig. 6 Tunicamycin-mediated block of folding selectively alters cysteine reactivity to TPE-MI. Data relates to Neuro-2a cells treated with TPE-MI (30 min; 200 μ M) after overnight tunicamycin pre-treatment (1 μ M) versus an untreated control. **a** Histogram of the peptide abundance ratios between tunicamycin-treated cells and control cells differentially labelled using SILAC. *** $P < 0.0001$ between two groups assessed by Wilcoxon matched-pairs signed rank test. See Supplementary Data 1 for full statistics. Differences indicate a tunicamycin-mediated loss of Cys residue-containing peptides attributable to increased TPE reaction. **b** Plot of peptide abundances (calculated from non-Cys peptide ratios) versus change of Cys reactivity (ratios of Cys peptides normalised to abundance by the non-cys peptides within the same protein) attributable to tunicamycin treatment. Shown are only peptides that display significant changes in Cys reactivity due to tunicamycin treatment ($P < 0.05$ measured by Student's *t*-test; details in Supplementary Data 2 and Supplementary Fig. 9). **c** Case study of peptides derived from ACTG. Shown are the mean $\pm$ s.d. peptide abundance ratios of five replicates (left) versus where the peptides map on the sequence of ACTG (right). Colours on bar graph are coordinated to the sequence map and cysteine residues are shown in red on the sequence.

The formation of visible protein aggregates that is associated with neurodegenerative disease has been previously suggested to reflect an adaptive, sequestration response to the stress exerted by mutant proteins^{38–40}. Our TPE-MI probe provides a unique tool to explore the relationship between protein aggregation and the backlog of unfolded (or misfolded) protein in the cell. To investigate this relationship, we applied our previously developed flow cytometry approach, Pulse Shape Analysis (PulSA), to separate cells with visible Httex1 aggregates (inclusion (i) population) from cells with unaggregated Httex1 protein (non-inclusion (ni) population)⁴¹. Classification of the cells into the ni and i groups revealed that, at the lower levels of expression, cells without visible inclusions exhibited with a much higher TPE-MI signal than cells with inclusions (Fig. 7b). Conversely the cells with 97Q Httex1 aggregates had a TPE-MI signal below even the 25Q counterpart at lower levels of expression. Only at the highest levels of expression did cells with visible inclusions become highly elevated in TPE-MI fluorescence to levels similar to cells lacking aggregates. Collectively, these data are consistent with the Httex1 monomer placing the greatest strain on proteostasis leading to the rest of the proteome becoming backlogged with unfolded protein. In contrast, the very low level of TPE-MI fluorescence after an inclusion has formed suggested that inclusion formation removes the backlog of unfolded protein load.

To further explore the role that mutant Huntingtin has in proteostasis imbalance, we examined TPE-MI reactivity of induced pluripotent stem cells derived from HD patients differentiated into primitive neural stem cells (pNSCs)⁴². A HD cell line encoding 180Q in the endogenous Huntingtin protein (ND36999) showed greater induction of TPE-MI reactivity than a 'wild-type' cell line with 33Q (ND36997) upon stressing proteostasis in the ER and autophagy pathways with thapsigargin (Fig. 7c). Because these two lines come from genetically different backgrounds, we also tested TPE-MI reactivity in the 180Q line gene-edited back to a 'wild-type' polyQ length (18Q) by CRISPR/Cas9 and homologous recombination⁴². Three independently corrected lines all showed reduced induction of TPE reactivity when challenged with thapsigargin. These results all suggest HD cells have a reduced ability to prevent the build-up of unfolded proteins under stress and provide a further demonstration of how TPE-MI can probe the breakdown of proteostasis dynamics in disease.

DHA promotes proteome unfolding in malaria parasites. Our last experiment was to test TPE-MI as a probe for proteome destabilization in a completely different biological setting. Dihydroartemisinin (DHA) is a frontline drug for the treatment of malaria⁴³. DHA has been postulated to be toxic to malaria

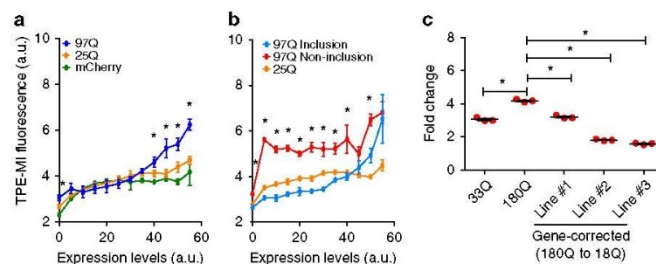


Fig. 7 Backlog of unfolded protein in Neuro-2a cells triggered by mutant Huntingtin protein. **a** Impact of Htt exon 1 (Httex1) expression on TPE-MI fluorescence comparing the wild-type polyQ length (25Q) with a severe mutant (97Q) as fusions to mCherry. Cells were transfected and analysed by flow cytometry with the data binned into different expression levels based on mCherry fluorescence. Data show $n = 3$ technical replicates; mean $\pm$ s.e.m. *represents significant differences between 25QHttex1 and 97QHttex1 ($P < 0.05$). **b** Reanalysis of the data in panel a by PulSA to separate the cells with visible Httex1 inclusions from those without. Data show $n = 3$ technical replicates; mean $\pm$ s.e.m. *represents significant differences between 97QHttex1 inclusion and 97QHttex1 non-inclusion populations. **c** TPE reports on stress in primitive neural stem cells derived from induced pluripotent stem cells from Huntington patients. Shown are Huntington disease human neural cell lines encoding variable polyQ lengths in the *Huntingtin* gene. The gene corrected lines were gene-edited by CRISPR-Cas9 technology. Cells were analysed by flow cytometry and show fold change in ER stress induced by 1 μ M thapsigargin relative to vehicle (dimethyl sulfoxide). Data represent $n = 3$ technical replicates, mean $\pm$ s.e.m. *represents significant differences. The details of the statistics in this figure are provided in Supplementary Data 1.

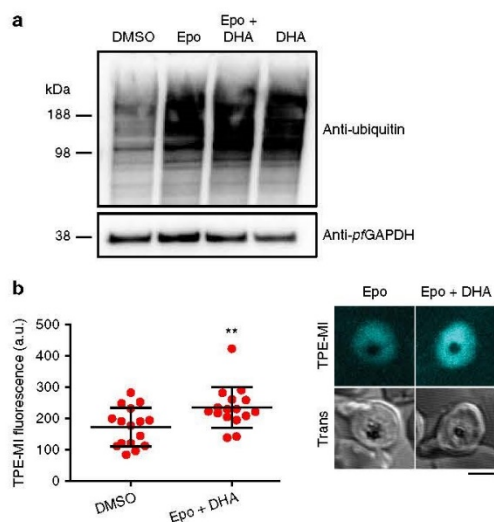


Fig. 8 TPE reports on accumulation of unfolded proteins in malaria parasites (*P. falciparum*) treated with dihydroartemisinin (DHA) and epoxomicin. **a** Western blot of trophozoite stage *P. falciparum* treated with DHA and the proteasome inhibitor, epoxomicin (Epo) or DMSO control, for 3 h. **b** TPE-MI fluorescence values from individual trophozoites within red blood cells, measured by confocal microscopy. (left) Bars represent the mean $\pm$ s.d. *represents a significant difference. The details of the statistics in this figure are provided in Supplementary Data 1. (right) A representative set of images. Scale bar, 3 μ m.

parasites by promiscuously damaging the proteome through free radical production, which leads to non-specific damage to proteins, resulting in a build-up of damaged and unfolded proteins and activation of stress responses^{44–47}. We tested whether we could use TPE-MI as a probe of this imbalance in proteostasis. *Plasmodium falciparum* trophozoite-infected red blood cells were

treated with DHA in the presence of the proteasome inhibitor, epoxomicin (to suppress degradation of damaged proteins). This treatment led to an increase in total ubiquitinated proteins as previously described (Fig. 8a)⁴⁷. DHA treatment led to significant elevations in TPE-MI reactivity over baseline conditions (Fig. 8b, c)—and also when compared to epoxomicin treatment alone (Supplementary Fig. 11). This result further supports the conclusion that DHA leads to a backlog of unfolded proteins as the basis for triggering the stress response.

Discussion

Here we describe the use of the fluorogenic dye TPE-MI to probe for unfolded proteins by assaying the unmasking of cysteine thiols that are buried in the folded state of globular protein domains. This is an important step for understanding proteostasis mechanisms because unfolded proteins are highly permissive to misfolding and forming aggregates, which are a key end-product feature of neurodegenerative diseases^{48, 49}. We validated the method with purified proteins and showed it can report on increases in unfolded protein load under diverse proteostatic stresses: heat shock, blockage of folding processing in the ER, hsp90 inhibition and free radical production.

Even though we expected TPE-MI to label surface-exposed cysteines which would contribute to the overall fluorescence yield, it was remarkable that we could detect a significant overall increase in fluorescence when we stressed proteome foldedness. This was especially apparent in the use of TPE-MI for proteomics-based analyses of unfolding. Our experiment was tailored for a simple assessment of the most abundant proteins in the proteome (by virtue of the 1-dimensional LC-fractionation step), but the method should be adaptable to broader (i.e., deeper) proteome coverage (e.g., by using 2-dimensional LC-based proteome fractionation methods), which offers the potential to systematically understand how the free cysteines change under proteome stresses. Furthermore, additional investigations of the TPE-MI concentration and time dependence of reactivity will likely capture a greater list of proteins of altered conformational status which could be most powerful for measuring the kinetic process of folding in live cells. The value of this capacity is highlighted by a recent study that used limited proteolysis on cell lysates to measure the thermodynamic stability of the proteome

upon chaotrope- and thermally mediated denaturation⁵⁰. Our method here provides in principle a capacity to extend this approach by measuring folding kinetics of proteins that contain free cysteine thiols in live cells. Hence, more broadly, we believe TPE-MI offers the capacity to perform both simple and rapid tests of proteostasis health in essentially any cell-based format—as well as more elaborate assays of folding of individual proteins containing cysteine residues.

Application of TPE-MI to probe proteostasis imbalance in HD cell models yielded insight into the early imbalance events that precede the formation of visible aggregates. In the induced pluripotent stem cell (iPSC)-derived neural cell lines, imbalances were reflected in an attenuated capacity of cells to suppress accumulation of unfolded proteins upon interference of proteostasis with thapsigargin. Prior studies have indicated problems in polyQ-expanded knock-in Huntington cell lines in energy production, which is concordant with a broader problem in proteostasis⁵¹. Other work has also implicated proteostasis imbalance as crucial in Huntington Disease (reviewed in ref. ⁵²).

One particularly interesting finding from our studies was the dramatic loss of TPE-MI reactivity in cells with inclusions. Prior work has suggested that misfolded proteins (which arise from unfolded proteins) can form two classes of visible aggregates for processing misfolded proteins called the IPOD and JUNQ^{39, 40}. While it is not clear whether the formation of IPOD or JUNQ structures alleviates proteostasis stress, prior studies have suggested that cells that form Httex1 inclusions, which partition into the IPOD structure, have a higher survival rate than cells retaining diffuse Httex1, when standardized to expression level³⁸. Other evidence to support this finding is that soluble mutant Httex1 is actively recognised as abnormal and preferentially targeted for degradation compared to the wild-type Httex1 counterpart^{53, 54}. In addition, it has been reported that ER stress pathways are activated by soluble pools of Httex1 before inclusion formation⁵⁵. We recently found that when Httex1 forms inclusions, the cells enter a state of quiescence which deactivates the trigger of apoptosis from soluble Httex1 and switches cell death to a slower necrotic mechanism⁵⁶. As inclusions matured there was a progressive coaggregation of prion-like proteins from the proteome⁵⁶. Hence our findings raise the interesting possibility that soluble Httex1 strain proteostasis before inclusions are formed, which promotes the accumulation of unfolded proteins from the broader proteome. Thus the backlogged unfolded proteome may become co-aggregated en masse into the inclusion as quiescence is established and mask many of the exposed cysteine residues.

In summary, our new approach to measure unfolded load provides insight into proteostasis capacity in cells and opens pathways to study mechanisms of protein quality control. In addition, the approach provides possibilities for the development of biomarkers for early stage diagnosis of neurodegenerative and possibly other diseases.

Methods

Materials. Tetrahydrofuran (THF, Labscan) was purified by distillation from sodium benzophenone ketyl under nitrogen prior to use. Maleic anhydride, sodium acetate, sodium carbonate, glutathione (GSH), guanidine hydrochloride (GuHCl), bovine β -lactoglobulin, yeast (*Saccharomyces cerevisiae*) enolase, bovine ubiquitin, NMM, tunicamycin and other reagents were purchased from Sigma-Aldrich and used as received. Human peroxiredoxin 3 was expressed and purified as described previously⁵⁷, and maintained in buffer with reducing agent tris(2-carboxyethyl) phosphine. Peroxiredoxin 3 was freshly desalted to remove reducing agents before use with a PD-10 column (GE Healthcare). *Renilla* luciferase vector was used from the Qiagen Signal Heat Shock Response Reporter (luc) Kit. Luciferase activity was measured using Dual-Luciferase Reporter Assay System (Promega), as per the manufacturers protocol. The Httex1 fusion constructs were prepared as described previously⁵⁸. All constructs were sequenced for verification.

Synthesis of TPE-MI. The precursor, 4-(1,2-diphenyl-2-(*p*-tolyl)vinyl)aniline (TPE-NH₂), was synthesised according to the previously reported procedure¹⁶. To a solution of TPE-NH₂ (0.361 g, 1.0 mmol) in dry THF (20 ml), maleic anhydride (0.196 g, 2.0 mmol) was added and stirred at room temperature for 12 h. The solvent of the resulting mixture was evaporated under reduced pressure and a yellow solid was obtained as an intermediate, which directly underwent further reaction without purification. Anhydrous sodium acetate (0.205 g, 2.5 mmol) was added into a solution of the intermediate in acetic anhydride (10 ml) and the mixture was stirred at 75 °C for 6 h. After cooling to ambient temperature, the reaction mixture was quenched with 20% Na₂CO₃ aqueous solution and extracted with dichloromethane three times. The organic layer was collected and washed by brine for three times. The crude product was concentrated and purified by silica-gel chromatography to furnish a light yellow solid (0.360 g). Yield = 82%. ¹H NMR (400 MHz, CDCl₃): δ (tetramethylsilane (TMS), p.p.m.) 7.12–7.02 (m, 14 H), 6.93–6.90 (m, 4 H), 6.81 (d, *J* = 3.2 Hz, 2 H), 2.27 and 2.26 (two singlets, 3 H). ¹³C NMR (100 MHz, CDCl₃): δ (TMS, p.p.m.) 168.83, 143.13, 142.92, 142.76, 141.07, 139.90, 139.68, 138.74, 135.77, 135.59, 133.53, 131.31, 131.27, 130.78, 130.74, 130.70, 130.67, 130.55, 128.61, 127.95, 127.75, 127.16, 127.13, 127.07, 126.97, 126.01, 125.82, 20.59, 20.56. High-resolution mass spectrometry (HRMS) (matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF), *m/z*): [M]⁺ calcd. for C₃₁H₂₃NO₂, 441.1729; found, 441.1734.

Synthesis of GSH-TPE-MI. TPE-MI was synthesised and characterised as previously reported¹⁶. In essence, to a solution of TPE-MI (5 mg) in THF (4 ml), glutathione (GSH, reduced, 5 mg) in water was added dropwise with stirring. The reaction mixture was further stirred at room temperature for 12 h and the solvent was evaporated under reduced pressure. The final product was obtained after purification using high-performance liquid chromatography (HPLC) and dried under vacuum as white powders (6.7 mg). Yield = 80%. HPLC (λ = 330 nm): methanol as eluent, retention time 1.441 min. HRMS (ESI, *m/z*): [M-H][−] calcd. for GSH-TPE-MI, 747.8359; found, 747.2462.

Synthesis of non-alkylating TPE analogue. The non-alkylating TPE analogue, 1-(*p*-bromophenyl)-2-(*p*-hydroxyethyl)-1,2-diphenylethylene was synthesised as described in ref. ⁵⁹. In essence, synthesis of 2-(4-(2-(4-bromophenyl)-1,2-diphenylvinyl)phenoxy)ethanol-1-ol (1): 4-(2-Hydroxyethoxy)benzophenone (2) was synthesised according to previous literature⁶⁰. To a solution of 2 (2 mmol, 484 mg), 4-bromobenzophenone (4 mmol, 1.04 g) and zinc powder (14.4 mmol, 941 mg) in dry THF, TiCl₄ (7.2 mmol, 0.79 ml) was added dropwise at −78 °C. Then the reaction mixture was warmed to ambient temperature and heated to reflux for 10–12 h. After cooling down, the reaction mixture was filtered and extracted with dichloromethane three times. The organic layer was combined, dried (MgSO₄) and concentrated under reduced pressure. The crude product was purified by silica gel chromatography with hexane/ethyl acetate (9:1) as eluent to afford the desired product (280 mg, 0.6 mmol) as pale yellow solid. Yield = 30%. ¹H NMR (400 MHz, CDCl₃): δ (TMS, p.p.m.) 7.25–7.19 (m, 2 H), 7.14–7.09 (m, 6 H), 7.04–6.99 (m, 4 H), 6.95–6.86 (m, 4 H), 6.69–6.64 (m, 2 H), 4.05–3.99 (m, 2 H), 3.94–3.91 (m, 2 H). ¹³C NMR (100 MHz, CDCl₃): δ (TMS, ppm) 156.67, 156.57, 142.94, 142.84, 142.74, 142.33, 142.24, 140.40, 138.31, 138.28, 135.65, 135.54, 132.36, 131.93, 131.90, 130.66, 130.64, 130.29, 130.17, 127.24, 127.22, 127.11, 127.03, 126.98, 126.05, 125.95, 125.89, 125.63, 119.64, 113.20, 113.03, 76.71, 76.39, 76.08, 68.34, 68.30, 60.82. HRMS (MALDI-TOF), *m/z* calcd. for C₂₈H₂₃BrO₂, 470.0881; found 470.0845 (M⁺).

Instrument used for compound characterisation. ¹H and ¹³C NMR spectra were measured on a Bruker ARX 400 NMR spectrometer using chloroform-*d* as the deuterated solvent with tetramethylsilane (TMS; δ = 0) as the internal standard. Mass spectrum of non-alkylating TPE analogue was run by a Waters[®] Micromass[®] MALDI micro MX[™] Mass Spectrometer operating on the reflectron mode with DCTB (*trans*-2-[3-(4-*tert*-Butylphenyl)-2-methyl-2-propenylidene]malononitrile) as matrix. Mass spectrum of GSH-TPE-MI was recorded on an Agilent Technologies 6520 Accurate-Mass Q-TOF LC/MS operating in an ESI negative ion mode. Ultraviolet-visible (UV-Vis) absorption spectra were measured on Cary 50 UV-Vis spectrometer. Steady-state fluorescence signals were recorded on a Cary Eclipse fluorimeter.

In vitro protein and lysate unfolding assays. Stock solutions of TPE-MI (1 mM) and proteins were prepared by dissolving appropriate amounts of the dye in dimethyl sulfoxide (DMSO) and the protein in PBS, respectively. In the unfolding experiments, protein (250 μ M for β -lactoglobulin and ubiquitin; 50 μ M for enolase; 5 μ M for peroxiredoxin; 0.5 mg ml^{−1} for the cell lysate) was incubated in 4.6 M GuHCl solution for 3 min before the addition of 50 μ M TPE-MI into the mixture. Time-course fluorescence intensity was recorded by using the excitation and emission wavelengths of 350 and 470 nm, respectively, and time interval of 1 s. GSH competition experiment was performed by mixing an appropriate amount of protein and GSH in 4.6 M GuHCl for 3 min. Fluorescence intensity at 470 nm was recorded immediately after the addition of 50 μ M TPE-MI into the mixture.

Cell lysate preparation. 7×10^5 HeLa cells were seeded in six-well plates. On the following day, cells were washed once in PBS and lysed by repeated pipetting in 300 μ l native lysis buffer per well (20 mM Tris pH 8, 2 mM $MgCl_2$, 1% Triton X-100, 1 EDTA-free Complete protease inhibitor tablet (Roche)/10 ml, 150 μ M NaCl, 20 Units ml^{-1} Benzonase (EMDMillipore)). Protein concentrations were determined by bicinchoninic acid assay using bovine serum albumin as mass standard.

Time-resolved fluorescence measurements. Fluorescence decay measurements were performed by the time-correlated single-photon counting technique using the frequency doubled output (380 nm) of a mode-locked and cavity-dumped Titanium:sapphire laser (Coherent Mira 900f). The repetition rate of the laser pulses was reduced to 5.4 MHz (APE PulseSwitch). Emission was collected under magic angle conditions, spectrally selected (470 nm) using a Jobin Yvon H20 monochromator and detected with a microchannel plate photomultiplier (Eldy EMI-132-1). Data were acquired using a Becker and Hickl SPC-150 TCSPC card and SPCM software, and analysed using the FAST (Fluorescence Analysis Software Technology) software (Edinburgh Photonics).

Mammalian cell line culturing and transfection. Neuro-2a, HEK293 and HeLa cell lines (from lab cultures originally obtained from ATCC) were used in this study and tested and cleared for mycoplasma. Cells were not tested for cross-contamination of other cell lines or misidentification. We justified their use in context of this in that the cell-line specific biology is not important to the general outcomes of the experiments. HeLa and HEK293 cells were maintained in Dulbecco's modified Eagles Medium (DMEM) supplemented with 1 mM glutamine, 100 Units ml^{-1} penicillin-streptomycin and 10% v/v foetal bovine serum. Cells were cultured at 37 °C in a humidified incubator with 5% atmospheric CO_2 . For microscopy experiments 7×10^4 HeLa cells were plated on eight-well μ -slides (ibidi). For flow cytometry, 2.8×10^5 HeLa cells were plated on 12-well plates. Neuro-2a cells were maintained in OptiMem (Life Technologies) supplemented with 10% v/v foetal calf serum, 1 mM glutamine, 100 units per ml penicillin and 100 μ g ml^{-1} streptomycin in a humidified incubator with 5% atmospheric CO_2 . For microscopy experiments 9×10^4 Neuro-2a cells were plated on 8-well μ -slides (ibidi). For flow cytometry, 4×10^5 Neuro-2a cells were plated on 12-well plates. Cells were transfected using Lipofectamine 2000 according to the manufacturer's instructions with 0.5 μ g DNA for the microscopy experiments and 1.6 μ g DNA for the flow cytometry. Media was refreshed daily after transfection.

hiPSC maintenance and differentiation. CAG33 human iPSC (hiPSCs) (ND36997) and CAG180 iPSCs (ND36999) harbouring *HTT* alleles with 33 and 180 CAG repeats, respectively, were obtained from the NINDS iPSC Repository at Coriell Institute. The expanded *HTT* allele in the CAG180 hiPSC line was corrected to 18 CAG repeats by homologous recombination to generate isogenic control hiPSC lines: isoHD-Corr-1, isoHD-Corr-2, and isoHD-Corr-3⁴². All hiPSCs were cultured on Matrigel-coated plates in mTeSR-1 medium (STEMCELL Technologies, #05850) in a cell culture incubator at 37 °C and 5% CO_2 . hiPSCs were induced into pNSCs according to a previously published protocol⁶¹. Briefly, hiPSCs at about 20% confluence were treated with N2B27 media (DMEM-F12/Neural Basal medium 1:1 with 1% N2, 2% B27, 1% pen/strep/ glutamine, 10 ng ml^{-1} hLIF, and 5 μ g ml^{-1} BSA) containing 3 μ M CHIR99021 (Tocris), 2 μ M SB431542 (Sigma) and 0.1 μ M compound E (Millipore) for the first 7 days. The culture was then split 1:3 for the next six passages using Accutase without compound E on Matrigel-coated plates. Cells at passage 5 were used for experiments.

TPE-MI cell staining. TPE-MI was dissolved in DMSO as 1 or 2 mM stocks. Stocks were kept at 4 °C in the dark. For experiments involving HeLa and Neuro2a cells, plated cells were rinsed with PBS and then treated with freshly diluted TPE-MI (50 μ M in PBS) for 30 min at 37 °C. TPE-MI solution was removed. For flow cytometry, cells were resuspended in PBS, pelleted by centrifugation (120 g for 6 min) and then resuspended in 250 μ l PBS in flow cytometry tubes. For imaging, cells were fixed on the plate with 4% (w/v) paraformaldehyde. For SDS-PAGE, the stained Neuro2a cells were lysed on the plate by addition of native lysis buffer and analysed by 12% acrylamide SDS-PAGE. Protein concentrations were determined by bicinchoninic acid assay with bovine serum albumin as mass standard. TPE-MI fluorescence was recorded on unstained gels with a Gel Imaging system (Bio-Rad). Then the gel was stained with Coomassie Blue for assay of total protein levels.

For experiments involving hiPSCs, pNSCs were treated with 1 μ M thapsigargin (Sigma) or DMSO for 16 hours. Treated cells were rinsed once with PBS and then incubated with freshly diluted TPE-MI (50 μ M in PBS) for 30 min at 37 °C. TPE-MI was removed and cells were detached using Accutase and filtered with 70 μ M cell screener for flow cytometry analysis.

Cytotoxicity test. 7×10^4 HeLa cells were plated onto 12-well plates. For the dose curves, TPE-MI was applied for 30 min. For the time dependent experiment, 100 μ M TPE-MI was applied for variable lengths of time. After the treatments, cells were detached in 400 μ l PBS and assessed for viability using the Countess Cell counter (Invitrogen) with Trypan blue staining.

NMM competition experiment. 4×10^5 Neuro-2a cells were plated on 12-well plates. The next day cells were washed with 1 ml PBS and then with PBS supplemented with NMM for 30 min. The media was replaced with 50 μ M TPE-MI for staining for 30 min. Cells were detached in 400 μ l PBS and assessed for fluorescence in a Cary Eclipse fluorimeter.

Tunicamycin and heat shock treatment. Tunicamycin was added to plated HeLa and Neuro2a cells before TPE-MI treatment. For heat shock, plated HeLa cells were placed in a water bath at 42 °C for variable time periods as indicated. To measure protein unfolding of *Renilla* luciferase, cells expressing Qiagen Signal Heat Shock Response Reporter (luc) plasmids were analysed for *Renilla* luminescence Dual-Luciferase Reporter Assay System (Promega), as per the manufacturers protocol. In this assay, *Renilla* luciferase was expressed via a constitutive CMV (cytomegalovirus) promoter and therefore changes in luminescence reflect loss of folding, rather than changes in expression. In the heat shock recovery experiments, media was changed after heat shock and cells returned to the 37 °C incubator.

Proteomics sample preparation. Neuro2a cells were cultured for 12 passages in SILAC-LYS6 (Catalogue Number: 89983, Thermo Fischer Scientific, San Jose, CA, USA) labelled media or unlabelled media. The SILAC-LYS6 labelled cells were treated with tunicamycin (1 μ g ml^{-1}) in T25 flasks overnight, then both SILAC-LYS6 labelled and unlabelled cells were treated with TPE-MI (200 μ M, 30 min). Cells were then detached using a cell scraper, lysed using RIPA buffer (50 mM Tris-HCl, 150 mM NaCl, Protease inhibitor, 1% v/v NP40, 2 mM EDTA and DNAase1), then centrifuged at 16,000 g for 10 min. The protein concentration in the supernatant was determined by a BCA assay (Catalogue number: 23225, Thermo Fischer Scientific) using BSA as the mass standard. Then, 50 μ g of the total protein in each supernatant (i.e., tunicamycin-treated and control) ($n = 5$) was combined in a 1:1 ratio and extracted using methanol-chloroform precipitation⁶². The protein pellet was solubilized in 100 μ l of 8 M urea, 50 mM triethylammonium bicarbonate, reduced using 10 mM tris(2-carboxyethyl)phosphine, pH 8.0, and alkylated with 10 mM iodoacetamide for 45 min, then subjected to trypsin digestion (2.5 μ g, 24 °C, overnight). The resultant peptides were desalted by solid-phase extraction, by acidification in 1% v/v formic acid; pre-washing the cartridge (Oasis HLB 1 cc Vac Cartridge, product number 186000383, Waters Corp., USA) with 1 ml of 80% v/v acetonitrile (ACN) containing 0.1% v/v trifluoroacetic acid (TFA) and washing with 1.2 ml of 0.1% v/v TFA three times by applying to a vacuum manifold; loading the sample on the cartridge and washing with 1.5 ml of 0.1% v/v TFA; eluting the sample with 0.8 ml of 80% v/v ACN containing 0.1% v/v TFA and collecting in 1.5 ml microcentrifuge tubes; then lyophilization by freeze drying (Virtis, SP Scientific). The peptides were resuspended in 100 μ l distilled water. A microBCA assay (Catalogue Number: 23235, Thermo Fischer Scientific) with BSA as the mass standard was performed in triplicate to quantify the peptides. Then, a 2 μ g aliquot of peptides was dried and resuspended in 15 μ l of 2% v/v ACN containing 0.05% v/v trifluoroacetic acid before analysis.

NanoESI-LC-MS/MS analysis. Samples were analysed by nanoESI-LC-MS/MS using a Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific) fitted with a nanoflow reversed-phase-HPLC (Ultimate 3000 RSLC, Dionex). The nano-LC system was equipped with an Acclaim Pepmap nano-trap column (Dionex—C18, 100 Å, 75 μ m $\times$ 2 cm) and an Acclaim Pepmap RSLC analytical column (Dionex—C18, 100 Å, 75 μ m $\times$ 50 cm). For each LC-MS/MS experiment, 1 μ l (0.135 μ g protein) of the peptide mix was loaded onto the enrichment (trap) column at an isocratic flow of 5 μ l min^{-1} of 3% CH_3CN containing 0.1% v/v formic acid for 5 min before the enrichment column was switched in-line with the analytical column. The eluents used for the LC were 0.1% v/v formic acid (solvent A) and 100% CH_3CN /0.1% formic acid v/v (solvent B). The gradient used (300 nl min^{-1}) was from 3–22% B in 90 min, 22–40% B in 10 min and 40–80% B in 5 min then maintained for 5 min before re-equilibration for 8 min at 3% B prior to the next analysis. All spectra were acquired in positive mode with full scan MS from m/z 400–1500 in the FT mode at 120,000 mass resolving power after accumulating to a target value 5.00e5 with maximum accumulation time of 50 ms. Lockmass of 445.12002 was used. Data-dependent HCD MS/MS of charge states > 1 was performed using a 3 s scan method, at a target value of 1.00e4, a maximum accumulation time of 60 ms, a normalised collision energy of 35%, an activation Q of 0.25, and at a 15,000 mass resolving power. Dynamic exclusion was used for 45 s.

Proteomic data analysis. Data analysis was carried out using Proteome Discoverer (version 2.1.0.81; Thermo Scientific) with the Mascot search engine (Matrix Science version 2.4.1) against the Swissprot *Mus Musculus* database (Version: 2015_07; 548,872 entries). The search was conducted with 20 p.p.m. MS tolerance, 0.6 Da MS/MS tolerance, with one missed cleavage allowed. The following modifications were allowed: methionine oxidation, N-terminal protein acetylation, N-terminal methionine cleavage and SILAC-LYS6 (variable); carbamidomethylcysteine (fixed). The false discovery rate maximum was set to 0.005% at the peptide identification level (actual was 0.005 for each replicate) and 1% at the protein identification level. Proteins were filtered for those containing at least one unique peptide in all five replicates. The common contaminant, Keratin, was excluded from the data set. Peptide quantitation was performed in Proteome

Discoverer v. 2.1.0.81 using the precursor ion quantifier node with the mass tolerance for isotope pattern multiplets set to 2 p.p.m., and two single peak/missing channels allowed. The protein abundance in each replicate was calculated by summation of the unique peptide abundances that were used for quantitation (heavy SILAC-LYS6 and light unlabelled derivatives). Quantified proteins were then filtered to those containing at least two unique peptides, one of which contained cysteine. The average peptide abundance ratio (i.e., tunicamycin-treated versus control) for the non-cysteine-containing peptides were then calculated for each protein. These values were then compared with the peptide abundance ratio of the corresponding cysteine-containing peptide(s) for each protein, in order to quantify the change in cysteine peptide abundance following TPE-MI treatment. Wilcoxon matched pair sign rank test was performed to calculate the statistical significance (Graphpad Prism, v.7) of cysteine depletion on TPE-MI treatment.

Malaria parasite experiments. Early trophozoite stage *P. falciparum*-infected red blood cells (21–24 h post invasion, 5% haematocrit) were incubated with 1 μ M DHA and 400 nM epoxomicin or with 400 nM epoxomicin alone in the presence of 0.1% DMSO (carrier solvent), for 3 h at 37 °C. Treated parasite cultures were attached to erythroagglutinin P(HA)-E-coated glass slides, incubated in 50 μ M TPE-MI for 30 min at 37 °C and fixed with 2% formaldehyde and 0.008% glutaraldehyde before imaging.

For the western blot experiment, red blood cells infected with trophozoite stage *P. falciparum* at 5% parasitemia and 5% haematocrit, were exposed to vehicle (0.1% v/v DMSO) or 400 nM epoxomicin and/or 1 μ M DHA for 3 h. Parasites were harvested for analysis of ubiquitinated proteins as previously described⁴⁷. The antibodies used were rabbit anti-ubiquitin (Dako-Z0458) and goat anti-rabbit IgG-peroxidase (Sigma-Aldrich-A0545), with dilutions 1:100 and 1:25,000, respectively.

Flow cytometry. Cells were analysed at a high flow rate in an LSRFortessa flow cytometer (BD Biosciences). 100,000–200,000 events were collected using a forward scatter threshold of 5000. Data were collected in pulse height, area and width parameters for each channel. For TPE-MI, data were collected with the 355 nm laser and a 450 $\pm$ 50 nm bandpass filter. For mCherry, data were collected with the 561 nm laser and 621 $\pm$ 20 nm bandpass filter. Flow cytometry data were analysed with FlowJo (Tree Star Inc.) and graphs were analysed in Prism 5. For PulSA, we analysed the data as described previously^{41, 54}, and used the non-aggregating protein samples to set the boundary for the ni gate. The gating strategies are shown in Supplementary Fig. 12a–f including TPE-MI signal gate.

For hiPSC experiments, cells were analysed using a BD FACSAria flow cytometer and 10,000 events were collected with a 355 nm laser. Cell parameters including pulse height, area and width were used for the analysis.

Imaging. After staining with TPE-MI, cells were fixed with 4% (v/v) paraformaldehyde in PBS for 15 min. For the cells treated with tunicamycin or heat shock, or examined for mCherry fluorescence, cells were imaged on a Zeiss LSM780 Confocal microscope for TPE-MI (excitation: 355 nm, emission: 445–500 nm) and mCherry (excitation: 561 nm, emission: 600–700 nm) using a $\times$ 40 objective lens. Cells that transfected with Httex1 were imaged 48 h after transfection.

For image quantitation, mean TPE-MI intensity within regions of interest were calculated by Fiji (ImageJ) in the cytosol of mammalian cells and within the plasmodium inside red blood cells of blood cells only containing a single parasite.

The mean TPE-MI signal, and corresponding Coomassie Blue pixel intensities, in SDS-PAGE images were measured on region of interests encompassing the entire lane for each sample.

Statistics. Statistical analyses were performed using Prism (Graphpad) software packages. The exact *P* values, raw values and statistical details are provided in Supplementary Data 1.

Data availability. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁵⁵ partner repository with the data set identifier PXD006527 and 10.6019/PXD006527. All other data are available from the corresponding author on reasonable request.

Received: 18 May 2016 Accepted: 12 June 2017
Published online: 07 September 2017

References

- Gershenson, A., Gierasch, L. M., Pastore, A. & Radford, S. E. Energy landscapes of functional proteins are inherently risky. *Nat. Chem. Biol.* **10**, 884–891 (2014).
- Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
- Brasemann, E., Chaney, J. L. & Clark, P. L. Folding the proteome. *Trends Biochem. Sci.* **38**, 337–344 (2013).
- Hutt, D. M. & Balch, W. E. Expanding proteostasis by membrane trafficking networks. *Cold Spring Harb. Perspect. Biol.* **5**, a013383 (2013).
- Roth, D. M. & Balch, W. E. Modeling general proteostasis: proteome balance in health and disease. *Curr. Opin. Cell Biol.* **23**, 126–134 (2011).
- Lindquist, S. The heat-shock response. *Annu. Rev. Biochem.* **55**, 1151–1191 (1986).
- Wang, M. & Kaufman, R. J. Protein misfolding in the endoplasmic reticulum as a conduit to human disease. *Nature* **529**, 326–335 (2016).
- Vidal, R. L., Matus, S., Bargsted, L. & Hetz, C. Targeting autophagy in neurodegenerative diseases. *Trends Pharmacol. Sci.* **35**, 583–591 (2014).
- McKinnon, C. & Tabrizi, S. J. The ubiquitin-proteasome system in neurodegeneration. *Antioxid. Redox Signal.* **21**, 2302–2321 (2014).
- Gidalevitz, T., Kikis, E. A. & Morimoto, R. I. A cellular perspective on conformational disease: the role of genetic background and proteostasis networks. *Curr. Opin. Struct. Biol.* **20**, 23–32 (2010).
- Gidalevitz, T., Ben-Zvi, A., Ho, K. H., Brignull, H. R. & Morimoto, R. I. Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* **311**, 1471–1474 (2006).
- Gupta, R. et al. Firefly luciferase mutants as sensors of proteome stress. *Nat. Methods* **8**, 879–U155 (2011).
- Liu, Y., Zhang, X., Chen, W., Tan, Y. L. & Kelly, J. W. Fluorescence turn-on folding sensor to monitor proteome stress in live cells. *J. Am. Chem. Soc.* **137**, 11303–11311 (2015).
- Marino, S. M. & Gladyshev, V. N. Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. *J. Mol. Biol.* **404**, 902–916 (2010).
- Sakai, K., Sakurai, K., Sakai, M., Hoshino, M. & Goto, Y. Conformation and stability of thiol-modified bovine beta-lactoglobulin. *Protein Sci.* **9**, 1719–1729 (2000).
- Liu, Y. et al. Simple biosensor with high selectivity and sensitivity: thiol-specific biomolecular probing and intracellular imaging by AIE fluorogen on a TLC plate through a thiol-ene click mechanism. *Chemistry* **16**, 8433–8438 (2010).
- Hong, Y., Lam, J. W. Y. & Tang, B. Z. Aggregation-induced emission. *Chem. Soc. Rev.* **40**, 5361–5388 (2011).
- Milo, R. What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays* **35**, 1050–1055 (2013).
- Hansen, R. E., Roth, D. & Winther, J. R. Quantifying the global cellular thiol-disulfide status. *Proc. Natl Acad. Sci. USA* **106**, 422–427 (2009).
- Ahn, Y. H., Lee, J. S. & Chang, Y. T. Combinatorial rosinamine library and application to in vivo glutathione probe. *J. Am. Chem. Soc.* **129**, 4510–4511 (2007).
- Su, D., Teoh, C. L., Sahu, S., Das, R. K. & Chang, Y.-T. Live cells imaging using a turn-on FRET-based BODIPY probe for biothiols. *Biomaterials* **35**, 6078–6085 (2014).
- Shibata, A., Furukawa, K., Abe, H., Tsuneda, S. & Ito, Y. Rhodamine-based fluorogenic probe for imaging biological thiol. *Bioorg. Med. Chem. Lett.* **18**, 2246–2249 (2008).
- Åkerfeldt, M., Morimoto, R. I. & Sistonen, L. Heat shock factors: integrators of cell stress, development and lifespan. *Nat. Rev. Mol. Cell Biol.* **11**, 545–555 (2010).
- Abayaya, K., Phillips, B. & Morimoto, R. I. Attenuation of the heat-shock response in HeLa-cells is mediated by the release of bound heat-shock transcription factor and is modulated by changes in growth and in heat-shock temperatures. *Genes Dev.* **5**, 2117–2127 (1991).
- Dorner, A. J. et al. The stress response in Chinese hamster ovary cells. Regulation of ERp72 and protein disulfide isomerase expression and secretion. *J. Biol. Chem.* **265**, 22029–22034 (1990).
- Marcu, M. G., Schulte, T. W. & Neckers, L. Novobiocin and related coumarins and depletion of heat shock protein 90-dependent signaling proteins. *J. Natl Cancer Inst.* **92**, 242–248 (2000).
- Conde, R., Belak, Z. R., Nair, M., O'Carroll, R. F. & Ovsenek, N. Modulation of Hsf1 activity by novobiocin and geldanamycin. *Biochem. Cell Biol.* **87**, 845–851 (2009).
- Ong, S. E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* **1**, 376–386 (2002).
- Wang, M. et al. PaxDb, a database of protein abundance averages across all three domains of life. *Mol. Cell Proteomics* **11**, 492–500 (2012).
- Gigliore, C., Boularot, A. & Meinel, T. Protein N-terminal methionine excision. *Cell Mol. Life Sci.* **61**, 1455–1474 (2004).
- Zhouravleva, G. et al. Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J.* **14**, 4065–4072 (1995).
- Pirczuk, H. et al. An in vivo control map for the eukaryotic mRNA translation machinery. *Mol. Syst. Biol.* **9**, 635 (2013).

33. Ito, K., Ebihara, K. & Nakamura, Y. The stretch of C-terminal acidic amino acids of translational release factor eRF1 is a primary binding site for eRF3 of fission yeast. *RNA* **4**, 958–972 (1998).
34. Ivanov, P. V., Gehring, N. H., Kunz, J. B., Hentze, M. W. & Kulozik, A. E. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J.* **27**, 736–747 (2008).
35. Park, S.-H. et al. PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone. *Cell* **154**, 134–145 (2013).
36. MacDonald, M. E. et al. A novel gene containing a trinucleotide repeat that is expanded and unstable on huntington's disease chromosomes. *Cell* **72**, 971–983 (1993).
37. Davies, S. W. et al. Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* **90**, 537–548 (1997).
38. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431**, 805–810 (2004).
39. Kopito, R. R. Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* **10**, 524 (2000).
40. Kaganovich, D., Kopito, R. & Frydman, J. Misfolded proteins partition between two distinct quality control compartments. *Nature* **454**, 1088–1095 (2008).
41. Ramdhan, Y. M. et al. Tracking protein aggregation and mislocalization in cells with flow cytometry. *Nat. Methods* **9**, 467–470 (2012).
42. Xu, X. et al. Reversal of phenotypic abnormalities by CRISPR/Cas9-mediated gene correction in Huntington disease patient-derived induced pluripotent stem cells. *Stem Cell Rep.* **8**, 619–633 (2017).
43. Eastman, R. T. & Fidock, D. A. Artemisinin-based combination therapies: a vital tool in efforts to eliminate malaria. *Nat. Rev. Microbiol.* **7**, 864–874 (2009).
44. Ismail, H. M. et al. Artemisinin activity-based probes identify multiple molecular targets within the asexual stage of the malaria parasites *Plasmodium falciparum* 3D7. *Proc. Natl Acad. Sci. USA* **113**, 2080–2085 (2016).
45. Wang, J. et al. Haem-activated promiscuous targeting of artemisinin in *Plasmodium falciparum*. *Nat. Commun.* **6**, 10111 (2015).
46. Klonis, N. et al. Altered temporal response of malaria parasites determines differential sensitivity to artemisinin. *Proc. Natl Acad. Sci. USA* **110**, 5157–5162 (2013).
47. Dogovski, C. et al. Targeting the cell stress response of *Plasmodium falciparum* to overcome artemisinin resistance. *PLoS Biol.* **13**, e1002132 (2015).
48. Jahn, T. R. & Radford, S. E. The Yin and Yang of protein folding. *FEBS J.* **272**, 5962–5970 (2005).
49. Radwan, M., Wood, R. J., Sui, X. & Hatters, D. M. When proteostasis goes bad: protein aggregation in the cell. *IUBMB Life* **69**, 49–54 (2017).
50. Leuenberger, P. et al. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **355**, eaai7825 (2017).
51. Seong, I. S. et al. HD CAG repeat implicates a dominant property of huntingtin in mitochondrial energy metabolism. *Hum. Mol. Genet.* **14**, 2871–2880 (2005).
52. Margulis, J. & Finkbeiner, S. Proteostasis in striatal cells and selective neurodegeneration in Huntington's disease. *Front. Cell Neurosci.* **8**, 218 (2014).
53. Tsvetkov, A. S. et al. Proteostasis of polyglutamine varies among neurons and predicts neurodegeneration. *Nat. Chem. Biol.* **9**, 586–592 (2013).
54. Ormsby, A. R., Ramdhan, Y. M., Mok, Y. F., Jovanoski, K. D. & Hatters, D. M. A platform to view huntingtin exon 1 aggregation flux in the cell reveals divergent influences from chaperones hsp40 and hsp70. *J. Biol. Chem.* **288**, 37192–37203 (2013).
55. Leitman, J., Ulrich Hartl, P. & Lederkremer, G. Z. Soluble forms of polyQ-expanded huntingtin rather than large aggregates cause endoplasmic reticulum stress. *Nat. Commun.* **4**, 2753 (2013).
56. Ramdhan, Y. M. et al. Huntingtin inclusions trigger cellular quiescence, deactivate apoptosis, and lead to delayed necrosis. *Cell Rep.* **19**, 919–927 (2017).
57. Phillips, A. J. et al. Peroxiredoxin is a versatile self-assembling tecton for protein nanotechnology. *Biomacromolecules* **15**, 1871–1881 (2014).
58. Polling, S. et al. Misfolded polyglutamine, polyalanine, and superoxide dismutase 1 aggregate via distinct pathways in the cell. *J. Biol. Chem.* **289**, 6669–6680 (2014).
59. Duan, X. P., Zeng, J., Lu, J. W. & Zhang, Z. B. Insights into the general and efficient cross McMurry reactions between ketones. *J. Org. Chem.* **71**, 9873–9876 (2006).
60. Kim, S. E., Wallat, J. D., Harker, E. C., Advincula, A. A. & Pokorski, J. K. Multifunctional and Spatially Controlled Bioconjugation to Melt Coextruded Nanofibers. *Polym. Chem.* **6**, 5683–5692 (2015).
61. Li, W. et al. Rapid induction and long-term self-renewal of primitive neural precursors from human embryonic stem cells by small molecule inhibitors. *Proc. Natl Acad. Sci. USA* **108**, 8299–8304 (2011).
62. Wessel, D. & Flugge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143 (1984).
63. Vizcaino, J. A. et al. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069 (2013).

Acknowledgements

We thank Peng Zeng and Sijie Chen from The University of Melbourne for technical support and helpful discussion. We also thank Michael Parker from St Vincents Institute of Medical Research, VIC, Australia for helpful discussion and feedback. We thank Juliet Gerrard from the University of Auckland, NZ, for provision of the human peroxiredoxin expression construct. This work was supported by grants to D.M.H. (National Health and Medical Research Council grants APP1049458, APP1049459 and APP1102059 and Australian Research Council (FT120100039 and DP170103093)) Y.H. (Australian Research Council DE170100058, Bruce Stone Fellowship from La Trobe University, Melbourne Neuroscience Institute Interdisciplinary Seed Funding (University of Melbourne), and McKenzie Fellowship from University of Melbourne) and grants to M.A.P. (Ministry of Education Singapore Tier 1 grant R-172-000-297-112, and Agency for Science, Technology and Research Strategic Positioning Fund for Genetic Orphan Diseases grant SPF2012/005).

Author contributions

D.M.H. and Y.H. conceived the ideas and were responsible for the overall project management. D.M.H., Y.H., L.T., M.A.P., J.B., G.E.R. and Z.C. designed the research. Z.C., N.S.M., M.R., R.J.W., J.B., T.A.S., Z.S., B.Z.T., Y.H., X.X., L. T., M.A.P. and D.M.H. performed experiments and/or interpreted the data. Z.C., Y.H. and D.M.H. wrote the manuscript with feedback from the other authors.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-00203-5.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

4.3 Conclusion and future directions

There is currently a distinct lack of methods for assessing the extent to which the proteome is folded in cells. Any issue with the management of proteostasis may result in a net change in the equilibrium of proteome folding towards unfolding and aggregated states. TPE-MI is a small cell permeable organic reporter probe that reacts with free cysteine thiols and measures the level of unfolded protein load, which is a proxy for the extent of proteome folding in cells. In chapter 4, the chemical dye was validated for probing unfolded protein states using multiple invitro models against various forms of proteome stresses and perturbations including clinically relevant HD models. Using proteomic approach, we identified and quantified proteins and peptides that are selectively unfolded in response to tunicamycin stress. Treatment with TPE-MI and evaluation of the absence of unmodified Cys residues in peptides indicated a comprehensive array of the most abundant proteins that demonstrated TPE-MI reactivity. Further examination of Cys-containing to non Cys-containing peptides identified a range of proteins with substantial increase in reactivity (median Cys/ Non-Cys ratio) to TPE-MI treatment indicating tunicamycin induced backlog of unfolded proteins. The method was also able to label newly synthesised peptides which are unable to fold upon tunicamycin treatment, which is a noteworthy finding of the study. The use of proteomic approaches to identify which proteins and peptides are selectively unfolded and the extent to which they are unfolding is a very useful approach to quantifying unfolding responses to stresses. The proteomic methodology hence developed will be advantageous to understand the propensity of unfolding of relevant specific proteins to different stressors.

The proteomics approach used in this study (Chapter 3) can also be used to establish a kinetic model to measure unfolding to specific stressors offering an omics view of unfolding and folding dynamics. The potential use of this biosensor technology to determine if proteostasis imbalances can be observed in a patient population will be quite important. Testing easily accessible primary tissues such as blood cells, biopsy and nasal brushing from HD patients and other neurological conditions involving unfolded proteins and protein aggregation would reveal its potential for biomarker development. The key promise of this research resides in its development as a marker for the early asymptomatic stage of the disease in which proteostasis imbalances are found, the assessment of which will be clinically useful. Understanding the baseline of unfolded protein load in these conditions and developing

models of perturbations to different stages of disease would be the critical first step in this direction.

4.4. Supplementary material

Supplementary materials for:

A thiol probe for measuring unfolded protein load and proteostasis in cells

Moore Z. Chen, Nagaraj S. Moily, Jessica L. Bridgford, Rebecca J. Wood, Mona Radwan, Trevor A. Smith, Zhegang Song, Ben Zhong Tang, Leann Tilley, Xiaohong Xu, Gavin E. Reid, Mahmoud A. Pouladi, Yuning Hong & Danny M. Hatters

Link for peer reviewed publication: <https://doi.org/10.1038/s41467-017-00203-5>

Link for supplemental information included in the publication: https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-017-00203-5/MediaObjects/41467_2017_203_MOESM1_ESM.pdf

Link for supplementary dataset 1 having detailed statistical summary of the data in the figures: https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-017-00203-5/MediaObjects/41467_2017_203_MOESM3_ESM.xlsx

Link for supplementary dataset 2 having detailed description of the proteins/peptides discovered by proteomics of TPE-MI treated tunicamycin stress and control samples: https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-017-00203-5/MediaObjects/41467_2017_203_MOESM4_ESM.xlsx

Data availability: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD006527 and 10.6019/PXD006527.

Chapter 5

Perspectives and Conclusion

HD is a neurodegenerative disease, that is dominantly inherited, fully penetrant, and is characterized by cognitive, motor and psychiatric disturbances. HD is caused by a CAG trinucleotide repeat expansion in the huntingtin gene on chromosome 4[4, 71]. Many studies have shown that changes in the striatum and the cortex occurs as disease progresses in the HD brain. Advances in the field of HD disease pathology and neuroimaging have also detailed comprehensive brain structural changes that occur with the progression of the disease and have contributed to a better understanding of the disease in recent years[56-58]. Advances in the management of HD have occurred recently and some drugs have undergone advanced clinical trials and testing. However currently there is no treatment for the disease and supportive management is the mainstay of treatment[4].

mHtt at the cellular level causes neuronal dysfunction and death by many mechanisms, including transcriptional dysregulation[2, 101], disruption of proteostasis[451, 452], mitochondrial dysfunction[453, 454] related to the toxicity of the mutant protein to name a few. The Htt protein is abnormally sticky because of excess glutamine residues which cause the protein to change its conformation and eventually aggregate[142]. The exact role that the various aggregation states play in disease, however, remains unclear. There are still many missing links in the HD pathology that need to be further explored. A better understanding of the disease-causing toxic mechanisms enables us to better treat and manage this devastating condition.

The overall aim of this thesis is to better understand the role of different aggregation states of mHtt protein in disease. In chapter 2 of the thesis, by using whole cell transcriptomic data from cells containing different aggregation state of mHtt and comparing them to wild-type Htt, we were able to define the cellular mechanism activated by mHtt aggregation states and correlate them with cellular health[168, 169]. There was ambiguity in the field with regard to the aggregation state specific toxicity. Some studies indicating that aggregation is toxic[170-175], whereas others indicating aggregation is a protective response to the toxic soluble mHtt[176, 455, 456]. Our study using longitudinal imaging showed that soluble mHtt aggregation state is the main cause of toxicity, which was in agreement with previous reports. Using transcriptomics on cells population of mHtt separated based on their aggregation states we were also able to define pre- and post-aggregation state specific transcriptional signatures. Amelioration of CREB signalling attributable to the toxic soluble form of mHtt was one of the most important mechanism causing neuronal dysfunction. We were able to validate by

specific CREB simulation improvement in HD cellular pathology validating our results. Previous studies have shown that CREB signalling is essential for neuronal development and survival, with disruption of CREB signalling leading to neurodegeneration[457]. Animal models have shown evidence of dysregulation of the CREB mediated CRE transcriptional pathway in HD[301, 458]. mHtt also facilitates transcription via a CRE-dependent mechanism *in vivo*[459] with mHtt aggregates sequester CREB activator CBP, downregulating the expression of CREB target genes[460, 461]. Hence CREB signalling dysfunction plays an important role in HD. Our study supports these finding with CREB related gene expression changes arising in soluble mHtt and propagated through aggregation. Though CREB related changes was the main mechanism driving cellular alterations identified, this study was also able to pinpoint additional gene hubs and define aggregation induced transcriptomic signatures which might modulate cell health both before and after aggregation. However, their potential role in modulating HD pathogenesis needs to be verified and validated in additional studies. Use of CRISPR-Cas9 inducible activation or knock out systems to further validate these changes or identify relevant mechanism of toxicity would be beneficial. The application of more relevant model systems like iPSC cells or HD organoids can potentially add value in identifying therapeutic targets for further study.

There are still many missing links in the HD pathology that need to be further explored. One interesting link we found is the spliceosome pathway. There was significant enrichment of the spliceosome pathway and the SR family of splicing factors linked to soluble mHtt aggregation state observed in our data. Our observation along with evidence of splicing alterations seen in HD brain and HD pathology[303, 348] necessitated further exploration of these changes. Our data showed that the key splicing factor SRSF2 showed significantly increased activity arising in soluble mHtt and continued with aggregation. This effect was related to the pathogenic soluble form of mHtt as verified with different expanded lengths of mHtt polyQ seen in the HD-hiPSCs. The spliceosome responsible for post transcriptional mRNA processing comprises of small nuclear RNA and other splicing proteins which are in constant stoichiometry inside the cell. In our study we show a dysregulation in snRNA levels arising in the soluble form of mHtt carried over in aggregation as also seen in HD-hiPSCs. This data taken together provides strong evidence of defective spliceosome integrity and function arising early in HD.

We then proceeded to identify and characterise the various alternative splicing events in different aggregation states of HD. Though we did not observe an enrichment of a particular

class of splicing event, extensive splicing alteration were seen in the different aggregation state of mHtt. Our splicing data was in concordance with previously established splicing changes in HD brain[303]. We also characterized four different splicing signatures and associated hub genes related to toxicity or protective response with regard to aggregation for further study. We identified *Dhx9*, *NF-kB1*, *Uba52*, *Racgap1* and others as key hubs involved in splicing networks. Interestingly CREB signalling links many of the hub genes identified, indicating that CRE mediated signalling may possibly play a role in dysregulated splicing. We also performed proteomic analysis of Htt aggregation states and identified protein changes associated with different aggregation states. A proteomic confirmation of splicing was also obtained by the identification of spliced RNA isoforms in the whole cell proteomics data. The conformation of RNA level splice changes with peptide level changes in the cell provides robust evidence of splicing defects in HD. Overall this study helps us better understand the disease pathology seen with soluble and with other aggregation states of HD. It will be interesting to further explore splicing changes by a sequencing study with higher sequencing depth (40-60 million reads per sample) or long read sequencing technologies available for complete isoform information. This combined with data independent swath MS proteomics will be ideal to identify translated splice isoforms and better understand verified splicing targets involved in HD pathology. The characterisation of mHttex1 mRNA associated proteins via an RNA pull down identified 36 spliceosome proteins bound to the mHttex1 RNA. The sequestration of one the RNA binding proteins PRPF8 resulted in mis splicing of CREB1, and overexpression rescued the HD pathology[304]. A detailed analysis of mHtt RNA binding proteins could provide strategies designed to augment CRE-mediated transcription or ameliorate splicing defects seen in HD. Similarly, the role of SR family of transcription factors needs to be elucidated. SRSF6 has been known to bind to Htt mRNA and produce aberrant splicing of the mHtt gene giving rise to toxic N-terminal mHtt fragments [342, 348]. Another SR protein SREK1 plays an important role in HD. Transgenic SREK1 mice showed transcriptomic changes similar to HD and when crossed with R6/1 HD mice, SREK1 overexpression attenuated the motor phenotype and striatal atrophy of HD mice[462]. Hence a detailed study of SR family of RNA binding proteins in HD would be useful. Transcription and post transcriptional RNA splicing are complex multimolecular processes that are highly coordinated and often functionally coupled[450, 463]. We have shown in this thesis that aggregation states produce distinct transcriptional and post transcriptional changes that arise in soluble mHtt and progress or are altered with

aggregation. These changes together indicate interlinked CREB signalling related defects in transcription and splicing which might be targets of further study.

HD pathology is complex, one of the key features of HD pathogenesis is an imbalance in proteostasis. Unfolded proteins accumulate in the cell in response to defective proteostasis leading to the protein aggregation. TPE-MI was used to measure and quantify the cellular unfolded protein load in response to various stressors including HD[464]. The TPE-MI proteomics helped identify the various unfolded proteins that expose cysteine residues and track the folding of individual proteins in response to tunicamycin induced ER stress. TPE-MI reactivity could be further utilised in kinetic studies to track protein folding over time in disease. TPE-MI dye is a promising tool in measuring unfolded protein load in protein aggregation disorders like HD which show increase in unfolded proteins with disease progression[465, 466]. It has potential for use in the clinic as a prognostic tool to measure disease progression instead of expensive MRI scans. However, its use in the clinic needs further study and validation. Overall this thesis provides a better overview of various aggregation related mechanisms in play in HD.

Bibliography

1. MacDonald, M.E., et al., *A Novel Gene Containing a Trinucleotide Repeat that is Expanded and Unstable on Huntington's Disease Chromosomes*. Cell, 1993. **72**(6): p. 971-983.
2. Saudou, F. and S. Humbert, *The Biology of Huntingtin*. Neuron, 2016. **89**(5): p. 910-26.
3. Martí, *RNA toxicity induced by expanded CAG repeats in Huntington's disease*. Brain pathology (Zurich, Switzerland), 2016. **26**(6).
4. McColgan, P. and S.J. Tabrizi, *Huntington's disease: a clinical review*. Eur J Neurol, 2018. **25**(1): p. 24-34.
5. Vale, T.C. and F. Cardoso, *Chorea: A Journey through History*. Tremor Other Hyperkinet Mov (N Y), 2015. **5**.
6. Huntington, G., *On Chorea*, in *The Medical and Surgical Reporter*. 1872, S.W. Butler: Philadelphia: p. 317-321.
7. Wexler, A., *A brief prehistory of Huntington's disease*. J Huntingtons Dis, 2013. **2**(3): p. 231-7.
8. Moily, N.S., et al., *Huntington's disease*. Monograph on Movement Disorders. The association of physicians of India: Indian College of Physicians, 2012: p. 213-34.
9. Bhattacharyya, K.B., *The story of George Huntington and his disease*, in *Ann Indian Acad Neurol*. 2016. p. 25-8.
10. Gusella, J.F., et al., *A polymorphic DNA marker genetically linked to Huntington's disease*. Nature, 1983. **306**(5940): p. 234-8.
11. Bates, G.P., *History of genetic disease: the molecular genetics of Huntington disease - a history*. Nat Rev Genet, 2005. **6**(10): p. 766-73.
12. Davies, S.W., et al., *Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation*. Cell, 1997. **90**(3): p. 537-48.
13. Scherzinger, E., et al., *Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo*. Cell, 1997. **90**(3): p. 549-58.
14. Cha, J.-H.J., et al., *Altered brain neurotransmitter receptors in transgenic mice expressing a portion of an abnormal human Huntington disease gene*. 1998.
15. Group, H.S., *A randomized, placebo-controlled trial of coenzyme Q10 and remacemide in Huntington's disease*. Neurology, 2001. **57**(3): p. 397-404.
16. Wild, E.J. and S.J. Tabrizi, *Therapies targeting DNA and RNA in Huntington's disease*. Lancet Neurol, 2017. **16**(10): p. 837-847.
17. Rodrigues, F.B. and E.J. Wild, *Huntington's Disease Clinical Trials Corner: February 2018*. J Huntingtons Dis, 2018. **7**(1): p. 89-98.
18. Mullard, A., *Pioneering antisense drug heads into pivotal trials for Huntington disease*. Nat Rev Drug Discov, 2019. **18**(3): p. 161-163.
19. Harper, P.S., *The epidemiology of Huntington's disease* | SpringerLink. 1992.
20. Pridmore, S.A., *The prevalence of Huntington's disease in Tasmania*. Med J Aust, 1990. **153**(3): p. 133-4.
21. Moily, N.S., et al., *Trinucleotide Repeats and Haplotypes at the Huntingtin Locus in an Indian Sample Overlaps with European Haplogroup A*. PLoS Curr, 2014. **6**.
22. Warby, S.C., et al., *HTT haplotypes contribute to differences in Huntington disease prevalence between Europe and East Asia*, in *Eur J Hum Genet*. 2011. p. 561-6.

23. Finkbeiner, S., *Huntington's Disease*. Cold Spring Harb Perspect Biol, 2011. **3**(6).
24. Rubinsztein, D.C., et al., *Normal CAG and CCG repeats in the Huntington's disease genes of Parkinson's disease patients*. Am J Med Genet, 1995. **60**(2): p. 109-10.
25. Langbehn, D.R., M. Hayden, and J.S. Paulsen, *CAG-Repeat Length and the Age of Onset in Huntington Disease (HD): A Review and Validation Study of Statistical Approaches*. Am J Med Genet B Neuropsychiatr Genet, 2010. **153B**(2): p. 397-408.
26. Lee, J.M., et al., *CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion*. Neurology, 2012. **78**(10): p. 690-5.
27. Squitieri, F., et al., *Juvenile Huntington's disease: does a dosage-effect pathogenic mechanism differ from the classical adult disease?* Mech Ageing Dev, 2006. **127**(2): p. 208-12.
28. Ross, C.A., et al., *Huntington disease: natural history, biomarkers and prospects for therapeutics*. Nat Rev Neurol, 2014. **10**(4): p. 204-16.
29. Myers, R.H., *Huntington's disease genetics*. NeuroRx, 2004. **1**(2): p. 255-62.
30. Aziz, N.A., et al., *Normal and mutant HTT interact to affect clinical severity and progression in Huntington disease*. Neurology, 2009. **73**(16): p. 1280-5.
31. L, D., et al., *Interaction of Normal and Expanded CAG Repeat Sizes Influences Age at Onset of Huntington Disease*. American journal of medical genetics. Part A, 2003. **119A**(3).
32. N, L. and T. SJ, *Huntington Disease: A Tale of Two Genes*. Neurology, 2009. **73**(16).
33. Slepko, N., et al., *Normal-repeat-length polyglutamine peptides accelerate aggregation nucleation and cytotoxicity of expanded polyglutamine proteins*. Proc Natl Acad Sci U S A, 2006. **103**(39): p. 14367-72.
34. A, B., et al., *Mutant Huntingtin Promotes the Fibrillogenesis of Wild-Type Huntingtin: A Potential Mechanism for Loss of Huntingtin Function in Huntington's Disease*. The Journal of biological chemistry, 2003. **278**(42).
35. XJ, L., F. M, and L. S, *Interacting Proteins as Genetic Modifiers of Huntington Disease*. Trends in genetics : TIG, 2007. **23**(11).
36. Trottier, Y., V. Biancalana, and J.L. Mandel, *Instability of CAG repeats in Huntington's disease: relation to parental transmission and age of onset*. J Med Genet, 1994. **31**(5): p. 377-82.
37. Swami, M., et al., *Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset*. Hum Mol Genet, 2009. **18**(16): p. 3039-47.
38. Gusella, J.F., M.E. MacDonald, and J.M. Lee, *Genetic modifiers of Huntington's disease*. Mov Disord, 2014. **29**(11): p. 1359-65.
39. Consortium, G.M.o.H.s.D.G.-H., *Identification of Genetic Factors That Modify Clinical Onset of Huntington's Disease*. Cell, 2015. **162**(3).
40. Moss, D.J.H., et al., *Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study*. Lancet Neurol, 2017. **16**(9): p. 701-711.
41. Pinto, R.M., et al., *Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches*. PLoS Genet, 2013. **9**(10): p. e1003930.
42. Massey, T.H. and L. Jones, *The central role of DNA damage and repair in CAG repeat diseases*, in *Dis Model Mech*. 2018.

43. Bettencourt, C., et al., *DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases*, in *Ann Neurol*. 2016. p. 983-90.
44. Goold, R., et al., *FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat*. *Human Molecular Genetics*, 2019. **28**(4): p. 650-661.
45. Larissa, A. and J.T. Epplen, *Genetic modifiers of Huntington's disease: beyond CAG*. *Future Neurology*, 2011. **7**(1).
46. Wexler, N.S., et al., *Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset*. *Proc Natl Acad Sci U S A*, 2004. **101**(10): p. 3498-503.
47. Mo, C., A.J. Hannan, and T. Renoir, *Environmental factors as modulators of neurodegeneration: insights from gene-environment interactions in Huntington's disease*. *Neurosci Biobehav Rev*, 2015. **52**: p. 178-92.
48. Lee, J.M., et al., *CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion*. *Neurology*, 2012. **78**(10): p. 690-5.
49. Siesling, S., et al., *Unified Huntington's disease rating scale: a follow up*. *Mov Disord*, 1998. **13**(6): p. 915-9.
50. Shoulson, I. and S. Fahn, *Huntington disease: clinical care and evaluation*. *Neurology*, 1979. **29**(1): p. 1-3.
51. Lee, J.M., et al., *CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion*, in *Neurology*. 2012. p. 690-5.
52. Rosenblatt, A., et al., *The association of CAG repeat length with clinical progression in Huntington disease*. *Neurology*, 2006. **66**(7): p. 1016-20.
53. Tabrizi, S.J., et al., *Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data*. *Lancet Neurol*, 2013. **12**(7): p. 637-49.
54. Coppen, E.M. and R.A.C. Roos, *Current Pharmacological Approaches to Reduce Chorea in Huntington's Disease*, in *Drugs*. 2017. p. 29-46.
55. Potkin, K.T. and S.G. Potkin, *New directions in therapeutics for Huntington disease*. *Future Neurol*, 2018. **13**(2): p. 101-21.
56. Rüb, U., et al., *Huntington's disease (HD): degeneration of select nuclei, widespread occurrence of neuronal nuclear and axonal inclusions in the brainstem*. *Brain Pathol*, 2014. **24**(3): p. 247-60.
57. Niccolini, F. and M. Politis, *Neuroimaging in Huntington's disease*. *World J Radiol*, 2014. **6**(6): p. 301-12.
58. Johnson, E.B., et al., *Dynamics of cortical degeneration over a decade in Huntington's Disease*. *bioRxiv*, 2019.
59. Rub, U., et al., *Huntington's disease (HD): the neuropathology of a multisystem neurodegenerative disorder of the human brain*. *Brain Pathol*, 2016. **26**(6): p. 726-740.
60. Walker, F.O., *Huntington's Disease*. *Semin Neurol*, 2007. **27**(2): p. 143-50.
61. Vonsattel, J.-P., et al., *Neuropathological Classification of Huntington's Disease*. *Journal of Neuropathology & Experimental Neurology*, 1985. **44**(6): p. 559-577.
62. Tabrizi, S.J., et al., *Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis*. *Lancet Neurol*, 2011. **10**(1): p. 31-42.

63. Furtado, S., et al., *Relationship between trinucleotide repeats and neuropathological changes in Huntington's disease*. Ann Neurol, 1996. **39**(1): p. 132-6.
64. Penney, J.B., Jr., et al., *CAG repeat number governs the development rate of pathology in Huntington's disease*. Ann Neurol, 1997. **41**(5): p. 689-92.
65. Reiner, A., et al., *Differential loss of striatal projection neurons in Huntington disease*. Proc Natl Acad Sci U S A, 1988. **85**(15): p. 5733-7.
66. Albin, R.L., et al., *Preferential loss of striato-external pallidal projection neurons in presymptomatic Huntington's disease*. Ann Neurol, 1992. **31**(4): p. 425-30.
67. Modregger, J., et al., *PACSIN 1 interacts with huntingtin and is absent from synaptic varicosities in presymptomatic Huntington's disease brains*. Hum Mol Genet, 2002. **11**(21): p. 2547-58.
68. DiProspero, N.A., et al., *Early changes in Huntington's disease patient brains involve alterations in cytoskeletal and synaptic elements*. J Neurocytol, 2004. **33**(5): p. 517-33.
69. Sapp, E., et al., *Early and progressive accumulation of reactive microglia in the Huntington disease brain*. J Neuropathol Exp Neurol, 2001. **60**(2): p. 161-72.
70. Myers, R.H., et al., *Decreased neuronal and increased oligodendroglial densities in Huntington's disease caudate nucleus*. J Neuropathol Exp Neurol, 1991. **50**(6): p. 729-42.
71. Group, H.S., *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, 1993. **72**(6): p. 971-83.
72. Lin, B., et al., *Differential 3' polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression*. Hum Mol Genet, 1993. **2**(10): p. 1541-5.
73. Hughes, A.C., et al., *Identification of novel alternative splicing events in the huntingtin gene and assessment of the functional consequences using structural protein homology modelling*. J Mol Biol, 2014. **426**(7): p. 1428-38.
74. Ruzo, A., et al., *Discovery of novel isoforms of huntingtin reveals a new hominid-specific exon*. PLoS One, 2015. **10**(5): p. e0127687.
75. Bates, G.P., et al., *Huntington disease*. Nature Reviews Disease Primers, Published online: 23 April 2015; | doi:10.1038/nrdp.2015.5, 2015.
76. Atwal, R.S., et al., *Huntingtin has a membrane association signal that can modulate huntingtin aggregation, nuclear entry and toxicity*. Hum Mol Genet, 2007. **16**(21): p. 2600-15.
77. DeGuire, S.M., et al., *N-terminal Huntingtin (Htt) phosphorylation is a molecular switch regulating Htt aggregation, helical conformation, internalization, and nuclear targeting*, in J Biol Chem. 2018. p. 18540-58.
78. Rockabrand, E., et al., *The first 17 amino acids of Huntingtin modulate its sub-cellular localization, aggregation and effects on calcium homeostasis*. Hum Mol Genet, 2007. **16**(1): p. 61-77.
79. Atwal, R.S., et al., *Kinase inhibitors modulate huntingtin cell localization and toxicity*. Nat Chem Biol, 2011. **7**(7): p. 453-60.
80. Guedes-Dias, P. and E.L.F. Holzbaur, *Huntingtin Fibrils Poke Membranes*. Cell, 2017. **171**(1): p. 32-33.
81. Albin, R.L., *Polyglutamine inclusion body toxicity*. Mov Disord, 2017. **32**(12): p. 1686.

82. Bauerlein, F.J.B., et al., *In Situ Architecture and Cellular Interactions of PolyQ Inclusions*. Cell, 2017. **171**(1): p. 179-187.e10.
83. Guo, Q., et al., *The cryo-electron microscopy structure of huntingtin*. Nature, 2018. **555**(7694): p. 117-120.
84. Bugg, C.W., et al., *Structural features and domain organization of huntingtin fibrils*. J Biol Chem, 2012. **287**(38): p. 31739-46.
85. Lin, H.K., et al., *Fibril polymorphism affects immobilized non-amyloid flanking domains of huntingtin exon1 rather than its polyglutamine core*. Nat Commun, 2017. **8**: p. 15462.
86. Mangiarini, L., et al., *Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice*. Cell, 1996. **87**(3): p. 493-506.
87. Palidwor, G.A., et al., *Detection of alpha-rod protein repeats using a neural network and application to huntingtin*. PLoS Comput Biol, 2009. **5**(3): p. e1000304.
88. Li, W., et al., *Expression and characterization of full-length human huntingtin, an elongated HEAT repeat protein*. J Biol Chem, 2006. **281**(23): p. 15916-22.
89. Vijayvargia, R., et al., *Huntingtin's spherical solenoid structure enables polyglutamine tract-dependent modulation of its structure and function*. Elife, 2016. **5**: p. e11184.
90. Qin, Y.W.F.L.Z.-H., *The role of post-translational modifications of huntingtin in the pathogenesis of Huntington's disease*. Neuroscience Bulletin, 2010: p. 153-162.
91. Martin, D.D., et al., *Identification of a post-translationally myristoylated autophagy-inducing domain released by caspase cleavage of huntingtin*. Hum Mol Genet, 2014. **23**(12): p. 3166-79.
92. Jeong, H., et al., *Acetylation Targets Mutant Huntingtin to Autophagosomes for Degradation*. Cell, 2009. **137**(1): p. 60-72.
93. Thompson, L.M., et al., *IKK phosphorylates Huntingtin and targets it for degradation by the proteasome and lysosome*. J Cell Biol, 2009. **187**(7): p. 1083-99.
94. Steffan, J.S., et al., *SUMO modification of Huntingtin and Huntington's disease pathology*. Science, 2004. **304**(5667): p. 100-4.
95. Yanai, A., et al., *Palmitoylation of huntingtin by HIP14 is essential for its trafficking and function*. Nat Neurosci, 2006. **9**(6): p. 824-31.
96. Ross, C.A. and S.J. Tabrizi, *Huntington's disease: from molecular pathogenesis to clinical treatment*. Lancet Neurol, 2011. **10**(1): p. 83-98.
97. Ehrnhoefer, D.E., L. Sutton, and M.R. Hayden, *Small Changes, Big Impact: Posttranslational Modifications and Function of Huntingtin in Huntington Disease*. The Neuroscientist, 2011. **17**(5).
98. Sharp, A.H., et al., *Widespread expression of Huntington's disease gene (IT15) protein product*. Neuron, 1995. **14**(5): p. 1065-74.
99. Landwehrmeyer, G.B., et al., *Huntington's disease gene: regional and cellular expression in brain of normal and affected individuals*. Ann Neurol, 1995. **37**(2): p. 218-30.
100. Zeitlin, S., et al., *Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue*. Nat Genet, 1995. **11**(2): p. 155-63.

101. Benn, C.L., et al., *Huntingtin Modulates Transcription, Occupies Gene Promoters In Vivo, and Binds Directly to DNA in a Polyglutamine-Dependent Manner*. J Neurosci, 2008. **28**(42): p. 10720-33.
102. S, E., et al., *Huntingtin-associated Protein 1 (HAP1) Interacts With the p150Glued Subunit of Dynactin*. Human molecular genetics, 1997. **6**(13).
103. JR, M., et al., *Interaction of Huntingtin-associated protein-1 With Kinesin Light Chain: Implications in Intracellular Trafficking in Neurons*. The Journal of biological chemistry, 2006. **281**(6).
104. D, Z., H. MV, and S. F, *Huntingtin's Function in Axonal Transport Is Conserved in Drosophila Melanogaster*. PloS one, 2013. **8**(3).
105. Caviston, J.P., et al., *Huntingtin facilitates dynein/dynactin-mediated vesicle transport*. PNAS, 2007.
106. PW, F., et al., *Huntingtin Interacts With a Family of WW Domain Proteins*. Human molecular genetics, 1998. **7**(9).
107. DA, S., et al., *Optineurin Links Myosin VI to the Golgi Complex and Is Involved in Golgi Organization and Exocytosis*. The Journal of cell biology, 2005. **169**(2).
108. Pal, A., et al., *Huntingtin–HAP40 complex is a novel Rab5 effector that regulates early endosome motility and is up-regulated in Huntington's disease*. J Cell Biol, 2006. **172**(4): p. 605-18.
109. Engqvist-Goldstein Å, E., et al., *The actin-binding protein Hip1R associates with clathrin during early stages of endocytosis and promotes clathrin assembly in vitro*. J Cell Biol, 2001. **154**(6): p. 1209-24.
110. Rigamonti, D., et al., *Huntingtin's neuroprotective activity occurs via inhibition of procaspase-9 processing*. J Biol Chem, 2001. **276**(18): p. 14545-8.
111. Cattaneo, E., C. Zuccato, and M. Tartari, *Normal huntingtin function: an alternative approach to Huntington's disease*. Nat Rev Neurosci, 2005. **6**(12): p. 919-30.
112. Zuccato, C., et al., *Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes*. Nature Genetics, 2003. **35**(1): p. 76-83.
113. LR, G., et al., *Huntingtin Controls Neurotrophic Support and Survival of Neurons by Enhancing BDNF Vesicular Transport Along Microtubules*. Cell, 2004. **118**(1).
114. HR, W. and H. F, *Neurotrophin-4/5 Promotes Survival and Differentiation of Rat Striatal Neurons Developing in Culture*. The European journal of neuroscience, 1994. **6**(11).
115. ZC, B., G. JA, and J. KR, *Early Striatal Dendrite Deficits Followed by Neuron Loss With Advanced Age in the Absence of Anterograde Cortical Brain-Derived Neurotrophic Factor*. The Journal of neuroscience : the official journal of the Society for Neuroscience, 2004. **24**(17).
116. Ross, C.A., *Huntington's disease: new paths to pathogenesis*. Cell, 2004. **118**(1): p. 4-7.
117. Elias, S., et al., *Huntingtin Regulates Mammary Stem Cell Division and Differentiation*. Stem Cell Reports, 2014. **2**(4): p. 491-506.
118. JD, G., et al., *Huntingtin Is Required for Mitotic Spindle Orientation and Mammalian Neurogenesis*. Neuron, 2010. **67**(3).
119. Park, I.-H., et al., *Reprogramming of human somatic cells to pluripotency with defined factors*. Nature, 2007. **451**(7175): p. 141-146.

120. Csobonyeiova, M., S. Polak, and L. Danisovic, *Recent Overview of the Use of iPSCs Huntington's Disease Modeling and Therapy*. Int J Mol Sci, 2020. **21**(6).
121. Ferrante, R.J., *Mouse models of Huntington's disease and methodological considerations for therapeutic trials*. Biochim Biophys Acta, 2009. **1792**(6): p. 506-20.
122. Schwarcz, R. and J.T. Coyle, *Striatal lesions with kainic acid: neurochemical characteristics*. Brain Res, 1977. **127**(2): p. 235-49.
123. Schwarcz, R., W.O. Whetsell, Jr., and R.M. Mangano, *Quinolinic acid: an endogenous metabolite that produces axon-sparing lesions in rat brain*. Science, 1983. **219**(4582): p. 316-8.
124. Damiano, M., et al., *Mitochondria in Huntington's disease*. Biochim Biophys Acta, 2010. **1802**(1): p. 52-61.
125. Faber, P.W., et al., *Polyglutamine-mediated dysfunction and apoptotic death of a Caenorhabditis elegans sensory neuron*. Proc Natl Acad Sci U S A, 1999. **96**(1): p. 179-84.
126. Lee, A.L., et al., *A new Caenorhabditis elegans model of human huntingtin 513 aggregation and toxicity in body wall muscles*. PLoS One, 2017. **12**(3): p. e0173644.
127. Xu, Z., et al., *Studying Polyglutamine Diseases in Drosophila*. Exp Neurol, 2015. **274**(0 0): p. 25-41.
128. Pouladi, M.A., A.J. Morton, and M.R. Hayden, *Choosing an animal model for the study of Huntington's disease*. Nat Rev Neurosci, 2013. **14**(10): p. 708-21.
129. Slow, E.J., et al., *Selective striatal neuronal loss in a YAC128 mouse model of Huntington disease*. Hum Mol Genet, 2003. **12**(13): p. 1555-67.
130. Hodgson, J.G., et al., *A YAC mouse model for Huntington's disease with full-length mutant huntingtin, cytoplasmic toxicity, and selective striatal neurodegeneration*. Neuron, 1999. **23**(1): p. 181-92.
131. Wheeler, V.C., et al., *Long glutamine tracts cause nuclear localization of a novel form of huntingtin in medium spiny striatal neurons in HdhQ92 and HdhQ111 knock-in mice*. Hum Mol Genet, 2000. **9**(4): p. 503-13.
132. Menalled, L.B., et al., *Early motor dysfunction and striosomal distribution of huntingtin microaggregates in Huntington's disease knock-in mice*. J Neurosci, 2002. **22**(18): p. 8266-76.
133. Meade, C.A., et al., *Cellular localization and development of neuronal intranuclear inclusions in striatal and cortical neurons in R6/2 transgenic mice*. J Comp Neurol, 2002. **449**(3): p. 241-69.
134. Carter, R.J., et al., *Characterization of progressive motor deficits in mice transgenic for the human Huntington's disease mutation*. J Neurosci, 1999. **19**(8): p. 3248-57.
135. DiFiglia, M., et al., *Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain*. Science, 1997. **277**(5334): p. 1990-3.
136. Sieradzan, K.A., et al., *Huntington's disease intranuclear inclusions contain truncated, ubiquitinated huntingtin protein*. Exp Neurol, 1999. **156**(1): p. 92-9.
137. Arrasate, M. and S. Finkbeiner, *Protein aggregates in Huntington's disease*. Exp Neurol, 2012. **238**(1): p. 1-11.
138. Arrasate, M., et al., *Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death*. Nature, 2004. **431**(7010): p. 805-10.

139. Sisodia, S.S., *Nuclear inclusions in glutamine repeat disorders: are they pernicious, coincidental, or beneficial?* Cell, 1998. **95**(1): p. 1-4.
140. Soto, C. and S. Pritzkow, *Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases*. Nat Neurosci, 2018. **21**(10): p. 1332-1340.
141. MD, B., et al., *Amyloid Nomenclature 2018: Recommendations by the International Society of Amyloidosis (ISA) Nomenclature Committee*. Amyloid : the international journal of experimental and clinical investigation : the official journal of the International Society of Amyloidosis, 2018. **25**(4).
142. JW, K., *The Alternative Conformations of Amyloidogenic Proteins and Their Multi-Step Assembly Pathways*. Current opinion in structural biology, 1998. **8**(1).
143. CA, G., et al., *Nuclear and Neuropil Aggregates in Huntington's Disease: Relationship to Neuropathology*. The Journal of neuroscience, 1999. **19**(7).
144. Sharma, D., et al., *Polyglutamine homopolymers having 8-45 residues form slablike beta-crystallite assemblies*. Proteins, 2005. **61**(2): p. 398-411.
145. D, C., et al., *Protein Aggregation in Cell Biology: An Aggregomics Perspective of Health and Disease*. Seminars in cell & developmental biology, 2018. **99**.
146. Scherzinger, E., et al., *Self-assembly of Polyglutamine-Containing Huntingtin Fragments Into Amyloid-Like Fibrils: Implications for Huntington's Disease Pathology*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(8).
147. Wetzel, R., *Physical chemistry of polyglutamine: intriguing tales of a monotonous sequence*. J Mol Biol, 2012. **421**(4-5): p. 466-90.
148. Chen, S., F.A. Ferrone, and R. Wetzel, *Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation*. Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11884-9.
149. Walters, R.H. and R.M. Murphy, *Aggregation kinetics of interrupted polyglutamine peptides*. J Mol Biol, 2011. **412**(3): p. 505-19.
150. LG, N., et al., *Identification of Novel Potentially Toxic Oligomers Formed in Vitro From Mammalian-Derived Expanded Huntingtin exon-1 Protein*. The Journal of biological chemistry, 2012. **287**(19).
151. Y, T., et al., *Detection of Polyglutamine Protein Oligomers in Cells by Fluorescence Correlation Spectroscopy*. The Journal of biological chemistry, 2007. **282**(33).
152. Y, N., et al., *Inhibition of Polyglutamine Protein Aggregation and Cell Death by Novel Peptides Identified by Phage Display Screening*. The Journal of biological chemistry, 2000. **275**(14).
153. MA, O., et al., *Tracking Mutant Huntingtin Aggregation Kinetics in Cells Reveals Three Major Populations That Include an Invariant Oligomer Pool*. The Journal of biological chemistry, 2010. **285**(28).
154. Takahashi, T., et al., *Soluble polyglutamine oligomers formed prior to inclusion body formation are cytotoxic*. Hum Mol Genet, 2008. **17**(3): p. 345-56.
155. Legleiter, J., et al., *Mutant huntingtin fragments form oligomers in a polyglutamine length-dependent manner in vitro and in vivo*. J Biol Chem, 2010. **285**(19): p. 14777-90.
156. Lyubchenko, Y.L., A.V. Krasnoslobodtsev, and S. Luca, *Fibrillogenesis of huntingtin and other glutamine containing proteins*. Subcell Biochem, 2012. **65**: p. 225-51.

157. Fink, A.L., *Protein aggregation: folding aggregates, inclusion bodies and amyloid*. Fold Des, 1998. **3**(1): p. R9-23.
158. M, T., et al., *Intra- And Intermolecular Beta-Pleated Sheet Formation in Glutamine-Repeat Inserted Myoglobin as a Model for Polyglutamine Diseases*. The Journal of biological chemistry, 2001. **276**(48).
159. V, V. and F. V., *The Route to Protein Aggregate Superstructures: Particulates and Amyloid-Like Spherulites*. FEBS letters, 2015. **589**(19 Pt A).
160. MR, K., D. KR, and D. AM, *Protein Aggregation: More Than Just Fibrils*. Biochemical Society transactions, 2009. **37**(Pt 4).
161. Saudou, F., et al., *Huntingtin Acts in the Nucleus to Induce Apoptosis but Death Does Not Correlate With the Formation of Intranuclear Inclusions*. Cell, 1998. **95**(1).
162. Hackam, A.S., et al., *The Influence of Huntingtin Protein Size on Nuclear Localization and Cellular Toxicity*. The Journal of cell biology, 1998. **141**(5).
163. Lunkes, A. and J.L. Mandel, *A Cellular Model That Recapitulates Major Pathogenic Steps of Huntington's Disease*. Human molecular genetics, 1998. **7**(9).
164. Ossato, G., et al., *A two-step path to inclusion formation of huntingtin peptides revealed by number and brightness analysis*. Biophys J, 2010. **98**(12): p. 3078-85.
165. Cisbani, G. and F. Cicchetti, *An in vitro perspective on the molecular mechanisms underlying mutant huntingtin protein toxicity*. Cell Death Dis, 2012. **3**: p. e382.
166. Hands, S.L. and A. Wytenbach, *Neurotoxic protein oligomerisation associated with polyglutamine diseases*. Acta Neuropathol, 2010. **120**(4): p. 419-37.
167. Bates, G., *Huntingtin aggregation and toxicity in Huntington's disease*. Lancet, 2003. **361**(9369): p. 1642-4.
168. Poirier, M.A., H. Jiang, and C.A. Ross, *A structure-based analysis of huntingtin mutant polyglutamine aggregation and toxicity: evidence for a compact beta-sheet structure*. Hum Mol Genet, 2005. **14**(6): p. 765-74.
169. Kaytor, M.D., K.D. Wilkinson, and S.T. Warren, *Modulating huntingtin half-life alters polyglutamine-dependent aggregate formation and cell toxicity*. J Neurochem, 2004. **89**(4): p. 962-73.
170. DM, H., *Protein Misfolding Inside Cells: The Case of Huntingtin and Huntington's Disease*. IUBMB life, 2008. **60**(11).
171. Y, C., et al., *Live-cell Imaging Reveals Divergent Intracellular Dynamics of Polyglutamine Disease Proteins and Supports a Sequestration Model of Pathogenesis*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(14).
172. JS, S., et al., *The Huntington's Disease Protein Interacts With p53 and CREB-binding Protein and Represses Transcription*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(12).
173. FC, N., et al., *Interference by Huntingtin and atrophin-1 With CBP-Mediated Transcription Leading to Cellular Toxicity*. Science, 2001. **291**(5512).
174. CC, H., et al., *Amyloid Formation by Mutant Huntingtin: Threshold, Progressivity and Recruitment of Normal Polyglutamine Proteins*. Somatic cell and molecular genetics, 1998. **24**(4).
175. A, K., et al., *Insoluble Detergent-Resistant Aggregates Form Between Pathological and Nonpathological Lengths of Polyglutamine in Mammalian*

- Cells. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(20).
176. Miller, J., et al., *Identifying polyglutamine protein species in situ that best predict neurodegeneration*. Nat Chem Biol, 2011. **7**(12): p. 925-34.
 177. RA, B., et al., *Pharmacological Promotion of Inclusion Formation: A Therapeutic Approach for Huntington's and Parkinson's Diseases*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(11).
 178. Ramdzan, Y.M., et al., *Huntingtin Inclusions Trigger Cellular Quiescence, Deactivate Apoptosis, and Lead to Delayed Necrosis*. Cell Rep, 2017. **19**(5): p. 919-927.
 179. Newcombe, E.A., et al., *Tadpole-like Conformations of Huntingtin Exon 1 Are Characterized by Conformational Heterogeneity that Persists regardless of Polyglutamine Length*. J Mol Biol, 2018. **430**(10): p. 1442-1458.
 180. Nagai, Y., et al., *A toxic monomeric conformer of the polyglutamine protein*. Nat Struct Mol Biol, 2007. **14**(4): p. 332-40.
 181. Sharma, S. and R. Taliyan, *Transcriptional dysregulation in Huntington's disease: The role of histone deacetylases*. Pharmacol Res, 2015. **100**: p. 157-69.
 182. Rubinsztein, D.C. and J. Carmichael, *Huntington's disease: molecular basis of neurodegeneration*. Expert Rev Mol Med, 2003. **5**(20): p. 1-21.
 183. Pogoda, A., et al., *Transcriptional dysregulation in Huntington's Disease. The role in pathogenesis and potency for pharmacological targeting*. Curr Med Chem, 2020.
 184. Neueder, A. and G.P. Bates, *A common gene expression signature in Huntington's disease patient brain regions*. BMC Med Genomics, 2014. **7**: p. 60.
 185. Labadorf, A., et al., *RNA Sequence Analysis of Human Huntington Disease Brain Reveals an Extensive Increase in Inflammatory and Developmental Gene Expression*. PLoS One, 2015. **10**(12): p. e0143563.
 186. Kumar, A., M. Vaish, and R.R. Ratan, *Transcriptional dysregulation in Huntington's disease: a failure of adaptive transcriptional homeostasis*. Drug Discov Today, 2014. **19**(7): p. 956-62.
 187. Kocerha, J., et al., *Longitudinal transcriptomic dysregulation in the peripheral blood of transgenic Huntington's disease monkeys*. BMC Neurosci, 2013. **14**: p. 88.
 188. Feyeux, M., et al., *Early transcriptional changes linked to naturally occurring Huntington's disease mutations in neural derivatives of human embryonic stem cells*. Hum Mol Genet, 2012. **21**(17): p. 3883-95.
 189. Cha, J.H., *Transcriptional dysregulation in Huntington's disease*. Trends Neurosci, 2000. **23**(9): p. 387-92.
 190. S, B., et al., *Gene Expression Profiling in the Post-Mortem Human Brain--No Cause for Dismay*. Journal of chemical neuroanatomy, 2001. **22**(1-2).
 191. R, L.-C., et al., *Polyglutamine and Transcription: Gene Expression Changes Shared by DRPLA and Huntington's Disease Mouse Models Reveal Context-Independent Effects*. Human molecular genetics, 2002. **11**(17).
 192. Hodges, A., et al., *Regional and cellular gene expression changes in human Huntington's disease brain*. Human Molecular Genetics, 2006. **15**(6): p. 965-977.

193. Perry, T.L., S. Hansen, and M. Kloster, *Huntington's chorea. Deficiency of gamma-aminobutyric acid in brain*. N Engl J Med, 1973. **288**(7): p. 337-42.
194. Vonsattel, J.P., et al., *Neuropathological classification of Huntington's disease*. J Neuropathol Exp Neurol, 1985. **44**(6): p. 559-77.
195. Burrus, C.J., et al., *Striatal Projection Neurons Require Huntingtin for Synaptic Connectivity and Survival*. Cell Rep, 2020. **30**(3): p. 642-657 e6.
196. Langfelder, P., et al., *Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice*. Nat Neurosci, 2016. **19**(4): p. 623-33.
197. Luthi-Carter, R., et al., *Decreased expression of striatal signaling genes in a mouse model of Huntington's disease*. Hum Mol Genet, 2000. **9**(9): p. 1259-71.
198. Ament, S.A., et al., *High resolution time-course mapping of early transcriptomic, molecular and cellular phenotypes in Huntington's disease CAG knock-in mice across multiple genetic backgrounds*. Hum Mol Genet, 2017. **26**(5): p. 913-922.
199. Agus, F., et al., *The caudate nucleus undergoes dramatic and unique transcriptional changes in human prodromal Huntington's disease brain*. BMC Med Genomics, 2019. **12**(1): p. 137.
200. Sugars, K.L. and D.C. Rubinsztein, *Transcriptional abnormalities in Huntington disease*. Trends Genet, 2003. **19**(5): p. 233-8.
201. H, O., *Polyglutamine diseases: a transcription disorder?* Cellular and Molecular Life Sciences : CMLS, 2003. **60**(7): p. 1427-1439.
202. SJ, A., F. RL, and E. PC, *Dopamine D1 and D2 Receptor Gene Expression in the Striatum in Huntington's Disease*. Annals of neurology, 1997. **42**(2).
203. SJ, A., et al., *Reduction in Enkephalin and Substance P Messenger RNA in the Striatum of Early Grade Huntington's Disease: A Detailed Cellular in Situ Hybridization Study*. Neuroscience, 1996. **72**(4).
204. PJ, N., et al., *Decreased Neuronal Nitric Oxide Synthase Messenger RNA and Somatostatin Messenger RNA in the Striatum of Huntington's Disease*. Neuroscience, 1996. **72**(4).
205. George J. Yohrling, I.a.J.-H.J.C., *Neurochemistry of Huntington's disease*. 2002: Oxford University Press.
206. TC, A., et al., *Huntington's Disease Progression. PET and Clinical Observations*. Brain : a journal of neurology, 1999. **122 (Pt 12)**.
207. JH, C., et al., *Altered Neurotransmitter Receptor Expression in Transgenic Mouse Models of Huntington's Disease*. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 1999. **354**(1386).
208. HT, O., *Microarrays and Polyglutamine Disorders: Reports From the Hereditary Disease Array Group*. Human molecular genetics, 2002. **11**(17).
209. EA, T., *Striatal Specificity of Gene Expression Dysregulation in Huntington's Disease*. Journal of neuroscience research, 2006. **84**(6).
210. Cha, J.H.J., *Transcriptional Signatures in Huntington's Disease*. Prog Neurobiol, 2007. **83**(4): p. 228-48.
211. CA, R. and T. SJ, *Huntington's Disease: From Molecular Pathogenesis to Clinical Treatment*. The Lancet. Neurology, 2011. **10**(1).
212. JS, S., et al., *Histone Deacetylase Inhibitors Arrest Polyglutamine-Dependent Neurodegeneration in Drosophila*. Nature, 2001. **413**(6857).

213. L, C., et al., *Transcriptional Repression of PGC-1alpha by Mutant Huntingtin Leads to Mitochondrial Dysfunction and Neurodegeneration*. Cell, 2006. **127**(1).
214. Seredenina, T. and R. Luthi-Carter, *What have we learned from gene expression profiles in Huntington's disease?* Neurobiol Dis, 2012. **45**(1): p. 83-98.
215. Hentosh, P., T. Sundin, and P. Mollica, *Poly-Q aggregate trap interference with DNA repair genes in Huntington disease cells* (736.10). 2014.
216. Moumne, L., S. Betuing, and J. Caboche, *Multiple Aspects of Gene Dysregulation in Huntington's Disease*. Front Neurol, 2013. **4**: p. 127.
217. G, S., et al., *Cellular Toxicity of Polyglutamine Expansion Proteins: Mechanism of Transcription Factor Deactivation*. Molecular cell, 2004. **15**(1).
218. Ruth Luthi-Carter, J.-H.J.C., *Mechanisms of transcriptional dysregulation in Huntington's disease*,. Clinical Neuroscience Research, 2003. **3**(3): p. 165-177.
219. JM, B., et al., *Aberrant Interactions of Transcriptional Repressor Proteins With the Huntington's Disease Gene Product, Huntingtin*. Human molecular genetics, 1999. **8**(9).
220. RN, F. and T. R., *Neurodegeneration. A Glutamine-Rich Trail Leads to Transcription Factors*. Science (New York, N.Y.), 2002. **296**(5576).
221. W, Z., et al., *In Vitro Analysis of Huntingtin-Mediated Transcriptional Repression Reveals Multiple Transcription Factor Targets*. Cell, 2005. **123**(7).
222. Steffan, J.S., et al., *Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in Drosophila*. Nature, 2001. **413**(6857): p. 739-743.
223. Buckley, N.J. and R. Johnson, *New insights into non-coding RNA networks in Huntington's disease*. Exp Neurol, 2011. **231**(2): p. 191-4.
224. Bithell, A., R. Johnson, and N.J. Buckley, *Transcriptional dysregulation of coding and non-coding genes in cellular models of Huntington's disease*. Biochem Soc Trans, 2009. **37**(Pt 6): p. 1270-5.
225. T, S., S. N, and L. GZ, *Protein Misfolding and ER Stress in Huntington's Disease*. Frontiers in molecular biosciences, 2019. **6**.
226. Chen, G., et al., *Brain-Derived Neurotrophic Factor Suppresses Tunicamycin-Induced Upregulation of CHOP in Neurons*. J Neurosci Res, 2007. **85**(8): p. 1674-84.
227. Lajoie, P. and E.L. Snapp, *Changes in BiP availability reveal hypersensitivity to acute endoplasmic reticulum stress in cells expressing mutant huntingtin*. J Cell Sci, 2011. **124**(19): p. 3332-43.
228. Smith, H.L. and G.R. Mallucci, *The unfolded protein response: mechanisms and therapy of neurodegeneration*. Brain, 2016. **139**(Pt 8): p. 2113-21.
229. Halliday, M. and G.R. Mallucci, *Targeting the unfolded protein response in neurodegeneration: A new approach to therapy*. Neuropharmacology, 2014. **76 Pt A**: p. 169-74.
230. Santos, L.E. and S.T. Ferreira, *Crosstalk between endoplasmic reticulum stress and brain inflammation in Alzheimer's disease*. Neuropharmacology, 2018. **136**(Pt B): p. 350-360.
231. Carnemolla, A., et al., *Rrs1 is involved in endoplasmic reticulum stress response in Huntington disease*. J Biol Chem, 2009. **284**(27): p. 18167-73.
232. Yang, H., et al., *Huntingtin Interacts with the Cue Domain of gp78 and Inhibits gp78 Binding to Ubiquitin and p97/VCP*. PLoS One, 2010. **5**(1).

233. Pereira, C.M.F., *Crosstalk between Endoplasmic Reticulum Stress and Protein Misfolding in Neurodegenerative Diseases*. ISRN Cell Biology, 2013. **2013**.
234. ML, D. and L. S, *Impaired ERAD and ER Stress Are Early and Specific Events in Polyglutamine Toxicity*. Genes & development, 2008. **22**(23).
235. Kalathur, R.K., et al., *The unfolded protein response and its potential role in Huntington's disease elucidated by a systems biology approach*. F1000Res, 2015. **4**: p. 103.
236. Sebastian, R.M. and M.D. Shoulders, *Chemical Biology Framework to Illuminate Proteostasis*. Annual Review of Biochemistry, 2020. **89**: p. 529-555.
237. Cohen, D.E., et al., *Sirt1 mediates neuroprotection from mutant huntingtin by activation of TORC1 and CREB transcriptional pathway*. Nat Med, 2012. **18**(1): p. 159-165.
238. Sj, et al., *Subcellular Localization And Formation Of Huntingtin Aggregates Correlates With Symptom Onset And Progression In A Huntington'S Disease Model*. Brain communications, 2020. **2**(2).
239. Jia, H., et al., *Selective histone deacetylase (HDAC) inhibition imparts beneficial effects in Huntington's disease mice: implications for the ubiquitin–proteasomal and autophagy systems*. Hum Mol Genet, 2012. **21**(24): p. 5280-93.
240. Jia, H., et al., *Histone deacetylase (HDAC) inhibitors targeting HDAC3 and HDAC1 ameliorate polyglutamine-elicited phenotypes in model systems of Huntington's disease*. Neurobiol Dis, 2012. **46**(2): p. 351-61.
241. Guan, Z., et al., *Integration of long-term-memory-related synaptic plasticity involves bidirectional regulation of gene expression and chromatin structure*. Cell, 2002. **111**(4): p. 483-93.
242. Devaiah, B.N., et al., *BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain*. Proc Natl Acad Sci U S A, 2012. **109**(18): p. 6927-32.
243. Chiang, C.-M., *Brd4 engagement from chromatin targeting to transcriptional regulation: selective contact with acetylated histone H3 and H4*. F1000Prime Rep, 2016. **1**(98).
244. Tropberger, P., et al., *Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer*. Cell, 2013. **152**(4): p. 859-72.
245. Devaiah, B.N., et al., *BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin*. Nat Struct Mol Biol, 2016. **23**: p. 540-548.
246. Houtkooper, R.H., et al., *The secret life of NAD⁺: an old metabolite controlling new metabolic signaling pathways*. Endocr Rev, 2010. **31**(2): p. 194-223.
247. Imai, S., *Nicotinamide phosphoribosyltransferase (Nampt): a link between NAD biology, metabolism, and diseases*. Curr Pharm Des, 2009. **15**(1): p. 20-8.
248. Zhou, M., et al., *Neuronal death induced by misfolded prion protein is due to NAD⁺ depletion and can be relieved in vitro and in vivo by NAD⁺ replenishment*. Brain, 2015. **138**(Pt 4): p. 992-1008.
249. Weidberg, H., et al., *LC3 and GATE-16/GABARAP subfamilies are both essential yet act differently in autophagosome biogenesis*. EMBO J, 2010. **29**(11): p. 1792-802.
250. Geetha, T., et al., *Sequestosome 1/p62: across diseases*. Biomarkers, 2012. **17**(2): p. 99-103.

251. Bjorkoy, G., et al., *p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death*. J Cell Biol, 2005. **171**(4): p. 603-14.
252. Bjorkoy, G., et al., *Monitoring autophagic degradation of p62/SQSTM1*. Methods Enzymol, 2009. **452**: p. 181-97.
253. Kedersha, N. and P. Anderson, *Stress granules: sites of mRNA triage that regulate mRNA stability and translatability*. Biochem Soc Trans, 2002. **30**(Pt 6): p. 963-9.
254. Li, Y.R., et al., *Stress granules as crucibles of ALS pathogenesis*. J Cell Biol, 2013. **201**(3): p. 361-72.
255. Viiri, J., et al., *Autophagy activation clears ELAVL1/HuR-mediated accumulation of SQSTM1/p62 during proteasomal inhibition in human retinal pigment epithelial cells*. PLoS One, 2013. **8**(7): p. e69563.
256. Srikantan, S. and M. Gorospe, *HuR function in disease*. Front Biosci, 2012. **17**: p. 189-205.
257. Pascale, A. and S. Govoni, *The complex world of post-transcriptional mechanisms: is their deregulation a common link for diseases? Focus on ELAV-like RNA-binding proteins*. Cell Mol Life Sci, 2012. **69**(4): p. 501-17.
258. Martin, S., et al., *Emerging extranuclear roles of protein SUMOylation in neuronal function and dysfunction*. Nat Rev Neurosci, 2007. **8**(12): p. 948-59.
259. O'Rourke, J.G., et al., *SUMO-2 and PIAS1 modulate insoluble mutant huntingtin protein accumulation*. Cell Rep, 2013. **4**(2): p. 362-75.
260. Huang, B., et al., *Brd4 Coactivates Transcriptional Activation of NF- κ B via Specific Binding to Acetylated RelA*. Mol Cell Biol, 2009. **29**(5): p. 1375-87.
261. Zhou, X., et al., *Ribosomal proteins: functions beyond the ribosome*. J Mol Cell Biol, 2015. **7**(2): p. 92-104.
262. Warner, J.R. and K.B. McIntosh, *How Common are Extra-ribosomal Functions of Ribosomal Proteins?* Mol Cell, 2009. **34**(1): p. 3-11.
263. Deisenroth, C. and Y. Zhang, *The Ribosomal Protein-Mdm2-p53 Pathway and Energy Metabolism: Bridging the Gap between Feast and Famine*. Genes Cancer, 2011. **2**(4): p. 392-403.
264. Pelletier, G., et al., *Competitive recruitment of CBP and Rb-HDAC regulates UBF acetylation and ribosomal transcription*. Mol Cell, 2000. **6**(5): p. 1059-66.
265. Satyanarayana, A. and P. Kaldis, *Mammalian cell-cycle regulation: several Cdk, numerous cyclins and diverse compensatory mechanisms*. Oncogene, 2009. **28**(33): p. 2925-2939.
266. Castedo, M., et al., *Cyclin-dependent kinase-1: linking apoptosis to cell cycle and mitotic catastrophe*. Cell Death Differ, 2002. **9**(12): p. 1287-93.
267. Ferri, K.F. and G. Kroemer, *Organelle-specific initiation of cell death pathways*. Nature Cell Biology, 2001. **3**(11).
268. Callen, E., et al., *The DNA Damage- and Transcription-Associated Protein Paxip1 Controls Thymocyte Development and Emigration*. Immunity, 2012. **37**(6): p. 971-985.
269. Munoz, I.M. and J. Rouse, *Control of histone methylation and genome stability by PTIP*. EMBO Rep, 2009. **10**(3): p. 239-45.
270. Enokido, Y., et al., *Mutant huntingtin impairs Ku70-mediated DNA repair*. J Cell Biol, 2010. **189**(3): p. 425-43.
271. Illuzzi, J., et al., *DNA breakage and induction of DNA damage response proteins precede the appearance of visible mutant huntingtin aggregates*. J Neurosci Res, 2009. **87**(3): p. 733-47.

272. Wu, J., et al., *PTIP regulates 53BP1 and SMC1 at the DNA damage sites*. J Biol Chem, 2009. **284**(27): p. 18078-84.
273. Xiao, H., et al., *A polyglutamine expansion disease protein sequesters PTIP to attenuate DNA repair and increase genomic instability*. Hum Mol Genet, 2012. **21**(19): p. 4225-36.
274. Ogryzko, V.V., et al., *The transcriptional coactivators p300 and CBP are histone acetyltransferases*. Cell, 1996. **87**(5): p. 953-9.
275. Vo, N. and R.H. Goodman, *CREB-binding protein and p300 in transcriptional regulation*. J Biol Chem, 2001. **276**(17): p. 13505-8.
276. Freedman, S.J., et al., *Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 α* . Proceedings of the National Academy of Sciences, 2002. **99**(8): p. 5367-5372.
277. Ryu, K.Y., et al., *The mouse polyubiquitin gene UbC is essential for fetal liver development, cell-cycle progression and stress tolerance*. Embo j, 2007. **26**(11): p. 2693-706.
278. Bennett, E.J., et al., *Global changes to the ubiquitin system in Huntington's disease*. Nature, 2007. **448**(7154): p. 704-8.
279. Trager, U., et al., *HTT-lowering reverses Huntington's disease immune dysfunction caused by NFkappaB pathway dysregulation*. Brain, 2014. **137**(Pt 3): p. 819-33.
280. Steffan, J.S., et al., *The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription*. Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6763-8.
281. An, W.G., et al., *Stabilization of wild-type p53 by hypoxia-inducible factor 1 α* . Nature, 1998. **392**(6674): p. 405-8.
282. Emanuele, M.J., et al., *Proliferating cell nuclear antigen (PCNA)-associated KIAA0101/PAF15 protein is a cell cycle-regulated anaphase-promoting complex/cyclosome substrate*. Proc Natl Acad Sci U S A, 2011. **108**(24): p. 9845-50.
283. Klein, J.A. and S.L. Ackerman, *Oxidative stress, cell cycle, and neurodegeneration*. J Clin Invest, 2003. **111**(6): p. 785-93.
284. Jeppesen, D.K., V.A. Bohr, and T. Stevnsner, *DNA Repair Deficiency in Neurodegeneration*. Prog Neurobiol, 2011. **94**(2): p. 166-200.
285. Yoshida, E., et al., *Functional association between CBP and HNF4 in trans-activation*. Biochem Biophys Res Commun, 1997. **241**(3): p. 664-9.
286. Dell, H. and M. Hadzopoulou-Cladaras, *CREB-binding protein is a transcriptional coactivator for hepatocyte nuclear factor-4 and enhances apolipoprotein gene expression*. J Biol Chem, 1999. **274**(13): p. 9013-21.
287. Block, R.C., et al., *Altered Cholesterol and Fatty Acid Metabolism in Huntington Disease*. J Clin Lipidol, 2010. **4**(1): p. 17-23.
288. Olah, J., et al., *Increased glucose metabolism and ATP level in brain tissue of Huntington's disease transgenic mice*. Febs j, 2008. **275**(19): p. 4740-55.
289. Riggelen, J.v., A. Yetil, and D.W. Felsher, *MYC as a regulator of ribosome biogenesis and protein synthesis*. Nat Rev Cancer, 2010. **10**(4): p. 301-309.
290. Vervoorts, J., et al., *Stimulation of c-MYC transcriptional activity and acetylation by recruitment of the cofactor CBP*. EMBO Rep, 2003. **4**(5): p. 484-90.
291. Lee, H., et al., *Persistently-activated Stat3 maintains constitutive NF-kB activity in tumors*. Cancer Cell, 2009. **15**(4): p. 283-93.

292. Grivennikov, S.I. and M. Karin, *Dangerous liaisons: STAT3 and NF-kappaB collaboration and crosstalk in cancer*. Cytokine Growth Factor Rev, 2010. **21**(1): p. 11-9.
293. Wang, R., P. Cherukuri, and J. Luo, *Activation of Stat3 sequence-specific DNA binding and transcription by p300/CREB-binding protein-mediated acetylation*. J Biol Chem, 2005. **280**(12): p. 11528-34.
294. Maxwell, P.H., et al., *The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis*. Nature, 1999. **399**(6733): p. 271-275.
295. Zhang, Z., et al., *Hypoxia Inducible Factor-1 as a Target for Neurodegenerative Diseases*. Curr Med Chem, 2011. **18**(28): p. 4335-43.
296. Arany, Z., et al., *An essential role for p300/CBP in the cellular response to hypoxia*. Proc Natl Acad Sci U S A, 1996. **93**(23): p. 12969-73.
297. Isaacs, J.S., et al., *Hsp90 regulates a von Hippel Lindau-independent hypoxia-inducible factor-1 alpha-degradative pathway*. J Biol Chem, 2002. **277**(33): p. 29936-44.
298. Luo, W., et al., *Heat shock protein 90 in neurodegenerative diseases*. Mol Neurodegener, 2010. **5**: p. 24.
299. Sipe, J.D., et al., *Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines*. Amyloid, 2016. **23**(4): p. 209-213.
300. Moily, N.S., et al., *Transcriptional profiles for distinct aggregation states of mutant Huntingtin exon 1 protein unmask new Huntington's disease pathways*. Mol Cell Neurosci, 2017. **83**: p. 103-112.
301. Choi, Y.S., et al., *CREB is a key regulator of striatal vulnerability in chemical and genetic models of Huntington's disease*. Neurobiol Dis, 2009. **36**(2): p. 259-68.
302. Wood, H., *Neurodegenerative disease: altered DNA methylation and RNA splicing could be key mechanisms in Huntington disease*. Nat Rev Neurol, 2013. **9**(3): p. 119.
303. Lin, L., et al., *Transcriptome sequencing reveals aberrant alternative splicing in Huntington's disease*. Hum Mol Genet, 2016. **25**(16): p. 3454-3466.
304. J, S., et al., *Deregulated Splicing Is a Major Mechanism of RNA-Induced Toxicity in Huntington's Disease*. Journal of molecular biology, 2019. **431**(9).
305. Cabrera, J.R. and J.J. Lucas, *MAP2 Splicing is Altered in Huntington's Disease*. Brain Pathol, 2017. **27**(2): p. 181-189.
306. Labadorf, A.T. and R.H. Myers, *Evidence of Extensive Alternative Splicing in Post Mortem Human Brain HTT Transcription by mRNA Sequencing*. PLoS One, 2015. **10**(10): p. e0141298.
307. Nik, S. and T.V. Bowman, *Splicing and neurodegeneration: Insights and mechanisms*. Wiley Interdiscip Rev RNA, 2019. **10**(4): p. e1532.
308. Carrocci, T.J. and K.M. Neugebauer, *Pre-mRNA Splicing in the Nuclear Landscape*. Cold Spring Harb Symp Quant Biol, 2020.
309. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing*. Nat Genet, 2008. **40**(12): p. 1413-5.
310. Liu, Y., et al., *Impact of Alternative Splicing on the Human Proteome*. Cell Rep, 2017. **20**(5): p. 1229-1241.
311. O, K., et al., *Function of Alternative Splicing*. Gene, 2013. **514**(1).

312. G, B. and C. JF, *Cellular Stress and RNA Splicing*. Trends in biochemical sciences, 2009. **34**(3).
313. Carew, N.T., et al., *Linking Endoplasmic Reticular Stress and Alternative Splicing*. Int J Mol Sci, 2018. **19**(12).
314. Wahl, M.C., C.L. Will, and R. Lührmann, *The spliceosome: design principles of a dynamic RNP machine*. Cell, 2009. **136**(4): p. 701-18.
315. Wahl, M.C. and R. Lührmann, *SnapShot: Spliceosome Dynamics I*. Cell, 2015. **161**(6): p. 1474-e1.
316. Wahl, M.C. and R. Lührmann, *SnapShot: Spliceosome Dynamics II*. Cell, 2015. **162**(2): p. 456-456.e1.
317. Wahl, M.C. and R. Lührmann, *SnapShot: Spliceosome Dynamics III*. Cell, 2015. **162**(3): p. 690-690.e1.
318. Matera, A.G. and Z. Wang, *A day in the life of the spliceosome*. Nat Rev Mol Cell Biol, 2014. **15**(2): p. 108-21.
319. Barash, Y., et al., *Deciphering the splicing code*. Nature, 2010. **465**(7294): p. 53-9.
320. Olthof, A.M., K.C. Hyatt, and R.N. Kanadia, *Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns*. BMC Genomics, 2019. **20**(1): p. 686.
321. Verma, B., et al., *Minor spliceosome and disease*. Semin Cell Dev Biol, 2018. **79**: p. 103-112.
322. Turunen, J.J., et al., *The significant other: splicing by the minor spliceosome*. Wiley Interdiscip Rev RNA, 2013. **4**(1): p. 61-76.
323. Madhani, H.D. and C. Guthrie, *A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome*. Cell, 1992. **71**(5): p. 803-17.
324. AJ, N. and N. C, *U5 snRNA Interacts With Exon Sequences at 5' and 3' Splice Sites*. Cell, 1992. **68**(4).
325. Fourmann, J.B., et al., *Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system*. Genes Dev, 2013. **27**(4): p. 413-28.
326. Shcherbakova, I., et al., *Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy*. Cell Rep, 2013. **5**(1): p. 151-65.
327. Matlin, A.J. and M.J. Moore, *Spliceosome assembly and composition*. Adv Exp Med Biol, 2007. **623**: p. 14-35.
328. Hoskins, A.A. and M.J. Moore, *The spliceosome: a flexible, reversible macromolecular machine*. Trends Biochem Sci, 2012. **37**(5): p. 179-88.
329. Will, C.L. and R. Lührmann, *Spliceosome Structure and Function*. Cold Spring Harb Perspect Biol, 2011. **3**(7).
330. Patel, S.B. and M. Bellini, *The assembly of a spliceosomal small nuclear ribonucleoprotein particle*. Nucleic Acids Res, 2008. **36**(20): p. 6482-93.
331. Zhou, Z. and X.D. Fu, *Regulation of splicing by SR proteins and SR protein-specific kinases*. Chromosoma, 2013. **122**(3): p. 191-207.
332. Shepard, P.J. and K.J. Hertel, *The SR protein family*. Genome Biol, 2009. **10**(10): p. 242.
333. Long, J.C. and J.F. Cáceres, *The SR protein family of splicing factors: master regulators of gene expression*. Biochem J, 2009. **417**(1): p. 15-27.
334. Spector, D.L., *Nuclear organization and gene expression*. Exp Cell Res, 1996. **229**(2): p. 189-97.

335. Spector, D.L., *Macromolecular domains within the cell nucleus*. Annu Rev Cell Biol, 1993. **9**: p. 265-315.
336. Ibrahim, E.C., et al., *Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers*. Proc Natl Acad Sci U S A, 2005. **102**(14): p. 5002-7.
337. Robberson, B.L., G.J. Cote, and S.M. Berget, *Exon definition may facilitate splice site selection in RNAs with multiple exons*. Mol Cell Biol, 1990. **10**(1): p. 84-94.
338. Fu, X.D. and T. Maniatis, *Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus*. Nature, 1990. **343**(6257): p. 437-41.
339. Fu, X.D. and T. Maniatis, *The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site*. Proc Natl Acad Sci U S A, 1992. **89**(5): p. 1725-9.
340. Fernández-Nogales, M., et al., *Faulty splicing and cytoskeleton abnormalities in Huntington's disease*. Brain Pathol, 2016. **26**(6): p. 772-778.
341. Schilling, J., et al., *Deregulated Splicing Is a Major Mechanism of RNA-Induced Toxicity in Huntington's Disease*. J Mol Biol, 2019. **431**(9): p. 1869-1877.
342. Neueder, A., et al., *Regulatory mechanisms of incomplete huntingtin mRNA splicing*. Nat Commun, 2018. **9**(1): p. 3955.
343. Tsoi, H., et al., *Perturbation of U2AF65/NXF1-mediated RNA nuclear export enhances RNA toxicity in polyQ diseases*. Hum Mol Genet, 2011. **20**(19): p. 3787-97.
344. Wojciechowska, M., et al., *RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders*. Nucleic Acids Res, 2014. **42**(19): p. 11849-64.
345. Neueder, A. and G.P. Bates, *RNA Related Pathology in Huntington's Disease*. Adv Exp Med Biol, 2018. **1049**: p. 85-101.
346. Bañez-Coronel, M., et al., *RAN Translation in Huntington Disease*. Neuron, 2015. **88**(4): p. 667-77.
347. Ratovitski, T., et al., *Huntingtin protein interactions altered by polyglutamine expansion as determined by quantitative proteomic analysis*. Cell Cycle, 2012. **11**(10): p. 2006-21.
348. Gipson, T.A., et al., *Aberrantly spliced HTT, a new player in Huntington's disease pathogenesis*. RNA Biol, 2013. **10**(11): p. 1647-52.
349. Ramdzan, Y.M., et al., *Tracking protein aggregation and mislocalization in cells with flow cytometry*. Nat Methods, 2012. **9**(5): p. 467-70.
350. Ramdzan, Y.M., et al., *Conformation sensors that distinguish monomeric proteins from oligomers in live cells*. Chem Biol, 2010. **17**(4): p. 371-9.
351. Xu, X., et al., *Reversal of Phenotypic Abnormalities by CRISPR/Cas9-Mediated Gene Correction in Huntington Disease Patient-Derived Induced Pluripotent Stem Cells*. Stem Cell Reports, 2017. **8**(3): p. 619-33.
352. W, L., et al., *Rapid Induction and Long-Term Self-Renewal of Primitive Neural Precursors From Human Embryonic Stem Cells by Small Molecule Inhibitors*. Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(20).
353. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner*. Bioinformatics, 2013. **29**(1): p. 15-21.

354. Shen, S., et al., *rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data*. Proc Natl Acad Sci U S A, 2014. **111**(51): p. E5593-601.
355. Hunt, S.E., et al., *Ensembl variation resources*. Database (Oxford), 2018. **2018**.
356. Kakaradov, B., et al., *Challenges in estimating percent inclusion of alternatively spliced junctions from RNA-seq data*. BMC Bioinformatics, 2012. **13 Suppl 6**: p. S11.
357. D, F.T.V. *maser: Mapping Alternative Splicing Events to pRoteins*. R package version 1.6.0. 2020; Available from: <https://github.com/DiogoVeiga/maser>.
358. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nat Protoc, 2009. **4**(1): p. 44-57.
359. Xia, J., M.J. Benner, and R.E. Hancock, *NetworkAnalyst--integrative approaches for protein-protein interaction network analysis and visual exploration*. Nucleic Acids Res, 2014. **42**(Web Server issue): p. W167-74.
360. Xia, J., E.E. Gill, and R.E. Hancock, *NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data*. Nat Protoc, 2015. **10**(6): p. 823-44.
361. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life*. Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
362. MA, K., et al., *Identification of Differentially Expressed Splice Variants by the Proteogenomic Pipeline Splicify*. Molecular & cellular proteomics : MCP, 2017. **16**(10).
363. Consortium, T.U., *UniProt: The Universal Protein Knowledgebase*. Nucleic acids research, 2017. **45**(D1).
364. J, C. and M. M, *MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-range Mass Accuracies and Proteome-Wide Protein Quantification*. Nature biotechnology, 2008. **26**(12).
365. Kametsky, L., et al., *Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software*. Bioinformatics, 2011. **27**(8): p. 1179-80.
366. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis*. Nature Methods, 2012. **9**: p. 676-682.
367. Zhong, X.Y., et al., *SR proteins in vertical integration of gene expression from transcription to RNA processing to translation*. Mol Cell, 2009. **35**(1): p. 1-10.
368. Staknis, D. and R. Reed, *SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex*. Mol Cell Biol, 1994. **14**(11): p. 7670-82.
369. Lee, Y. and D.C. Rio, *Mechanisms and Regulation of Alternative Pre-mRNA Splicing*. Annu Rev Biochem, 2015. **84**: p. 291-323.
370. Fica, S.M., et al., *RNA catalyses nuclear pre-mRNA splicing*. Nature, 2013. **503**(7475): p. 229-34.
371. Tsuiji, H., et al., *Spliceosome integrity is defective in the motor neuron diseases ALS and SMA*. EMBO Mol Med, 2013. **5**(2): p. 221-34.
372. Arakelian, E. and F. Kfoury, *Predicting Percent Spliced In (PSI) in Alternative Splicing Using Deep Networks*. 2021.

373. *Hypergeometric p-value calculator*. Available from: <https://systems.crump.ucla.edu/hypergeometric/>.
374. Pierre, S., et al., *Capturing adenylyl cyclases as potential drug targets*. Nat Rev Drug Discov, 2009. **8**(4): p. 321-35.
375. K, S., P. S., and P. TB, *Protein Associated With Myc (PAM) Is a Potent Inhibitor of Adenylyl Cyclases*. The Journal of biological chemistry, 2001. **276**(50).
376. OM, S., J. B., and M. U, *A Dominant Role for the Raf-MEK Pathway in Forskolin, 12-O-tetradecanoyl-phorbol Acetate, and Platelet-Derived Growth Factor-Induced CREB (cAMP-responsive Element-Binding Protein) Activation, Uncoupled From Serine 133 Phosphorylation in NIH 3T3 Cells*. Molecular endocrinology (Baltimore, Md.), 1999. **13**(7).
377. B, T., et al., *Forkhead Box Protein p1 Is a Transcriptional Repressor of Immune Signaling in the CNS: Implications for Transcriptional Dysregulation in Huntington Disease*. Human molecular genetics, 2012. **21**(14).
378. M, C., et al., *Bioinformatics Analysis of Ras Homologue Enriched in the Striatum, a Potential Target for Huntington's Disease Therapy*. International journal of molecular medicine, 2019. **44**(6).
379. Ogata, H., et al., *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Nucleic Acids Res, 1999. **27**(1): p. 29-34.
380. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic Acids Res, 2000. **28**(1): p. 27-30.
381. Tsirigotis, M., et al., *Activation of p38MAPK contributes to expanded polyglutamine-induced cytotoxicity*. PLoS One, 2008. **3**(5): p. e2130.
382. Ratovitski, T., et al., *PRMT5- mediated symmetric arginine dimethylation is attenuated by mutant huntingtin and is impaired in Huntington's disease (HD)*. Cell Cycle, 2015. **14**(11): p. 1716-29.
383. Xia, J., M.J. Benner, and R.E.W. Hancock, *NetworkAnalyst - integrative approaches for protein-protein interaction network analysis and visual exploration*. 2014.
384. Slepnev, V.I. and P. De Camilli, *Accessory factors in clathrin-dependent synaptic vesicle endocytosis*. Nat Rev Neurosci, 2000. **1**(3): p. 161-72.
385. Rao, D.S., et al., *Huntingtin interacting protein 1 Is a clathrin coat binding protein required for differentiation of late spermatogenic progenitors*. Mol Cell Biol, 2001. **21**(22): p. 7796-806.
386. Regan-Klapisz, E., et al., *Ubiquilin recruits Eps15 into ubiquitin-rich cytoplasmic aggregates via a UIM-UBL interaction*. J Cell Sci, 2005. **118**(Pt 19): p. 4437-50.
387. Umrani, M.R., et al., *Connexins and microRNAs: Interlinked players in regulating islet function?* Islets, 2017. **9**(5): p. 99-108.
388. Pardo, P.S., et al., *MRG15, a novel chromodomain protein, is present in two distinct multiprotein complexes involved in transcriptional activation*. J Biol Chem, 2002. **277**(52): p. 50860-6.
389. Hayakawa, T., et al., *MRG15 binds directly to PALB2 and stimulates homology-directed repair of chromosomal breaks*. J Cell Sci, 2010. **123**(Pt 7): p. 1124-30.
390. S, Z. and G. F, *Nuclear DNA Helicase II Unwinds Both DNA and RNA*. Biochemistry, 1994. **33**(13).
391. Aratani, S., et al., *Dual Roles of RNA Helicase A in CREB-Dependent Transcription*. Mol Cell Biol, 2001. **21**(14): p. 4460-9.

392. A, S., et al., *Control of Cognition and Adaptive Behavior by the GLP/G9a Epigenetic Suppressor Complex*. Neuron, 2009. **64**(5).
393. Hervás-Corpión, I., et al., *Early alteration of epigenetic-related transcription in Huntington's disease mouse models*. Scientific Reports, 2018. **8**(1): p. 1-14.
394. A, K., et al., *Activation of the I kappa B Kinase Complex and Nuclear factor-kappa B Contributes to Mutant Huntingtin Neurotoxicity*. The Journal of neuroscience : the official journal of the Society for Neuroscience, 2004. **24**(37).
395. Mattson and S. Camandola, *NF-κB in neuronal plasticity and neurodegenerative disorders*. J Clin Invest, 2001. **107**(3): p. 247-54.
396. ZH, Q., et al., *Nuclear Factor kappa B Nuclear Translocation Upregulates c-Myc and p53 Expression During NMDA Receptor-Mediated Apoptosis in Rat Striatum*. The Journal of neuroscience : the official journal of the Society for Neuroscience, 1999. **19**(10).
397. RH, L., et al., *Nuclear Factor kappa B Is a Critical Determinant in N-methyl-D-aspartate Receptor-Mediated Neuroprotection*. Journal of neurochemistry, 2001. **78**(2).
398. Haack, K.K., A.K. Mitra, and I.H. Zucker, *NF-κB and CREB Are Required for Angiotensin II Type 1 Receptor Upregulation in Neurons*. PloS one, 2013. **8**(11).
399. Fan, Y., et al., *Regulation of programmed cell death by NF-kappaB and its role in tumorigenesis and therapy*. Adv Exp Med Biol, 2008. **615**: p. 223-50.
400. Oberst, A., *Death in the fast lane: what's next for necroptosis?* FEBS J, 2016. **283**(14): p. 2616-25.
401. Luo, J.L., H. Kamata, and M. Karin, *IKK/NF-kappaB signaling: balancing life and death--a new approach to cancer therapy*. J Clin Invest, 2005. **115**(10): p. 2625-32.
402. Khosravi-Far, R. and E. White, *Programmed Cell Death in Cancer Progression and Therapy*. 2008: Springer, Dordrecht.
403. AL, E. and T. CB, *Death by design: apoptosis, necrosis and autophagy*. Current opinion in cell biology, 2004. **16**(6).
404. Yalamanchili, N., et al., *Distinct Cell Stress Responses Induced by ATP Restriction in Quiescent Human Fibroblasts*. Front Genet, 2016. **7**: p. 171.
405. Lieberthal, W., S.A. Menza, and J.S. Levine, *Graded ATP depletion can cause necrosis or apoptosis of cultured mouse proximal tubular cells*. Am J Physiol, 1998. **274**(2): p. F315-27.
406. Leeman, J.G., RaThomas D., *Alternative splicing in the NF-κB signaling pathway*. Gene, 2008. **423**(2): p. 97-107.
407. RJ, G. and G. S, *Alternative splicing of RNA transcripts encoded by the murine p105 NF-kappa B gene generates I kappa B gamma isoforms with different inhibitory activities*. Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(10).
408. Khavari, P.A., et al., *BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription*. Nature, 1993. **366**(6451): p. 170-4.
409. Hubert, L., et al., *Topoisomerase 1 and single-strand break repair modulate transcription-induced CAG repeat contraction in human cells*. Mol Cell Biol, 2011. **31**(15): p. 3105-12.
410. Quinti, L., et al., *Evaluation of histone deacetylases as drug targets in Huntington's disease models. Study of HDACs in brain tissues from R6/2 and*

- CAG140 knock-in HD mouse models and human patients and in a neuronal HD cell model. PLoS Curr*, 2010. **2**.
411. Hockly, E., et al., *Suberoylanilide hydroxamic acid, a histone deacetylase inhibitor, ameliorates motor deficits in a mouse model of Huntington's disease. Proc Natl Acad Sci U S A*, 2003. **100**(4): p. 2041-6.
 412. Ferrante, R.J., et al., *Histone deacetylase inhibition by sodium butyrate chemotherapy ameliorates the neurodegenerative phenotype in Huntington's disease mice. J Neurosci*, 2003. **23**(28): p. 9418-27.
 413. Pallos, J., et al., *Inhibition of specific HDACs and sirtuins suppresses pathogenesis in a Drosophila model of Huntington's disease. Hum Mol Genet*, 2008. **17**(23): p. 3767-75.
 414. Bates, E.A., et al., *Differential contributions of Caenorhabditis elegans histone deacetylases to huntingtin polyglutamine toxicity. J Neurosci*, 2006. **26**(10): p. 2830-8.
 415. Gray, S.G., *Targeting Huntington's disease through histone deacetylases. Clin Epigenetics*, 2011. **2**(2): p. 257-77.
 416. Gray, S.G., *Targeting histone deacetylases for the treatment of Huntington's disease. CNS Neurosci Ther*, 2010. **16**(6): p. 348-61.
 417. Kobayashi, M., et al., *The ubiquitin hybrid gene UBA52 regulates ubiquitination of ribosome and sustains embryonic development. Sci Rep*, 2016. **6**: p. 36780.
 418. Kimura, Y. and K. Tanaka, *Regulatory mechanisms involved in the control of ubiquitin homeostasis. J Biochem*, 2010. **147**(6): p. 793-8.
 419. Mao, J., et al., *Ubiquitin A-52 residue ribosomal protein fusion product 1 (Uba52) is essential for preimplantation embryo development. Biol Open*, 2018. **7**(10).
 420. E, J., et al., *The proteasome in health and disease. Current pharmaceutical design*, 2013. **19**(6).
 421. I, A., S. T., and W. DH, *Protein quality control and elimination of protein waste: the role of the ubiquitin-proteasome system. Biochimica et biophysica acta*, 2014. **1843**(1).
 422. M, T., et al., *Protein aggregation and degradation mechanisms in neurodegenerative diseases. American journal of neurodegenerative disease*, 2013. **2**(1).
 423. Kepchia, D., et al., *Diverse proteins aggregate in mild cognitive impairment and Alzheimer's disease brain. Alzheimer's Research & Therapy*, 2020. **12**(1): p. 1-20.
 424. Touré, A., et al., *MgcRacGAP, a new human GTPase-activating protein for Rac and Cdc42 similar to Drosophila rotundRacGAP gene product, is expressed in male germ cells. J Biol Chem*, 1998. **273**(11): p. 6019-23.
 425. Tourette, C., et al., *A Large Scale Huntingtin Protein Interaction Network Implicates Rho GTPase Signaling Pathways in Huntington Disease. J Biol Chem*, 2014. **289**(10): p. 6709-26.
 426. TR, S. and L. DA, *Rho family GTPases: key players in neuronal development, neuronal survival, and neurodegeneration. Frontiers in cellular neuroscience*, 2014. **8**.
 427. Ma, M.W., et al., *NADPH oxidase in brain injury and neurodegenerative disorders. Molecular Neurodegeneration*, 2017. **12**(1): p. 1-28.

428. Droppelmann, C.A., et al., *The emerging role of guanine nucleotide exchange factors in ALS and other neurodegenerative diseases*. Front Cell Neurosci, 2014. **8**: p. 282.
429. Radha, V., et al., *The guanine nucleotide exchange factor, C3G regulates differentiation and survival of human neuroblastoma cells*. J Neurochem, 2008. **107**(5): p. 1424-35.
430. DA, L. and L. FA, *Diverse roles of Rho family GTPases in neuronal development, survival, and death*. Frontiers in bioscience : a journal and virtual library, 2008. **13**.
431. CA, D., et al., *The emerging role of guanine nucleotide exchange factors in ALS and other neurodegenerative diseases*. Frontiers in cellular neuroscience, 2014. **8**.
432. Pereira, A.L., et al., *Mammalian CLASP1 and CLASP2 cooperate to ensure mitotic fidelity by regulating spindle and kinetochore function*. Mol Biol Cell, 2006. **17**(10): p. 4526-42.
433. Kapur, M., C.E. Monaghan, and S.L. Ackerman, *Regulation of mRNA Translation in Neurons-A Matter of Life and Death*. Neuron, 2017. **96**(3): p. 616-637.
434. PK, C., et al., *Mutations in TRNT1 cause congenital sideroblastic anemia with immunodeficiency, fevers, and developmental delay (SIFD)*. Blood, 2014. **124**(18).
435. Sasarman, F., et al., *The 3' addition of CCA to mitochondrial tRNA^{Ser}(AGY) is specifically impaired in patients with mutations in the tRNA nucleotidyl transferase TRNT1*. Hum Mol Genet, 2015. **24**(10): p. 2841-7.
436. DeLuca, A.P., et al., *Hypomorphic mutations in TRNT1 cause retinitis pigmentosa with erythrocytic microcytosis*. Hum Mol Genet, 2016. **25**(1): p. 44-56.
437. Giannelou, A., et al., *Aberrant tRNA processing causes an autoinflammatory syndrome responsive to TNF inhibitors*. Ann Rheum Dis, 2018. **77**(4): p. 612-619.
438. Sharma, T.P., et al., *Patient-specific induced pluripotent stem cells to evaluate the pathophysiology of TRNT1-associated Retinitis pigmentosa*. Stem Cell Res, 2017. **21**: p. 58-70.
439. Liwak-Muir, U., et al., *Impaired activity of CCA-adding enzyme TRNT1 impacts OXPHOS complexes and cellular respiration in SIFD patient-derived fibroblasts*. Orphanet J Rare Dis, 2016. **11**(1): p. 79.
440. Kampinga, H.H. and E.A. Craig, *The HSP70 chaperone machinery: J proteins as drivers of functional specificity*. Nat Rev Mol Cell Biol, 2010. **11**(8): p. 579-92.
441. Kim, Y.E., et al., *Soluble Oligomers of PolyQ-Expanded Huntingtin Target a Multiplicity of Key Cellular Factors*. Mol Cell, 2016. **63**(6): p. 951-64.
442. X, S., et al., *Widespread remodeling of proteome solubility in response to different protein homeostasis stresses*. Proceedings of the National Academy of Sciences of the United States of America, 2020. **117**(5).
443. Radwan, M., et al., *Immiscible inclusion bodies formed by polyglutamine and poly(glycine-alanine) are enriched with distinct proteomes but converge in proteins that are risk factors for disease and involved in protein degradation*. PLoS One, 2020. **15**(8): p. e0233247.

444. Park, S.H., et al., *PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone*. Cell, 2013. **154**(1): p. 134-45.
445. Tan, W., W. Wang, and Q. Ma, *Physiological and Pathological Function of Serine/Arginine-Rich Splicing Factor 4 and Related Diseases*. Biomed Res Int, 2018. **2018**: p. 3819719.
446. Sapra, A.K., et al., *SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo*. Mol Cell, 2009. **34**(2): p. 179-90.
447. YC, H., et al., *Tau-Mediated Disruption of the Spliceosome Triggers Cryptic RNA Splicing and Neurodegeneration in Alzheimer's Disease*. Cell reports, 2019. **29**(2).
448. H, J., et al., *Sirt1 mediates neuroprotection from mutant huntingtin by activation of the TORC1 and CREB transcriptional pathway*. Nature medicine, 2011. **18**(1).
449. RK, C., et al., *Transducer of regulated CREB-binding proteins (TORCs) transcription and function is impaired in Huntington's disease*. Human molecular genetics, 2012. **21**(15).
450. Amelio, A.L., M. Caputi, and M.D. Conkright, *Bipartite functions of the CREB co-activators selectively direct alternative splicing or transcriptional activation*. EMBO J, 2009. **28**(18): p. 2733-47.
451. RJ, H. and T. YF, *Proteostasis in Huntington's disease: disease mechanisms and therapeutic opportunities*. Acta pharmacologica Sinica, 2018. **39**(5).
452. TR, S., et al., *Targeting the proteostasis network in Huntington's disease*. Ageing research reviews, 2019. **49**.
453. C, C., et al., *Mitochondrial Dysfunction in Huntington's Disease*. Advances in experimental medicine and biology, 2018. **1049**.
454. TA, I., M. EA, and G.-P. R, *Mitochondrial Dysfunction in Huntington's Disease; Interplay Between HSF1, p53 and PGC-1 α Transcription Factors*. Frontiers in cellular neuroscience, 2019. **13**.
455. Arrasate, M., et al., *Inclusion Body Formation Reduces Levels of Mutant Huntingtin and the Risk of Neuronal Death*. Nature, 2004. **431**(7010).
456. F, S., et al., *Huntingtin Acts in the Nucleus to Induce Apoptosis but Death Does Not Correlate With the Formation of Intranuclear Inclusions*. Cell, 1998. **95**(1).
457. T, M., et al., *Disruption of CREB function in brain leads to neurodegeneration*. Nature genetics, 2002. **31**(1).
458. Abd-Elrahman, K.S. and S.S.G. Ferguson, *Modulation of mTOR and CREB pathways following mGluR5 blockade contribute to improved Huntington's pathology in zQ 175 mice*. Molecular Brain, 2019. **12**(1): p. 1-9.
459. Obrietan, K. and K.R. Hoyt, *CRE-Mediated Transcription Is Increased in Huntington's Disease Transgenic Mice*. 2004.
460. FC, N., et al., *Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity*. Science (New York, N.Y.), 2001. **291**(5512).
461. Steffan, J.S., et al., *The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription*. 2000.
462. IH, H., et al., *Pathogenic SREK1 decrease in Huntington's disease lowers TAF1 mimicking X-linked dystonia parkinsonism*. Brain : a journal of neurology, 2020. **143**(7).

- 463. Kornblihtt, A.R., et al., *Multiple links between transcription and splicing*. RNA, 2004. **10**(10): p. 1489-98.
- 464. Chen, M.Z., et al., *A thiol probe for measuring unfolded protein load and proteostasis in cells*. Nature Communications, 2017. **8**(1): p. 474.
- 465. JJ, Y., et al., *Walking the tightrope: proteostasis and neurodegenerative disease*. Journal of neurochemistry, 2016. **137**(4).
- 466. A, K., et al., *Cellular Proteostasis in Neurodegeneration*. Molecular neurobiology, 2019. **56**(5).