THE UNIVERSITY
of EDINBURGH

The effects of mutations in the eEF1A2 gene in mouse gene expression profiles and identification of potential markers for motor neuron degeneration.

Vesa Qarkaxhija

MSc Genetics and Molecular Medicine.

University of Edinburgh

2017

Declaration page

I declare that this thesis is my own work unless stated otherwise and any inclusions of contributions or publications from others have been clearly acknowledged. This thesis has not been submitted for any other degree or professional qualification.

Vesa Qarkaxhija

Acknowledgments

I would like to thank Cathy Abbott, whose guidance and support has made this year both incredibly enlightening and enjoyable. I have learnt an inexpressible amount from her and could not have asked for a more wonderful supervisor.

I would also like to thank the Abbott lab group, who have also taught me so much, in the lab and out, whom I had many a titillating conversations and debates with that made my time here most entertaining.

My family and friends both old and new who had enriched this experience and accompanied me on many Scottish adventures during the course of my masters.

Contents:

Abstract:

The elongation factor 1 alpha (eEF1A) exists in mammals as two highly conserved isoforms: eEF1A1 and eEF1A2 which share 98% amino acid sequence similarity. When bound with GTP, both forms recruit aminoacylated-tRNA for delivery to the ribosome during translation elongation. eEF1A1 is expressed ubiquitously during development and is downregulated in mature neurones, cardiomyocytes and myocytes. Downregulation is observed concurrently with eEF1A2 expression increasing in the terminally differentiated cells. This shift in expression may be resultant of non-canonical roles that can differ between isoforms, and although eEF1A1 is well characterised, less is known about eEF1A2. Given the tissue-specific nature of this shift, it suggests that eEF1A2 may be involved in the development of neurodegeneration. eEF1A2 in humans has been implicated in severe neurodevelopmental disorders, in which sufferers can display symptoms of repeated seizures, intellectual disability and autism. However, patients carry differing mutations in eEF1A2 and each case can present varied severity of symptoms. To explore the effects that mutations in eEF1A2 have, two mouse lines were generated using CRISPR/Cas9; a mutation that was found in humans, D252H and a deletion that arose in the founders, Del.22.ex3. Homozygous (-/-) mice displayed a severe neurodegenerative phenotype. In Del.22.ex3, eEF1A2 is absent in homozygotes, whereas in D252H, mice express eEF1A2 but the protein is impaired or non-functional. An analysis of the founder mice identified mosaic alleles, some had incorporated the target mutation but a range of insertions and deletions were also present. The expression of eEF1A2 was observed to be reduced across the mosaic mice. The extent of neuronal damage that loss of functioning eEF1A2 may cause was investigated by immunohistochemistry. Identification of biomarkers for prognostic purposes for potential therapies of motor neuron degeneration was conducted by a bottom up proteomic approach. Label-free quantitative mass spectrometry was used to define the proteome of spinal cords from homozygotes and wild types for comparative study and identified potential biomarkers. In complement, an analysis on microarray data from wasted mice spinal cords identified differentially expressed genes. Some of these supported proteins of interest as being significantly differentially regulated, whilst not being confounded by varying protein turnover rates or stability. Proteins and genes that were significantly differentially expressed underwent gene ontology enrichment analysis exploring which pathways and functions were over-represented to better understand pathogenesis, some of which demonstrated affiliation with neuronal disorders and cell metabolism. Understanding the loss of eEF1A2 and its neuronal degeneration phenotype, the affected protein and genetic expression patterns across the spinal cord has elucidated proteins enriched for particular pathways, and provided possible prognostic benchmarks for future therapeutic development. However these finding are only preliminary and more penetrating study is required into the differences of expression profiles between healthy and diseased mice with more replicates, as well as establishing whether the changes observed are within the translationally impaired motor neurons or glial cells.

Lay Summary:

Mutations in a gene named eEF1A2 have been found in humans that suffer from epilepsy, autism and intellectual disabilities. When mice were genetically engineered to have mutations in eEF1A2, they developed severe neurological disorders. This thesis aims to identify how such mutations affect mice. It has been recognised that mutations in this gene result in neuronal degeneration in mice. To examine the extent of this damage I visualised a molecule (glial fibrillary acidic protein) whose presence or lack of is reflective of the degree of damage in neurons, across different thin slices of the spinal cord to learn if the damage is widespread or progressive. In the development of therapies for neuronal degeneration there is a need for biological markers that can be indicative of the disease progress. To achieve this, I analysed the abundances of different proteins found in spinal cords of diseased and healthy mice by a technique called mass spectrometry. This technique can identify distinct proteins in a system by comparing certain physical characteristics of parts of the protein, whilst informing on the quantity present. As well as this I compared the level of gene expression to investigate how active different genes were, and if this correlated with the changes in protein quantity. This comparative study between healthy mice and those with mutated eEF1A2 genes has illustrated the degree of neuronal damage in their spinal cords as well as how the mutations affected the expression of their genes. This has uncovered potential biological markers that can be used in monitor disease progression. However these findings are only preliminary and further investigation into these differences are required.

List of Abbreviations:

eEF1A	Eukaryotic elongation factor – 1 Alpha (Protein)
eEF1A1	Eukaryotic elongation factor – 1 Alpha 1 (Protein)
eEF1A2	Eukaryotic elongation factor – 1 Alpha 2 (Protein)
Eef1a	Eukaryotic elongation factor – 1 Alpha (Gene)
Eef1a1	Eukaryotic elongation factor – 1 Alpha 1 (Gene)
Eef1a2	Eukaryotic elongation factor – 1 Alpha 2 (Gene)
eEF1	Eukaryotic elongation factor – 1
GTP	Guanosine triphosphate
GDP	Guanosine diphosphate
eEF1B	Eukaryotic translation elongation factor 1B complex
eEF1Bα	Eukaryotic translation elongation factor 1B alpha subunit
eEF1Bδ	Eukaryotic translation elongation factor 1B delta subunit
eEF1Bγ	Eukaryotic translation elongation factor 1B gamma subunit
RNA	Ribonucleic acid
Met-tRNA^{Met}	Methylated methionyl-tRNA
tRNA	Transfer Ribonucleic acid
HSR	Heat shock response
HSP	Heat Shock Protein
HSF1	Heat Shock Factor-1
HSR1	Heat Shock RNA-1
HIV	Human immunodeficiency virus
F-actin	Filamentous actin
CaM	Ca^{2+}/calmodulin
Prdx-1	Peroxiredoxin-1
TAT	Trans-activator of transcription
CRISPR	clustered regularly interspaced short palindromic repeats
HDR	homology directed repair
NHEJ	non-homologous end joining
PAM	protospacer adjacent motifs
DNA	Deoxyribonucleic acid
mRNA	Messenger ribonucleic acid
PD	Parkinson's disease
AD	Alzheimer's disease
CUS	Chronic unpredictable stress
MND	Motor neuron disease.

GFAP	glial fibrillary acidic protein
wst	*wasted*
Nf	Neurofilament
SOD1	Cu/Zn superoxide dismutase
CSF	Cerebrospinal fluid
Cas9	CRISPR assiciated protein 9
LFQ-MS	Label free quantitative Mass spectrometry
LFQ Intensity	Label free quantitative Intensity
SAM	Significance analysis of microarray
bp	base pairs
H+E	Haemotoxylin and Eosin
indel	Insertions and deletions
CNS	Central nervous system
cDNA	complementary Deoxyribonucleic acid
Chrna4	Neuronal acetylcholine receptor subunit alpha-4
Chac1	ChaC glutathione specific gamma-glutamylcyclotransferase 1
Kif5C	Kinesin family member 5C
Srgap3	SLIT-ROBO Rho GTPase-activating protein 3
Tardbp	TAR DNA Binding Protein
Add1	Adducin 1
Ndufa2	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 2
Ndufa7	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 7
Fgf12	Fibroblast growth factor 12
GSTK1	Glutathione S-transferase kappa 1

Chapter 1: Introduction

1.1Translation and Eukaryotic Elongation factor Alpha (eEF1A).

The path to gene expression involves complex synthetic activity in the cell, a series of highly regulated stages that ensure translational fidelity. This may be broken down into three distinct activities: initiation, elongation and termination. Initiation, through the employment of multiple initiation factors, assembles the ribosome to a precise site on the mRNA. At the ribosomal P-site, the initiation codon is paired with Met-tRNA^{Met}, allowing for advancement into the elongation phase (Jackson *et al.* 2010).

Elongation is a series of steps repeated until aminoacylated tRNAs are polymerized into a growing polypeptide chain. Translational elongation involves elongation factors which vastly improve translational efficiency and accuracy (Alberts, B. *et al.* 2008). One such factor is eEF1, a GTP-dependent pentamer (Marco, *et al.* 2004). The *eEF1A* subunit protein binds to GTP (Guanosine triphosphate) and aminoacylated tRNA to deliver it the acceptor site of the ribosome where they form a ternary complex and position optimally for peptide bond formation (Raven *et al.*, 2014). Most incorrectly matched tRNAs preferentially dissociate. However once the cognate anti-codon is detected, the ribosome triggers hydrolysis of the GTP molecule inducing conformational changes that result in *eEF1A* releasing the aminoacylated tRNA and *eEF1A*-GDP dissociating from the ribosome. In order to restart the cycle, *EEF1A* requires binding of GTP and *eEF1B* the guanine-nucleotide exchange factor, which consists of three subunits *EF1Bα, EF1Bδ* and *EF1Bγ*, catalyzes and reactivates it with GTP for further functioning (Fig.1).

The culmination of the chain occurs when a stop codon signals termination and release factors facilitate the dissociation of mRNA and disassembly of the ribosome.

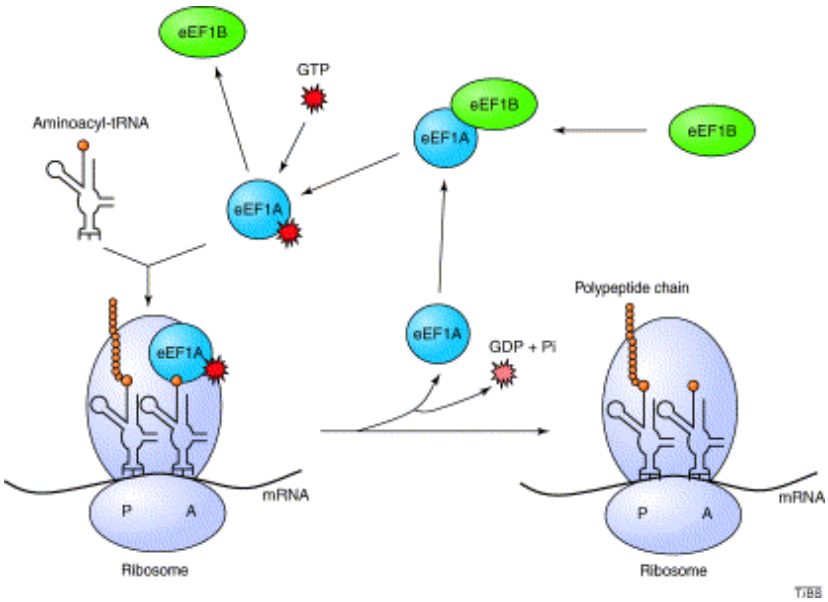


Fig.1: Translation elongation. Eukaryotic elongation factor 1A (*EEF1A*), when complexed with GTP, delivers the aminoacylated tRNA to the A-site of the ribosome. When appropriate codon–anticodon recognition occurs, GTP is hydrolyzed and *EEF1A*–GDP is released from the ribosome. *EEF1A* then interacts with *EEF1B*, thereby promoting the exchange of the bound GDP for GTP to regenerate active *EEF1A*–GTP. The cartoon is schematic and not to scale. Taken from (Abbott and Proud, 2004).

1.2 eEF1A isoforms.

eEF1A is one of the most abundant proteins in cells, making 1-3% of total protein content (Abbas *et al.*, 2015), and 3-5% of total protein in brain (Lee *et al.*, 1993). It exists in two isoforms; eEF1A1and eEF1A2 and is encoded by separate gene loci in humans, located at 6q14 (*EEF1A1*) and 20q13.3 (*EEF1A2*) respectively (Lund, *et al.*, 1996). However, the encoded proteins share 92% peptide sequence identity and a further 98% sequence similarity (Soares *et al.* 2009). Both isoforms are highly conserved across different species, which is expected given their crucial role in translation.

Alignments of *EEF1A2* and *EEF1A1* mRNA in vertebrates show high conservation across the transcripts, with most species displaying conservation in most base pairs. As demonstrated in subsections of *EEF1A2* and *EEF1A1* sequences (Fig.2, Fig.4). Full alignments can be found in Supplementary information. Base pairs in Fig.2 and Fig.4 that do not show conservation across all seven compared species, more often than not, are only divergent in one or two species, the most deviating in both isoforms being *Danio rerio*. This is also reflected in the phylogenetic trees displaying small evolutionary distances between the species with *Danio rerio* placing as the most evolutionarily distant in the context of *EEF1A1* and *EEF1A2* (Fig.3, Fig.5).

```
Danio          GGGACATCTCAGGCGGATTGTGCTGTCTTAATTGTAGCGGCTGGAGTGGGTGAATTTGAG   523
Oryctolagus    GGCACGTCCCAGGCGGACTGCGCGGTGCTCATCGTGGCCGCGGGCGTGGGCGAGTTTGAG   396
Homo           GGTACATCCCAGGCGGACTGCGCAGTGCTGATCGTGGCGGCGGGCGTGGGCGAGTTCGAG   701
Mus            GGCACATCCCAGGCGGACTGCGCAGTGCTGATCGTGGCAGCCGGTGTGGGCGAGTTTGAG   739
Bos            GGCACATCCCAGGCGGACTGCGCCGTGCTGATTGTGGCCGCAGGAGTGGGTGAGTTCGAG   475
Xenopus        GGAACCTCTCAGGCAGACTGTGCAGTGCTGATAGTGGCAGCCGGAGTGGGTGAGTTTGAA   407
Gallus         GGCACATCCCAGGCTGACTGCGCCGTACTGATTGTTGCTGCCGGTGTCGGTGAGTTTGAA   1172
               ** ** ** ***** ** ** ** **  * ** ** ** ** ** ** ** ** ** **

Danio          GCAGGCATTTCCAAAAATGGCCAAACAAGGGAACACGCCCTGCTGGCCTACACACTTGGT   583
Oryctolagus    GCCGGCATCTCCAAGAACGGGCAGACGCGGGAGCACGCGCTGCTGGCCTACACGCTGGGC   456
Homo           GCGGGCATCTCCAAGAATGGGCAGACGCGGGAGCATGCCCTGCTGGCCTACACGCTGGGT   761
Mus            GCGGGCATCTCCAAGAACGGGCAAACCCGGGAACACGCACTCCTGGCCTACACTCTGGGT   799
Bos            GCAGGCATCTCCAAGAATGGGCAGACCCGGGAGCACGCGCTGCTGGCCTACACGCTGGGC   535
Xenopus        GCTGGCATCTCCAAGAATGGACAGACCCGTGAACATGCCCTCCTGGCTTACACCCTCGGA   467
Gallus         GCCGGCATCTCCAAGAACGGGCAGACCCGTGAGCACGCCCTGCTGGCTTACACCCTGGGG   1232
               ** ***** ***** ** ** ** **  * ** ** ** ** ***** ***** ** **

Danio          GTCAAGCAACTGATCGTAGCCGTCAACAAGATGGACTCCACCGAGCCTTCCTACAGTGAG   643
Oryctolagus    GTGAAGCAGCTCATCGTGGGCGTCAACAAGATGGACTCCACAGAGCCGGCCTACAGCGAG   516
Homo           GTGAAGCAGCTCATCGTGGGCGTGAACAAAATGGACTCCACAGAGCCGGCCTACAGCGAG   821
Mus            GTGAAGCAGCTCATTGTGGGTGTCAACAAGATGGACTCCACGGAACCAGCCTACAGCGAG   859
Bos            GTGAAGCAGCTCATCGTGGGGGTGAACAAGATGGACTCCACGGAGCCCGCCTACAGCGAG   595
Xenopus        GTCAAACAGCTCATCGTGGGAATCAATAAAATGGACTCCACCGAGCCTCCCTACAGCGAG   527
Gallus         GTGAAGCAGCTCATTGTGGGCATCAACAAGATGGATTCCACGGAGCCTGCATACAGCGAG   1292
               ** ** ** ** ** ** *   * ** ** ***** ***** ** **  * ***** ***
```

Fig.2: Conservation of EEF1A2: Clustal alignments of *EEF1A2*, mRNA transcripts acquired from GenBank of vertebrate species: *Mus Musculus* (NM_007906.2); *Homo sapien* (NM_001958.3); *Danio rerio* (NM_00100237); *Xenopus tropicalis* (NM_001011418); *Gallus gallus* (NM_001032398.3); *Oryctolagus cuniculus* (NM_001082031.1); *Bos tarus* (NM_001037464). * denotes site conserved in all species, whilst number to the right is the position of the sequence.

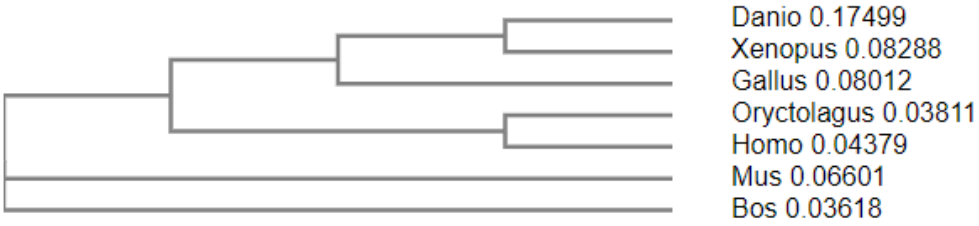


Fig.3: Phylogenetic tree of EEF1A2: Neighbour-joining tree without distance corrections of EEF1A2, mRNA transcripts acquired from GenBank of vertebrate species: *Mus Musculus* (NM_007906.2); *Homo sapien* (NM_001958.3); *Danio rerio* (NM_00100237); *Xenopus tropicalis* (NM_001011418); *Gallus gallus* (NM_001032398.3); *Oryctolagus cuniculus* (NM_001082031.1); *Bos tarus* (NM_001037464). Numbers specify evolutionary distance of each species.

```
Danio          CTCCACTGAGCCCCCTTACAGCCAGGCTCGTTTTGAGGAAATCACCAAGGAAGTCAGCGC   559
Xenopus        TTCAACTGAACCCCCATACAGCCAGAAAAGATATGAGGAAATCGTTAAGGAAGTCAGCAC   572
Gallus         TTCCACTGAGCCACCTTACAGCCAGAAGAGATACGAAGAGATCGTCAAGGAAGTCAGCAC   585
Bos            TTCCACTGAGCCACCCTATAGCCAGAAGAGATACGAAGAAATTGTTAAGGAAGTCAGCAC   715
Mus            TTCCACCGAGCCACCATACAGTCAGAAGAGATACGAGGAAATCGTTAAGGAAGTCAGCAC   647
Homo           TTCCACTGAGCCACCCTACAGCCAGAAGAGATATGAGGAAATTGTTAAGGAAGTCAGCAC   590
Oryctolagus    TTCCACTGAGCCACCCTACAGCCAGAAGAGATACGAGGAAATCGTTAAGGAAGTCAGCAC   570
                ** ** ** ** ** ** ** ***    * *  ** ** **    *********** *

Danio          ATACATCAAGAAGATCGGCTACAACCCTGCCAGTGTTGCCTTCGTCCCAATTTCAGGATG   619
Xenopus        ATACATCAAGAAGATTGGTTACAACCCTGATACTGTTGCCTTTGTACCTATTTCTGGATG   632
Gallus         TTACATCAAGAAAATTGGCTACAACCCAGACACTGTAGCTTTTGTGCCAATCTCTGGTTG   645
Bos            CTATATTAAGAAAATTGGCTACAACCCCGACACAGTAGCATTTGTGCCAATTTCTGGCTG   775
Mus            CTACATTAAGAAAATTGGCTACAACCCTGACACAGTAGCATTTGTGCCAATTTCTGGTTG   707
Homo           TTACATTAAGAAAATTGGCTACAACCCCGACACAGTAGCATTTGTGCCAATTTCTGGTTG   650
Oryctolagus    CTACATTAAGAAAATTGGCTACAACCCTGACACAGTAGCATTTGTGCCAATTTCTGGTTG   630
                ** ** ***** ** ** ******** *  *  ** ** ** ** ** ** ** ** **

Danio          GCACGGTGACAACATGCTGGAGGCCAGCTCAAACATGGGCTGGTTCAAGGGATGGAAGAT   679
Xenopus        GAACGGTGACAACATGCTTGAGCCCAGCGCCAATATGCCTTGGTTTAAGGGGTGGAAAAT   692
Gallus         GAACGGGGACAACATGCTGGAGCCTAGCTCTAACATGCCCTGGTTCAAGGGATGGAAGGT   705
Bos            GAATGGTGACAACATGCTAGAACCAAGTGCTAATATGCCATGGTTCAAGGGATGGAAAGT   835
Mus            GAATGGTGACAACATGCTGGAGCCAAGTGCTAATATGCCTTGGTTCAAGGGATGGAAAGT   767
Homo           GAATGGTGACAACATGCTGGAGCCAAGTGCTAACATGCCTTGGTTCAAGGGATGGAAAGT   710
Oryctolagus    GAACGGTGACAACATGCTGGAGCCAAGTGCTAATATGCCGTGGTTCAAGGGATGGAAAGT   690
                * * ** ********** **  * **  * ** ***   ***** ***** *****  *
```

Fig.4: Conservation of EEF1A1: Clustal alignments of *EEF1A1*, mRNA transcripts acquired from GenBank of vertebrate species: *Mus Musculus* (NM_010106.2); *Homo sapien* (NM_001402.5); *Danio rerio* (AY422992.1); *Xenopus tropicalis* (BC157768.1); *Gallus gallus* (NM_001321516.1); *Oryctolagus cuniculus* (NM_001082339.1); *Bos tarus* (NM_174535.2). * denotes site conserved in all species, whilst number to the right is the position of the sequence.

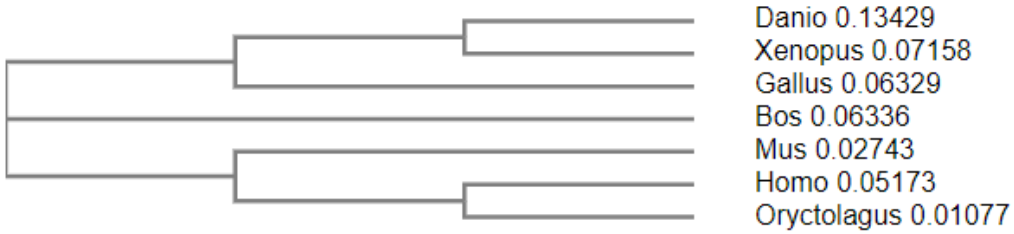


Fig.5: Phylogenetic tree of EEF1A1: Neighbour-joining tree without distance corrections of EEF1A2, mRNA transcripts acquired from GenBank of vertebrate species: *Mus Musculus* (NM_010106.2); *Homo sapien* (NM_001402.5); *Danio rerio* (AY422992.1); *Xenopus tropicalis* (BC157768.1); *Gallus gallus* (NM_001321516.1); *Oryctolagus cuniculus* (NM_001082339.1); *Bos tarus* (NM_174535.2). Numbers specify evolutionary distance of each species.

The folded 3D structure of eEF1A has three domains (Fig.6). Domain I has been identified to bind to GDP, whilst the aminoacetylated-tRNA binds to domain II (Li *et al*, 2013). The variations of amino acids between the eEF1A1 and eEF1A2 are located across all three domains (Fig.6), however most appear to congregate in two regions in the tertiary structure. One clusters within domain I whilst the other cluster stretches across domain II and III. These differences however are not located near the eEF1Bα interface nor domain-domain junctions, which could be the reason behind the structural integrity being conserved between the two isoforms (Soares, *et al.* 2009).

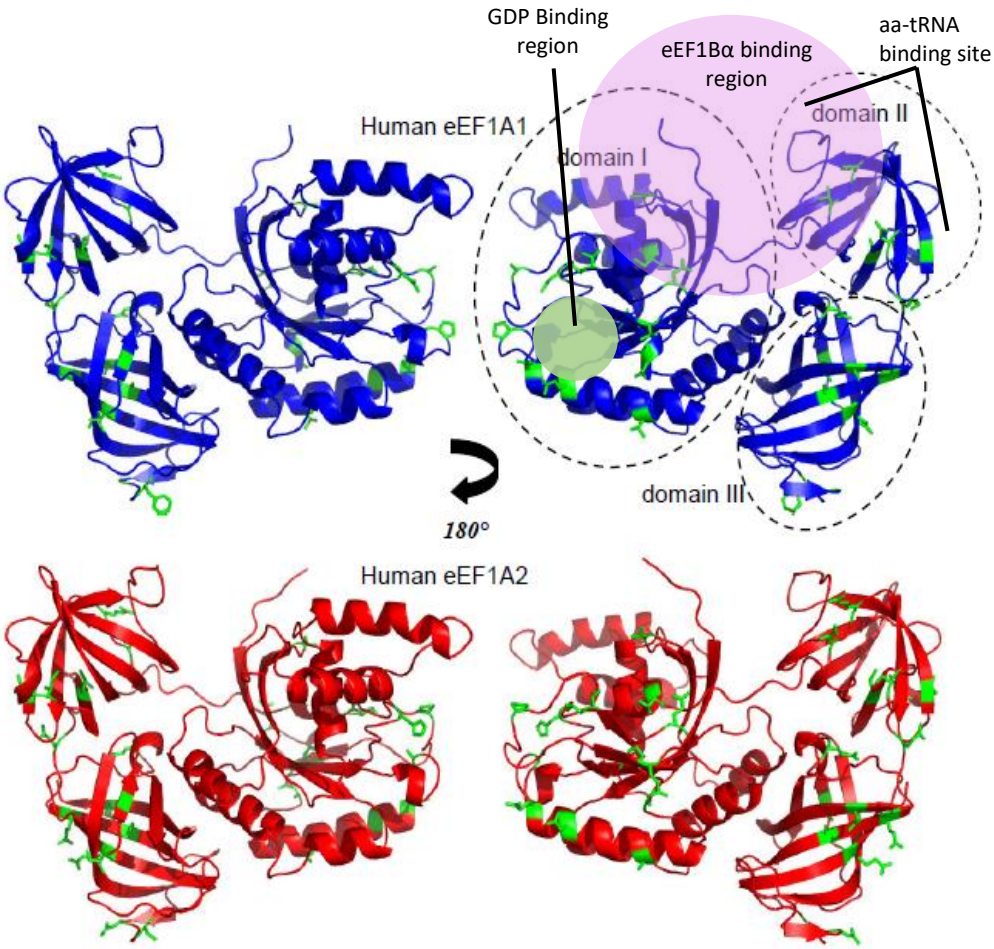


Fig.6: Human eEF1A1 and eEF1A2 structure: Cartoon schematic representations of the 3D-models from two views rotated by 180° along the y-axis. (Above) eEF1A1 coloured in blue with the domains outlined and the binding regions of GDP, eEF1B and aminoacylated-tRNA respectively highlighted. (Below) eEF1A2 coloured in red. The location of variant side chains between the two isoforms are coloured in green. Image taken from (Soares, *et al.* 2009), with added binding regions information from (Li, *et al.* 2013).

Notably the variants were found to show differing affinities for GTP and GDP; with eEF1A1 showing equal affinities for both as the ratio of bound GTP and GDP was 0.82, whereas eEF1A2 scored 1.50, demonstrating a great affinity for GDP (Kahns *et al.,* 1998). However, this does not seem to have a significant influence on their GTPase activity, as release of GDP was not rate limiting (Kahns *et al.,* 1998).

The lack of drastic 3D structural divergence and restricted variation that does not appear to affect binding sites greatly might imply that there little difference in the isoforms function in the cell. However this is not the case, as outlined later, and further study into the tertiary structures revealed the clusters as enriched for post-translational modifications such as phosphorylation and methylation (Soares and Abbott, 2013). This supports the possibility that isoforms are enacted upon differently, allowing them to deviate function from one another.

The *EEF1A* variants have been observed to be expressed differentially across specific tissues and at different stages of development. *EEF1A1* is ubiquitously expressed during development across a range of species; S.*cerevisiae*, M.*racemosus*, A.*salina*, D.*melanogaster* and X.*laevis*. Yet it has been shown in mice and rats to decline in terminally differentiated neurons, myocytes and cardiomyocytes concomitantly with *EEF1A2* RNA and protein expression increasing in mature cells (Lee *et al.*, 1993, Khalyfa *et al.,* 2001, Pan *et al.,* 2004). However, in transgenic studies the eEF1A1 promoter is still active in mature neurons (Stanley *et al.,* 2013). In neurons translation is controlled selectively, adjusting the proteome at specific subcellular locations which at the synapse maintain synaptic plasticity. Translation is regulated at the axonal growth cone and appears enhanced at proximal dendrites and more infrequent in distal dendrites (Wu *et al.,* 2016).

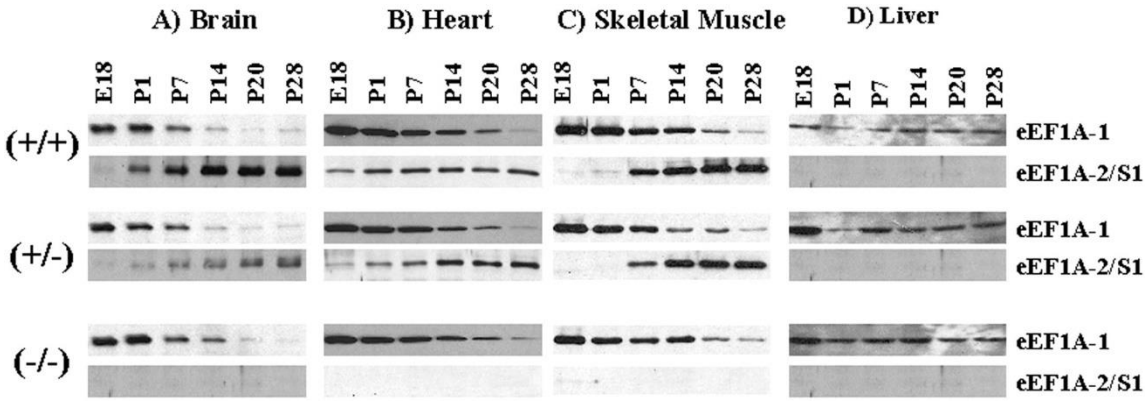


Fig.7: Differential expression of eEF1A-1 and eEF1A-2/S1 proteins during development of wild-type, heterozygous, and mutant mice. Mouse tissues from Brain (A), Heart (B), Skeletal muscle (C), and liver (D) obtained at ages of embryonic day 18 (E18) and postnatal (P) days 1-, 7-, 14-, 20- and 28 days old were used for immunoblotting assays for eEF1A1 and eEF1A2/S1 from wildtype (+/+), heterozygous (+/-), and mutant (-/-) mice. Taken from (Khalyfa *et al.,* 2001).

eEF1A2 expression patterning is unique among translation factors as it is tissue-specific, being solely expressed in neurons and muscle (Doig *et al.*, 2013). *EEF1A2* expression in humans has also been observed in the adrenal gland, pancreatic islet cells and at low rates, squamous epithelial cells in the oesophagus and oral mucosa, as well as the seminal vesicles (Fig.8). In contrast, *EEF1A1* was found across a wide range of tissues at far higher reads than *EEF1A2*. RNA expression of *EEF1A1* appears lower in tissues that report higher RNA expression of *EEF1A2* in humans (Fig.9), much like observations made in other species.

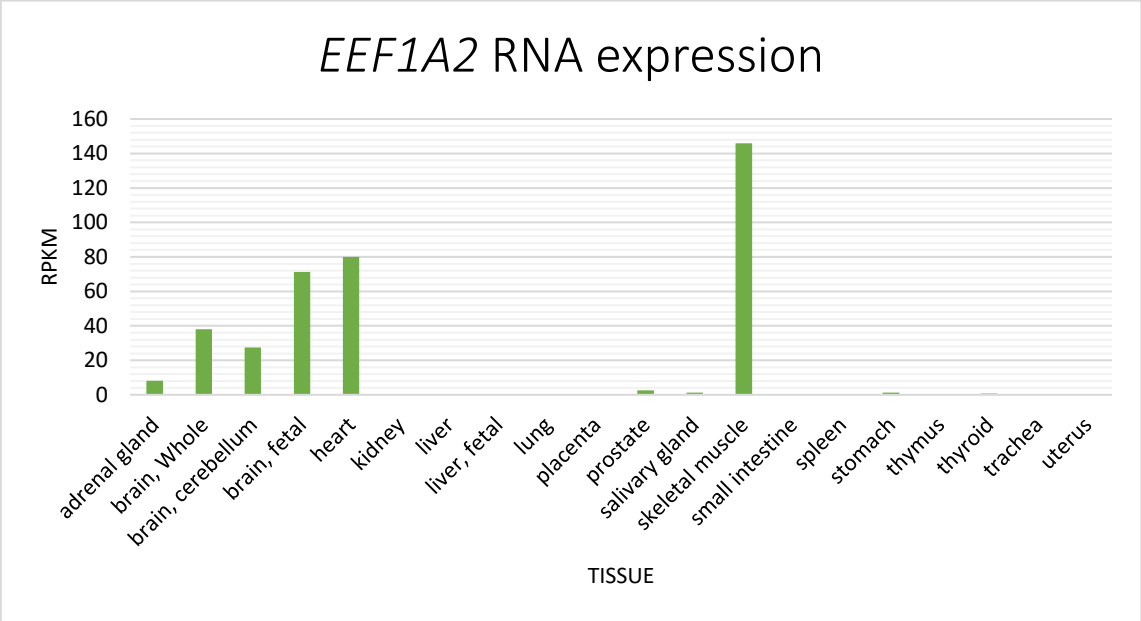


Fig.8: RNA expression overview of EEF1A2 in Humans: RNA expression data of *EEF1A2* in 20 human tissue samples. Expression levels measured in Reads Per Kilobase of transcript per Million mapped reads (RPKM). Skeletal muscle present high levels of *EEF1A2*, whereas neuronal cells and heart muscle medium levels of expression and glandular, squamous epithelial cells and seminal vesicles are low. Image generated from (NCBI Bioproject data, 2017).

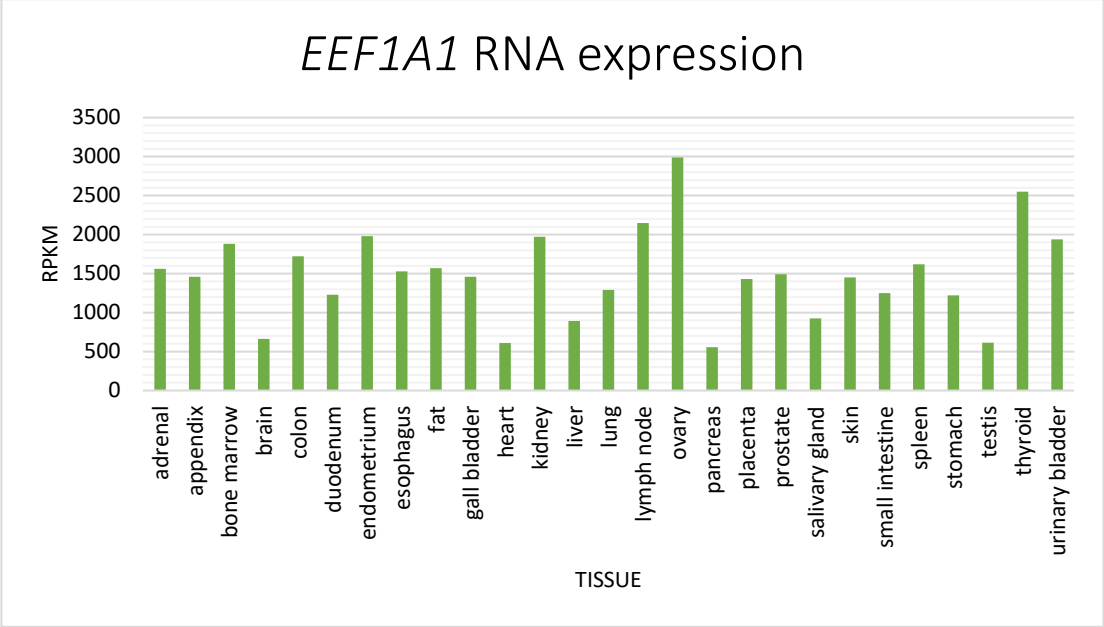


Fig.9: RNA expression overview of EEF1A1 in Humans: RNA expression data of *EEF1A1* in 20 human tissue samples. Expression levels measured in Reads Per Kilobase of transcript per Million mapped reads (RPKM). Most tissues present high levels of expression. Image generated from (NCBI Bioproject data, 2017).

eEF1A has a demonstrated involvement in a range of functions other than its translational role. Despite the conservation between the two isoforms, in some cases they have been identified as behaving differently in terms of their non-canonical roles. Although eEF1A1 is well characterised, less is known about eEF1A2. The variations between isoforms may be credited to the differential non-canonical roles, their respective interactions with other non-translational molecules.

eEF1A has been shown to be involved with the cell's response to stress. When cells undergo stress that leads to proteotoxic environments such as extreme heat or oxidative stress, proteins are prone to misfolding and aggregating. In response to this, a highly conserved mechanism to maintain proteostasis in cells is enacted: the synthesis of heat shock proteins (HSPs) (Akerfelt *et al.*2010). These function as molecular chaperones that work to restore proteostasis. eEF1A plays a role in activating the transcription of heat shock factor-1 (HSF1) when cells experience thermal stress. It forms a complex with heat shock RNA-1 (HSR1) which facilitates the trimerization of HSF1 or stability of it (Shamovsky *et al.*, 2006), which in turn promotes the transcription of HSP by binding to HSP mRNA stabilizing it for nuclear export to the ribosomes (Vera *et al.* 2014). However, eEF1A1 also acts as a repressor of transcription when the stress response is inactive by binding to the promoter region of HSPs. When cells experience stress, a direct effect is a reduction in translation rates, which releases eEF1A for functioning in the heat shock response (HSR) (Shamovsky *et al.*, 2006), as found when cells infected with bacteria

inducing the heat shock response (HSR) showed an increase in HSF1/eEF1A complexes (Shen *et al.,* 2009). Knock down of ~70% of eEF1A1 was sufficient to generate a deficiency in several HSPs in mouse and human cells. However when eEF1A2 was knocked down and levels of HSP70 were monitored, there was no such decrease, implying that eEF1A2 alone cannot support the heat shock response (Vera *et al.* 2014), therefore as eEF1A2 replaces eEF1A1, the cells response to stress is impaired. As motor neurons cannot express eEF1A1 they cannot mount the heat shock response and as a result are more vulnerable to stress and disease (Shamovsky *et al.,* 2006).

eEF1A in plants and trout has also been implicated in the cold shock response (Ejiri, 2002). During the cold response in plants eEF1A transcripts are detected at higher levels (Filipowicz and Hohn, 1996) and in trout cold acclimatisation appears to improve translational functions. However these are predominantly correlation studies but support the associations of cell stress response with eEF1A expression.

A common outcome of cells too damaged or impaired is programmed cell death; apoptosis. Therefore if a cell struggles to maintain homeostasis, particular pathways are initiated resulting in elimination of that cell. eEF1A became understood as an element coordinating apoptosis when initial studies identified increased levels of eEF1A to be associated with more expeditious cell death (Duttaroy *et al.,* 1998) and that this increase is facilitated posttranscriptionally (Chen *et al.,* 2000). Yet as further research was conducted into the separate isoforms, eEF1A1 was revealed to be proapoptotic whereas eEF1A2 behaved protectively and was suggested to be anti-apoptotic (Ruest *et al.,* 2002). eEF1A2 became implicated in the regulation of Caspase 3 and this is potentially the mechanism behind its involvement in cell survival. Ruest *et al.*. 2007, speculated that the isoforms promote or protect against apoptosis respectively by differential translation efficiency; eEF1A1 may display affinity for translation of proapoptotic genes mRNA and/or repress those that are prosurvival, whilst eEF1A2 conducts itself contrariwise. Additionally i*n vivo* and *in vitro* eEF1A2 is observed to interact with Peroxiredoxin-1 (Prdx-1), which when cotransfected into cells displayed higher apoptotic resistance than when either were transfected alone. The increase in eEF1A2/Prdx-1 also correlated with an increase in the apoptosis suppressive element Akt as well as seeing a decrease in Caspase 3 and 8, proteins that act proapoptotically (Chang and Wang, 2007). However, no recent studies have further researched this and eEF1A2's role in apoptosis as outlined should be treated with caution.

EEF1A mRNA levels are escalated in proliferating cells, more so than other elongation factors (Condeelis, 1995). *In vitro* both isoforms bind F-actin when dimerised which results in actin bundling, conversely monomers are unable to bundle F-actin. Monomerization of either eEF1A1 or eEF1A2 is dependent upon Ca^{2+}/calmodulin (CaM) binding, upon this F-actin filaments loosen (Bunei *et al.,* 2006). Interestingly, eEF1A2 has presented a reduced ability to bind with calmodulin (Novosylna *et al.,* 2017) which may suggest that tissues expressing solely eEF1A2 do not exercise as much cytoskeletal

remodelling or be required to constrain it, a concept which also aligns with the fact that it is only in terminally differentiated cells that eEF1A2 is upregulated. The involvement of eEF1A1 in F-actin detachment may serve as one reason for its aforementioned increased expression in apoptotic cells, as apoptosis can be induced by microtubule severing factors (Ruest *et al.*, 2002).

In addition to the roles mentioned above, eEF1A has been implicated in other non-canonical roles. Preliminary findings suggest a model in which it may regulate protein degradation (Mateyak and Kinzy, 2010), as well as nuclear export, as it has been found in the nucleus and shown to be associated with the export of tRNA species in yeast (Mateyak and Kinzy, 2010, Khacho *et al.*, 2008).

Although having a crucial canonical roles to play in the cell, eEF1A demonstrates many differing capabilities in other pathways that too, that are vital to the cell and decisive of its fate. The study of isoforms separately has been growing as it is no longer hindered by antibody limitations, previously unable to discern between each variant in specific tissues (Newbury *et al.* 2007). However, observing the expression of variants postnatally in neuronal cells is still complicated as glial cells express high levels of eEF1A1, making a whole tissue analysis impossible. Furthermore, studies that probe for the roles of eEF1A1 and eEF1A2 separately are identifying differing, and in some cases, opposing functions. Therefore the developmental switch from eEF1A1 to eEF1A2 in long-lived, terminally differentiated neurons, myocytes and cardiomyocytes can be postulated as driven by these differing non-canonical roles as these cells require less actin remodelling and protection against apoptosis.

As mentioned before, *EEF1A1* and *EEF1A2* are highly conserved isoforms and differences in sequence do not drastically change tertiary structure or function in theory. However given their contrasting expression patterns and 'moonlighting' functions and the small localised variations that are enriched for post-translational modification, *EEF1A1* and *EEF1A2* are non-redundant and vital in their differing properties.

1.3 Translation factors and *EEF1A* in disease.

The essentialness of translation factors for protein synthesis and survival means that mutations that lead to dysregulated or dysfunctional proteins can have wide reaching consequences across the cell as a fundamental cell mechanism is impeded. A range of diseases are known to develop because of inherited or *de novo* mutations that impair translational machinery (Scheper *et al.*,2007, Nakajima *et al.*, 2015, Ejiri 2002, Bottley and Kondrashov, 2013). Many studies have identified abnormal expression of eEF1A in disease but have not entirely outlined the involvement of eEF1A as of yet. Given the tissue-specific nature of this switch, the isoforms' significant canonical and non-canonical

function and the high degree of conservation across species, *EEF1A* isoforms are strongly implicated in the development of disease.

EEF1A mRNA expression was observed as upregulated in skeletal muscle tissue in humans and rodent models with Type 2 and Type 1 diabetes respectively, whilst subunits *EF1Bδ* and *EF1Bγ* remained unchanged. Insulin treatment was then able to reverse this (Reynet and Kahn, 2001), intriguingly insulin is known to regulate the elongation by altering phosphorylation of eEF1A isoforms. In addition, defective HSP function is associated with diabetes (Atalay *et al.,* 2009). This may allude to a symptom of diabetes - uncontrolled oxidative stress in cells, resulting in dysregulation of eEF1A1 and in turn impaired stress response. However the location of measured *EEF1A* implies that it is the *EEF1A2* isoform being enhanced. Also, notably, genome-wide linkage analysis has identified a diabetes susceptibility locus at 20q13.3 (*EEF1A2*s' gene locus) (Rotimi *et al.,* 2004).

1.4 eEF1A isoforms in Neurological disease.

eEF1A isoforms are involved in a range of neurological disorders, either directly or via observed changes in expression levels. The tissue-specific nature of eEF1A2 suggests that it may be involved in the development of neurodegeneration. Its inability to take part in the heat shock response suggests that in the cells expressing solely eEF1A2, the reaction to cell stress and proteotoxicity might be impaired. Included in these stressors is oxidative stress, a condition induced by an imbalance between reactive oxygen species and antioxidant defences leading to damage of DNA structure, cell membrane and protein structure and function from oxidation. Not only are neuronal cells vulnerable to oxidative damage because of their high demand for oxygen and abundance of peroxidisable substrates (known to induce cell death (Whittemore *et al.*,1995)), the sole expression of the eEF1A2 isoform could weaken them further. Notably, particular types of neuronal groups can be more susceptible to oxidative stress (Gandhi and Abramov, 2012). Oxidative stress is a key factor in neurodegenerative pathophysiology. Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD) pathogenesis has not been completely outlined, but it is proposed that the misfolded, dysfunctional proteins aggregating to toxic levels because of oxidative damage are in part responsible for disease development and progression (Kim *et al.,* 2015, Kumar and Ratan, 2016). Human brain tissue from patients diagnosed with these diseases have found decreased eEF1A protein levels (Garcia-Esparcia *et al.,* 2015).

Dysregulation of eEF1A has also been identified in less severe neurological disorders; depression and anxiety. Through a proteomic study of Zebrafish experiencing chronic unpredictable stress (CUS) to induce and anxiety and mood disorder, the downregulation

of eEF1A in brain was detected. This again was determined to be indicative of neuroprotective roles (Chakravarty *et al.,* 2013). This downregulation of eEF1A may be due to the fact that long-term potentiation stimulates eEF1A synthesis (Panayiotis *et al.,* 2005), yet CUS has been shown to impair long-term potentiation and in turn synaptic plasticity (Alfarez *et al.,* 2003).

There has been a growing consensus that many of the previous idiopathic cases of neurological disorders such as neurodevelopmental and neurodegenerative actually have a genetic basis Macleod and Appleton, 2007, Kaufman *et al.,* 2010), despite many having more complex modes of inheritance that are not entirely known; with 70-80% not being attributed to acquired conditions (Hidlebrand *et al.,* 2013). However one can conject that the elusive nature of the cause may be due to a limit in current diagnostic tests.

Growing evidence has revealed associations between particular genes and the emergence of severe neurodevelopmental disorders and an epilepsy phenotype (Myers and Mefford, 2015). A meta-analysis of genome wide association studies has identified statistically significant loci and implicated genes in forms of epilepsy (International League against Epilepsy Consortium on Complex Epilepsies, 2015). Facilitated by exome sequencing and advances in technologies, more data is being collated and genetic studies have identified that *de novo* heterozygous missense mutations in *EEF1A2* are associated with disease (see Table 1) (Lam *et al.,* 2016) and two cases of homozygous missense mutation (Cao *et al.,* 2017). In humans, manifesting from early childhood, mutations in EEF1A2 have been implicated in severe neurodevelopmental disorders, in which sufferers can display symptoms of repeated seizures, intellectual disability, and autism. In some severe cases no purposeful movement and patients are wheelchair bound and/or have respiratory issues.

Various different mutations in eEF1A2 were reported in affected children that were not present in the parents, except in the case of the P33L mutation (Table 1.). These are predicted to have a damaging effect on the protein, especially as eEF1A2 is under excessive selective constraint (Samocha *et al.,* 2014), and mutations identified in humans are also conserved in the orthologues of evolutionary distant species (e.g yeast).

Mutation	Developmental delay	Reference
A92T	Global developmental delay	Lopes *et al.,* 2016
D252H	Developmental delay	Nakajima *et al*., 2015
D252H	Global developmental delay	DDD
D91N	Non-Verbal, global developmenatal delay	Lam *et al.*, 2016
E122K	Non-Verbal, gross motor delay	Lam *et al.*, 2016
E122K	Non-Verbal, developmenatal delay	Inui *et al.*, 2015
E122K	Non-Verbal, developmenatal delay	Inui *et al.*, 2015
E122K	Motor delay	Nakajima *et al*., 2015
E124K	Significant language delay	Lam *et al.*, 2016
F98L	Global developmental delay	Lam *et al.*, 2016
G70S	Global developmental delay	Lam *et al.*, 2016
G70S	Non-Verbal	Veeramah *et al.*2013
G70S	Global developmental delay	de Ligt *et al.*,2012.
G70S	Not reported	de Kovel *et al.*, 2016
G70S	Global developmental delay	Yang *et al.* 2014
I71L	Global developmental delay	Lam *et al.*, 2016
P33L	Global developmental delay	Cao *et al.*, 2017
P33L	Global developmental delay	Cao *et al.*, 2017
R266W	Not reported	DDD
R382H	Not reported	Iossifov *et al.*, 2014
R423C	Global developmental delay	Lam *et al.*, 2016
T432M	Not reported	DDD

Table 1: Mutations found in eEF1A2. The various amino acid mutations found in human cases and their respective developmental delays.

The location of some of the different eEF1A2 mutations identified are illustrated in Fig.10. These mutations have been reported as located near regions of importance for eEF1A functioning. Notably, the least severe phenotype mutation (E124K) is more distally located from eEF1A2's binding sites.

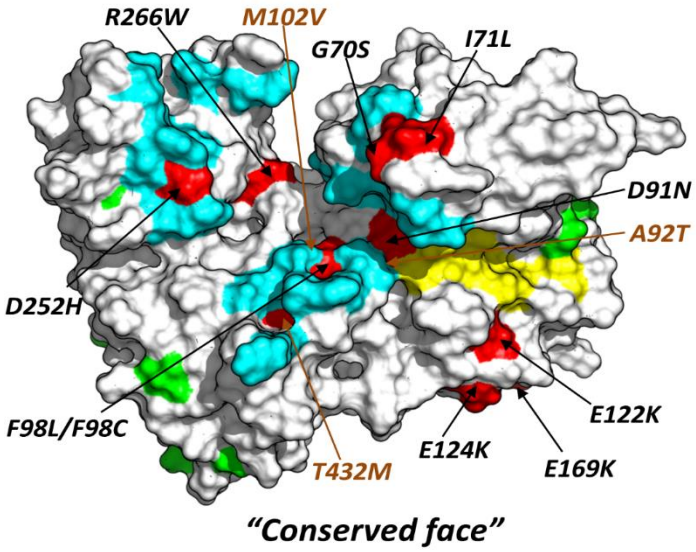


Fig.10: Mutations in eEF1A2. Location of reported missense mutations in humans in the protein structure (Red). eEF1β binding site (Blue). GTP binding site (Yellow) and variable amino acids between eEF1A isoforms (Green). Model and annotations by Soares.

1.5 Motor neuron disease and eEF1A2.

It has been previously seen that mice absent of eEF1A2 expression appear to develop aggressive early onset motor neurodegeneration (Chambers *et al.*, 1998) with evidence of distinct vacuolation of the motor neurons in the spinal cord, alongside neuromuscular junctions deteriorating signal transmission, progressive retraction of motor endplates (Newbery *et al.* 2005). Motor neuron disease (MND) encompasses a range of disorders; amyotrophic lateral sclerosis (ALS), hereditary spastic paraplegia (HSP) and spinal muscular atrophy to name a few, differentiated by the selective regions of motor neurons affected. The most common form is ALS, in which the both the upper motor neurons (motor cortex in the brain) and lower motor neurons (brain stem and spinal cord) are affected. Cases predominantly occur sporadically with only ~10% being identified as familial with primarily autosomal dominant inheritance (Boylan, 2015). Characterised by the premature degeneration and death of motor neurons in the brain and spinal cord, the loss of these cells lead to muscular atrophy, weakness and paralysis of voluntary muscles.

Death of upper motor neurons results in spasticity and hyper-excitability of reflexes, whereas death in the lower motor neurons are responsible for muscle atrophy and paralysis (Dadon-Nachum *et al.* 2011). The disease is progressive with rate and pattern varying greatly between patients but leading eventually to death, usually due to respiratory paralysis. Paralysis manifests itself focally initially then advances in a pattern suggesting that the spread of degeneration occurs through contagious pools of motor neurons (Pasinelli, and Brown, 2006). Given the heterogeneity of motor neuron disease, the pathophysiology is still not entirely understood. The pathology of neuronal damage is well outlined, on the other hand, involving axonal swelling, accumulation of phosphorylated neurofilaments, deposition of inclusions (spheroids) and ubiquitinated material in these axons as well as the activation and proliferation of astrocytes and microglia in reactive gliosis. The pathological observations however, give little or no reference to the stage of the disease and are limited in describing the mechanistic workings behind the cellular distress. The causes are undefined with several theories postulated, nevertheless researchers have identified various cellular processes influenced such as protein misfolding, excessive excitatory tone, altered axonal transport and activation of proteases and nucleases (Pasinelli, and Brown, 2006).

A mechanism preceding the death of neurons is the dying-back phenomenon, in which distal axons degenerate and progresses towards the cell body. This 'die-back' of axons has been observed in the motor neurons of ALS mouse models (Dadon-Nachum *et al.* 2011, Fischer *et al.* 2004). When considering that the death of motor neurons occurs in a disseminated fashion, it could mean that initial stages of the disease may start with the dying-back of axons that perhaps eventually initiate cell death. Motor neuron degeneration in the spinal cords of mice has been linked to null mutations in eEF1A2, with similarities in that the damage is preceded by axonal degeneration and the phenotype is of neurogenic origin, as restoration of eEF1A2 expression in muscle failed to rescue any of the phenotypic aberrations (Newbury *et al.,* 2005, Doig *et al.*, 2013, Murray *et al.* 2008).

This loss of eEF1A2 expression in mice has triggered dying-back neuropathy (Murray *et al.* 2008). eEF1A2's role in this maybe due to its cytoskeletal remodelling function as there have been links made with axonal damage and impaired microtubule assembly (Bommel *et al.,* 2002). In the cells in which dying-back occurs, only the eEF1A2 isoform is expressed and this has reduced actin bundling capabilities in comparison to the eEF1A1 isoform.

As of yet there is no cure for MND, there are only measures for management of the disorder and a single drug approved, Riluzole, which delays advancement of disease marginally (MND Association, 2017)). The lack of understanding combined with homogeneity of initial symptoms with other less severe disorders, has meant that diagnosis is limited to the interpretation of physiological symptoms, electro-diagnostic and in some cases muscle biopsies upon manifestation of symptoms (NHS, 2017). In many cases this leads to delays in diagnosis or early misdiagnosis which can be stressful to patients

uncertain of their wellbeing and fate. Hence, to develop therapies to target progression of MND, understanding why selective regions of motor neurons are deteriorating and why the advancement is patterned is crucial. Therapies for diseases that are progressive in nature, often work best if they are enacted during initial stages of disease development and before symptoms worsen. This is why research into non-invasive diagnosis and prognosis at a quantifiable biological level is paramount. However, preceding this is the ability to confidently diagnose and examine prognosis from acquired tissue in humans and in animal models.

1.6 Mouse Models:

An unprecedented spontaneous recessive mutation arose in a HRS/J stock mice from Jackson laboratory. This was subsequently found to be a 15.8-kb deletion that encompassed the promoter region and first non-coding exon of *Eef1a2* and no other gene, resulting in mice with complete loss of eEF1A2 expression. These mice were termed *wasted* (*wst*), their phenotype being characterised by ataxia, weight loss, progressive paralysis (Chambers *et al.*, 1998). The loss of eEF1A2 expression appears to result directly to the development of the *wasted* phenotype (Newbery *et al.* 2007). As mentioned previously, the phenotype displays a very aggressive and early onset motor neuron degeneration, demonstrating vacuolation of the motor neurons in the spinal cord, as well as neuromuscular junctions deteriorating signal transmission, progressive retraction of motor endplates (Newbery *et al.* 2005) and reactive gliosis (Abbott *et al.*, 2009). These changes are preceded by axonal and somatic degeneration ((Murray *et al.* 2008). (Doig *et al.*, 2013) identified the phenotype as being of neurogenic origin, as restoration of eEF1A2 expression in muscle failed to rescue any of the phenotypic aberrations. However it is only mice homozygous (*wst/*wst) for the deletion that develop the phenotype. Heterozygous mice, despite having approximately 50% reduced eEF1A2 protein expression, do not develop the *wasted* phenotype, they were instead observed to have normal neuromuscular functioning, and are indistinguishable from the wild type mice at both a physical and pathological level (Griffiths *et al.*, 2012). The onset of physical deterioration is observed at ~21 days where the expression of *Eef1a1* is downregulated – whilst independently *Eef1a2* is reaching peak expression (Khalyfa *et al.*, 2001, Pan *et al.*, 2004). Survival rates of *wasted* mice do not exceed ~28 days (Davis *et al.*, 2017). However onset of the neuronal degeneration is detectable by 17 days postnatal on the basis of enhanced GFAP staining Table.2 in spinal cords, and the retraction and denervation of motor endplates in thoracic skeletal muscle (Newbury *et al.*, 2005).

Age (days)	Region	*wst/wst*	*+/wst*
19	Cervical	48	2
	Thoracic	27	3
	Lumbar	7	4
24	Cervical	43	7
	Thoracic	37	2
	Lumbar	22	0
28	Cervical	37	0
	Thoracic	31	0
	Lumbar	4	0
29	Cervical	148	0
	Thoracic	52	3
	Lumbar	16	0

Table.2: Mean Number of GFAP-Positive Cells in the Grey Matter of the Spinal Cord at Cervical, Thoracic and Lumbar regions in *wst/wst* and *+/wst* mice at ages 19,24,28 and 29. Taken from (Newbery *et al.* 2005).

The pathology of motor neuron degeneration appears as a progressive rostrocaudal gradient, with more motor neuron deterioration occurring initially at the cervical level. There is speculation as to whether these changes work as a cascade and eEF1A1 is switched off progressively which would explain why caudal areas do not seem to be as affected (Newbury *et al.* 2005).

The *wasted* phenotype shows great similarity with the pathology of MND and has been suggested as a potential model for therapeutic study. Previous work in the Abbott group has concerned the development of various mouse lines to recapitulate particular *EEF1A2* mutations that have been reported on page 22.

The most common mutation, G70S, when genetically engineered into the purebred C57BL/6 mice was found to result in a non-functional protein, resulting in motor neuron degeneration and in some cases sudden unexplained deaths and audiogenic seizures in mice completely null for eEF1A2(Davis *et al.*, 2017).

The D252H mutation was also recreated in mice using CRISPR/Cas9. This line of mice displayed similar phenotypes to that of *wasted* in homozygotes with heterozygotes remaining unaffected. Onset of physiological symptoms, akin to *wasted* occurs at ~21 days where the switch is said to be near complete. Initial study show homozygotes express eEF1A2 but that the protein is impaired or non-functional. However this line remains largely uncharacterised and the nature of eEF1A2 expression and pathology has yet to be fully understood and will be outlined subsequently.

A 22 base pair deletion within exon 3 of *Eef1a2* that arose from CRISPR/Cas9 mutagenesis resulted in a null mutation, this was bred into a line labelled Del.22.Ex3. The mice within this line present a phenotype so severe that mice do not survive much longer after the onset of disease (~ 21-23 days) suffering from the aforementioned symptoms with the addition of fatal seizures in some cases of homozygous mice. This limits their use as potential models for MND as the mice would be at risk of spontaneously dying in addition to being unethical. However they remain largely uncharacterized, yet again, this phenotype is only present with mice homozygous for the deletion.

The genetically engineered lines have shown to be robust and reproducible and relatively easy to develop with CRISPR/Cas9 techniques making them reliable to study aspects of MND and probe for potential biomarkers. Additionally, there is a lack of disparity between the females and male, thus reducing further variation (which can be a problem with other current models) (Perrin, 2014). Unlike the SOD1 and C9ORF72 mouse models, among others, (see supplementary information) (Chew *et al.*, 2014), the phenotype manifests very early. This is beneficial for studying biomarkers as changes resultant from therapies can be detected earlier and robustly unlike the current models (e.g.SOD1) in which to establish if there is any delay in death, it can be over 100 days. Although the *Eef1a2* mutation has not been observed in human MND cases, the model would be validated for its use in cell based experimentations and observations of therapies. However, the effects of the various mutations in *Eef1a2* have in these mice must first be interpreted.

1.7 Biomarkers for Motor neuron degeneration.

Currently there are no clinically used biomarkers for MND and a great deal of research is being conducted to identify prospective biomarkers for diagnostic and prognostic purposes. Initial symptoms being in common with other disorders complicates diagnosis, which often can only be made confidently after significant progression of the disease and thus left to human interpretation, creating demand for molecular markers indicative of MND. The repeated failures of clinical trials are also an issue; vast sums of resources, time and effort are exhausted. Pertinent biomarkers can indicate better the potential of a

therapy in models before advancing to clinical trials, as well as improve the efficiency of clinical trials. The search for biomarkers is increasingly becoming data-driven, with the development of molecular (proteins, genes or metabolic products) neurophysiological (changes in upper and lower motor neurons) and neuroimaging markers, achieved by proteomic, genomic and metabolomics studies.

Genetic molecular markers can be complex to identify as cases of disease development may be polygenic. Studies in proteins as potential markers have shown greater success with some reporting over 90% sensitivity for distinguishing ALS sufferers from healthy controls. Levels of the Nf light chain protein (the light chain of neurofilaments – cytoskeletal proteins of neurons that are released subsequent to neuronal damage) in blood samples were more than 20-fold higher in ALS 49 patients (Gaiottino *et al.*, 2013). Proteomic studies have identified a superior marker: Tau. Tau proteins stabilise microtubules and are found in abundance in neuronal cells. Dysfunctional tau proteins have been associated with neurological disorders such as Alzheimer's and Parkinson's disease. Hyperphosphorylation of Tau leads to protein aggregation and sequestration of other cytoskeletal proteins. Patients experiencing abated disease progression demonstrated reduced levels of Tau and phosphorylated neurofilament heavy chain during drug trials for ALS (Levine *et al.*, 2010). The associated oxidative damage with motor neuron degeneration made biochemical markers of the oxidative stress response a promising focus area. However contradicting results have transpired; molecules involved in the glutathione pathway and Cu/Zn superoxide dismutase (SOD1) were observed to be reduced in the erythrocytes of ALS patients (Babu *et al.*, 2008), but were also found increased in the same cells (Tuncel *et al.*, 2006, Cova *et al.*, 2010) and in cerebrospinal fluid (CSF) (Boll *et al.* 2003, Kokic *et al.* 2005).

Although there are promising biomarkers being developed, they have yet to be translated into a clinical setting. With contradictory results emerging and most research being geared towards biomarkers for primary end points, seeking further potential biochemical markers that can act as surrogate end points is of great value. In addition to this, investigating possible biomarkers for MND can inform our understanding of the disease's enigmatic mechanisms and the identification of pathogenesis. This information in turn may lend itself towards further possible targets for novel therapies.

1.8 Project Aims:

The aims of this research project are to characterise the effects of mutations in *Eef1a2* in mice at a genetic and protein level, alongside identifying potential markers that are indicative of motor neuron degeneration through use of 'wet' laboratory work in conjunction with computational biology.
I aimed to understand how the loss of eEF1A2 and its neuronal degeneration phenotype affects protein and genetic expression pattern, as well as the pathways and biochemistry involved with its loss. I also aimed to provide possible prognostic benchmarks for future therapeutic development.
The research has been carried out predominantly on two lines of mice; the D252H line and Del.22.Ex.3. eEF1A2 expression in these mice was identified as well as the extent of neurodegeneration experienced by using immunohistochemistry for markers of neurodegeneration. To probe for biomarkers, a bottom up proteomics approach was applied. A comparative study of the differences in protein expression between eEF1A2 null and wildtype mice was conducted. In complement to this, previous RNA expression data was also analysed. Proteins and genes that were significantly differentially expressed were then investigated further for biological significance.

An outline of the steps taken to achieve these aims:

1. Characterisation of founder mice from D252H line, the mutations resulting from CRISPR experiments and their effects on eEF1A2 expression.
2. Pathology of spinal cords from these mice at the cervical, thoracic and lumbar regions probing for glial fibrillary acidic protein (GFAP) to observe any gliosis, and whether neurodegeneration occurs as a progressive rostrocaudal gradient
3. Quantitative analysis of eEF1A2 null and wild type spinal cord proteome by Label-free Quantitative Mass spectrometry (LFQ-MS).
4. Microarray data analysis of spinal cord RNA from wasted mice and wild type mice.
5. Gene ontology enrichment analysis of proteins and genes significantly differentially expressed and regulated, exploring which pathways and functions were over-represented.

Chapter 2: Materials and methods.

2.1: Clustal Alignments and Phylogenetic trees.

For EEF1A2, mRNA transcripts acquired from GenBank. Accession numbers of each species: Mus Musculus (NM_007906.2); Homo sapien (NM_001958.3); Danio rerio (NM_00100237); Xenopus tropicalis (NM_001011418); Gallus gallus (NM_001032398.3); Oryctolagus cuniculus (NM_001082031.1); Bos tarus (NM_001037464). Clustal alignments of EEF1A1, mRNA transcripts acquired from GenBank of vertebrate species: Mus Musculus (NM_010106.2); Homo sapien (NM_001402.5); Danio rerio (AY422992.1); Gallus gallus (NM_001321516.1); Oryctolagus cuniculus (NM_001082339.1); Bos tarus (NM_174535.2). However for Xenopus tropicalis (BC157768.1), the sequence was from cDNA clone MGC:184686.

Alignments and Neighbour-joining tree without distance corrections were used Clustal Omega, version 1.2.4. Multiple sequence alignment tool with input parameters: Output guide tree: false. Output distance matrix: false. Dealign input sequences: false. mBed-like clustering guide tree: true. mBed-like clustering iteration: true. Number of iterations: 0. Maximum guide tree iterations: -1. Maximum HMM iterations: -1. Output order: Aligned, Sequence Type: RNA.

2.2: Genotyping.

2.2.1: Genomic DNA extraction.

For genotyping ear notches acquired at 14 days from the founder mice underwent the sodium hydroxide method of DNA extraction. 300µl of 15mM NaOH was added to ear notches then heated at 100°C for 10 minutes and vortexed before adding 25µl of Tris 1M at pH8 the sample was then stored at -20°C.

2.2.2: Nested PCR.

A nested PCR for D252H founder mice to amplify a region of exon 5 of the Eef1a2 gene. The 1st round of PCR reaction was performed using 1x Taq PCR buffer, 2mM $MgCl_2$, 0.2µM dNTPs, 0.4µM of primers, 5% 1,2 propanediol and 1U of Taq polymerase. With primer sequences: mD252H 1F (5'- AGGCTACCCCTTAGGCAGGT-3') and mD252H 1R (TGAACAAATGGTAGGTGGGAGG). On a program of Denaturation at 95°C for 3min, 20 cycles of (denaturation) 95°C for 30s, (annealing) 60°C for 30s and (extension) 72°C for 1min, the final extension at 72°C for 5 minutes then held at 10°C. The PCR products were then diluted 1:10 and 1µl was used in the 2nd round of PCR reaction. With primer sequences: mD252H Spare 1F (ATTTGTAAGTGGTGGGGGCA) and mD252H

1R (GTCCCTAGCTTGTGGCTGAG). On a program of Denaturation at 95°C for 3min, 20 cycles of 95°C for 30s, 67°C for 30s and 72°C for 45S, the final extension at 72°C for 5 minutes then held at 10°C. Products were then visualized by electrophoresis on a 1.5% agarose gel.

2.2.3: TOPO Cloning.

TOPO cloning was conducted by mixing 0.5µl of PCR product was with 1µl salt solution, 3.5µl dH_2O and 1µl TOPO® Vector at room temperature for 30min, placed on ice briefly before transforming. 2µl of TOPO reaction was mixed into cells and heat shocked for 30s at 42°C then transferred to ice. 250µl of SOC was added to the transformed cells which was then shaken at 37°C, 200rpm for an hour. This was then spread on L-ampicillin plates and left to incubate overnight at 37°C. Colonies with incorporated gene were identified by the blue-white test and picked to be sequenced. Clones were sequenced by Sanger sequencing using 3130 or 3730 Genetic Analyser (Applied Biosystems).

2.3: Protein expression.

2.3.1: Sample preparation.

Muscle tissue of founder mice were prepared by adding 10µl of 0.32M sucrose with protease inhibitor per 1mg of tissue. Samples were then homogenised mechanically for 1 minute in bead beating tubes with 1.4mm large ceramic beads. These were then centrifuged at 10000g for 15minutes at 4°C, the pellet was then discarded. Protein concentrations were quantified using the Thermo Fischer Pierce BCA Protein Assay kit. Concentrations were then equalised and equal volumes of Laemmli loading buffer was added to each sample. These were then heated at 95°C for 5minutes before adding 1M DTT at 10% v/v of sample and stored at -20°C.

2.3.2: Protein separation.

A 10% separating gel was used due to the proteins size of 50kDa and was composed of 1.5M Tris at a pH of 8.8, 30% acrylamide, 20% SDS, TEMED and 25% Ammonium Pisulphate. Whilst a 4.3% stacked gel consisted of 0.5M Tris-HCL pH6.8, 30% acrylamide, 20% SDS, TEMED and 25% Ammonium Pisulphate. Samples were then mixed with Laemmli loading buffer and a total of 15 µl was deposited into the wells, alongside 5 µl Full range Rainbow ladder. This was run at 120V for 2 hours.

2.3.3: Protein transfer, immunoblotting and quantification.

Hybond-P membranes were transferred by electrophoresis on a stir plate at 100V for 1 hour in a cold room. The membranes were blocked overnight in Licor Odyssey buffer. Followed by incubation with primary AbCAM eEF1A2 1:1000 dilution in blocking solution and GAPDH 1:2000 dilution for 1 hour. After a series of washes with 2% PBS-Tween the secondary anitbodies were applied: For AbCAM a Licor anti-rabbit and for

GAPDH, Licor anti-mouse, both diluted at 1:5000 in blocking solution and incubated for 1 hour. Membranes were visualized using the Licor Transilluminator.Visualised membranes were then analysed using Image J Lite Studio version 5.2. The signal intensities of the bands generated at 50kDa and 37kDa to quantify the levels of eEF1A2 and GFAP respectively were measured. Readings from the eEF1A2 bands were normalised to those of the GAPDH bands for each sample.

2.4: Immunohistochemistry.

2.4.1: Sample preparation.

The spinal cords of mice from the Del.22.Ex.3 and D252H lines were extracted for immunohistochemistry. For the Del.22.Ex.3 line three repeats of each homozygote, heterozygote and wildtype were prepared, whereas for the D252H only a singular mouse was used from each genotype. These were a mixture of females and males. The spinal cords were extracted by ejection using a 1.1x50mm needle. Mice were partially skinned dorsally and decapitated. A transverse incision was made through the lower lumber spine cranial of the iliac crest muscles. The needle attached to PBSx1 filled syringe was inserted into the exposed spinal canal, ejected whole and immediately frozen on dry ice. Spinal cords where then submerged in formalin for fixation for 24hrs. The formalin was exchanged for 10% EDTA pH7.4 changed weekly for three weeks and then submerged in formalin for 24hrs. The spinal cords were then processed by the University of Edinburgh Pathology Histology service using a Leica tissue processor ASP 300S. Paraffin embedded tissue were then sectioned by the Pathology Histology service into 4µm sections and mounted on Superfrost plus slides in preparation for immunohistochemistry.

2.4.1: Section staining.

For staining with GFAP the paraffin embedded spinal cords were first deparaffinised by emersion in xylene (2x5 minutes) followed by rehydration in 100% ethanol (2x5 minutes) and 70% ethanol (2x5 minutes). To remove any remaining residuals the slides were washed in running water for 5 minutes. Antigen retrieval was done by treating slides with proteinade K at RTM for 10 minutes and washed for 5 minutes. Slides were submerged in 3% hydrogen peroxide for 10 minutes and washed in water for 5 minutes before being washed with PBS for 5 minutes. The slides were blocked with 100µl of 1:5 diluted goat blocking serum in PBS for 10 minutes. The primary antibody Dako GFAP rabbit 1:500 diluted in PBS was applied overnight at 4°C. After washes in PBS, slides were incubated with the secondary antibody (dako goat anti-rabbit diluted 1:500 in PBS) for 30 minutes then washed before incubating with 3 drops of Strept ABC for 30 minutes. Slides were washed with PBS the treated with DAB (diaminobenzidene, Abcam) for 10 minutes for visualisation of staining. The slides were washed in water and counterstained with haematoxylin for 5 minutes, washed in water, then differentiated in Blue in lithium carbonate for 5 seconds and washed. To dehydrate the slides they were sequentially

submerged for 5 minutes in 70% ethanol, 100% ethanol and xylene then finally mounted using DPX.

2.5: Proteomics.

2.5.1: Sample extraction and preparation.

The spinal cords of mice from the Del.22.Ex.3 line were extracted for mass spectrometry. Six mice: three homozygotes and three wildtypes from the same litter were extracted at 21 days. Null.1, Null.2, Null.3, Wt.2 and Wt.3 were all males whilst Wt.1 was female. The spinal cords were extracted by ejection using a 1.1x50mm needle. Mice were partially skinned dorsally and decapitated. A transverse incision was made through the lower lumber spine cranial of the iliac crest muscles. The needle attached to PBSx1 filled syringe was inserted into the exposed spinal canal, ejected whole and immediately frozen on dry ice. Spinal cords were washed twice in pre-cooled 1xPBS and 500µl of cold lysis buffer (RIPA buffer (50mM Tris pH 7.5, 150mM NaCl, 0.5% (w/v) Sodium Deoxycholate, 1% (v/v) NP-40 (Igepal), 0.1% SDS (v/v) with protease inhibitor) was added per 10mg of tissue. This was homogenised mechanically for 2 minutes before sonication using Covaris E220 Sonicator (PIP 90W, Duty factor 20%, Cycles per burst 200) for 180s, at 6°C. Lysate was cleared by centrifugation at 12000rpm for 20 minutes at 4°C. These were then submitted for mass spectrometry.

2.5.2: Mass spectrometry.

*LFQ-MS/MS was conducted by using Filter Aided Sample Preparation (*FASP) combining each prepared samples with 200µl UA (8M urea, 100mM tris) in a Vivaon 500 30,000 (R) (Sartorius VN01H22). All centrifugation steps were performed at 14,000x g. The filter was then washed twice with 200µl UA. 100µl 50mM iodoacetamide in UA was applied to the samples and incubated in the dark for 30 minutes to alkylate. Post spinning, this was followed by two washes with UA and another two washes with 50mM ammonium bicarbonate. 100µg trypsin (Life Technologies 90058) in 2ml 50mM ammonium bicarbonate was prepared on ice and 40µl added to each filter. After overnight incubation in a wet 37°C chamber, samples were acidified by addition of 5µl 10% trifluoroacetic acid, pH check by spotting onto pH paper, and peptide concentration estimated using a NanoDrop. 10µg of the resulting peptide solution was loaded onto an activated (20µl methanol), equilibrated (100µl 0.1% TFA) C18 StAGE tip, and washed with 100µl 0.1% TFA. The bound peptides were eluted into a Protein LoBind 1.5ml tube with 20ul 80% ACN 0.1% TFA and concentrated to less than 4ul in a vacuum concentrator. The final volume was adjusted to 6µl with 0.1% TFA.

Online LC was performed using a Dionex RSLC Nano. Following the C18 clean-up, 5µg peptides were injected onto a C18 packed emitter and eluted over a gradient of 2%-80% ACN in 120 minutes, with 0.1% TFA throughout. Eluting peptides were ionised at +2kV

before data-dependent analysis on a Thermo Q-Exactive Plus. MS1 was acquired with mz range 300-1650 and resolution 70,000, and top 12 ions were selected for fragmentation with normalised collision energy of 26, and an exclusion window of 30 seconds. MS2 were collected with resolution 17,500. The AGC targets for MS1 and MS2 were 3e6 and 5e4 respectively, and all spectra were acquired with 1 microscan and without lockmass.

2.5.3: Statistical tests.

The data was analysed using MaxQuant in conjunction with uniprot fasta database, with match between runs (MS/MS not required). LFQ with 1 peptide required, and statistical analyses performed in R. Contaminants were removed from database along with proteins with <1 unique peptide. Proteins with a median of zero across LFQ Intensities were removed from the dataset. Any remaining missing values were imputed by MNAR is left-censoring. P-values were calculated by pooled variance, two-tailed t-test in R and fold changes based on means of the three samples from each group.

2.6: Microarray.

The microarray data was conducted by Andy Sims and exact protocol is unknown.

2.6.1: Sample extraction and preparation.

The spinal cords of *wasted* mice were extracted at 21 days old from six homozygotes and six wildtype.

2.6.2: Statistical tests.

The p-values for volcano plot were calculated by two-tailed student's t-test in R and fold changes based on means of the three samples from each group. The Significance Analysis of Microarrays (SAM) was conducted in R.

2.7: Gene Ontology analysis.

Differentially expressed proteins as identified as significantly downregulated and upregulated by pooled variance, two-tailed t-test in R. As well as genes: identified as significantly downregulated and upregulated by SAM. These lists were separately subjected to PANTHER classification system; version 12.0 released 2017-07-10 (http://www.pantherdb.org/). GO terms enriched for molecular functions and pathways were compiled into pie charts.

Chapter 3: Characterisation of the D252H mutation in eEF1A2 in mice.

3.1 Development of mouse lines.

Transgenic founder mice were generated by others by injecting CRISPR/Cas9 into fertilised mouse oocytes of the C57BL/6 inbred line. The microbial clustered regularly interspaced short palindromic repeats (CRISPR) and its associated RNA-guided Cas9 nuclease is an immune response of prokaryotes against viruses which has been harnessed to mediate genomic engineering (Marraffini, 2015). Non-coding RNA otherwise referred to as guide RNA (gRNA) directs Cas9 nuclease to induce double stranded breaks at specific sites which include protospacer adjacent motifs (PAM), a sequence located 5' of the target DNA. The damaged DNA then undergoes one of two repair mechanisms, the non-homologous end joining DNA repair pathway (NHEJ) or the homology directed repair (HDR) pathway. NHEJ repair is error prone and often results in insertions and/or deletions of varying lengths, whereas HDR is more precise in introducing mutations and insertions from the donor templates (Ran *et al.*, 2013). However an issue when the HDR and NHEJ pathways are enacted and Cas9 nuclease may continue to operate and cleave at intact PAM sites. In order to avoid this and because the PAM site was located in a coding region, a silent mutation was incorporated into the PAM site of the donor template, thus protecting it against Cas9. The donor template would also include the targeted mutation. Located in exon 5 of the *Eef1a2* gene Exon 5, it would convert 252 Aspartic acid into a Histidine, which was shown to be damaging as human cases suffer from intellectual disability and autism, it was also predicted to be damaging (using the PolyPhen-2 tool) and affect protein functioning, as the site overlaps with *Eef1β* binding site; this may impede *Eef1a2*'s regenerative ability to its active GTP-bound state.

The CRISPR D252H line was developed from breeding on from two founders heterozygous for the D252H mutation. The CRISPR D252H line has been shown to be robust and reproducible and relatively easy to develop with CRISPR/Cas9 techniques making it a practical model. This mutation was also presented in human cases, as outlined in the introductions. I characterised the remaining mice from the experiment to evaluate how successful the CRISPR experiment was and analysing the resultant mutations.

Despite the progression in transgenesis and widespread practice of CRISPR gene editing technology, there still remain limitations and complications in the resulting genome. Issues arose with the founder mice displaying mosaicism and range of insertions and deletions whilst not always incorporating the intended PAM site and targeted mutation. This may have been because the DNA repair pathway selected for by the cell was the imprecise, error prone NHEJ. However, a possible reason for the resultant mosaicism,

upon CRISPR/Cas9 injection the cell continued to divide whilst the system remained active, or repair pathways were not activated until DNA replication causing different cells to carry varying mutations and/or insertions and deletions.

Following the CRISPR/Cas9 experiment characterisation of the founder mice was required. Therefore I genotyped the mice by TOPO cloning and sequencing as previous sequencing results had overlaying peaks that needed to be separated in order to recover specific alleles. Alongside examining their respective expression of eEF1A2 through western blotting in order to determine the effects of any deletions on expression levels.

3.2 Genotyping.

As initial sequencing identified the founder mice as mosaic with overlapping sub-peaks, an allele sensitive method was required to delimit the varying point mutations, insertions or deletions incorporated into each mouse. I carried out nested PCR with a 678 bp final product region surrounding the D252H mutation and encompassing PAM sites located from exon 5 of the *Eef1a2* gene, followed by TOPO cloning and Sanger sequencing. This allowed me to analyse the products of each allele found in each mouse to allow clear reads of each allele.

Three out of the twelve mice genotyped appeared mosaic having more than two alleles (table 3), mice #14, #16 and #18. Mouse #14 had three alleles none with CRISPR induced mutations to include the target by HDR: a wildtype allele, an 11bp deletion and an insertion of 34bp, both of which caused the reading frame to shift. Mouse #16 was also mosaic with a wildtype allele, a 1bp insertion in another as well as the D252H mutation and PAM site mutations G→C and C→T respectively in the third allele. The other mosaic mouse #18 had incorporated a missense mutation, changing the proline into a leucine, in the second nucleotide of the PAM site. In this same allele downstream of the target sites a large 113bp insertion was found. Another two alleles had a 6bp deletion, one of which demonstrated a T→C mutation downstream of the D252 site. The phenotype of #18 phenotype was not comprehensively reported but it had normal weight gain, which suggests there was no deterioration phenotype, this may have been either because the mutations experienced were non-harmful, or that the presence of the wildtype allele was sufficient to resist the development of the phenotype, as seen in non-mosaic heterozygotes. Mouse #10 was recorded as having a phenotype of tremor, absence seizures, periods of ataxia/gait changes/low movement interspersed with periods of more normal movement and a lack of tremor. On one allele a 16bp insertion was found and on the other allele a 5bp deletion that encompassed a PAM site. Evidently these indels were enough to cause a phenotype so severe that for ethical reasons, led to an early culling. Although these changes did not cause a frameshift, one cannot eliminate the possibility of the resultant

eEF1A2 protein being either non-functional, ablated or another allele that was predominant in the brain. Mouse #11 appeared to have successfully incorporated a C→T mutation in the PAM site and the D252H mutation. It also had a 1bp deletion just prior to D252H resulting in a frameshift. Its other allele was wildtype. This mouse had developed a slight tremor and moved slower than its littermates but gained weight normally. Mouse #12 too had one wild type allele but on the other I saw a 26bp insertion. Mouse #15 had wild-type sequence except for the target D252H site (G→C) and at one of the PAM sites (G→C), along with a wild type allele. Although the #15.2 clone (see supplementary information) there is an S in the position of the D252 site, this denotes a strong possibility that the nucleotide is either a Guanine or a Cytosine, and since this clone did not have any other mutations, it is most likely also a wild type allele. Mouse #17 had a wildtype allele and another with the largest deletion at 37bp between the PAM sites which appears to have deleted the D252 site. The one allele identified in mouse #21 did not incorporate any of the desired mutations but had a 16bp insertion, which was a repeat of the base pairs surrounding the original D252 location (6bp upstream and 2bp downstream), causing a frameshift. Using the Provean software, a tool developed to predict whether protein sequence variation would affect the resultant proteins function, the effect of this insertion was revealed to be deleterious, scoring -9.31. The threshold value for whether a variant is deleterious is ≤ -2.5, any higher and the variant would be predicted to have a neutral effect (Choi *et al.*, 2012).

Mouse #23 was wildtype and unaffected by CRISPR in the region sequenced. The alleles in mouse #24 had a 19bp deletion and 18bp insertion respectively. The insertion has changed the reading frame of the resulting protein which would strongly suggest that the eEF1A2 protein is non-functional. Across the topo clones from #24 some point mutations were observed in what had been thought to be a wildtype allele. The possible causes were that Cas9 had cut here incidentally, that the Taq error rate had introduced these changes into the PCR products before cloning, These reasons may have also been behind the T→C mutation in one of #18s alleles, which would have reduced its apparent level of mosaicism. Mouse #25 had not incorporated any of the targeted mutations, but had an insertion of 16bp just preceding the D252 location causing a frameshift, whereas the other allele has a 6bp deletion.

Some mice experienced a disruption in the reading frame, these frameshifts usually result in abnormal protein products that can be truncated or misfolded.

Mouse	***Alleles***	***D252H***	***PAM***	***Mutation (x3, Frameshift causing)***	***Phenotype***
10	2	-	Deleted	-	Severe
11	2	G→C	C→T	Frameshift	Normal
12	2	-	-	-	Normal
14	3	-	-	Frameshift	Normal
15	2	G→C	G→C	-	Moderately slowed movement
16	3	G→C	C→T	Frameshift	Normal
17	2	Deleted	-	-	Normal
18	4	-	C→T	-	Normal
21	1	-	-	Frameshift	Normal
23	1	-	-	-	Normal
24	2	-	-	Frameshift	Normal
25	2	-	-	Frameshift	Normal

Table 3. Resultant alleles and mutations from CRISPR/Cas9 experiment: Mutations seen in the targeted for regions and whether a frameshift was observed, as well as the number of alleles in each mouse.

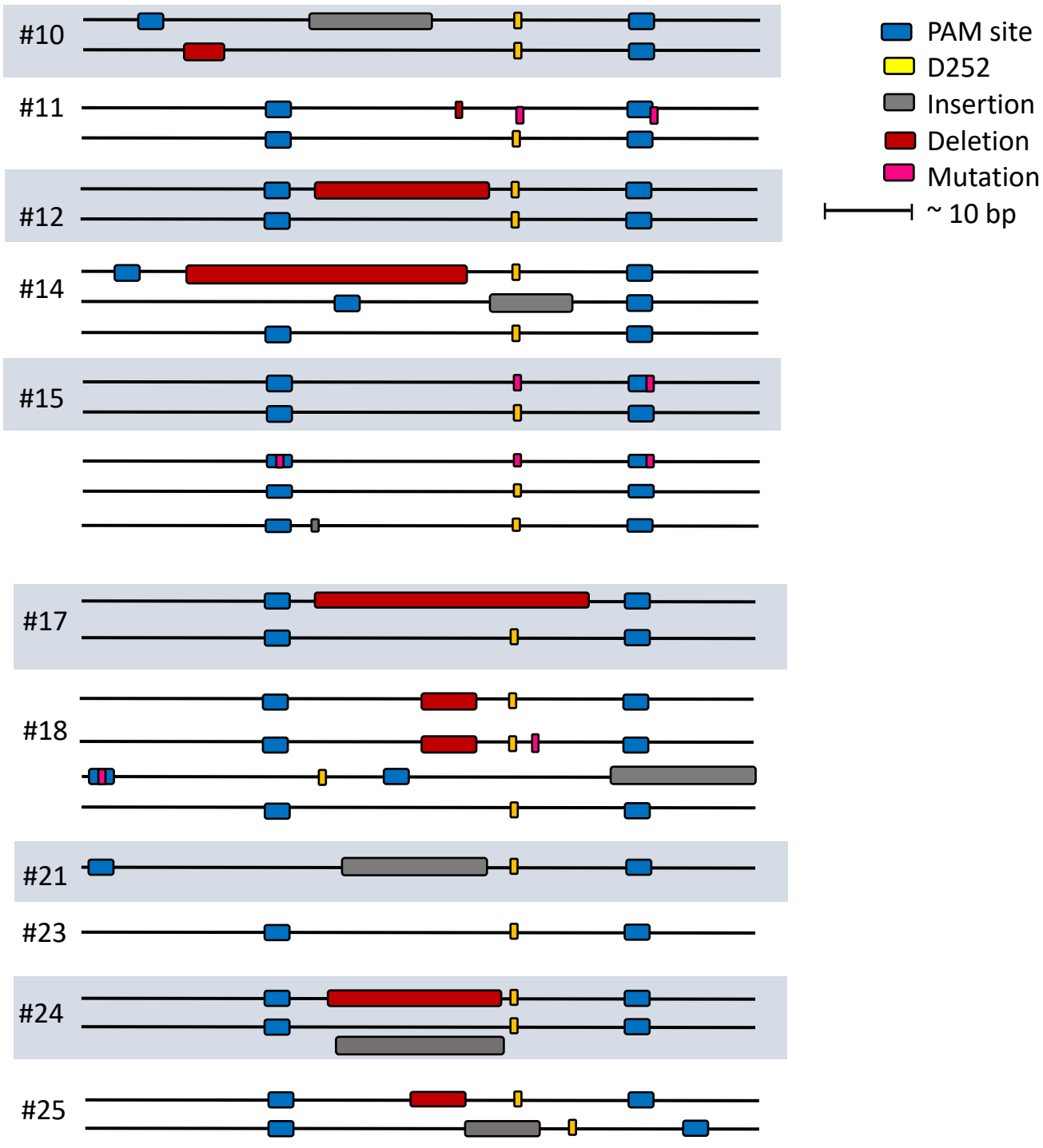


Fig.11. D252H CRISPR experiment mice alleles: Graphical representation of average of 120bp length of sequence excerpt of the alleles found in D252H founder mice showing the relative sizes of deletions (Red) and insertions (Green), PAM sites (Blue), D252 location (Yellow) and mutations (Pink).

3.3 Protein Expression.

To investigate the degree of effect the differing mutations in *Eef1a2* had on its protein expression, western blot analysis on the muscle tissue of the D252H founder mice was conducted. The mice were 60 days at age of culling therefore should be expressing eEF1A2 alone, as myocytes would be terminally differentiated and the switch from eEF1A1 to eEF1A2 complete.

Antibodies selecting for eEF1A2 were used and normalised with the protein levels Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) a housekeeping gene expressed ubiquitously that is commonly employed as loading control. Signals were present at the correct sizing of 50kDa and 37kDa respectively.

The founder mice exhibit varying expression levels of eEF1A2, this is most likely due to the variation of mutations found observed across them, and the varying degrees of mosaicism. Unfortunately tissue from mice #10, #15 and #16 were not collected as #10 had been euthanised and mice #15 and #16 were not culled and bred on from, therefore these mice were not included in the protein expression assay.

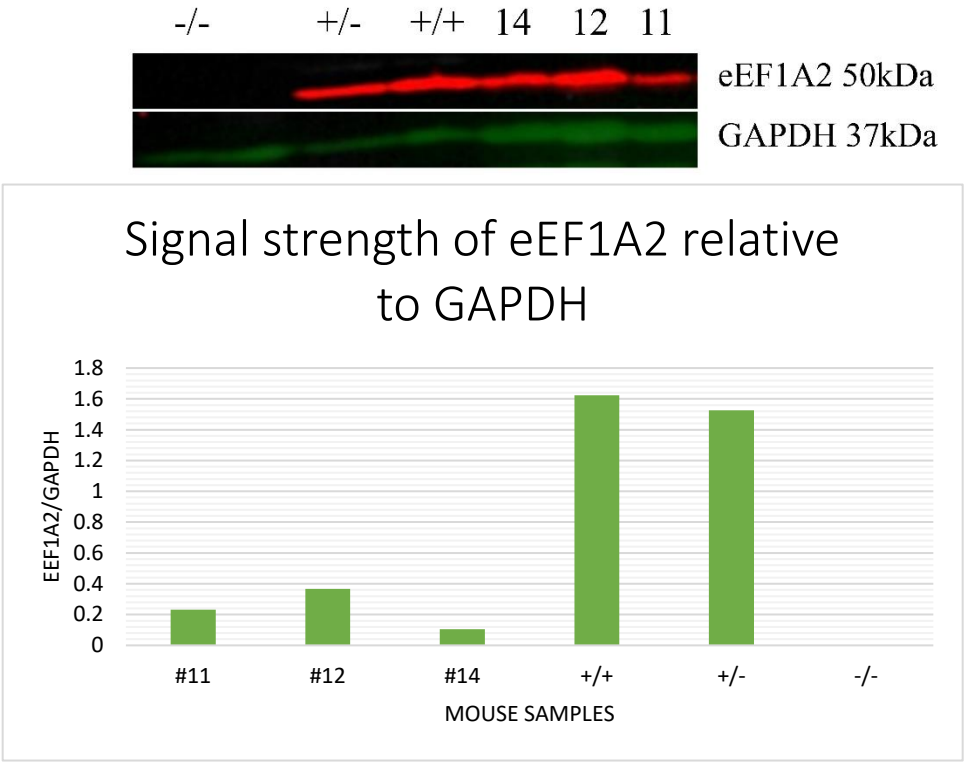


Fig 12. Expression of eEF1A2 in D252H founder mice and signal strengths of eEF1A2 relative to GAPDH: Western blotting results of eEF1A2 (red band) observed at 50kDa and relative GAPDH expression (green band) observed at 37kDa of mice #11, #12 and #14, alongside a *wasted* homozygote, heterozygous and wild type mouse. Founder mice demonstrate abated expression of eEF1A2 whilst the wildtype and heterozygous significantly more upon statistical analysis.

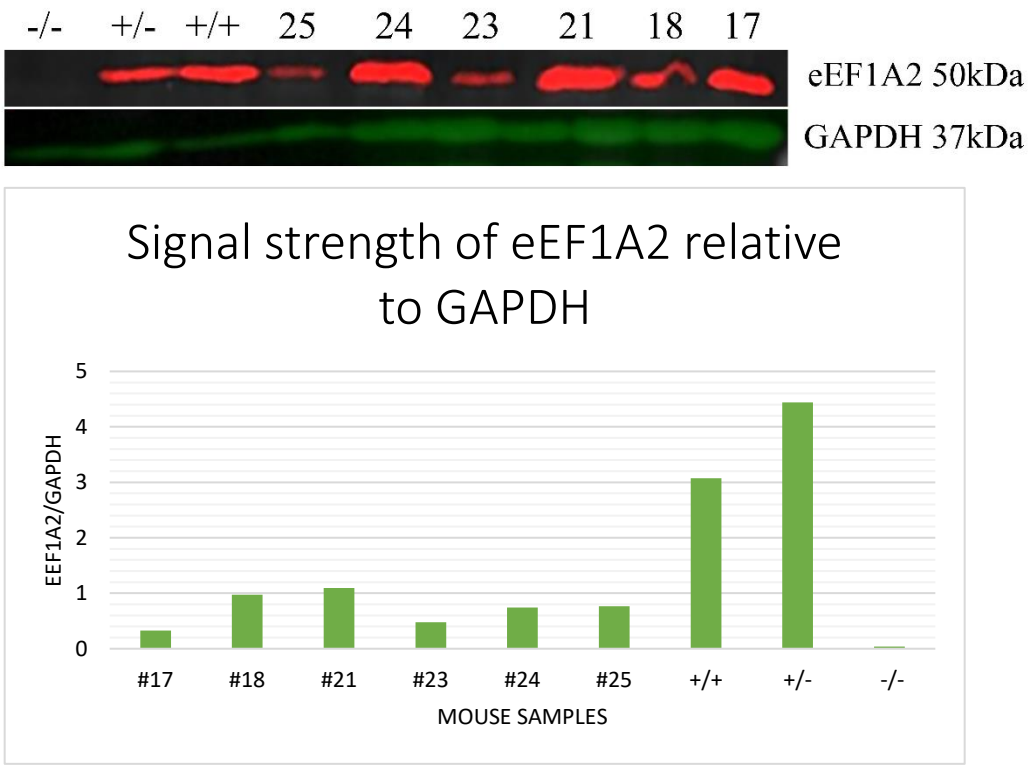


Fig 13. Expression of eEF1A2 in D252H founder mice and signal strengths of eEF1A2 relative to GAPDH: Western blotting results of eEF1A2 (red band) observed at 50kDa and relative GAPDH expression (green band) observed at 37kDa of mice #17, #18, #21, #23, #24 and #25 alongside a *wasted* homozygote, heterozygous and wild type mouse. Founder mice demonstrate abated expression of eEF1A2 whilst the wildtype and heterozygous significantly more upon statistical analysis.

Mouse #11 had incorporated the desired D252H mutation. Supporting the suspicion that the mutation impairs eEF1A2 expression, there were lowered levels of eEF1A2. However the protein was still detected which may have been due to the wildtype allele expressing, although it did not demonstrate nearly as much expression as in the control heterozygote. Notably, the mice bred on from the D252H #15 and #16 expressed eEF1A2 but it appeared non-functional, therefore the reduced expression in #11 may be attributed to the 1bp deletion and not the D252H mutation which is the most likely explanation as this too was predicted as damaging by Provean analysis. Mouse #12 also had reduced eEF1A2 expression. This was not surprising considering the large deletion found in one of its alleles, but again it had a wildtype allele yet still did not express as much as the control heterozygote. The mosaic mouse #14 experienced mutations that caused frameshifts in two of its alleles, ensuing from indels, explaining why it had ablated expression as the protein may have not been able to fold correctly or be truncated. Mouse #17 had the largest deletion observed that eliminated the site of the D252 mutation, this is reflected in its

reduced eEF1A2 expression. Mouse #18 also had reduced expression, which can be attributed to the missense mutation and deletions impairing the stability of the resultant protein and preventing it from being detected. However it still retained a wild type allele that might be responsible for the expression detected. Mouse #21 was reported as having a single allele with an insertion and presented ~60% less eEF1A2 expression than its comparative wild type control. Mouse #23 surprisingly despite being noted as wildtype showed far lower levels of eEF1A2 expression than controls and many of the other mice who had a range of mutations and/or indels. Mice #24 and #25 exhibited similar levels of eEF1A2 that were significantly lower than the wildtype expression. Although #24 had a wildtype allele whereas #25 consisted of two with indels. However this may be because of undetected mosaicism; the tissues used for analysis in the western blot were different from the tissue used in the preparation of DNA analysis therefore the cells may have sampled from a different population with different eEF1A2 sequence.

The controls used were acquired from *wasted* mice; the homozygotes had no or insignificant amounts of eEF1A2 in contrast to the wildtype which showed high levels of expression. Unpredictably however, the heterozygous control had either equal or excessive eEF1A2 levels which is surprising considering *wasted* homozygotes are reported as having the expected ~50% of normal expression. Most founder mice retained a wildtype allele, despite this, eEF1A2 expression parallel to the expected heterozygous level was not demonstrated. Reasons for this may have been that the ages of the controls and founder mice were different, therefore expression of eEF1A2 may have varied in the older controls. The sex of the controls were also unknown therefore it may have been an effecting factor however this is less likely as previous work done on the mice reveal little to no difference in eEF1A2 expression between males and females. Therefore the comparisons made and conclusions drawn upon in regards to the control heterozygote are questionable. They were from different sexes and various ages which may have affected expression of eEF1A2.

A repeat of the western blot using alternative controls was conducted to ensure reproducibility and found some discrepancies between the expression patterns but the majority of samples were observed to be expressing at similar levels. Mouse #12 saw an increase in expression in the repeat experiment reaching nearly 50% eEF1A2 levels compared to the wildtype. This is more in line with the identified genotype as it had reserved a wildtype allele and mimics, as mentioned previously, the *wasted* heterozygous expression profile. Mice #21 and #25 also demonstrated enhanced eEF1A2 levels in the repeat. These findings suggest for both mice that the indels borne were not damaging to the protein. On the other hand, this supports the postulation that the mice have more alleles that were not detected.

The D252H line that was established from these founder mice used only #15 and #16, breeding them with wildtypes. Similar to *wasted*, homozygotes manifest symptoms of ataxia, weight loss and seizures, whereas the heterozygotes do not develop any phenotype.

D252H mice appear to express eEF1A2, but the phenotypes suggest that the protein is non-functional.

3.4 Discussion.

The D252H mutation in eEF1A2 has been identified in humans with the two cases showing symptoms of global developmental delay, one of which has also been reported as being non-verbal whilst the other is too young to report fully on (Nakajima *et al.*, 2015). To explore the role of eEF1A2 in the development of the aberrant neurological phenotype the mutation was recreated in mice. A CRISPR/Cas9 experiment caused a range of mutations in the C57BL/6 mouse fertilised oocytes in attempts to generate the D252H mutation. The resultant founder mice revealed a range of mutations including large indels when genotyping and also revealed that some mice were mosaic. Their respective eEF1A2 expression was also analysed by western blotting, which had shown that among most of the founders, expression was reduced.

After genotyping it was clear that few mice had incorporated the desired mutation which may have been a consequence of the cell employing the NHEJ mechanism as opposed to the HDR with constructed repair templates with the mutations. This is a common issue with CRISPR/Cas9 experiments as it is dependent on the efficiency of the cell to utilize the HDR pathway. The various differing mutations observed in the founders in some cases would cause a frameshift in the amino acid sequence. These often have damaging effects upon the resultant protein as they alter the stop codon in the sequence either by introducing it prematurely causing a truncated protein, or later to cause an abnormally long protein which can have an effect on the proteins overall structure. These changes cause non-functional proteins that can have dominant negative effect or deleterious gain of function activity. In cases with premature stop codons the nonsense mediated decay pathway is activated. This is a cellular surveillance mechanism that can recognise these premature stop codons and essentially mark them for degradation by the nonsense mediated decay complexes. However the pathway model suggests that it is only applied when the premature stop codon is located 50-55 nucleotides upstream from an exon junction complex (EJC), which are assembled predominantly at exon-exon junctions (Hug *et al.*, 2016).

As only a specific region was sequenced, it cannot be certain that there were no off target effects from the CRISPR/Cas9 experiment in the founders. The induced mutations that did not necessarily cause a frameshift can still cause issues with the resultant protein. Important regions of the sequence may have been lost or amino acids change that impede the proteins functioning. In the case of the mosaic mouse #18, one of the alleles had a missense mutation changing the proline amino acid into a leucine. Proline is a cyclic nonpolar aliphatic amino acid that has lower configurational entropy due to the pyrrolidine

ring, it has been repeatedly reported that when mutated into another amino acid protein stability is decreased (Ge and Pan, 2009,Suzuki *et al.*, 1987, Gray *et al.*, 1996). Although the degree of destabilisation tends to be marginal, it may have had an effect on eEF1A2, however this is unlikely. Leucine is a nonpolar aminio acid that can pass the blood-brain barrier more rapidly than other amino acids. It serves as a donor of amino acid groups for glutamate synthesis (a neurotransmitter), which is usually maintained at low intrasynaptic concentrations to minimize excitotoxicity in neurons (Yudkoff *et al.*, 2005). However, this may not be the case for mouse #18.

Nearly all founders retained a wildtype allele, meaning that most were heterozygous for a type of mutation in eEF1A2. These mice, apart from #10 and #15, did not demonstrate any physical phenotype and as the expression of eEF1A2 in these mice was significantly lower, there was no evidence of compensation by this allele. However, it does suggest for haplosufficiency; the wildtype allele was able to produce enough functional protein to prevent the diseased phenotype. This appeared to be the case in the *wasted* line heterozygotes, which showed 50% reduced eEF1A2 expression but no *wasted* phenotype (Griffiths *et al.*, 2012). The levels of eEF1A2 detected below 50% in the D252H mice survived for up to 60 days without a diseased phenotype. This observation can have repercussions in the development of therapies for patients with neuronal developmental disorders. It suggests that only low levels of eEF1A2 is required to prevent development of disease which is a promising factor for possible gene therapy as it would not need as much compensation for the therapy to be effective. It also suggests that the missense mutation does not entirely result in loss of function.

A limitation of the genotyping methods used was at the TOPO cloning stages in which the detection of alleles is dependent on whether or not they are transfected and grown successfully. In all cases an abundance of colonies were grown and a predicted to be sufficient number were sequenced, but there is always a possibility of excluding alleles unintentionally. Furthermore, an issue with drawing conclusions on the connection between the genotypes of #21 and #23 on eEF1A2 expression is that from the multiple colonies, the only clear sequencing results were obtained from two clones for #21 and a singular clone from #23 TOPO cloning (see supplementary information). The fact that the Provean software predicted #21 as having deleterious effects, yet this not being reflected in either the eEF1A2 expression or the phenotype of the mouse supports the idea that some alleles may have remained undetected. TOPO cloning may have failed to manifest all alleles present and some may have been lost through incoherent sequencing, this offers an explanation for some inconsistencies in eEF1A2 expression and *Eef1a2* alleles, although the genotype and protein expression correlates for most of the founder mice. The discrepancies between the western results and the genotype may have been because of undetected mosaicism. There is no assurance that tissues used for analysis in the western blot had the same eEF1A2 sequence as that of the tissues used in DNA analysis, therefore

the cells may have sampled from different populations with different levels or forms of eEF1A2.

A study into the expression of RNA by qPCR to measure the gene expression levels would have been beneficial as it would have identified any differences in transcription or translation and possible protein degradation. The greatest issue with results from the western blotting were the discrepancies between the eEF1A2 expression levels of the *wasted* heterozygous tissue used and the levels recorded in another study (Griffiths *et al.*, 2012). The heterozygous eEF1A2 levels I detected were either similar to or higher than the wildtype expression levels. Upon further examination of the signals without normalising to GAPDH there was still no drastic improvement in the heterozygous controls demonstrating the expected 50% eEF1A2 expression, however *wasted* mice do not have the same genetic background as the D252H mice, this may also have some effect on the expression of eEF1A2. A repeat using alternative *wasted* heterozygous samples that were age and sex matched may clarify this issue, or studying whether there is a degree of variation across *wasted* heterozygous mice. However, it may be an issue with normalisation with GAPDH an alternative method to normalise would be using the total protein concentration in the lane. The issue with the heterozygous control confounded identifying the degree of eEF1A2 reduction in these mice as it would be inconclusive to draw comparisons with eEF1A2 null heterozygous and the various heterozygous mutations seen in the founder mice.

Chapter 4: Analysis of eEF1A2 null and Wildtype spinal cord quantitative proteome and neuronal damage.

To identify potential biomarkers for MND, a comparative quantitative analysis into the proteome of the spinal cords of eEF1A2 null and wildtype mice was conducted using Label free Mass spectrometry (LFQ-MS) to identify the proteins present and their relative abundances. The results from this experiment were processed and statistically analysed to identify proteins that were identified as differentially expressed between the groups. Significant findings were probed further for biological significance by gene ontology enrichment analysis exploring which pathways and functions were over-represented, in order to better understand pathogenesis. In addition to this analysis, the extent of neuronal damage in the spinal cords of the eEF1A2 nulls, as well as the D252H homozygous and heterozygous mice was analysed by immunohistochemistry (IHC). This would evidence the degree of neuronal degeneration experienced with differing levels of eEF1A2 and identify how severe the pathology of MND is in eEF1A2 mutant mice. Understanding the loss of eEF1A2 and its neuronal degeneration phenotype, the resultant affected protein expression within the spinal cord has elucidated proteins enriched for particular pathways and provided possible prognostic benchmarks for future therapeutic development.

4.1 Biological specimens.

Spinal cords from three wild type mice and three eEF1A2 null mice from one litter of the Del.22.Ex.3 line were subjected to proteome analysis. A comparative study between these two groups is likely to yield a difference associated with neuronal degeneration. The absence of eEF1A2 expression and severity of symptoms observed in the homozygotes of this line suggests that there may be differential expression of a protein or group of proteins that are distinctive and would act as biomarkers of neuronal degeneration. Whereas mice heterozygous for eEF1A2 null mutations have displayed reduced expression of the protein, disease phenotype and pathology has not presented itself in live mice or spinal cords, so this genotype was therefore not included in the analysis.

Spinal cords were extracted from 21 day old mice, a time-point in which the switch between eEF1A isoforms is more or less completed, with only trace amounts of eEF1A1 being expressed (Chambers *et al.*, 1996). As mentioned before, this is concurrent with the onset of symptoms of motor neuron degeneration associated with eEF1A2 null genotypes (Newbury *et al.* 2005).

The spinal cord would be appropriate, as changes in protein expression will be monitored across glial and motor neuron cells. Whole spinal cord analysis is beneficial due to

observations of changes made in glial cells being more telling as they would not be affected directly by the mutation and downregulation of protein synthesis given that glial cells only expresses the eEF1A1 isoform (Newbury *et al* 2007). Whereas motor neurons expression only eEF1A2 will display a quantity of changes in expression, predominantly resultant of downregulation of protein synthesis. Ergo p21 spinal cords should display changes affected by mutated eEF1A2 but before differences in protein expression can be heavily influenced by impaired translation; if analysis was conducted upon later time-points any resulting differences may be confounded by the loss of translation.

4.2 Establishing the extent of neurodegeneration in the new homozygous mice.

A study into the pathology of the developing aberrant neurological phenotype of mice from the D252H and Del.22.Ex.3 lines was conducted using immunohistochemistry to establish the degree of neurodegeneration seen in the new homozygous mutant mice. Previous work on the *wasted* mice revealed that neurons in the spinal cord experienced vacuolation, progressive retraction of nerves from motor endplates, as well as evidence of gliosis(Newbury *et al.* 2005, Abbott *et al.,* 2009). This pathology appears as a progressive rostrocaudal gradient. With more motor neuron deterioration occurring initially at the cervical level, there is speculation as to whether these changes work as a cascade and eEF1A1 is switched off progressively, which would explain why caudal areas do not seem to be as affected (Newbury *et al.* 2005). The gliosis reaction of glial cells upon damage in neuronal cells is commonly observed in central nervous system (CNS) injury including motor neuron diseases (Ince *et al.,*1998, Burda and Sofroniew, 2014). Reactive gliosis is a response that characterised by the accumulation of glial filaments, a constituent of which is glial fibrillary acidic protein (GFAP). As a result, GFAP has been commonly employed as a biomarker of neuronal damage (O'Callaghan and Sriram, 2005). The spinal cords of homozygous and heterozygous Del.22.Ex.3 and D252H at the ages of 21 days were stained using this to demonstrate, if any, pathological changes between homozygotes, heterozygotes and wildtypes that may be connected to the neurological phenotype. In order to visualize the levels of neuronal degeneration and to examine if, as observed in the *wasted* line, degeneration is progressive in the Del.22.Ex.3 and D252H lines, the spinal cords of wild type, heterozygous and homozygous mice at cervical, thoracic and lumbar sections were stained for GFAP. These experiments were an important part of the characterisation of the new lines of mice prior to carrying out proteomic analyses of spinal cords.

Colin Smith (Professor of Neuropathology) analysed H+E staining of homozygote, heterozygote and wildtype spinal cords of D252H and Del.22.Ex.3 mice. He identified unambiguously that homozygote mice experience neuronal degeneration in their spinal

cords whilst heterozygotes appears unaffected when compared to the wildtype. In sections from homozygous mice there were clear signs of vacuolation along with various indicators of the evolution of neuronal degeneration; abnormal nuclei that have lost their nucleoli which then deteriorate further losing cell structure and disintegration of the cell membrane, eventually leading to a dead neuron with no definitive cellular structure. The cervical regions had greater degrees of vacuolation and neuronal degeneration, with the lumbar regions demonstrating less degeneration. Therefore this analysis supports the case, as recorded in the *wasted* mice, that the switch between isoforms works progressively rostrocaudally in this eEF1A2 null line, however this was only an observational conclusion. The D252H homozygote showed similar neuron degeneration in all regions in comparison to the Del.22.Ex.3 homozygote, whilst heterozygotes in both lines showed no damage.

GFAP staining in Del.22.Ex.3 spinal cords (Fig.14) was markedly increased in homozygous mice and not prolific in wildtype and heterozygous mice. A high abundance of GFAP is observed in the eEF1A2 null sections across all regions. An apparent decrease in GFAP can be detected from the cervical through to lumbar region. This is in line with the neuronal degeneration witnessed in the H+E stains, again supporting the concept that eEF1A1 is switched off in neurons in a rostrocaudal fashion. However GFAP is also present in the heterozygous and wildtype spinal cord. The heterozygote displays a surprising amount of GFAP whereas the H+E stains revealed no neuronal damage. This finding is also in contradiction with the findings in heterozygous *wasted* mice which were recorded to have rare amounts of GFAP (Newbury *et al.* 2005). It is anomalous in its nature as the mouse also did not manifest any phenotype indicative of the damage. The wildtype retained the least amount of GFAP staining although there was still some present.

The D252H mice exhibited similar results, with the homozygote showing a similar rostrocaudal pattern of decreasing GFAP present (Fig.15). There appeared to be less GFAP detected in the D252H homozygote in comparison to the Del.22.Ex.3 homozygote. This is curious on account of the D252H mice having a more severe phenotype than the Del.22.Ex.3 mice (personal communication). The heterozygote unmistakably showed a degree of GFAP, which did not correlate with the findings from the H+E staining's reporting no degeneration. Such a degree of neuronal damage was unexpected in the heterozygote as they demonstrate no phenotype. The wildtype displayed more GFAP than anticipated, yet it was still significantly less overall than the homozygote and heterozygote.

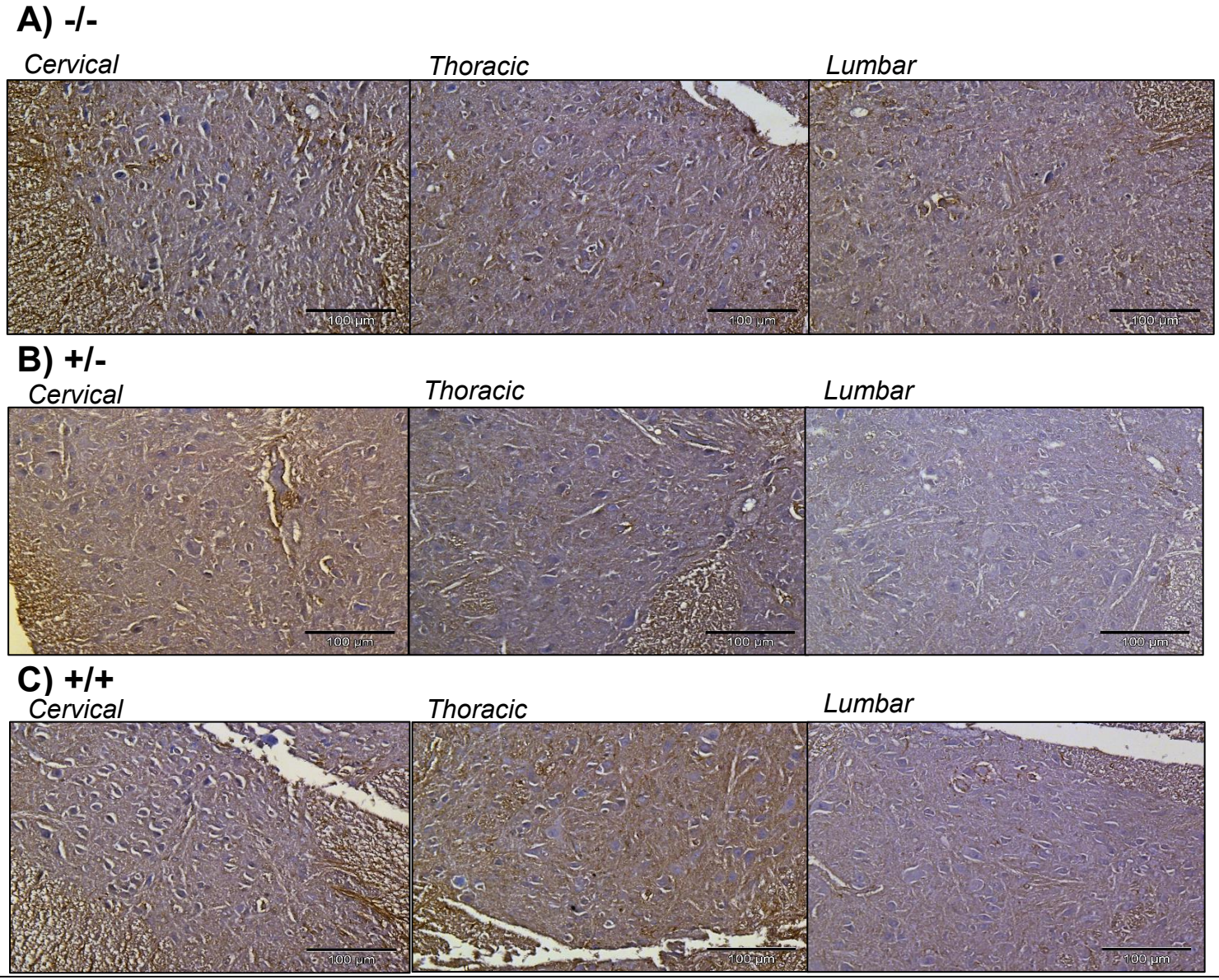
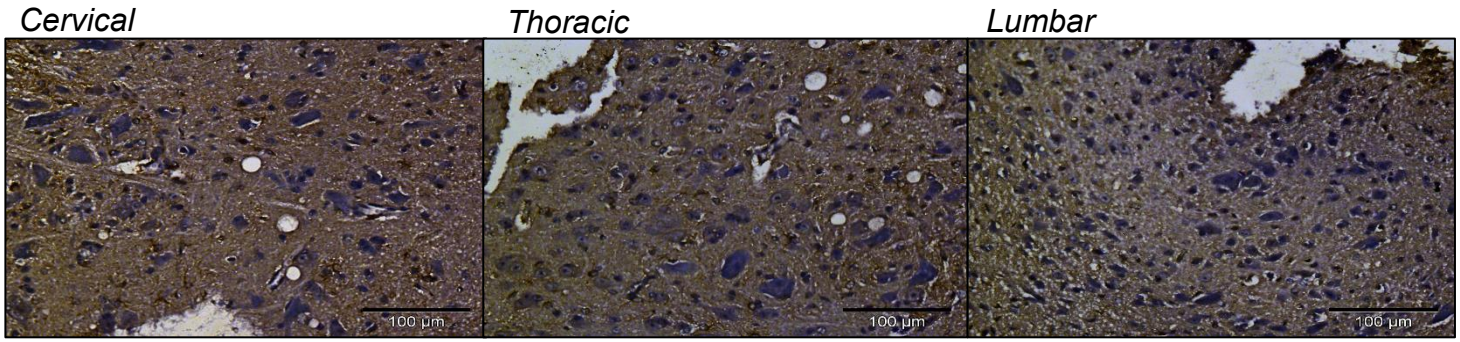
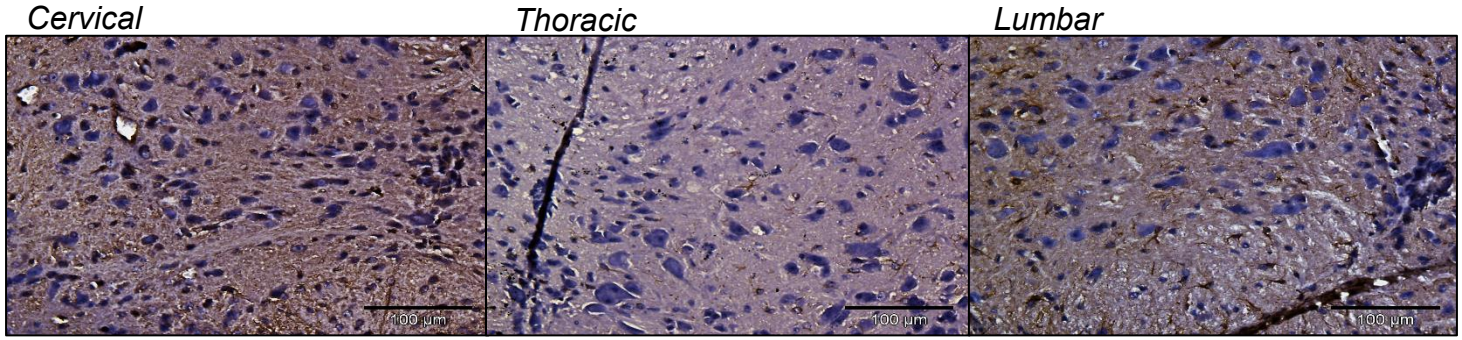


Fig.14. GFAP stained Del.22.Ex.3 spinal cords. Spinal cords of Del.22.Ex.3 stained with GFAP (Brown markings) imaged at x40 magnification. Homozygous, (-/-) B) Heterozygous (+/-) and C) wildtype (+/+).Spinal cords were compared at the cervical, thoracic and lumbar regions. The (-/-) demonstrates high levels of GFAP as a rostrocaudal gradient, with cervical and thoracic regions showing vacuolation. The (+/-) exhibits a degree of GFAP staining as well as (+/+) but significantly less than the (-/-).

A) -/-



B) +/-



C) +/+

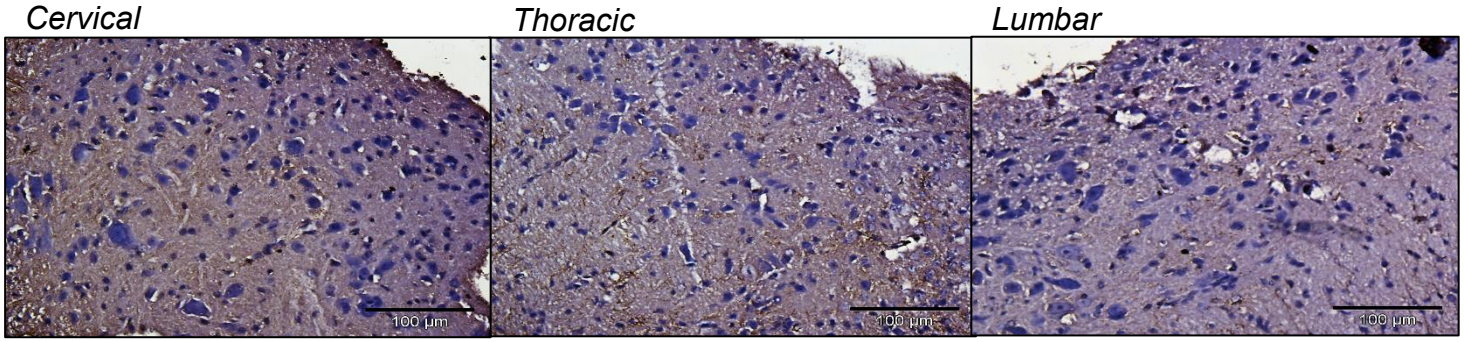


Fig.15. GFAP stained Del.22.Ex.3 spinal cords. Spinal cords of Del.22.Ex.3 stained with GFAP (Brown markings) imaged at x40 magnification. Homozygous, (-/-) B) Heterozygous (+/-) and C) wildtype (+/+).Spinal cords were compared at the cervical, thoracic and lumbar regions. The (-/-) demonstrates high levels of GFAP as a rostrocaudal gradient, with cervical and thoracic regions showing vacuolation. The (+/-) exhibits a degree of GFAP staining as well as (+/+) but significantly less than the (-/-).

4.3 Mass Spectrometry.

4.3.1 Quantitative Label-free Mass spectrometry.

To illustrate comprehensively the proteome of two different states for comparative studies, a bottom up proteomic approach by label free quantitative mass spectrometry has emerged as an adept approach for absolute quantification of proteins. This approach can ascertain protein profiles by characterizing peptides in proteolytic digests and quantifying each protein (Resing et al. 2005). It is a method that can provide efficient sequence coverage in protein identification and is also able to examine post-translational modifications (Latosinska et al. 2015, Wang et al. 2012, Bantsheff et al. 2012, Witze et al., 2007). It does however require robust analytical configuration of ultra-High performance or High performance liquid chromatography for mass spectrometry. LFQ-MS is ideal for exploring proteins that are differentially expressed between wild type and eEF1A2 nulls as one can retroactively quantify proteins, therefore the initial hypothesis is not limited by narrowed choice of select proteins of interest. This study is only a preliminary analysis, however if taken further, LFQ-MS can also facilitate large-scale projects.

Its limitation lies predominantly with its inability to detect low abundance proteins, where it lacks sensitivity in comparison to labelled methods (Hamacher, *et al*. 2011), which can be a substantial issue if proteins are significantly downregulated below the level of detection. There is also an enhanced possibility of introducing technical variance when compared to other quantitative proteomic techniques. Biological samples are kept separate during processing and analysis, and can unintentionally experience subtle differences that may effect and distort the detected proteome and respective abundances, leading to misleading results. This also reduces the reproducibility of the experiment, requiring replications for confident results. Another limitation is that it is not suitable for enriched samples, however this is not applicable to the biological samples used in this study. Nonetheless, facilitated by the advancements in instruments and software it has become increasingly employed for Biomarker discovery studies as it is comparatively low cost and requires less resources in terms of sample preparation (Wang *et al.* 2012). Although there is not a standard method of analysis, LFQ-MS newfound popularity has allowed for range of software and protocol to be developed for use accurate protein profiling.

The results from Label free Mass spectrometry (LFQ-MS) underwent data processing; to interpret the raw data from the mass spectrometry (MS) from each fraction and multiple runs of samples into quantifiable protein abundance a series of steps were taken. Peptides signals were normalized and matched into proteins utilizing MaxQuant software in conjunction with Uniprot fasta database, filtered out for contaminants and imputating missing values before udergoing statistical analysis. Of this, any significant findings were then probed further for biological significance by functional analysis. LFQ-MS and data processing utilizing MaxQuant software; acquisition of raw mass spectrometry data,

protein identification and abundance profiling was done by the Mass Spectrometry services at the Institute of Genomics and Molecular Medicine.

4.3.2 Peptide identification.

Proteins from the biological samples are digested into peptides which are broken down further into fragment ions by tandem Mass spectrometry. These are then accelerated in the mass spectrometer with heavier ions 'flying' slower than lighter ones, the system measures the mass/charge ration of an ion and using this information reconstructs the peptide sequence. Protein identities are then inferred from the peptide sequences when they are matched to protein groups by organism specific sequence database searches (these databases also include contaminants and reverse sequences). A limitation of this approach is that the same peptide sequence can be present in multiple different protein or different isoforms, especially in higher eukaryotes having high degrees of sequence homology. This can cause dubiousness in protein identifications.

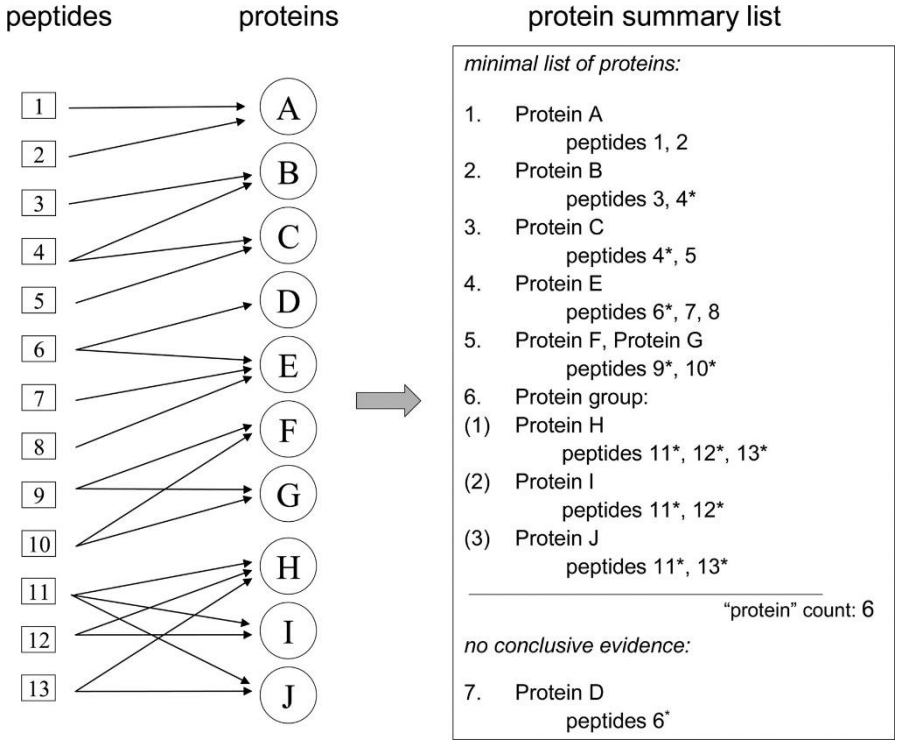


Fig.16: A simplified example of how proteins are inferred. Peptides are assigned to all corresponding proteins and an Occam's razor philosophy of deriving the minimal list of proteins that can explain all observed peptides. Shared peptides are marked with and asterisk. Proteins that are impossible to differentiate on the basis of identified peptides are collapsed into a single entry (As shown by F and G), or presented as a group (H,I and J). Proteins that cannot be conclusively identified do not contribute towards total protein count, but are still shown. Image take from (Nesvizhskii and Aebersold, 2005).

Peptide sequences that are unique to a singular protein in the proteome are termed 'Unique peptides', these are often used to determine the confidence in the proteins' identification. It is common practice for a 'two-peptide' rule to be observed, by only including proteins that have a minimum of 2 unique peptides, as this would reduce the rate of false-positives. However, this two-peptide bias has not been theoretically proven as superior. A study into the performance of this rule against including proteins with singular unique peptide for analysis, identified that the larger set of protein identifications are generated from the single-peptide approach, than from the two-peptide, that are still reliable (Gupta and Pevzner, 2009).

4.3.3 Determination of protein abundances.

Mass spectrometry is inherently non-quantitative as equal amounts of different peptides can generate ions with differing signal intensities due to ionisation efficiency varying and competition with other analytes. Therefore the proxy used for generated for abundance is crucial to true changes in expression being reflected in the analysis. Advances made in the field of proteomics has allowed for experimental data from the signals detected to be interpreted proficiently into data that is representative of protein abundance.

MaxQuant has implemented a novel approach to building accurate abundance profiles for each protein across samples, permitting its use for comparisons between the diseased and wild type groups. Although there are alternative methods of defining protein abundance such as spectral counting, the approach outlined is more accurate, has a higher dynamic range and if required is capable of quantification of post-translational modifications. Termed "LFQ Intensity", it is upheld as representative of a quantifiable presence of a protein, facilitating the assessment of which proteins may be upregulated or downregulated in the eEF1A2 null mice. This approach is based entirely upon the experimental dataset acquired with no external standards. It can overcome the complex issues of normalizing multiple runs across multiple fractions of samples, whilst also allowing for greater amounts of data points to be included by increasing the rate of identification, thus maximizing quantification of proteins. Normalizing the intensities of the peptide ions across fractions of each samples and runs, alongside selecting for peptide signals that should optimally determine the protein signals. The normalization step is crucial when having to draw conclusions from multiple runs and fractionated samples; eliminating signals that were resultant of inevitable variations from technical measurements and random effects from biological samples allows for better deductions to be made upon the proteome dynamics between the two groups and determine biologically significant differences. It is a stage to reduce technical bias introduced such as carry-over and drifts in ionisation and detector efficiencies (America and Cordewener, 2008).

Fractionation of samples for this experiment complicates normalization due to the introduction of variance within each fraction, as well as the total peptide ion signal, which

is usually necessary to normalize,which is split across runs of each fraction. Meaning that any normalization coefficients cannot be classified. Hence the concept of "Delayed normalization" was developed, in which firstly the intensities are summed up with normalization coefficients as free variables, followed by determination of the quantities by global optimization procedures that would achieve the least overall proteome variation.

Approaching the issue by attempting to equalize total signals by adjusting the normalization coefficients for each fraction can introduce errors if there was a divergent average for a particular run. In response to this, the notion that the majority of the proteins change minimally between conditions allows for the use of average behaviour to be a standard.

Peptide ion signals, as aforementioned are summed up across the fractions without normalization coefficients (Nj). The Nj factors are then determined in a nonlinear optimization model which minimizes overall changes for all peptides across all samples. Like this, the total intensity of a peptide ion is defined as:

$$Ip, a(N) = \sum_{k} N_{run(k)} XIC_k$$

Where p is the peptide, a, is the sample and k the runs over all isotope patterns for p. With XIC, in this case, the area of cross section at retention time when maximum intensity is reached. This quantity is then used as the sum of the squared logarithmic changes in all samples for all peptides.

$$H(N) = \sum_{P \in peptides} \sum_{(a.b) \in Samplepairs} \left| log \frac{Ip, a}{Ip, b} \right|^2$$

To achieve the least amount of differential regulation for the most amount of proteins, the $H(N)$ value is then minimized with respect to Nj by Levenberg-Marquardt optimization. As such the peptide ion intensities were normalized across the different runs of samples and their fractions.

However to generate the abundance profile, a selection of which peptide ions to contribute to this intensity must also be chosen. MaxQuant does this by selecting peptide species present between samples and then calculating the ratio using the intensities (Cox *et al.*, 2014). This initial pair-wise ratio is defined as the median to protect against outliers, is followed by determination of the other pair-wise ratios, in this analysis the minimal number of 1 peptide ratios for a given protein ratio was considered valid. The resulting matrix corresponds to the underlying abundance profile across samples and undergoes a least-squares analysis to reconstruct the abundance profile so as to satisfy the individual protein ratios in the matrix based on the sum of the squared differences expressed as:

$$\sum_{(j,k)\in Valid\ pairs} \left(logr_{j,k} - logI_{j,} + logI_k\right)^2$$

Each profile is rescaled to the cumulative intensity across the samples, thus preserving the total summed intensity for a protein over all samples.

4.4 Data Analysis.

An experimental issue arose during mass spectrometry in which tandem digestion was reduced to digestion only with trypsin. The preparation for spinal cord extracts unfortunately was possibly too gentle and did not break down the proteins enough. It would have benefitted from a harsher protocol. Trypsin is a serine protease with high proteolytic activity, high cleavage specificity and resulting peptides have optimal size (700-1500 Da) and charge for mass spectrometry analysis. However, complete digestion does not always occur as tightly folded proteins resist trypsin digestion and trypsin is inefficient in cleaving at lysine residues. To address these issues, tandem digestion with Lys-C drastically improves the efficiency of digestion as it compensates for trypsin's missed cleavage of lysine residues (Giansanti *et al.*, 2016). This ensures higher sequence coverage which can discriminate closely related protein isoforms, as well as aid precise protein identification and quantification. In the absence of Lys-C incomplete digestion is suspected and MaxQuant analysis of peptide and parent protein identification works on the assumption of complete digestion. Despite this limiting factor, data analysis was conducted as proteomic studies were initially conducted using trypsin as a singular enzyme.

The contaminants and reverse sequences as identified by MaxQuant using Uniprot fasta database in the dataset were removed alongside identified proteins with a unique peptide <0 for the reasons mentioned in section 4.3.2.

Missing values can encumber statistical analysis and in LFQ-MS experiments missing values are not uncommon. It has been reported that as many as 70-90% of proteins or peptides harbour at least one missing value (Lazar *et al.*, 2016). Due to LFQ-MS known issues with detecting peptides and proteins at abundances that border the level of sensitivity of the instrument, missing values are classed as 'missing not at random' (MNAR), hence why a single-value approach was taken with the assumption that missing values are present yet below the level of detection. A mechanism of MNAR is left-censoring; a condition when the missing values pertain to the lower intensities, and that intensity distribution of the dataset is truncated on the left side. Upon imputation, a left tailed skew emerges across the distribution of intensity. Proteins with a median of 0 (missing value) across the 3 samples for each group were eliminated from the dataset, as

otherwise the confidence of correct detection lies with a singular value. In each sample the lowest value for intensity was assigned to the remaining missing values.

However, even after using an imputation that promotes left tailed skewness, the dataset maintained relatively normal distribution. The Shapiro Wilks test tests for normality by examining if the dataset is not normally distributed. The null hypothesis is that the dataset follows normal distribution, therefore if the p value is less than 0.05, the data is not normally distributed, in addition to the giving a measure of 'closeness' to an expected normal distribution of a W value. The eEF1A2 null sample means when tested had a W value of 0.99, as this nears 1 it is indicative of normal distribution, although the p value from this test was observed to be $1.42x10^{-11}$; this can be attributed to the large sample size which has the ability to detect even small deviations from normality. The wildtype means also generated a W value of nearing 1 (0.99), and a p value of $8.9x10^{-12}$, therefore both protein populations are suggested to be normally distributed as the tests conducted do not indicate that they are not. However, because the Shapiro Wilk test is subject to bias in rejecting the null hypothesis in the case of large sample sizes, further analysis by other means are required to support its finding that the eEF1A2 nulls and wildtype means are normally distributed.

A Q-Q plot can demonstrate how well the data compares to that of normal distribution. It is a probability plot that is used to compare two distributions by plotting quantiles or estimates of the quantiles of two distributions against each other, the pattern of the plot is telling as to how they compare. A Normal Q-Q plot was generated by plotting the quantiles of either the eEF1A2 null means or wildtype means with the quantiles of a standard normal distribution (Fig.17.). Therefore the distribution of the means can be compared with that of normal distribution to observe if they too fit normality. In Fig.17 the majority of the points fall within the line of normality and although there are bends towards the end tails, it is not so severe so as to describe extreme values that would not fit a normal distribution

pattern. Both the Shapiro Wilks test and the Q-Q plot suggest the data is normally distributed.

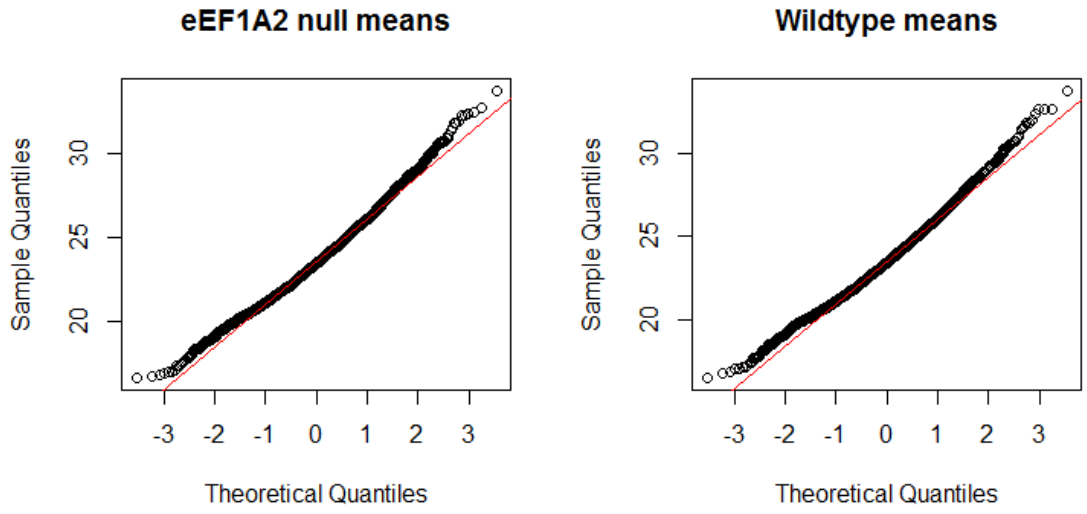


Fig.17. Q-Q plots assessing normal distribution. Q-Q plots using the quantiles of normally distributed data (x-axis) was plotted against the logarithmic eEF1A2 null means (Left) and the logarithmic Wildtype means (Right) quantiles (y-axis). The individual points belong to the eEF1A2 null and Wildtype means. The red line passes through the first and third quantile, this is a robust approach for estimating the parameters of normal distribution. The majority of the datasets points are positioned in approximately the same line as that of normal distribution. The tails diverge somewhat from normal distribution, however their departure from the line are not indicative of non-normality.

4.4.1 Expression profiles:

To identify if the proteome of wildtype and Del.22.Ex.3 mice is distinguished from one another a cluster analysis was done by complete linkage method for hierarchal clustering. This class of cluster analysis functions by initially assigning each observation a separate cluster then examines all the distances between the individual and pairs the closest two clusters together. The following observations then join it or not depending on the distance between the cluster and observation. The complete linkage method of hierarchal clustering uses the maximum distance between object and the objects within a cluster when forming clusters. This analysis identified no particular dissimilarity in overall protein abundances between the wild type mice group and eEF1A2 null group. Usually minimal or no changes in majority of the proteins is expected between healthy and mutated/diseased groups. The groupings were neither made based on gender as female and male mice were not clustered separately. Given the lack of distinguishing variable between clusters and relatively small distances between clusters it can be concluded that eEF1A2 nulls at p21 do not have a differing overall proteome from the wildtypes.

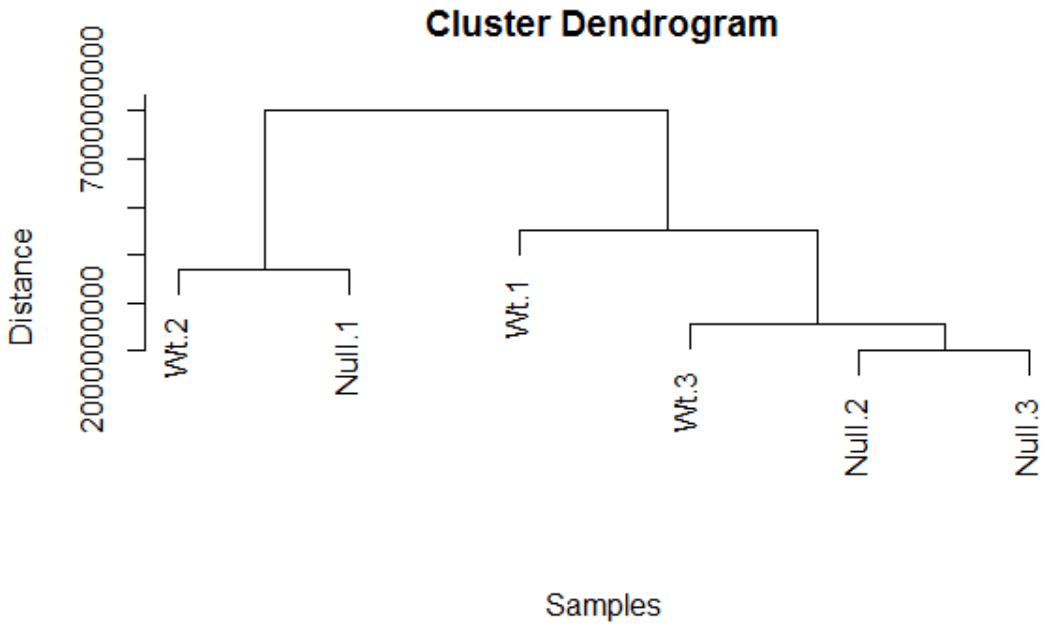


Fig.18. Cluster dendrogram by complete linkage for hierarchal clustering. Wild type mice samples denoted as Wt.1, Wt.2 and Wt.3. whilst mice with eEF1A2 null mutation are denoted as Null.1, Null.2 and Null.3. Samples are represented on the X axis and are connected by decreasing similarity along the Y axis.

This lack of differences are also seen in Fig.19. showing that the means of each sample are very similar. Even the outliers from the normal distribution do not great distinction between samples and even less so between the wildtype and nulls.

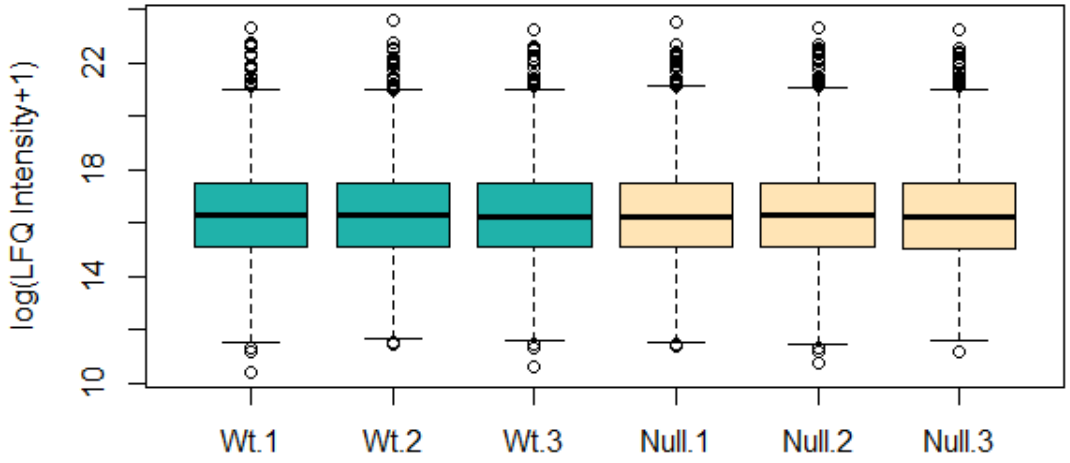


Fig.19. Box and whisker of protein profiles of each wildtype and null samples. The log transformed protein expression of each sample was plotted with 50% of the data centered around the median and the 1st and 2nd quartile being the lower and higher whiskers respectively. Observations plotted outside the whiskers are extended beyond the 5% and 95% points of normal distribution. Features of box and whisker plot are outlined in (Chambers *et al.* 1987). Wildtype samples denoted as Wt.1, Wt.2 and Wt.3. (Green) and eEF1A2 nulls as Null.1, Null.2 and Null.3 (Cream).

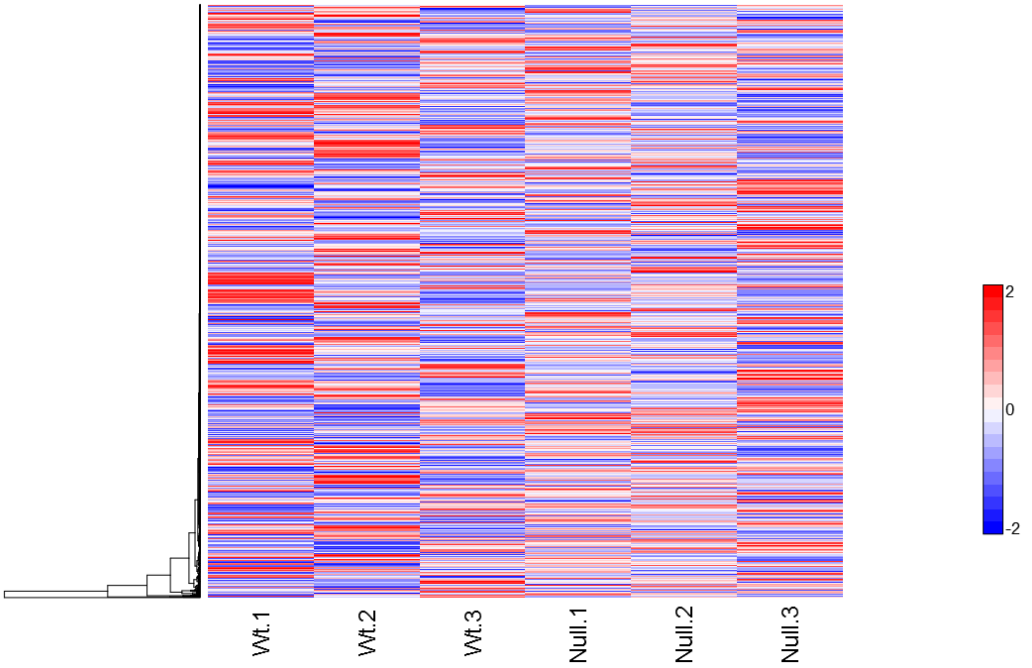


Fig.20. Heatmap of protein expression patterns. Abundance of individual proteins detected by mass spectrometry (rows) and samples Wt.1, Wt.2 and Wt.3. (Wild type) and Null.1, Null.2 and Null.3 (eEF1A2 null) expression profiles (columns) were organised by complete linkage hierarchical clustering Euclidean distance.

A further visualisation of the protein abundance across all samples to search for quantitative patterns is shown in Fig.20. A data-matrix heatmap graphically represents the numerical data by pseudo-colouring protein expression and arranging the proteins rows and sample columns so that similar profiles are closer and it is easier to interpret patterns and trends from the dataset. The resultant heatmap however shows no quantitative pattern in either individual protein expression nor the samples, the latter being expected as the cluster analysis revealed no conspicuous groupings. It is clear in the absence of distinct blocks of colour in the heatmap that the individual proteins share similar expression profiles and there is great variability across the samples.

To examine if proteins were, and the degree at which they were, overexpressed or under-expressed, relative changes in the protein abundance, the logarithmic fold change for each gene was calculated:

$$Log2\left|\frac{eEF1A2\ \text{LFQ intensity mean}}{\text{Wildtype}\ \ \text{LFQ intensity mean}}\right|$$

This is one of the simplest approaches to identify differentially expressed proteins is by measuring the fold change ratio of the protein between the sample means. The higher the fold change ratio the larger the difference in expression between groups. However, in its simplicity it is reduced to measuring the difference in expression at the surface value, it is limited in assessing the significance of a difference in the presence of biological and experimental variation. It has been shown to perform poorly when variability in the data is high (Murie *et al.* 2009). Therefore a simple fold change measure may not account for real differences, the resultant proteins deemed as differential based on the fold changes may contain a high rate of false positives. For this reason, t tests would also be a beneficial measure of significant differential expression. T tests use more robust central tendency and dispersion estimates to adjudge significant difference in expression, than relying on the use of means alone which can be affected by the degree of deviation.

T tests can measure the observed pairwise differences in individual proteins between replicated samples of wildtype and null and evaluate the confidence of these. The p-values for each protein across the groups were calculated using pooled variance, two-tailed t-test. This was appropriate for the dataset as it meets all the assumptions made for it; both means follow normal distribution and homoscedasticity. The p-values for each protein were plotted Fig.21 to illustrate how the test behaved across the data and whether the null hypothesis rejected (that there was no significant difference in gene expression between diseased and healthy groups) or accepted.

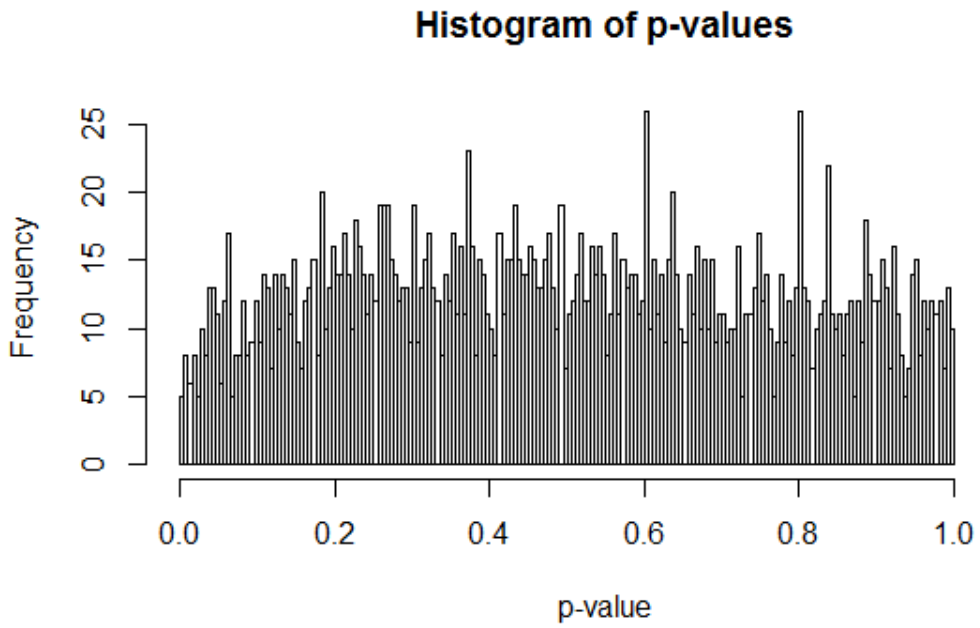


Fig.21. Histogram of the frequency of p-values. P-values calculated from two tailed t-test of protein expression levels acquired by LFQ-MS in wildtype and eEF1A2 mice.

Fig.21. depicts no statistically significant changes in protein abundance between the wildtype group and eEF1A2 null group. The uniform distribution of p-values indicate that the null hypothesis is accepted; the majority of the proteins are not present at different abundances between the groups. This was expected given the results emerging from the previous cluster analysis, heatmap and plots, there appears to be very little differential protein expression between the two groups. However, there may be specific proteins that are differentially expressed and these may identified by visualising both the fold changes and p-values in a volcano plot.

Any meaningful changes in the proteins between the spinal cords of wildtype mice and eEF1A2 null mice and the proportion of these in the dataset are represented in Fig.22.

Fig. 22. Volcano plot comparing gene expression between wildtype and eEF1A2 null mice. The Log2 Fold changes against –Log10 p-values of each genes expression between p21 wildtype and eEF1A2 null mice. Non-significant values are shown in grey. Proteins with fold changes >1.5 (Pink), p-values ≤0.05 (Blue). Proteins with fold changes >1.5 and) a p-value ≤0.05 (Green). Proteins with fold changes >-1.5 and) a p-value ≤0.05 (Yellow).

The threshold for a significant p value was set at 0.05 and a fold change of 1.5. The majority of the proteins were not significantly different between the two groups. However ~6% of proteins were significantly differentially expressed, the majority of which were upregulated in the Del.22.Ex.3 mice, 60% of the significantly differentially expressed proteins (see supplementary information). Six proteins experienced a fold change greater than 1.5, five of which were downregulated in eEF1A2 nulls (fibroblast growth factor 12 (*Fgf12*), Glutathione S-Transferase Kappa 1, Tubulin Folding Cofactor E Like, biquinol-Cytochrome C Reductase Hinge and Ubiquitin Like 4A and upregulated epoxide hydrolase 1. Such few changes are in line with the observations made prior to this analysis. Out of these proteins only F*g*f12 was calculated as statistically significant, and was present at a 2.4 fold lower level in the eEF1A2 nulls. Very few proteins were detected as differentially expressed between the wildtype and eEF2A2 nulls.

4.5 Gene Ontology analysis:

Proteins identified as significantly differentially expressed by the t tests conducted were then used in gene ontology enrichments analysis so as to explore if particular pathways and functions were over-represented. This would aid in understanding how the loss of eEF1A2 affects the proteome in spinal cords and perhaps pathogenesis of the MND developed in homozygous Del.22.Ex.3 mice. Gene ontology analysis was conducted on separate lists from the data analysis output; proteins that were significantly downregulated in eEF1A2 nulls and upregulated in eEF1A2 nulls (see supplementary information). The reason why proteins with a fold increase greater 1.5 were not included was because there were too few, making enrichment analysis redundant. Gene ontology enrichment was done using PANTHER (Protein ANalysis THrough Evolutionary Relationships). This is a classification system (Mi *et al.,* 2013) that was designed for proteins and their genes to facilitate high-throughput analysis. Proteins are classified according to:

- Molecular function: the function of the protein by itself or with directly interacting proteins at a biochemical level.
- Pathway: similar to biological process, but a pathway also explicitly specifies the relationships between the interacting molecules.

The molecular function enriched in list of proteins that were significantly downregulated in the spinal cords of the Del.22.Ex.3 mice is shown in Fig.23. The majority of proteins were enriched for binding when classified by their molecular function. Upon further inspection many of these proteins were in fact related to actin remodelling and transportation. The second most common enrichment term was catalytic activity, these included proteins that were involved in hydrolase, transferase and oxidoreductase activity among others. The proteins that were enriched for structural molecular activity were all

linked to actin remodelling. The transporter activity was compromised of proteins involved in transmembrane transportation.

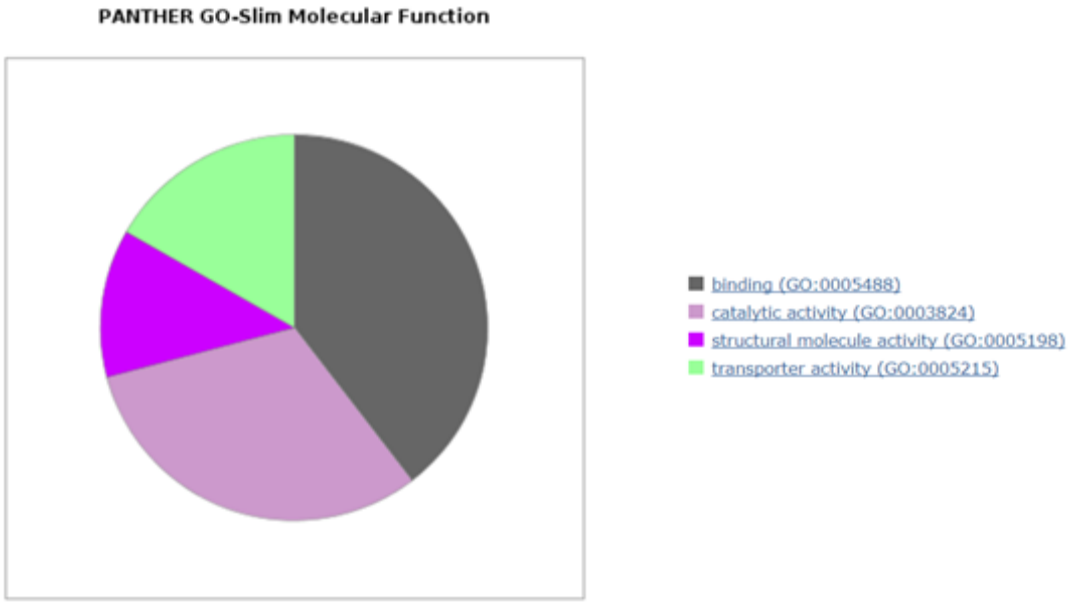


Fig.23. PANTHER molecular function enrichment analysis pie chart of proteins downregulated in eEF1A2 null mice. The molecular functions of proteins identified as downregulated in the spinal cords of Del.22.Ex.3 mice were enriched for binding, catalytic activity, structural molecule activity and transporter activity.

Fig.24 highlights pathways that were over-represented, many of which have been implicated in the development of neuronal degeneration, such as Dync1i2, a subunit of the cytoplasmic dynein complex functioning as a transporter for axonal transport, described further in Chapter 6. However, for each pathway identified, only one protein is included in the list of proteins enriched for these particular pathways. The downregulated proteins have been enriched for functions and pathways that have been linked to eEF1A2 non-canonical roles, as well as neuronal damage.

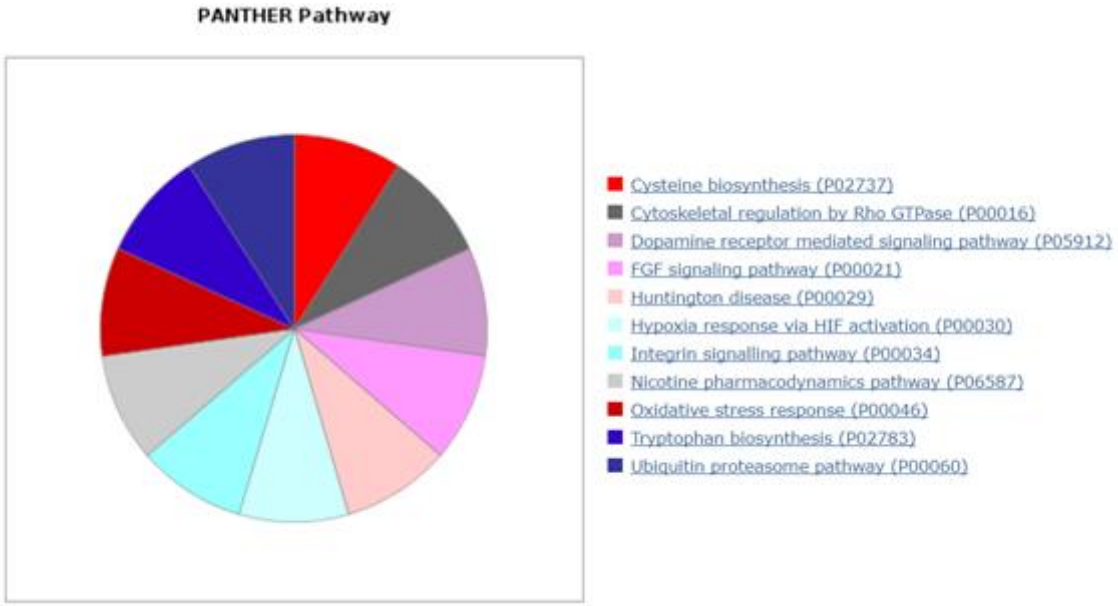


Fig.24. PANTHER Pathway enrichment analysis pie chart of proteins downregulated in eEF1A2 null mice. The proteins identified as downregulated in the spinal cords of Del.22.Ex.3 mice were involved in a range of pathways.

Proteins identified as differentially expressed by statistical analysis that were upregulated in spinal cords of Del.22.Ex.3 were used in Fig.25. Many of the molecular functions enriched for were shared with those of downregulated proteins. Protein binding made up the majority of the binding activity function but these proteins were enriched more so for G-proteins and GDP associated proteins. The catalytic activity category had the addition of ligase and lyase activity. Proteins enriched for structural molecule activity however were involved predominantly structural constituents of the ribosome, whereas the transporter activity proteins were also associated with transmembrane transportation. The identification of translation regulator activity is negligible as it is the differential expression of eEF1A that is responsible for this category being identified.

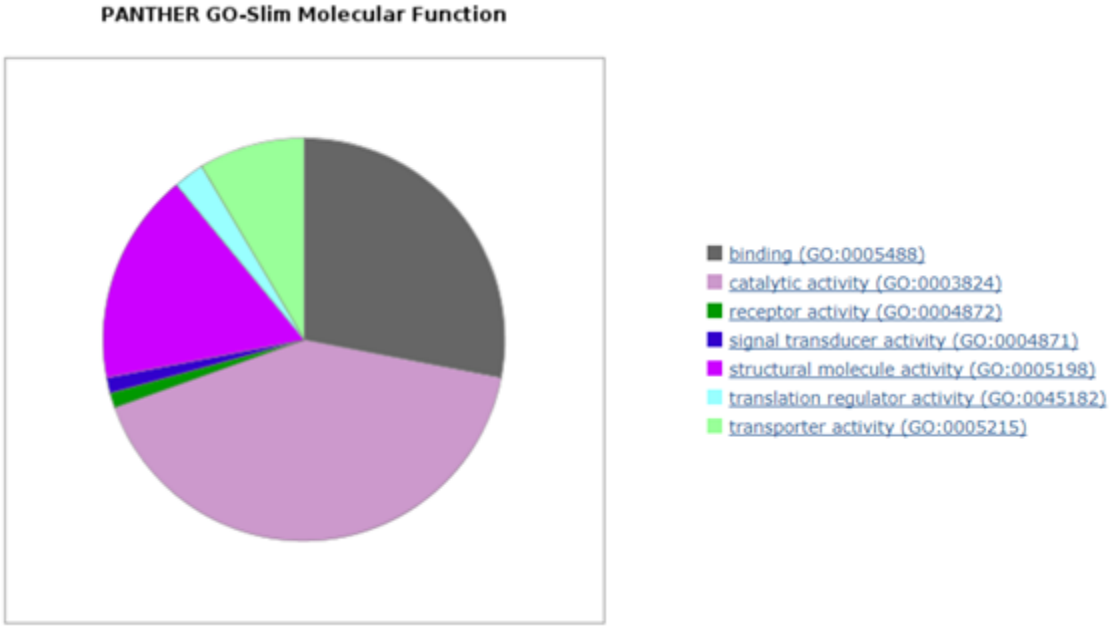


Fig.25. PANTHER molecular function enrichment analysis pie chart of proteins upregulated in eEF1A2 null mice. The molecular functions of proteins identified as upregulated in the spinal cords of Del.22.Ex.3 mice were enriched for binding, catalytic activity, receptor activity, signal transducer activity, structural molecule activity, translation regulator activity and transporter activity.

Many more pathways were identified from the proteins that were upregulated in eEF1A2 null mice, some of which have been identified as possible contributors to the pathogenesis of neuronal degeneration and the development of disease. These were Alzheimer's, Huntington's and Parkinson's. The proteins enriched for Alzheimer's and Huntington's disease pathways were predominantly related to actin. As described in Chapter 1.4, the eEF1A2 isoform has reduced cytoskeleton remodelling, so the upregulation of these genes may lead to abnormal actin bundling, which in neurons may also be linked with synaptic function (Cingolani and Goda, 2008). The Parkinson's pathway had also proteins related to actin with addition the protein Cytoplasmic FMR1-interacting protein 2 (Cyfip2), which has been connected with apoptosis as it is p53-inducible (Jackson *et al.,* 2007), it has also been observed as upregulated in patients suffering from fragile X syndrome (Hoeffer *et*

al., 2012). This has similar phenotypes to patients who carry heterozygous missense eEF1A2 mutations; intellectual disability and autism. The upregulation of this may in part be behind some of the neuronal cell death in the eEF1A2 null mice. Many of the other pathways enriched for included this protein as well as ADP-Ribosylation factors, these are GTP-binding factors some of which can act as regulators for intracellular traffic. Excessive expression of such protein can lead to neuronal damage (Skaper, 2003, Lai *et al.*, 2017), whilst inhibition of this has shown to improve axon regeneration (Byrne *et al.*, 2016). Therefore upregulation of ADP-Ribosylation factors may be responsible for the axonal degeneration seen in the *wasted* mice which too are eEF1A2 nulls.

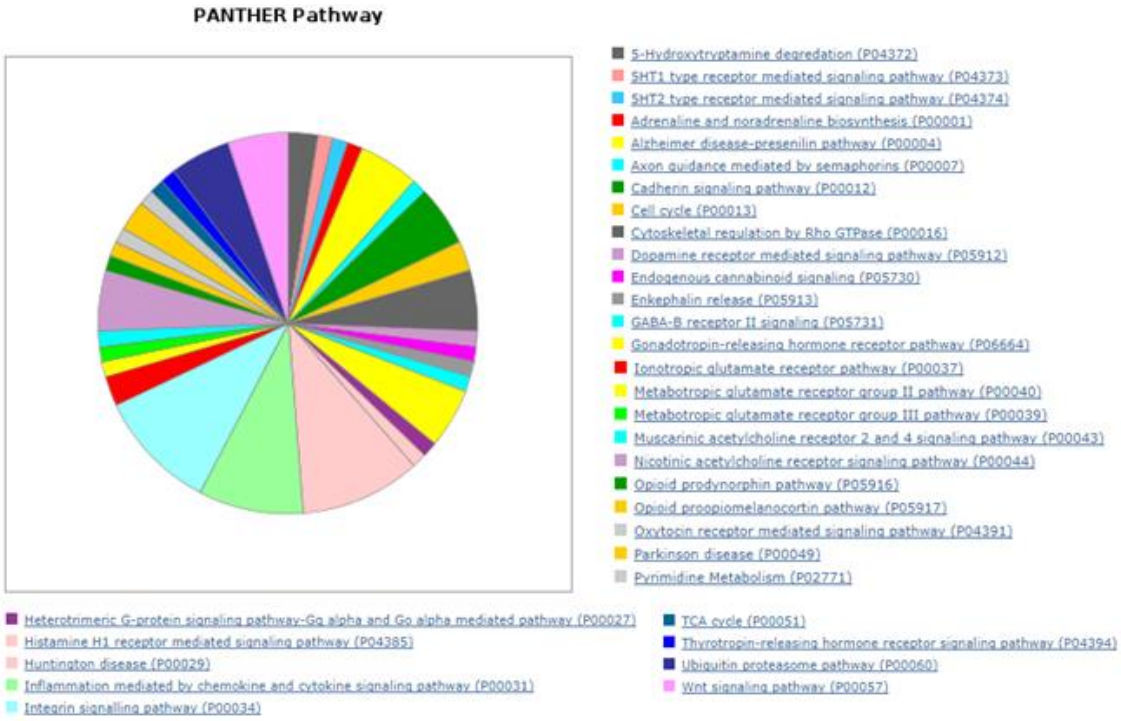


Fig.26. PANTHER Pathway enrichment analysis pie chart of proteins upregulated in eEF1A2 null mice. The proteins identified as upregulated in the spinal cords of Del.22.Ex.3 mice were involved in a range of pathways.

The gene ontology enrichment analysis has highlighted that the proteins both downregulated and upregulated are involved in binding, catalytic and transporter activity functions in the cell and many of these were actin related proteins, suggesting that cytoskeleton remodelling and the transportation may be the cause of pathogenesis of neuronal damage. Proteins upregulated in the eEF1A2 nulls are associated with pathways that when dysregulated cause neuronal damage. Hence the proteins identified as significantly differentially expressed between eEF1A2 null and wildtype mice have been demonstrated to be contributors to neuronal damage and speculated to be part of the pathogenesis of neuronal disorders in other studies. Similar pathology may be occurring the D252H eEF1A2 mutated mice, as the loss of eEF1A2 functioning in homozygotes will almost certainly lead them to develop neuronal damage.

4.6 Discussion

IHC probing for markers of neuronal degeneration by GFAP and H+E staining on the D252H and Del.22.Ex.3 lines established the extent of neurodegeneration whilst the label-free quantitative mass spectrometry and a range of statistical analysis defined the proteome of spinal cords from homozygotes and wild types for comparative study. This revealed there was little overall difference between them. However, some proteins were differentially expressed and may be potential biomarkers for MND as the gene ontology analysis showed that some are involved in neuronal pathways that, when dysregulated, are implicated in neuronal degeneration.

The H+E staining Colin Smith described and IHC evidenced neuronal degeneration by GFAP staining and that in both Del.22.Ex.3 and D252H homozygotes there was a greater presence of degeneration than the heterozygotes and wildtypes. Notably, degeneration in homozygotes appeared as a rostrocaudal gradient which agreed with the observations made by (Newbury *et al.* 2005), identifying a similar progression in *wasted* mice by measuring levels of damage at different ages. They suggested that the switch between isoforms works as a cascade; eEF1A1 is switched off progressively initiating at rostral regions. The spinal cords analysed were extracted at p21 when the switch between isoforms should be nearing completion (Khalyfa *et al.,* 2001). The D252H mice have a more severe phenotype than the Del.22.Ex.3 mice, as another student in the Abbott group identified D252H struggling to gain weight in comparison, however no seizures were recorded in the D252H whereas in some cases Del.22.Ex3 mice did experience them. This suggests that the D252H missense mutation causes a gain of function. The D252H mice still express eEF1A2 as a seemingly non-functional form impairing translation. Yet there is the possibility that the D252H mutation may have residual functioning but be greatly impaired. These speculations are based upon the conclusion that the mutation is loss of function, however this is contradicted by phenotypic observations on the mice (conducted by Laura Kaminioti-Dumont, an honours student in the lab group). The D252H homozygous mice showed higher phenotypic scores measuring the sum performance from four tests: ledge test, hindlimb clasping, gait and kyphosis, where a higher score represents impaired performance. As well as this they displayed a lack of weight gain from 15 days in contrast to homozygous Del.22.Ex.3 mice (representing a completely null phenotype) which continued to gain weight until 21 days. It is also worth noting that for ethical reasons the D252H mice are culled before we are able to observe the extent of their degeneration, so they may too, eventually develop seizures. Based upon these findings, the mutation would be gain of function, hence in addition to impaired translation, it may have gained other toxic functions.

The degree of neuronal damage as detected by GFAP was unprecedented in D252H heterozygotes especially as the H+E stain showed no neuronal damage. It contradicts entirely the phenotype witnessed in both lines; the heterozygote showed no signs of abnormality. Heterozygous *wasted* mice show no neurodegeneration in their phenotype nor pathology (Newbury *et al.* 2005), therefore the detection of GFAP in Del.22.Ex.3 heterozygotes (another eEF1A2 null) is highly unlikely. There was also detection of GFAP in the wildtypes. Considering the distribution of GFAP in the both the heterozygotes and wildtypes was irregular across all sections, it is more likely that there was an abundance of non-specific staining. This confounds the results from GFAP. The GFAP expression seen in neuronal tissues can be dependent on the abundance of astrocytes present (Garman, 2010), if there was a variation of these between the different genotypes, it may explain some of discrepancies. Nonetheless, the H+E stains identify unequivocally that mice homozygous for the mutations in eEF1A2 develop a rostrocaudal gradient of neurodegeneration. However further study in a larger cohort of mice from each genotype would confirm this and perhaps with statistical analysis of the degree of neurodegeneration.

The application of LFQ-MS is apt for quantifying the proteins within the spinal cords of Del.22.Ex.3 and wildtype mice. However, due to incomplete digestion the identified proteins and their respective abundances may not be representative of their presence and the differential expression detected between wildtype and eEF1A2 nulls may not be true changes of expression. However these proteins upon further investigation were revealed to have biological involvement in neuronal functioning and have been implicated in MND or other types of neuronal disorders (outlined in Chapter 6). The gene ontology analysis highlighted some of these proteins as enriched for pathways which dysregulation of has been implicated in neuronal damage, as well as molecular functions that are associated with previously studied non-canonical roles of eEF1A isoforms. This implies that the pathology of neuronal damage developed in the homozygous eEF1A2 null mice is in part due to dysregulation of actin related proteins and these may act as potential biomarkers for MND and neuronal death.

A major challenge of proteomic studies is the limitation in the number of replicates available, acquiring biological samples is dependent on many different factors such as availability of diseased samples and controls or accessibility of tissues for analysis. This is of particular importance in LFQ-MS as the separate treatment of samples can introduce variation, therefore replication is required to identify true changes. For this study a limiting factor was the availability of homozygous mice at p21 from the Del.22.Ex.3 line within the timeline of the project. However this was not a dominating issue, as the nature of the study was explorative, which has demonstrated little in proteome differences between the wildtype and eEF1A2 null mice spinal cords. The nature of acquiring the biological samples for experimentation also meant that the mice were unavailable for subsequent study of disease progression, although this does not present a problem for this

study as phenotype is homogeneous among Del.22.Ex.3 homozygotes. However, studying a larger cohort of homozygote Del.22.Ex.3 mice will be necessary for identifying truly differential expression and ensuring the potential biomarkers are reproducible. One of the greatest limitations to LFQ-MS is its inefficiency in detecting proteins with low abundances. The proteome of a sample may not be identified in its entirety and as of yet it is most likely that proteomics studies fail to do this. This issue means not all proteins that were present in the samples were included in the output of proteins detected, hindering in the discovery of biomarkers. However given that for a biomarker to be practical it must be reproducible, proteins with low levels may not necessarily be promising. Especially if the biomarker is for prognostic use in model organisms as it may not always be possible to collect enough tissue for low abundance proteins to be measured. However, there are other methods that can support low level protein detection, yet these are more targeted and biased to proteins suspected of being differentially expressed. As this study was explorative by nature this could not have been done.

Additionally, LFQ-MS is constrained by the potential to misidentify proteins. This can be done at different stages of analysis, including initially from the resultant spectrum from the mass spectrometer which can have a great deal of noise and unexplained peaks. These can often be attributed to fragmentation events where small molecular groups are separated from peptides during the fragmentation stage, or these may be contaminants that enter the mass spectrometer (Noble and MacCoss, 2012). The noise detected is often reduced by using technical replicates so as to reduce its influence in downstream analysis. Other issues that are to be dealt with are the background subtraction, outlier detection and signal distribution normalization (Sellers and Miecznikowski, 2010). Assigning the peaks generated can also be challenging, the various software available for analysing raw data such as Progenesis, Peaks Q and MaxQuant can differ modestly in feature detection. Feature detection is the detection and quantification of peaks in the spectral data, a crucial stage which can cause perpetuated misidentifications, false positives, false negatives and missing values. There has been a range of software and algorithms developed to combat the dilemmas and confusion of assigning homogeneous peptides to individual proteins. MaxQuant performs admirably among these, yet the variation in protein identifications is dependent on which software is implemented. Isoforms also present a challenge to protein identification, as isoforms can share sequence homology meaning peptide fragments may not be confidently assigned to the correct isoforms, hence the abundance of proteins with isoforms must not be assumed to be belonging to the specific isoform necessarily. Hence proteins identified must not be assumed to be true nor their respective abundances without validation.

As proteomic studies are relatively new, there is yet no definite standards of practice, or unequivocally better methods for data pre-processing and analysis the large datasets generated. A stage at which proteomics data processing can vary, generating differing outcomes is the imputation of missing values. A range of algorithms have been developed

so as to deal with missing values making downstream data analysis feasible. The imputation method applied; assuming that the proteins are present but below the level of detection, hence assigned the lowest detected value of that sample. There have not been many comparative studies in the assessment of imputation methods in LFQ datasets but those that have been carried out have identified no considerably superior method of imputation in this or in other types of mass spectrometry experiments (Webb-Roberston *et al.,* 2015, Miecznikowski *et al.,* 2010, Karpievitch *et al.,* 2012, Sandra *et al.,* 2017). The single value method chosen does lead to a left hand skew of the data but this did not influence the downstream analysis as the overall distribution was normal. However, it may introduce more false positives, which is of particular interest in this analysis as the proper corrections have not been placed and the proteins identified as differentially expressed have not been validated thoroughly. Alternatives that could have been used that do not rely heavily on assumptions are the Random Forest imputation, K nearest neighbours and local least squares. These would have instead integrated information from across the sample to impute the missing values and preserve better the variance and distribution of the dataset, in addition to reducing the amount of false positives by not increasing the differences between the groups. As this is caused by the imputation method utilized it will mean that the difference between the groups, which is already not conspicuous, may have even less differentially expressed proteins and that the findings must be taken with caution.

There is an assortment of statistical tests and variations of these alongside arbitrary cut-off values for judging significance. Hence, the same data can be interpreted multifariously, with different conclusions being drawn on which proteins are differentially expressed. Just as a p-value of <0.05 is conventional to use despite being an arbitrary value, fold change cut-offs appear to be set at 1.5 or 2 for proteomics studies (Ting *et al.,* 2009). The lower value of 1.5 is less stringent and allows for more potential proteins to be identified as being differentially expressed. A limitation of measuring differential expression by fold change is the assumption that all proteins have the same variance, which is not necessarily the case. As mass spectrometry can struggle to identify proteins bordering the level of detection, low abundance proteins have great variability. The implications of a fold change at a protein level can vary depending on its abundance, fold changes in lowly expressed protein may not have a great impact upon the proteins function within the cell, whereas proteins that are highly abundant may alter the dynamics of the cell even if the fold change is observed to be identical to that of the lowly expressed protein. Therefore pursuing the biological significance of proteins is required to inform whether the proteins with fold changes have consequences that alter the cells normal function, or for the case of biomarker discovery, act as an indicator of disease and/or disease progression.

One of the major concerns with the analysis conducted is the lack of control for false positives. Some of the differentially expressed proteins have a chance of failing and being false statistically significant changes. Multiple testing corrections are necessary to adjust

p-values to account for the occurrence of false positives as t test analysis assesses the significance of each protein independent of one another. There are many corrections that can attend to this issue and reduce the amount of false findings such as the Bonferroni or Hommel corrections. However these are considered quite stringent for large datasets and the false discovery rate (FDR) approach is usually more suitable for the dataset resulting from LFQ-MS experiments. This would have functioned by identifying false discoveries (here a discovery meaning a statistically significant finding) the level of which deemed acceptable set usually at 5%. Therefore any significant finding would have a 5% chance of being a false positive. This approach differs from others as it doesn't start with the assumption that no differences exist between the groups. The application of FDR correction would have reported proteins that more likely to be truly significant and would have been most suitable given the size of the dataset and low number of biological replicates, as the rate of false discoveries may have been reduced with more replicates to an extent. However, the distribution of the p-value from the resultant t-tests (Fig.21) being uniform would most likely have led to larger FDR values. This would have meant that the significant findings are in fact false positives. In addition to including more replicates to identify false positives, increasing the p-value cutoff would have reduced the stringency of the tests as 0.05 is an arbitrary cutoff value that is predominantly established by common practise, however this would not necessarily mean that the differences are significant and the best method to examine significance would be through validation steps and increasing the cohort of mice included in analysis by repeating LFQ-MS with more samples.

As differences between the two groups have failed to manifest with great certainty the proteins that have been identified as of interest and potential markers must be taken with caution as the rigorous analysis into their differential expression has not been conducted. The small sample size also means there is low statistical power for identifying differential expression. Although the students t test is the most common approach when assessing differences between two groups the develops small sample size error estimation methods (Murie *e al.,* 2009). The students t test power is reduced also in its underlying assumption that each observation is independent of one another, which decreases the overall measurement of variance causing an overestimation of statistical significance as the samples are in replicate (Ting *et al.,* 2009). Other significance tests that may have performed better is the Empirical Bayes test which outperforms other tests in identifying true differences and reduced number of false positives overall (Murie *e al.,* 2009). However variations in tests for statistical significance have shown to generate differing outcomes (Ting *et al.,* 2009). A form of an Empirical Bayes test with appropriate corrections may have led to the discovery of truly significant differences within the dataset but even with this change the lack of standardization in data analysis causes each proteomic study to generate results based on individual interpretation. Although studies are attempting to modulate the approach to analysing significantly differential expression, validation of results is a crucial step in determining true discoveries.

It would be interesting to attempt different methods of analysis to compare how the data performs. However this would be better suited to a more complete dataset given the experimental issues. The sample preparation appeared to not be harsh enough to break down the proteins, a revised protocol would be necessary to ensure such an issue does not arise again.

In the expression analysis the lack of distinction between the groups, despite the eEF1A2 nulls being translationally impaired may be compounded by the ages of the mice; at p21 the switch between eEF1A1 and eEF1A2 is only just being completed, meaning there may still be functional eEF1A1 proteins maintaining regular translation. However if the study was to be conducted at any later time points it would be confounded by the loss of translation. The proteins detected and their relative abundances would become dependent in part on their stability and turnover rates once translations is reduced drastically. The clustering of the samples was not found to be due to other factors such as sex whilst their ages are consistent and all wildtypes and Del.22.Ex.3 were from the same litter. To confirm this later stage westerns may be conducted with later stage mice where it would normalise to other proteins.

The analysis into the pathology revealed that homozygotes have a great degree of neuronal damage in their spinal cords that manifests as a rostrocaudal gradient whereas the wildtype and heterozygous does not show such neuronal damage. The effects of the mutations when present as homozygotes in mice results in non-functional proteins that cause a severe phenotype. Whereas heterozygotes do no not appear to manifest this phenotype as there is no haploinsufficiency. However these observations were made on a relatively small sample size, therefore it be beneficial to use a larger cohort of mice.
If there was more time the proteomics experiment would require a repeat because of the issues with incomplete digestion, along with including more biological replicates to increase statistical confidence and utilize the appropriate statistical tests to reduce false discoveries. It will also be important to validate the most differentially expressed proteins that also have extensive literature supporting an association with MND, so as to confidently state their potential as a biomarker for prognosis of MND in the development of therapies.

Chapter 5: RNA expression profiles of wasted mice and wild type mice.

The transcriptional activity in a biological sample can be assessed by microarray experiments that allow for genome-wide expression changes between healthy and diseased groups to be measured. This is an unbiased, systematic approach that facilitates the discovery of novel functional roles of genes, potential diagnostic or prognostic biomarkers and implicate potential pathways and mechanisms that cause disease development.

Microarrays are microscopic slides that consist of a large range of 'spots' that contain multiple identical strands of DNA from the same gene. Genes from the experimental samples are converted into complementary DNA (cDNA) with fluorescent tags, these hybridize to their matching sequences on the spots. The strength of the binding will depend upon how complementary the cDNA is to gene sequences on each spot. The tags inform on the levels of hybridization in comparison to a reference sample. These spots are arranged in a particular order to allow for detection of genes that are being expressed (Scitable by nature, 2017).

Having described the translational activity of eEF1A2 null mice and identified proteins expressed differentially as potential markers of motor neuron degeneration, it would be intriguing to see if these changes are reflected at the genomic level, which would strengthen confidence in the conclusions derived from the proteomic study.

The correlation between RNA and protein expression can be influenced by many elements. One of the most obvious, and in this study overtly present, are factors that impact translational activity which would cause discrepancies between RNA and protein expression detected. Alongside translational efficiency that vary between RNA. Samples collected at p21 meant that the effects being seen would not as of yet be affected drastically by the lack of translation (as eEF1A1 expression is abated) but that the levels of protein observed would have become dependent on their respective turnover rates and stability. Microarray data would not be confounded by this as the RNA levels are not affected.

The microarray experiment and subsequent data pre-processing, as well as significance analysis of microarray (SAM) generating a list of 500 genes with significant differences in expression was conducted by Andy Sims, a previous member of the Abbott lab. Further data analysis as described in this chapter was done so as to identify genes of interest and compare RNA expression with protein.

5.1 Biological specimens.

Spinal cords from six wild type mice and six homozygous *wasted* mice underwent microarray analysis. Taken at p21, the timepoint in which severe phenotypes develop and eEF1A2 expression is enhanced, whilst eEF1A1 is reduced to only trace amounts in neurons but expressed in glial cells. Samples were collected at the same time-point (p21) and tissue as for proteomic analysis for appropriate comparison.

5.2 Expression profiles.

To visualise whether overall gene expression was distinctive for each group, a cluster analysis was done by complete linkage method for hierarchal clustering calculating the pairwise distances between samples based on the gene expression profiles of each sample

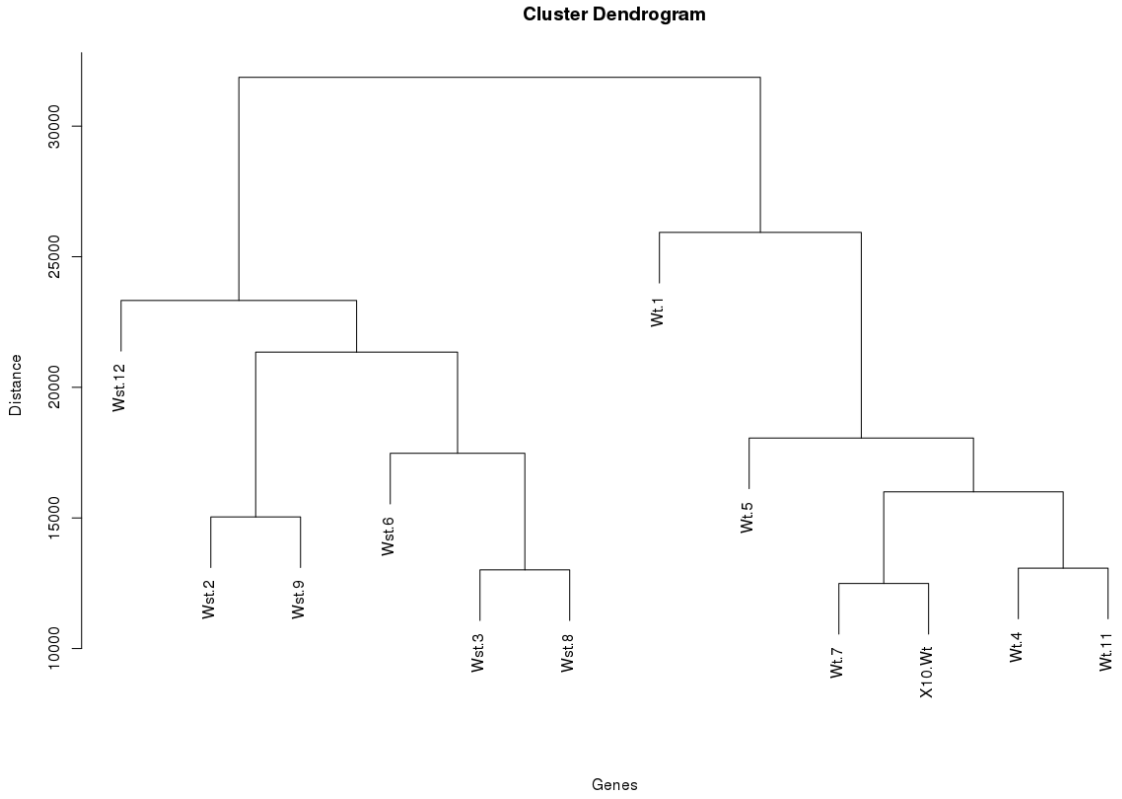


Fig.27: Cluster dendrogram of wasted and wildtype mice gene expression profiles. Complete linkage for hierarchal clustering for Wild type mice samples denoted as Wt.1, Wt.4, Wt.5, Wt.7, Wt.X10 and Wt.11), whilst *wasted* mice with eEF1A2 null mutation are denoted as Wst.2, Wst.3, Wst.6, Wst.8, Wst.9 and Wst.12. Generated in R program. Samples are represented on the X axis and are connected by decreasing similarity along the Y axis.

Fig.27. Unlike in the proteomic analysis, the wildtype and *wasted* form distinguishing clusters indicating that eEF1A2 loss alters overall gene expression in spinal cords. Samples Wt.1 and Wst.12 appear to have a closer gene expression pattern between the two groups but are nonetheless distinctly separate. A possible reason for the wildtype and eEF1A2 nulls forming separate clusters in the microarray data and not mass spectrometry could be resultant of experimental differences and not reflective biological factors at play such as impaired translation.

The heatmap in Fig.28 shows the relative expression of a random subset of 6000 genes out of >30000 in the microarray. It further clarified that the gene expression between wildtype and *wasted* samples differ greatly and be easily distinguished from one another. The *wasted* mice show a distinctively reduced overall expression when compared to the wildtype. There is also no clear set of genes that are conspicuously upregulated or downregulated, however this is only a subset chosen at random so patterns in gene expression may not have manifested.

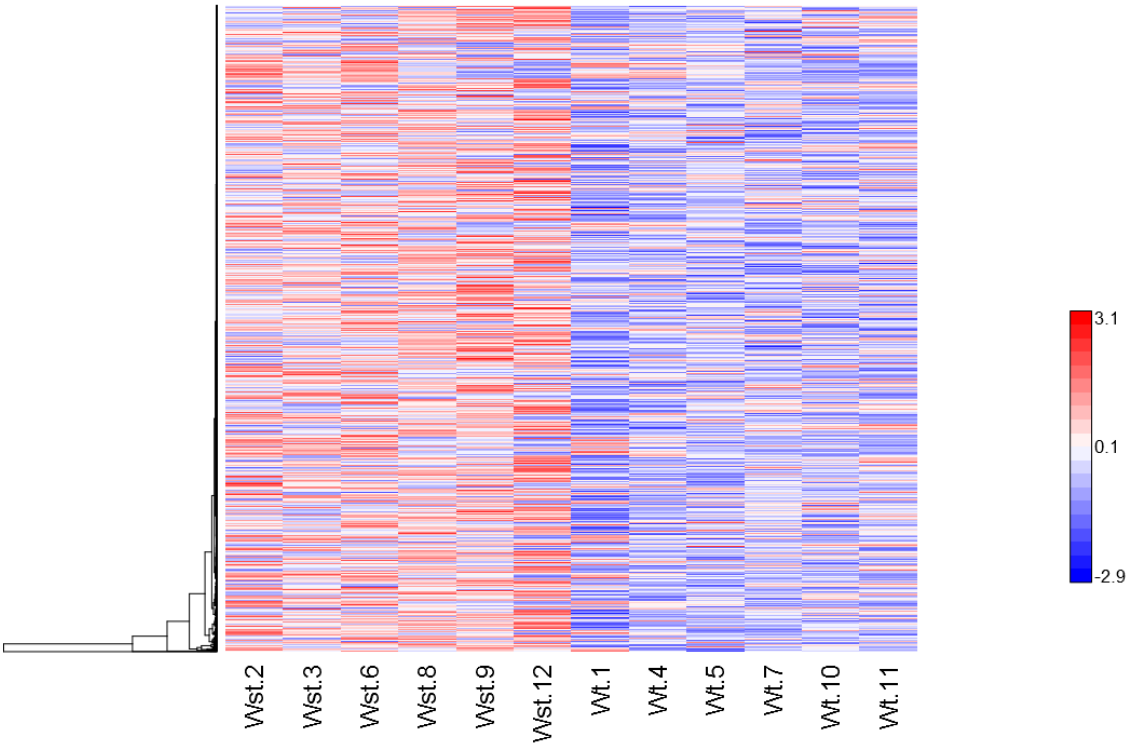


Fig.28. Heatmap of gene expression patterns. Expression of individual genes from microarray experiment (rows) and samples Wt.1, Wt.4, Wt.5, Wt.7, Wt.X10 and Wt.11. (Wild type) and Wst.2, Wst.3, Wst.6, Wst.8, Wst.9 and Wst.12. (*wasted*) expression profiles (columns) were organised by complete linkage hierarchical clustering Euclidean distance.

The p-values for each gene were calculated by two-tailed t-test and the resulting histogram for frequency of p-values distributed between 0-1 was plotted to illustrate how the test behaved across the data and whether the null hypothesis rejected (that there was no significant difference in gene expression between diseased and healthy groups) or accepted (Fig.29). The anti-conservative histogram demonstrated that the alternative hypothesis was accepted for a proportion of the genes. The following right hand uniform

distribution indicates many of the genes were not differentially expressed between the two groups but the protruding left hand peaks are confirmation that some are.

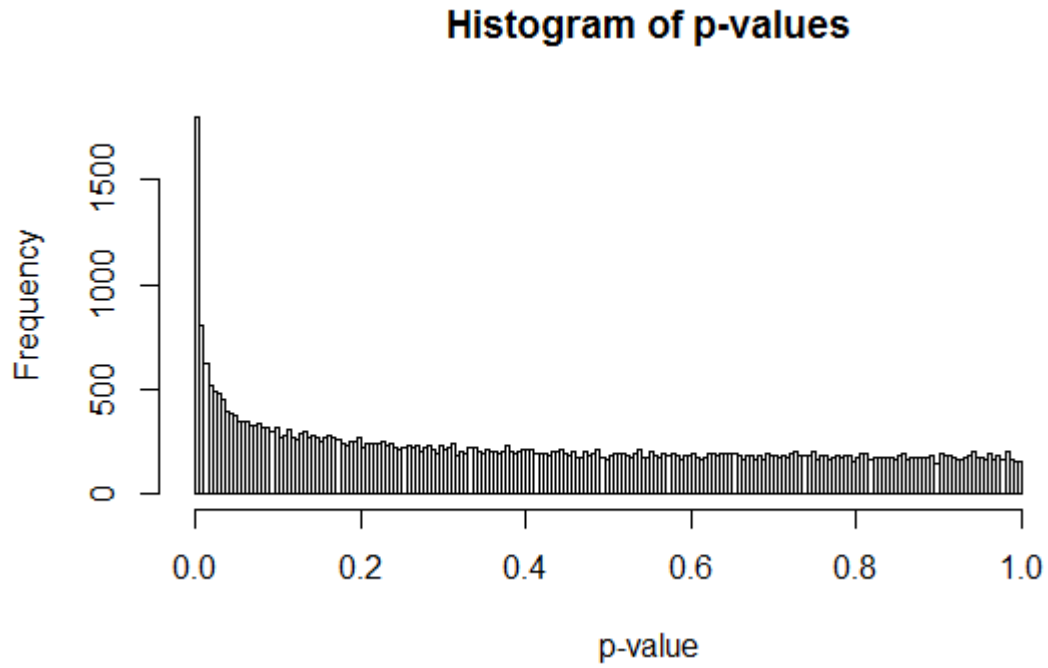


Fig.29. Histogram of the frequency of p-values. P-values calculated from two tailed t-test of gene expression levels acquired by microarray in wildtype and *wasted* mice.

To examine genes that were, overexpressed or under-expressed relative changes in the RNA abundance, the logarithmic fold change for each gene was calculated:

$$Log2\left|\frac{wasted\ \text{RNA mean}}{\text{Wildtype RNA mean}}\right|$$

As mentioned previously in Chapter 4, the fold change only functions well if the variance of the data is not high. The variance of the *wasted* means and wildtype means individually have very high variability, however when comparing their variability with one another, they appeared to share similar variance with the ratio ~1.

Fig. 30. Volcano plot comparing gene expression between wildtype and wasted. The Log2 Fold changes against –Log10 p-values of each genes expression between p21 wildtype and *wasted* mice. Non-significant values are shown in grey. Genes with fold changes >1.5 (Pink), p-values ≤0.05 (Blue). Genes with fold changes >1.5 and) a p-value ≤0.05 (Green). Genes with fold changes >-1.5 and) a p-value ≤0.05 (Yellow).

The volcano plot visualises any meaningful changes between the gene expressions in the spinal cords of wildtype mice with wasted mice and can quickly show the proportion of genes identified as either significant and/or experience a fold change. The threshold for a significant p value was set at 0.05. What is more argued is the threshold set for what is deemed as a consequential fold change. For the purposes of this study it was set at a ratio change of 1.5.

A great number of genes were identified as significantly differentially expressed, 4573, however this made up only 10% of the total genes measured. Out of these, three genes showed to have a greater fold change of 1.5; three of the points measuring to have fold changes are all *Eef1a2*, Neuronal acetylcholine receptor subunit alpha-4 (Chrna4) and ChaC glutathione specific gamma-glutamylcyclotransferase 1 (Chac1). Additionally, Transthyretin (Ttr) had a 2.2 fold increased abundance in the *wasted* mice but was not observed to be statistically significant in regards to its p-value. Assessing differentially expressed genes based on p-value is acceptable, however another computational tool has been developed specifically for microarrays large data output: SAM. SAM identified genes whose expression is significantly related to the response variable, diseased and healthy. It tackles the multiple testing problem that occurs with t tests and works upon the q values as a measure of significance. It functions by computing a statistic for each gene measuring the strength of the relationship between the genes expression and the response variable permuting these to determine if relationship is significant. SAM imputes any

missing values by a K-Nearest Neighbour algorithm normalization. A list of 500 genes (See supplementary information) was output from SAM as possessing this significance. The genes described as differentially expressed in SAM identified 132 genes that were not deemed of significance by p-value analysis alone (Fig.29).

5.3 Gene Ontology analysis.

The genes that were listed as differentially expressed by SAM were underwent gene ontology analysis using the PANTHER analysis tool. This highlighted the molecular functions and pathways enriched for in the eEF1A2 null mice by either upregulation or downregulation of genes in the spinal cords. Exploring the molecular functions and pathways by this analysis may also insinuate pathogenesis in the mice. Genes that showed a fold change greater than 1.5 were not included as they were too few to search for functional and pathway enrichment terms.

The molecular function enriched in list of genes that were significantly downregulated in the spinal cords of *wasted* mice is shown in Fig.31. The majority of genes were enriched for catalytic activity when classified by their molecular function, which included genes that were involved in hydrolase, transferase predominantly as well as enzyme regulation and oxidoreductase activity among others. The second most enriched for term was binding, which predominantly encompassed genes associated with protein and nucleic acid binding. Genes enriched for calcium ion binding were also included these were calmodulin binding proteins. Structural activity terms were enriched for and genes associated with actin binding were downregulated in the *wasted* mice. In addition to genes that are associated with neuronal receptors as well as signal transduction and transporter activity.

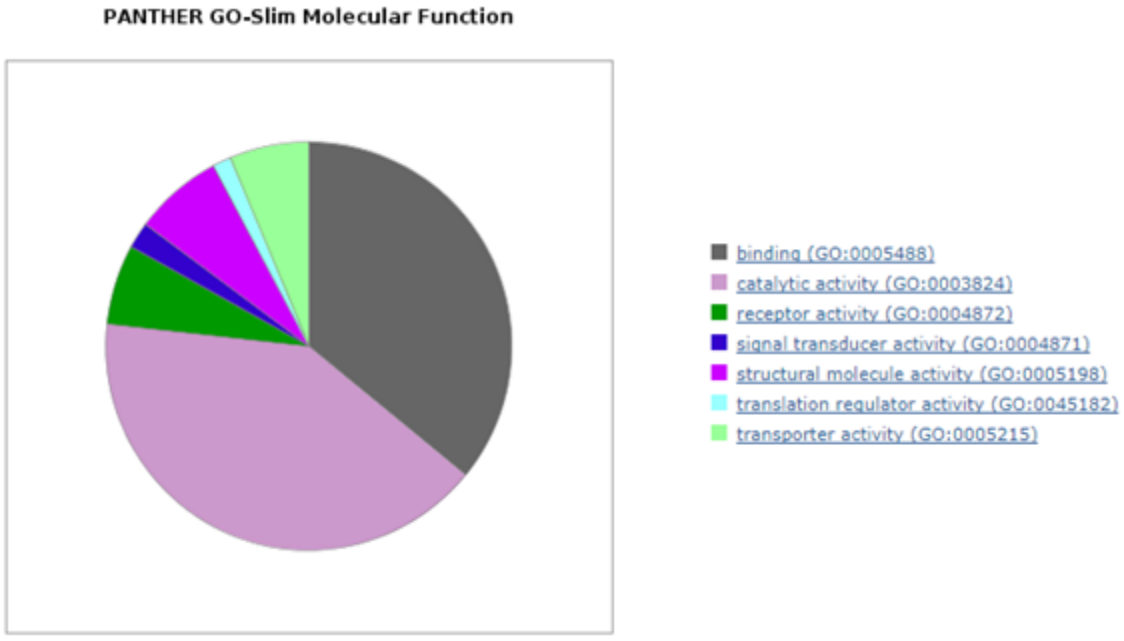


Fig.31. PANTHER molecular function enrichment analysis pie chart of genes downregulated in eEF1A2 null mice. The molecular functions of genes identified as downregulated in the spinal cords of *wasted* mice were enriched for catalytic activity, binding, structural molecule activity, receptor activity, transporter activity, signal transducer activity and translation regulator activity.

The pathways that were over-represented in genes significantly downregulated in *wasted* mice are identified in Fig.32. Of these, the greatest number of genes were enriched for Integrin and Wnt signalling. Genes involved in the Integrin signalling pathway were associated with structural and G-proteins. There were many other pathways that were enriched for and the same genes were involved in multiple pathways, for example the structural genes found in the integrin pathway also appeared as Huntington disease. Across the pathways many of genes involved were part of the Rho family of GTPases, which play a role in cytoskeletal dynamics.

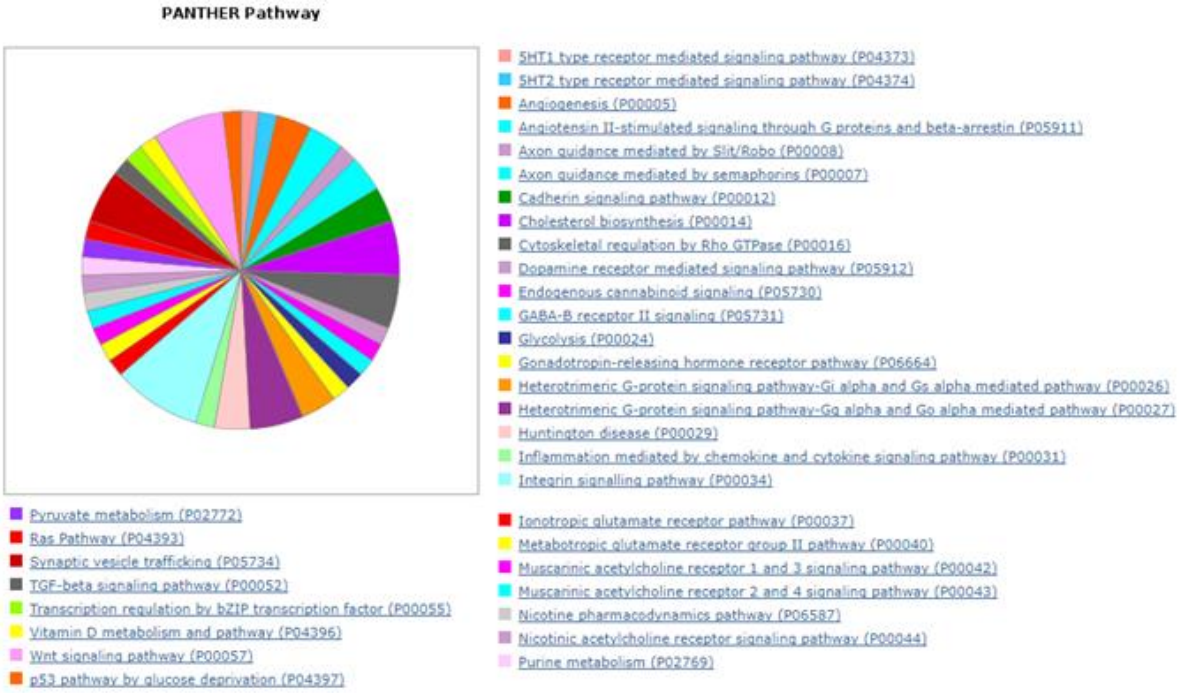


Fig.32. PANTHER Pathway enrichment analysis pie chart of genes downregulated in eEF1A2 null mice. The genes identified as downregulated in the spinal cords of *wasted* mice were involved in a range of pathways.

Genes differentially expressed by SAM that were upregulated in the spinal cords of *wasted* mice were used in Fig.33 to identify the molecular functions that were enriched. Many of these were shared with those of downregulated proteins, however more were enriched for catalytic activity in upregulated proteins.

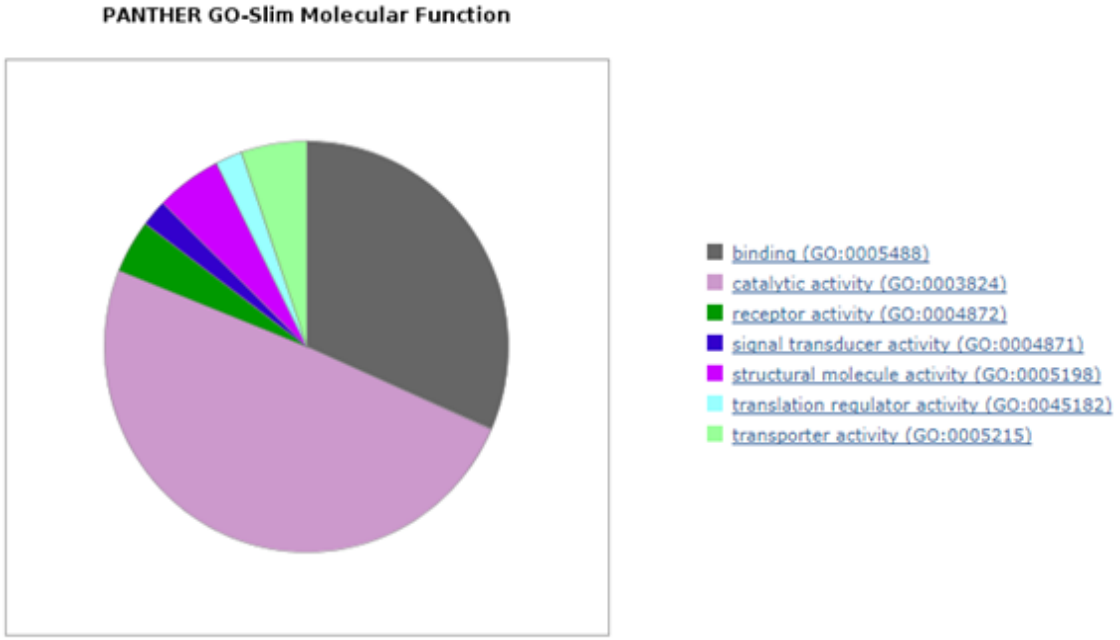


Fig.33. PANTHER molecular function enrichment analysis pie chart of proteins upregulated in eEF1A2 null mice. The molecular functions of proteins identified as upregulated in the spinal cords of *wasted* mice were enriched for binding, catalytic activity, receptor activity, signal transducer activity, structural molecule activity, translation regulator activity and transporter activity.

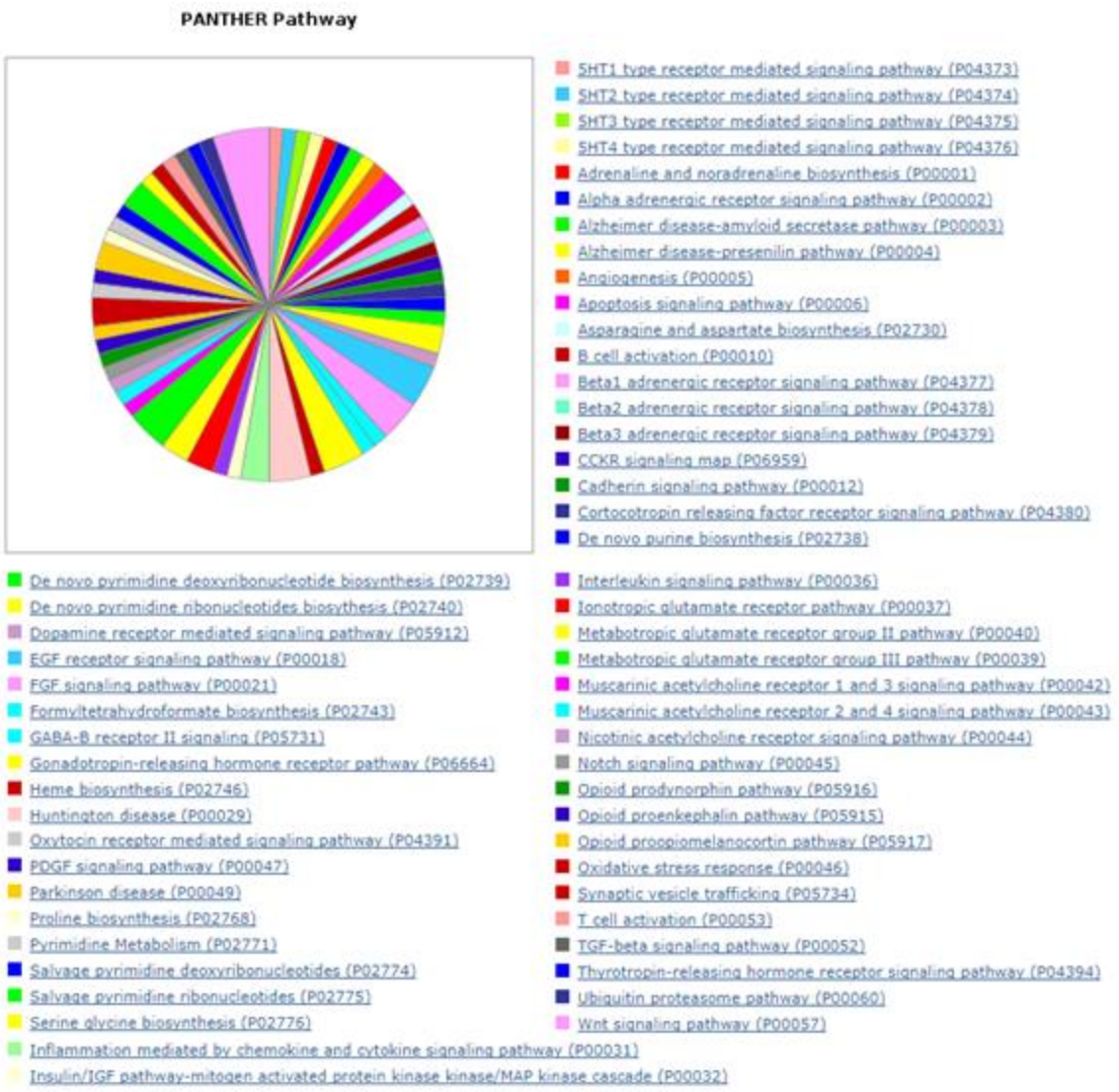


Fig.34. PANTHER Pathway enrichment analysis pie chart of genes upregulated in eEF1A2 null mice. The genes identified as upregulated in the spinal cords of *wasted* mice were involved in a range of pathways.

There were far more pathways enriched for in the upregulated genes of *wasted* than there were in the downregulated genes, suggesting genes upregulated had more diverse functions. Although there were genes represented in multiple pathways there was still more diversity among the upregulated genes. Similar to the pathways for downregulated genes, the Rho family of GTPases was represented in the upregulated genes too. Some of these were involved in neuronal diseases such as Alzheimer's, Huntington's and Parkinson's. The majority of the genes that were included in these pathways were also seen to be upregulated in these diseases upon examination and associated with the stress response. One as such was the heat shock protein family A (Hsp70) member 9 (Hspa9) gene, which part of the heat shock stress response primarily localized in the mitochondria

(Refseq, 2010). In addition to this the oxidative stress response pathway was also enriched for and the genes involved were not the same as those identified in the neuronal disease pathways. The FGF signaling pathway was also represented, this has been linked to early-onset epilepsy (Guella *et al.,* 2016, Al-Mehmadi *et al.,* 2016), a symptom that is also present in some of the cases with eEF1A2 mutations (Table.1). The GO analysis on the pathways enriched for in significantly upregulated genes found that particular ones that may contribute to the development of the aforementioned diseases that can share pathology with MND, however the oxidative stress response is also upregulated may be a response to the neuronal damage occurring in the *wasted* mice.

GO enrichment analysis of significantly differentially expressed genes of those both downregulated and upregulated are involved in predominantly catalytic and binding activity. Examining the molecular function of these genes revealed many associated with actin binding and calcium ion binding in both downregulated and upregulated genes.This is of note as a non-canonical role of eEF1A isoforms has been demonstrated to remodel cytoskeleton. Upon dimerization the isoforms can bind to F-actin resulting in actin bundling. Yet the F-actin is loosened when eEF1A isoforms bind to Ca^{2+}/calmodulin (Bunei *et al.,* 2006). The pathways that these gene lists may be involved in the development of neuronal damage when dysregulated. The genes that are differentially expressed between *wasted* and wildtype mice have shown connections to neuronal degeneration and the non-canonical roles of eEF1A2.

5.4 Discussion

The microarray results have identified many differentially expressed genes through significance analysis and few more by observing genes with fold changes. Given that the dataset analysis demonstrates good quality and that there are clear differences in overall gene expression between the spinal cords wildtypes and *wasted* mice at p21. Both the hierarchal cluster analysis and heat map clearly discern each group. For class comparison studies as such, the number of biological replicates is deemed appropriate for explorative purposes. Six samples per group is sufficient enough to distinguish true differences between the conditions. It also facilitates meaningful permutation tests such as SAM. The technical replicates of particular genes on the array itself demonstrated little variance between them. However the addition of more samples would increase the statistical power of the analysis.

The thresholds chosen were the same as those applied in the analysis of the LFQ-MS data in Chapter 4.4.1. Just as in proteomic studies the thresholds are set as deemed appropriate for the dataset. In this case as it is exploring for possible biomarkers, a lower threshold can reveal several results that may be promising. Another study identifies that using this

less stringent criteria was beneficial and found more GO terms associated with the diseased group (Dalman *et al*., 2012), but this being said validation of these results will also be key in identifying significantly differentially expressed genes. As mentioned previously probing for differential expression by using fold changes is limited when assessing significance in datasets with high variability (Murie *et al.,* 2009), which is the case in the microarray dataset. Hence why SAM was utilized. It overcomes the multiple testing problem and copes well with the large datasets output from microarray experiments and controls for false discoveries. However, there have been critiques as to the performance of SAM due to its FDR estimations (Zhang, 2009, Jeanmougin *et al.,* 2010). The FDR is estimated by analysing the permutations of the genes, which has a tendency to overestimate the level of FDRs (Hirakawa *et al.* 2007). Its bias lies in the lack of strict assumptions made upon the data, it has also been critiqued for overestimating the number of false positives in by estimating the distribution of non-differentially expressed genes as more dispersed based on the permutations than the data truly is (Hirakawa *et al.* 2007). So although SAM can cope with analysis of larger datasets it still struggles to impose appropriate corrections. Other studies have proposed alternatives to SAM (Jeanmougin *et al.,* 2010), but the majority of microarray analysis to identify differentially expressed genes is conducted using SAM.

The GO analysis demonstrated that some of the differentially expressed genes in the *wasted* mice were associated with neuronal degeneration. There is a possibility that the reason for these genes upregulation is activation of pathways that protect against the damages developing. The Wnt signalling that was enriched for in the pathways has been linked to a neuroprotective role in Alzheimer's disease (Inestrosa and Toledo, 2008). In addition, genes for the cell stress response were upregulated, this is more suggestive of a response to the neuronal damage and not dysregulation of it necessarily.

The microarray data revealed many significantly differentially expressed genes between the *wasted* and wildtype mice, which led to strongly discriminate genetic profiles (Fig.27 and Fig.28). However differentially expressed genes would still be required to be validated by western blot or ELISA, especially considering SAMs critiques on FDR correction. Some of these genes revealed themselves to be associated with neuronal damage, yet whether these changes are resultant of pathogenesis or the cell response is uncertain.

Chapter 6: Proteins and genes in the context of MND.

This data-driven search for differentially expressed proteins has identified a few proteins of interest. The proteomic study identified differences in abundance between the eEF1A2 nulls and wildtypes, which have been implicated in the development of motor neuron degeneration. Fgf12, Dync1i2, Kif5c and Dnajc5 have presented themselves as of biological significance. They all were detected as downregulated in the eEF1A2 nulls except for Fgf12 which was upregulated. Although these changes were significant at a protein level they were not all seen to change as significantly at a genomic level from the microarray data, it was only Kif5c that showed significance in both datasets. However the functions and pathways of these select proteins were heavily represented in GO analysis in the microarray data. The proteins were chosen based on their differential expression, their functional role being connected to that of eEF1A2s non-canonical roles and their emergence as also being connected with MND pathology.

6.1 Fibroblast growth factor 12 (Fgf12).

Abundance of Fgf12 was detected as downregulated in the eEF1A2 null mice. It demonstrated the greatest fold change at -2.4 and returned as significantly differentially expressed by statistical tests. It is also known as FHF1 and encodes a small cytosolic protein that interacts with the cytoplasmic tails of voltage gated sodium channels (Siekierska *et al.,* 2016). It promotes neuronal excitability by elevating the voltage dependence of the neuronal sodium channel SCN8A fast and long-term inactivation. Fibroblast-growth-factor homologous factors (FHFs) compose a family of 4 proteins in many vertebrate species. Developing and mature neurons show the highest expression as seen in Fig.35. Expression starts during embryogenesis after post-mitotic neurons emerge at different neuronal sites (Goldfarb 2005). FHF1 knockout mouse have shown moderate muscle weakness, however when knocked along with anther FHF the phenotype becomes more pronounced and mice show ataxia and gait impairments, seemingly similar to eEF1A2 null homozygotes. However, there was no reports on abnormality in the nervous system detected by immunohistochemistry (Goldfarb, 2005).

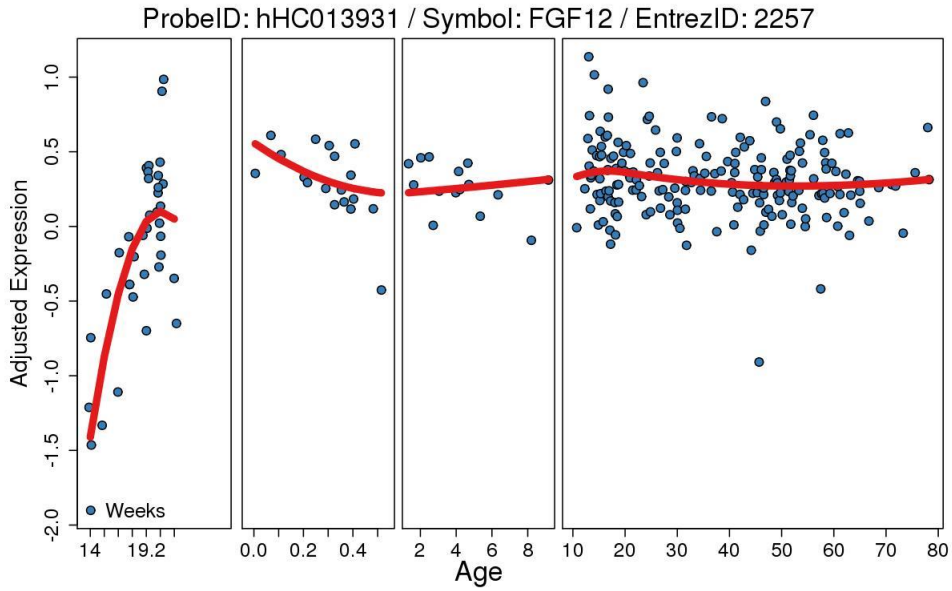


Fig.35. mRNA expression of Fgf12 in human brain. The mRNA expression in postmortum human brains at different ages. There is a sharp increase initial weeks after birth, then levels remain high until old age.

Mutations in Fgf12 have been recorded in humans with early-onset epilepsy. De novo mutations and patients with inherited mutations have been identified and is rarely verified to be the cause of epilepsy (Guella *et al.*, 2016, Al-Mehmedi *et al.*, 2016). The cases have also presented epileptic encephalopathy, which has also been a phenotype presented in humans with eEF1A2 mutations (Lam *et al.*, 2016). The mutation responsible for these 5 cases was shown to be a toxic gain of function mutation (Siekierska *et al.*, 2016) which is of interest as Fgf12 was observed to be upregulated. The remarkable similar phenotype in human cases and mice, as well as the fact that it is upregulated at both the protein and genomic levels shows potential for being a biomarker. However the lack of pathology manifesting in the nervous system along with the pattern of expression, which is markedly similar to eEF1A2 expression (as detected from the same database, see supplementary information) does cast doubt on whether Fgf12 would act as a biomarker for MND.

6.1 Cytoplasmic dynein 1 intermediate chain 2 (Dync1i2).

Dync1i2 expression was observed to be significantly reduced in the Del.22.Ex.3 homozygotes. However it was not seen to reduce as drastically at a genomic level in the *wasted* mice reporting a fold change of -1. Nonetheless the functional role of this protein implicates it in the motor neuron degeneration phenotype observed in the mice.

Dync1i2 is one of the several non-catalytic accessory components located in the tail domain of the cytoplasmic dynein 1 complex. This large multi-subunit complex formed

by homodimerization is the most abundant from the two cytoplasmic dynein complexes, it is responsible for the retrograde transport in neurons. It acts as an adenosine triphosphate (ATP) driven motor, moving along microtubules to transporting essential signals in vesicles and organelles from distal site to the cell body. It is also involved in spindle-pole organisation, nuclear migration during mitosis and the positioning and functioning of the endoplasmic reticulum, Golgi apparatus and nucleus (Pfister *et al*., 2006). The cytoplasmic dynein 1 complex links dynein to cargos and adapter proteins to regulate dynein function.

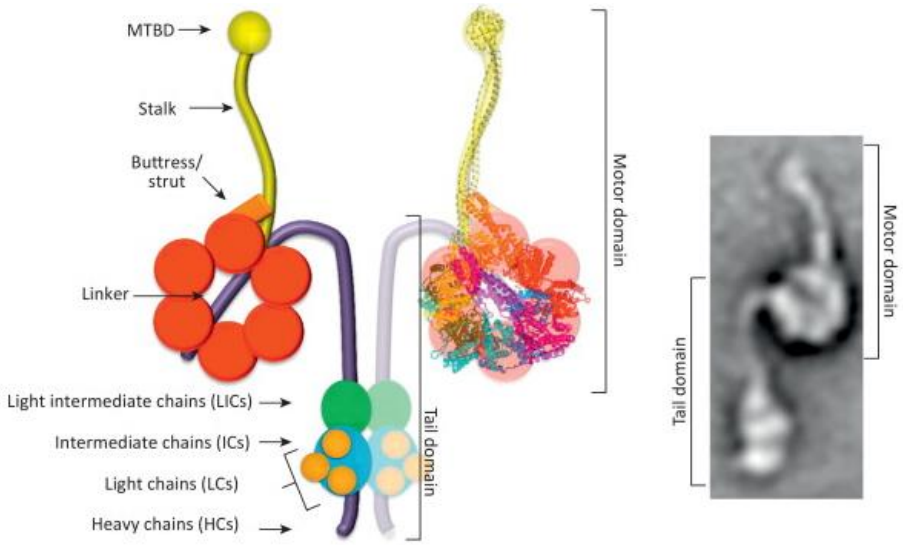


Fig.36. The cytoplasmic dynein complex. (Left). A model of the motor domain built from yeast cytoplasmic dynein and the mouse microtubule-binding domain (MTBD), overlapped with the schematic of the dynein HC in its apo or post-power stroke form. Conformational changes driven by ATP hydrolysis in the motor domain, alter the relative position of the stem and the tail/linker, are hypothesised to lead to the power stroke and progression on microtubules. The Heavy chains (purple) contain the six AAA ATPase domains (in red), the stalk region (light yellow), the buttress (orange), and the linker region. Heavy chains are associated with light intermediate chains (Dync1li1 and Dync1li2) (green), intermediate chains (in cyan), and light chains (dark yellow). (Right) The electron micrograph of an isolated molecule of monomeric dynein from *Chlamydomonas reinhardtii* flagella in its pre-power stroke form. Diagram taken from (Schiavo *et al.* 2013).

The intermediate chains (Dync1i1 and Dync1i2) are distinctly clustered separately in phylogenetic trees based on sequence analysis. Dync1i2 is present in vertebrate species only (Schiavo *et al.,* 2013). Inferred from its primary sequence, Dync1i2 is thought to be post-translationally modified by phosphorylation (Hughes *et al.* 1995). Its phosphorylated form is found only in the slow component of axonal transport whereas its unphosphorylated form has been found in both slow and fast axonal transport.

Axonal transport has been implicated as a possible contributor to motor neuron degeneration. Impaired axonal retrograde transport is identified as one of the earliest pathological events in the SOD1 mouse a popular model for MND. Mutations in the components of cytoplasmic dynein complex have recapitulated several characteristics of MND in mice (Hafezparast *et al.*, 2003, Munch *et al.*, 2004). Intriguingly, aberrations in facial motor neuron migration, many diverged prematurely and assembled at a more anterior position as they do not dislocate in the direction of axon extension to their correct destination (Hafezparast *et al.,* 2003). Facial abnormalities have also been seen in some of the human cases with *EEF1A2* mutations, therefore there may be a link between the impairment of axonal transport as a result of eEF1A2 function being disrupted. The connection between axonal transport and MND development has also been made in humans. In a family with autosomal dominant form of lower motor neuron degeneration was identified with a mutation in the largest subunit (p150) of dynactin (Puls *et al.* 2003), a complex which links dynein to the cargo for transport (Vaughan and Vallee, 1993). The disruption of axonal transport would result in signals and other vital molecules to not be delivered to their appropriate locations. This can cause an accumulation of proteins which is regarded as having a central role in MND development, however the mechanisms are still not fully understood (Reynaud 2010), but the pathology is clear that it is characteristic in motor neuron degeneration. The expression of Dync1i2 is upregulated during cell differentiation and neurite extension upon stimulation of the nerve growth factor (NGF) (Angelastro *et al.,*2000). NGF is a neurotrophic factor that is transport retrogradely and expressed throughout adult life (Conner *et al.,* 2009). It has been demonstrated as an essential factor for maintaining synaptic plasticity (Vivar *et al.,* 2013). The delivery of NGF is facilitated by the actin network, therefore regulation of this network is crucial in sustaining synaptic plasticity.

In some mice from the aforementioned study (Hafezparast *et al.,* 2003), the age-related neurodegeneration may have been caused by a constantly reduced supply of NGF.

The downregulation of Dync1i2 in the proteome of eEF1A2 nulls may be indicative of the mechanism of moto neuron degeneration observed in the mice. As eEF1A has been implicated in microtubule reorganisation, or lack of, in the case of eEF1A2 mutations, hence it may disrupt the networks dynamics or as eEF1A2 is non-functional, Dync1i2 is downregulated, impairing retrograde transport causing neuronal degeneration due to toxic protein accumulation. Thus, reducing synaptic plasticity and being responsible for the retraction of synapse observed in eEF1A2 null mice. However if this is the case, Dync1i2 may be a potential biomarker, its biochemistry and already established connections with motor neuron degeneration, in which it is seen as one of the earliest pathologies can indicate whether accumulation of misfolded proteins is occurring.

6.2 DnaJ homolog subfamily C member 5 (Dnajc5).

In the proteomic study Dnajc5 was downregulated in eEF1A2 null mice. Dnajc5, otherwise known as cysteine string protein (CSP) is part of a group of proteins belonging to the conserved J protein family, also known as Heat shock protein 40kD (Hsp40) or (DnaJ) that are co-chaperones for maintaining proteostatis in cells by assisting in protein folding, assembly and stability and prevent the toxic aggregation of proteins. Dnajc5 domains follow the organisation of DnaJ C-class Hsp40 co-chaperones; it consists of a J domain and a C-terminal cysteine motif of a stretch of 25 cysteine residues (Patel *et al.*, 2016). Dnajc5 is expressed at higher levels in neuronal tissues (NCBI BioProject, 2017) and restricted to the presynaptic termini (Zinsmaier *et al*., 1990, Kohan *et al*., 1995).

As a synaptic vesicle protein and molecular chaperone they function in membrane trafficking and protein folding by interacting with Hsp70 (heat shock protein 70kD), a chaperone involved in protein folding through its J domain. As well as Heat shock cognate 70 (Hsc70), which is the main responder as in neuronal cells Hsp70 induction is reduced. In response to cell stress, Dnajc5 stimulates the ATPase activity of Hsc70 (Fernandez-Chacon *et al.*, 2004). Dnajc5 and Hsc70 then forms a trimeric complex with small glutamine-rich tetratricopeptide repeat-containing protein (SGT), recruiting them the synaptic vesicle surface (Tobaben *et al.*,2001). This interaction maximises Hsc70 ATPase

activity to catalyse the refolding of denatured proteins into its native state, with nucleotide exchange factor (NEF) stimulating the exchange of ATP-ADP (Gorenberg *et al.*, 2017).

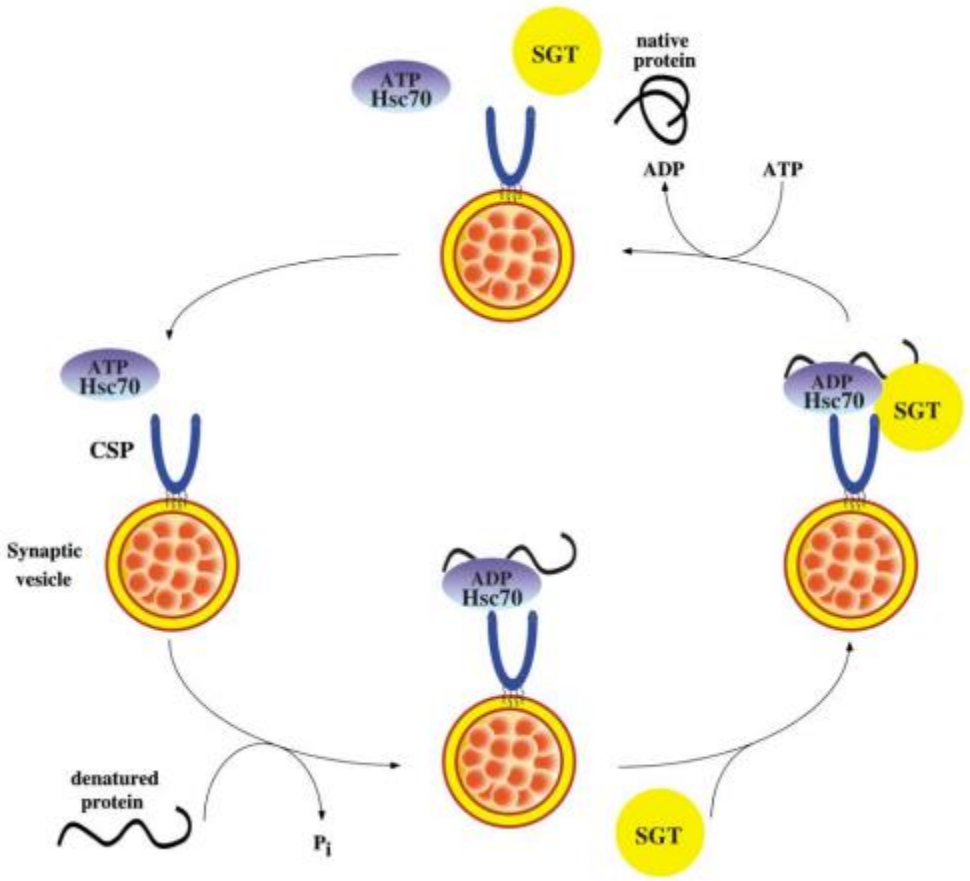


Fig.37 Model Showing the Association-Dissociation Cycle of the CSP/Hsc70/SGT Complex on the Synaptic Vesicle. Dnajc5 (here denoted as CSP) recruits Hsc70 and SGT at the synaptic vesicle surface. This complex in the presence of ATP and available substrate this protein complex dissociates. As a consequence of ATP hydrolysis and chaperone catalysis, unfolded protein substrates in the vicinity of the synaptic vesicle surface are refolded and reactivated. Taken from (Tobaben *et al.,* 2001).

Mutations in the Dnajc5 gene in humans has been identified in adult onset neuronal ceroid lipofuscinosis (ANCL), by whole exome sequencing, linkage analysis and candidate gene resequencing (Burgoyne and Morgan, 2015). ANCL is a rare hereditary neurodegenerative disease that is often misdiagnosed due to its broad clinical variability, common symptoms include generalised epilepsy, ataxia and progressive dementia (Cadeiux-Dion *et al.,* 2013).

Knock out of Dnajc5 has been shown to result in a neurodegenerative phenotype in mice, flies and *C.elegans* (Gorenberg and Chandra, 2017, Rozas *et al.*, 2012, Zinsmaier *et al.*, 1994, Burgoyne and Morgan, 2015). Dnajc5 nulls in flies were embryonic lethal for 95%, survivors went on to develop progressive sluggishness, uncoordinated movements, high temperature paralysis and premature death The phenotype was attributed to neurotransmitter deficits, defects in Ca^{2+} dynamics and the progressive deterioration of Dnajc5 null synapses. This could have been resultant from the accumulation of misfolded proteins at presynaptic termini that led to the debilitated neurotransmitter release. (Gorenberg and Chandra, 2017). When Dnajc5 was knocked out in mice however they appeared normal at birth. It was only at ~20 days that they displayed progressive neurodegeneration and eventual death, which are the similar timelines for the eEF1A2 nulls. It was recorded that there was age-dependent synapse degeneration, selective vulnerability of synaptotagmin-2^{+} GABAergic cells to deterioration and deficits in neurotransmission and activity-dependent loss of synapses (Rozas *et al.*, 2012, García-Junco-Clemente *et al.*, 2010). Further analysis upon the mice revealed impaired synaptic vesicle recycling at the neuromuscular junctions of motor neurons, decreasing the abundance of releasable vesicles thus reducing exocytosis and the neurons excitory capacity (Rozas *et al.*, 2012). Notably α-synuclein, a presynaptic neuronal protein associated with Parkinson's disease, when overexpressed in the mice can rescue the phenotype fully, whilst knocking it out worsens the phenotype (Chandra *et al.*, 2005). Studies done with the mutant α-synuclein^{ha30p} in the Dnajc5 knock out mice found it improved synaptic organisation, synaptic vesicle content and protected against neurodegeneration in the mice (Ruiz *et al.*, 2014). The α-synuclein is thought to contribute to neuronal disease such as Parkinson's by aggregating to toxic levels, and is already being probed in itself as a biomarker or give way to the identification of novel therapeutic targets (Stefanis, 2012).

A decrease in Dnajc5 was found by quantitative mass spectrometry, in the frontal cortex of ~40% in post-mortem brains of Alzheimer patients when compared with age-matched controls (Zhang *et al.*, 2012). A decrease of ~30% in Dnajc5 was found in the eEF1A2 nulls. Although this difference is not as radical as other potential biomarkers for MND, it has informed upon a key mechanism of neuronal deterioration occurring in the Del.22.Ex.3 mice and has shown Dnajc5 to be involved in neurodegeneration in another line of mice.

There is already evidence for the vulnerability of motor neurons to cellular stress damage (Shamovsky *et al.*, 2006), in addition to the impaired heat shock response as eEF1A2 replaces eEF1A1 (Vera *et al*., 2014). Dnajc5's role in recruiting key heat shock response proteins in neurons suggests it contributes to neuronal degeneration. Furthermore the reduced expression of Dnajc5 in differing model organisms has been demonstrated to induce a neurodegenerative phenotype that is improved by already established MND associated protein. Although the mechanisms of Dnajc5 have not entirely been clarified,

it unequivocally is involved in the deterioration of neuronal cells and impaired excitory capacity of motor neurons. Given the aforementioned reasons, it would be a promising biomarker.

6.3 Kinesin heavy chain isoform 5C (Kif5C).

Kinesin heavy chain isoform 5C (Kif5c) is a subunit of kinesin, it belongs to the class of N-1 kinesins (Miki *et al.,* 2001). Kinesins are the motors that mediate anterograde axonal transport along microtubules. The kinesin-1 family (Kif5) within this includes three isoforms; Kif5a, Kif5b and Kif5c. Kif5c is expressed selectively in neurons, enriched at motor neurons, among cranial nerves and the spinal cord (Kanai *et al.,* 2000, Aizawa *et al.,*1998). It is upregulated in the differentiated motor neurons of 2 week old mice and older. Kif5C null mice were viable and demonstrated little phenotype apart from smaller brain size and relative loss of motor neurons to sensory neurons. This would suggest Kif5C plays a role in the maintenance of motor neurons (Kanai *et al.,* 2000). Kif5C is the primary motor for mitochondrial transport in neurons (Pilling *et al.,* 2006). It has also been crucial in reducing mitochondria from clustering near cell centres (Tanaka *et al.,* 1998). Kif5 acts a tetramer, assembling with another heavy chain and two light chain kinesins. It contains an ATPase at its N-terminal and C-terminal for binding the cargo for transport or the cargo adaptor. It is through adaptor proteins that Kif5C attaches to mitochondria, this coupling ensures targeted trafficking and regulation of transport.

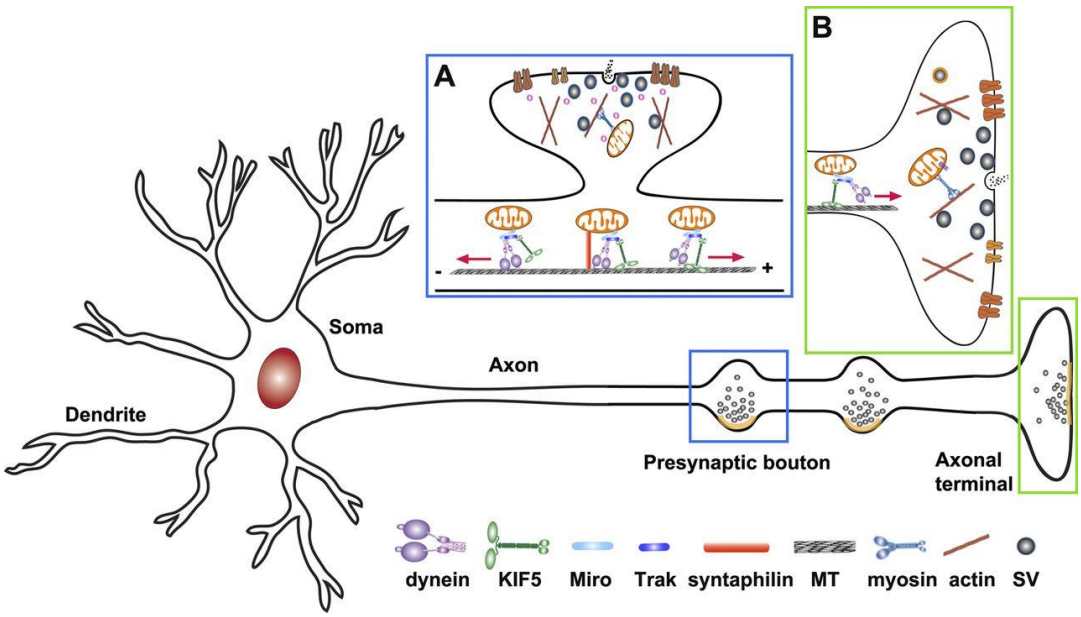


Fig.38. Mitochondrial trafficking and anchoring in neurons. The transporting of mitochondria to A) the presynaptic bouton and B) the axon terminal along microtubules. The polarity of the microtubules directs the movement of mitochondria whilst dynein or kinesins conduct the activity. In axons the (+) ends is oriented towards the axonal terminal whilst the (-) ends lead to the soma. Kif5 mediate anterograde transport whereas dynein retrograde, both proteins can facilitate the transportation of mitochondria with the motor adaptor Trak proteins. Myosin motors are thought to drive short-range mitochondrial movement at the presynaptic terminals. The mitochondria can then be recruited into station pools by dynamic anchoring interactions between syntaphillin and microtubules or other unknown actin-based anchoring receptors. Diagram taken from (Sheng 2014).

The efficient and appropriate transport of mitochondria in synapses is crucial in maintaining cell survival, especially in neurons that have high metabolic demands. Synaptic mitochondria regulate neurotransmission by maintaining Ca^{2+} buffering capacity in neurons, it is also incites short-term synaptic plasticity by uptake of presynaptic Ca^{2+}transients educed by a sequence of action potential and releasing after stimulation (Levy *et al.,* 2003). This study also found that in mice developed impaired learning and hippocampal synaptic plasticity when deficient in mitochondrial voltage-dependent anion channels, a pore on the outer membrane of mitochondria that facilitate the diffusion of small molecules including Ca^{2+}. The neurodegenerative phenotype observed in the Del.22.Ex.3 mice may have been caused in part by impaired kinesin activity as it was significantly reduced in the null mice. Impaired transport of mitochondria in neurons can cause a range of damage if not delivered to the appropriate location or if damaged mitochondria are not removed efficiently. The region deficient in mitochondria may not receive enough ATP, which in synapses can be fatal as their high-energy demands are not met (Le Masson *et al.,* 2014). It also means that the Ca^{2+} buffering capacity in the neuron is impeded and thus the calcium homeostasis. Further damage may be inflicted by the

release of reactive oxygen species (ROS). Mitochondria are the largest producer in the cell of ROS and mitochondrial-derived ROS (mROS) production is regulated by intracellular Ca^{2+} level. Higher concentrations of Ca^{2+} results in an increase of mROS (Beckhauser *et al.,* 2016). Toxic levels of ROS can cause oxidative stress in cells that can lead to neurodegeneration.

It is evident that Kif5C has a role in maintaining neuronal survival by ensuring efficient transport of mitochondria to synapses. The issues that arise from dysfunction of this mechanism causes oxidative stress and reduced synaptic plasticity in the affected neurons, which appears to be predominantly terminally-differentiated motor neurons based upon the expression pattern of Kif5C. Dysfunctional mitochondria has long been associated with neurodegenerative pathology (Court and Coleman, 2012). Mutations in kinesins, including Kif5C have been identified in patients suffering from severe ID, microcephaly, epilepsy and cortical malformation. The mutation was recapitulated in rat hippocampal neurons and found the mutated protein to accumulate in the cell body and show abridged movement as it was detected at reduced levels in distal dendrites. The mutation also led to a decreased miniature excitatory postsynaptic currents (mEPSC), which implied that the excitatory capacity of the neuron is altered (Willemsem *et al.,* 2014). Aberrant Kif5C expression can lead to a disruption of the neurons homeostatic plasticity, which often precedes neuronal death. So although Kif5C has not been probed as a potential biomarker previously it has been heavily implicated in the development of neurodevelopmental disorders and neurodegenerative disease and is significantly differentially expressed at both a genomic and proteomic level.

6.4 Discussion.

The differences between the wildtype and eEF1A2 null protein expression has led to the discovery of proteins shown to be implicated in neuronal degeneration and disease. This has alluded to the pathology of the neurodegeneration in the Del.22.Ex.3 mice, they are collectively indicative of impaired axonal transport and oxidative damage in motor neurons. Additionally, they have be observed to be differentially expressed in human cases and/or considered the underlying mutation of neurodevelopmental disorders. Although the differences in expression between wildtypes and eEF1A2 nulls are not as radical for these proteins as other researched biomarkers are momentarily, this was a small scale preliminary study and more is required to validate these differences. However, their biochemical functions and causative role in the development of neurodegenerative phenotypes in model organisms and human cases are supportive of their potential as markers of MND.

Chapter 7: Discussion.

eEF1A2 is unique in its nature as being the sole translation factor whose expression is tissue-dependant. It is highly conserved across the animal kingdom and one of the most abundant proteins in cells. All this alludes to its essentiality, that becomes ever clearer as more is learnt about the isoforms non-canonical roles that are critical in cell functioning and survival.

eEF1A2 has also repeatedly been implicated in neurological disorders in humans and demonstrated sufficient to cause motor neuron degenerative phenotypes in mice either through its loss of expression or non-functioning protein forms. The phenotypes observed can vary drastically on a case by case basis that is regarded to be due to the various different mutations reported in sufferers. One of these mutations, D252H, was recreated in mice so as to discern the effects the mutation has on protein expression and manifesting pathology. The CRISPR/Cas9 experiment generated mice with a range of different mutations, insertions, deletions and mosaicism that showed a variation of protein expression profiles.

The proteomic study was conducted upon the Del.22.Ex.3 mice which had no eEF1A2 expression. This line was compared with wildtypes and the study found several proteins that were differentially expressed that have been associated with neuronal damage and therefore potential biomarkers for MND. The microarray data analysed had confirmed some of these changes and involvement in neuronal damage as present at a transcriptomic level too, supporting their validity as markers of the motor neuron degeneration phenotype and not necessarily the impaired translation.

7.1 eEF1A2 and protein expression:

The CRISPR/Cas9 experiment to develop the D252H line resulted in a range of mutations in the founder mice. Initial genotyping revealed multiple alleles that could not be unpicked, which when separated by TOPO cloning revealed mosaicism in particular mice and varied insertions and deletions, with most mice retaining a wildtype allele. The most severely affected mouse (#10) had no wildtype allele, the insertion and deletion generated in its genotype had probably caused ablated or non-functional eEF1A2 protein, although this was unable to be measured as the mouse had had to be euthanized. The other mice had no visible phenotype and this was most likely due to the retention of a wildtype allele. However upon protein expression analysis it was revealed that the expression of eEF1A2 was depleted in mice with mutations in at least one allele, yet there were some discrepancies between the genotype and phenotype when compared with the expected protein expression that may have been resultant of the TOPO cloning experiment and sequencing failing to manifest all alleles present in the mice. The results do however

suggest that low levels of eEF1A2 are sufficient to resist the development of neuronal degeneration. This suggests that the missense mutation does not entirely result in loss of function. It is also a promising prospect for gene therapy to compensate for either the loss of or dysfunctional eEF1A2 protein, as the success rate of integration of corrected protein can vary with often only low levels being seen in apparently unaffected mice. What is notable however is that although mice homozygous for mutations in eEF1A2 present a diseased phenotype, the cases found in humans are all heterozygous apart from two siblings (see Table 1). This may mean that mice have more tolerance for the dysfunctional eEF1A2.

7.2 MND in mice and eEF1A2:

The use of mouse models has advanced many differing fields in biology, its heavy usage has also resulted in it undergoing great scrutiny in the efficiency of research, ethical aspects and ability for results to be extrapolated to humans. Much research conducted in MND for the development of therapeutics fails upon trials conducted on humans and there is a deficit of translational research for neuronal diseases. This has led to a great waste of resources and time. Therefore identifying biomarkers in mice may aid in the prognosis of therapies to be made more accurately before they progress onto humans. If the biomarkers identified in mice were to ever be translated into humans, it would have to be identified in tissues that can be tested non-invasively. This presents a challenge as currently most of the indicators of damage have been found in the areas of damage themselves, such as brain and spinal cord. Although the brain tissue is unsuitable for collection until post-mortem, the extraction of spinal fluid is practised but exceedingly dangerous and disquieting for the patients. What is also notable in the context of this research project is that the mice seem to have a degree of tolerance in comparison with humans for eEF1A2 mutations as humans heterozygous for eEF1A2 mutations are severely affected but only homozygous mice develop a phenotype. Although the Del.22.Ex.3 and D252H lines are not suitable as a model organisms for MND as their phenotype cannot be determined, in the Del.22.Ex.3 because of their seizures it would be unethical or even impossible to keep them alive. However, the MND phenotype in the lines make them still appropriate organisms for searching for potential biomarkers that can act as tools for prognosis in the development of therapies and perhaps in future diagnostics.

In the established D252H and Del.22.Ex.3 lines sections of the spinal cords were analysed to observe the extent of neuronal damage that is resultant from the missense mutation and complete loss of eEF1A2. Another eEF1A2 null line, *wasted*, demonstrated neuronal degeneration occurred rostrocaudally in homozygous mice, which would suggest that eEF1A1 is switched off progressively (Newbury *et al.* 2005). The analysis of different regions of the spinal cords in Del.22.Ex.3 and D252H mice revealed that cervical regions showed greater neuronal damage than the more caudal area in mice homozygous for mutated *Eef1a2* in agreement with the findings in *wasted* mice. The H+E stains analysed by Colin Smith also unequivocally demonstrated this. The GFAP staining also showed this gradient in Del.22.Ex.3 homozygotes and marginally so in D252H homozygous mice

which had overall a lesser degree of GFAP staining. GFAP in the heterozygotes however were not consistent with a rostrocaudal gradient of GFAP manifesting in the D252H and showed no discernible pattern in the Del.22.Ex.3 heterozygotes. This degree of neuronal damage as detected by GFAP was unprecedented as it contradicts the phenotype observed in all heterozygote mice from these lines. In addition, the *wasted* heterozygotes also show no neuronal damage. A possible reason for the GFAP appearing as a gradient may be down to variations in staining, it may be that there is no pattern of gliosis in the heterozygotes but that in one D252H mouse it was the case. However, a larger cohort of mice is required to support the identification of a rostrocaudal gradient in homozygotes and clarify the ambiguity surrounding the heterozygotes. In addition to this, the analysis was made by observations alone and is subject to interpretation, to ensure confidence in the analysis using a statistical approach to comparing sections would be more effective. This could be done by foci counts in stains or employ additional tests for neurodegeneration such as the use of the styryl dye FMI-43 which would visualize the neuromuscular junctions and the rate of exocytosis and endocytosis (Amaral *et al.* 2011). Observations of this would also be interesting as some of the differentially expressed proteins that were downregulated in eEF1A2 nulls were associated with reduced exocytosis.

The spinal cords were used for a biomarker search by monitoring protein expression changes across glial and motor neuron cells. By analyzing the whole spinal cord, observations of differential expression in glial cells would not be affected directly by the mutation and downregulation of protein synthesis given that glial cells only express the eEF1A1 isoform (Newbury *et al* 2007). Motor neurons however express only eEF1A2 and would be expected to display numerous changes in expression, predominantly resulting from downregulation of protein synthesis. Spinal cords were extracted from 21 day old mice, a time-point in which the switch between eEF1A isoforms is more or less completed, with only trace amounts of eEF1A1 being expressed (Chambers *et al.*, 1996). As mentioned before, concurrent with the onset of symptoms of motor neuron degeneration associated with eEF1A2 null genotypes (Newbury *et al.* 2005). Therefore the studied spinal cords should display changes affected by the loss of eEF1A2 but before differences in protein expression can be heavily influenced by impaired translation; if analysis was conducted upon later timepoints any resulting differences may be confounded by the loss of translation.

7.3 The search for biomarkers:

The demand for biomarkers for MND is increasing as there is an absence of diagnostic tools in early stages of the disorder that are efficient in diagnosis. The failures of novel therapies to be translated into humans after successful trials in model organisms also highlights the need for biomarkers for prognostic purposes. It is speculated that the failures lie in part because of lack of rigorous testing of preclinical therapies in the model

organisms (Perrin, 2014). Therefore to ensure the treatment is effective, the model organisms must also be tested on a biochemical level which can be facilitated by the use of biomarkers. However biomarker discovery remains challenging because of the absence of standardization of approaches in proteomics and genomic studies, it is also greatly hindered at the data analysis stage, especially in regards to proteomics as the technology and analysis is still being developed.

To search for potential biomarkers for MND, a comparative quantitative analysis into the proteome of the spinal cords of eEF1A2 nulls and wildtypes was conducted using LFQ-MS to identify the proteins present and their abundances. Statistical analysis of the spinal cord proteomes of Del.22.Ex.3 mice did not reveal much difference between that of the wildtypes. The overall protein expression profiles of each sample did not vary greatly from one another, to the extent that the diseased group were not discernible from the wildtypes. The cluster analysis (Fig.18) and heatmap (Fig.20) demonstrated this visually. The samples were from age matched littermates and prepared for mass spectrometry with the same reagent and protocol for homogenisation, inspection of their clustering was not dependent on sex either therefore the clustering cannot have resulted from these factors. From this it must be accepted that there are no proteome differences between the spinal cords at p21 of Del.22.Ex.3 mice and wildtypes. However this conclusion is weakened somewhat by the issue that arose in the LFQ-MS experiment. The deficiency of Lys-C in the peptide digestion stage resulted in incomplete digestion as digestion by trypsin alone is not efficient enough for tightly folded proteins and lysine residues (Giansanti *et al.,* 2016). As mentioned before this can lead to reduced sequence coverage that can discriminate closely related protein isoforms, as well as aid precise protein identification and quantification. This may impede analysis as the MaxQuant software that processed the experimental data works on the assumption of high digestion efficiency. The failure of tandem digestion may have caused missed proteins in the biological system or led to the misidentification of certain proteins, although initially, proteomic studies used only trypsin for digestion. Nonetheless the data was analysed with the acknowledgement that only trypsin was used for digestion. It must be made clear that the proteome of the spinal cord may include proteins that are potentially differentially expressed, but not identified due to lack of efficient digestion; missing out sites of post-translational modifications, protein segments or even subsets of proteins. The reliability of LFQ-MS is also impeded by its inefficiency to detect proteins with low abundances. The proteome of a sample may not be identified in its entirety, limiting the range of proteins that can be possible biomarkers. The pre-processing of the data for analysis had also removed many of these low abundance proteins as proteins that returned a 0 in 2/3 biological replicates were removed. Whether these proteins were expressed below the level of detection, or false identifications is unknown, but to rely upon a single value to represent the abundance would be dubious. The missing values remaining were imputated by assuming that they were below the level of detection of the mass spectrometer instrumentation. This can often result in a skew of data points towards the zero, however it did not disturb the overall distribution of the dataset. Although there are different methods of imputation, a superior one has not been revealed (Webb-Roberston *et al.,*2015, Miecznikowski *et al.,* 2010, Karpievitch *et al.,* 2012, Sandra *et al.,* 2017). This may be because different imputation

approaches and algorithms lend themselves to particular attributes of a dataset; the variance, distribution and size. Therefore how much influence utilizing a different imputation method on the output in downstream analysis is questionable. It would be of interest to see how variations in imputation of missing values affects the amount of potential biomarkers discovered. A repeat of the experiment with tandem digestion and the inclusion of more biological replicates could increase the amount of proteins detected and improve correct identification as well as reduce the degree of proteins eliminated entirely because of a median of 0 among their respective groups. However, given that for a biomarker to be practical it must be reproducible, proteins with low levels in both groups may not necessarily be promising. Especially if the biomarker is for prognosis use in model organisms, as it may not always be possible to collect enough tissue for low abundance proteins to be repeatedly measured.

Two approaches were used to search for proteins with differential expression; fold changes and the student's t test. Examining proteins that have fold change is the most simplistic approach to determining differential expression, but is somewhat limited in this simplicity. It uses the sample means alone and measures changes of expression between the two groups superficially. It did however identify proteins that may be involved in the development of MND in the mice. The functional roles of these implicate them in the developing neuronal degeneration phenotype, Gstk1 being involved in the oxidative stress response and Fgf12 a member of the fibroblast growth factor family which dysregulation of has been found in neurological disease (Hensel *et al.,* 2016). Additionally mutations in Fgf12 were found in humans that demonstrated early-onset epilepsy (Guella *et al.,* 2016, Al-Mehmadi *et al.,* 2016), a symptom that is also present in some of the cases with eEF1A2 mutations (Table.1). The t tests also identified certain proteins of interest, although these were not the most significantly differentiated proteins, they have been heavily implicated in the development of neurodegeneration and connected with eEF1A isoforms non-canonical roles. However, differences between the two groups have failed to manifest with great certainty the proteins that have been identified as of interest and potential markers must be taken with caution. The absence of standard procedure for analysing quantitative mass spectrometry has also meant that different methods of analysis can lead to differing interpretations of the results. The analysis of the data revealed the proteins Kif5c, Dnajc5 and Dync1i2 to all be significantly downregulated and Fgf12 upregulated in the eEF1A2 null mice. Although the student's t test is a common approach when assessing differences between two groups results in small sample size error estimation methods (Murie *e al.,* 2009), it can also overestimate statistical significance (Ting *et al.,* 2009). If this was the case in these findings, it reduces further confidence in the proteins identified as differentially expressed. A major caveat of the data analysis is the lack of corrections made for false discoveries. Without this, some of the differentially expressed proteins have a chance of being false statistically significant changes. The large amount of data requires multiple testing corrections to be made to adjust p-values to account for the occurrence of false positives as t test analysis conducted assesses the significance of each protein independent of one another. Many can be quite stringent for

large datasets, but in proteomics the use of FDR is usually the approach chosen. This would have reduced the amount of discoveries of differentially expressed proteins by 5% and reporting proteins that more are likely to be truly significantly differentially expressed. Given the size of the dataset and low number of biological differences the rate of false discoveries may be a substantial issue resulting in the detection of differentially expressed proteins that cannot be reproduced as they that are not truly significant, thus encumbering the search for biomarkers as these proteins are pursued for further analysis instead of actually differentially expressed proteins that would be indicative of MND. This filtration is an important step in determining biomarkers but this was just an explorative project. Additionally, the distribution of p-values was uniform, suggesting there were little to no significantly differentially expressed proteins. Hence if FDR correction was to be applied it would most likely remove all significant hits. For this reason, the proteins identified as possible novel markers of MND cannot be assumed as truly indicative of the disease until validation.

The small sample size is also a limitation in assessing differential expression. There is low statistical power with just 3 biological replicates for each group. A repeat of the experiment with the addition of more replicates may better identify false positives as well as have great protein coverage revealing more potential biomarkers that are more likely to be indicative of MND. Also, using other statistical tests may be more sensitive to truly differentially expressed changes and better biomarkers. Yet the most important stage is validation of potential findings. The lack of standardization in data analysis causes each proteomic study to generate results based on individual interpretation. Although studies are attempting to modulate the approach to analysing significantly differential expression, validation of results is a crucial step in determining true discoveries.

Previous microarray data analysed by Andy Sims on p21 spinal cords of *wasted* mice was analysed to investigate genes that were differentially expressed between eEF1A2 nulls and wildtype mice. This was done as a complement to the proteomics research, having outlined the translational activity of eEF1A2 nulls from the Del.22.Ex.3 line and identified differentially expressed proteins that demonstrated potential as biomarkers, it would be intriguing to see if these changes were reflected at the transcriptomic level, which would add power to the findings from the proteomics discoveries.
The microarray data demonstrated the gene expression profiles of *wasted* and wildtype mice are strongly distinguished from each other. It revealed many significantly differentially expressed genes between the *wasted* and wildtype mice by SAM and fold change analysis. These genes however cannot be accepted as truly differentially expressed until validated by western blot. The GO analysis showed that many of the genes were associated with neuronal damage. Some of the findings in the microarray data supported the proteomics results and the proteins of interest as being significantly differentially regulated, whilst not being confounded by varying protein turnover rates or stability. Many more genes were identified as being statistically significant than in the proteomics study, however this may be due to the fact that the microarray observed far more genes than the LFQ-MS detected with the microarray data being 18 fold larger than the proteome

data. The comparisons made between the microarray and proteomics data is limited in part due to the differing backgrounds of the mice. However they were age matched and had ablated expression of eEF1A2.

The GO analysis revealed that the molecular functions enriched for across all significantly differentially expressed genes and proteins were predominantly those involved in binding, catalytic and structural activity. However the structural activity as less represented in the microarray data. The pathways enriched for differed slightly between the proteins and genes; there were far more pathways presented by the mRNA analysis, yet this may be in part because there were far more transcripts included than proteins and not necessarily that there were more diverse set of genes differentially expressed in eEF1A2 nulls. Both the proteins and trancripts of eEF1A2 nulls were represented in pathways of neuronal disorders (Alzheimer's, Huntington's and Parkinson's) along with Wnt signalling which has shown to be linked to a neuroprotective role in Alzheimer's disease (Inestrosa and Toledo, 2008). Although Fgf12 was not identified as significantly differentially expressed in the microarray data, the FGF signalling pathway was enriched for. In addition to these connections with neuronal damage, the proteins and genes demonstrated enrichment for structural and stress response. These were predominantly actin and oxidative stress associated proteins and genes which is notable as the eEF1A isoforms have been shown to have roles in the regulation of these in cells. The dysregulation of actin associated genes and resultant proteins in cells may have led to abnormal actin bundling, which in neurons may also be linked with synaptic function (Cingolani and Goda, 2008). The proteins of interest Dync1i2 and Kif5c are reliant upon the cytoskeleton for efficient axonal transport that it maintains. Dnajc5 is involved in the stress response and directly interacts with Hsp70, which contributed to the pathways in the microarray data GO analysis.

Chapter 8: Conclusion.

Mutations in eEF1A2 have disastrous effects on mice homozygous for the mutation. They suffered from a severe neurodegenerative phenotype whilst mice heterozygous for eEF1A2 mutations remained unaffected. The founder mice from the CRISPR/Cas9 experiment mice incorporated the desired D252H mutation, but also presented a varied amount of large insertions and deletions that were strongly suggested to be damaging to the resultant protein's functioning. The CRISPR/Cas9 experiment also induced mosaicism in some of the mice. The expression of eEF1A2 across the founder mice was predominantly reduced, which correlated with the genotyping results that were predictive of the mutations being damaging. This would suggest haplosufficiency in the founders, even in those that had incorporated the D252H mutation, #15 and #16. Despite having significantly reduced eEF1A2, mice that retained wildtype alleles did not manifest the neurodegenerative phenotype. The neuronal damage in the D252H mice as well as in an eEF1A2 null line (Del.22.Ex.3) that was visualized by GFAP in the spinal cords showed the extent of the phenotype on a pathological level. The homozygotes, as speculated, demonstrated the greatest degree of neuronal damage both by IHC with GFAP and H+E stains when compared with heterozygotes and wildtypes. The neuronal degeneration also presented itself as a rostrocaudal gradient in homozygotes from both lines. However the strength of these findings is weakened when the heterozygous mice are examined, despite showing no phenotype, there is consistent GFAP staining which means that the findings are not altogether concrete and would need further investigation into whether this is the case. Nonetheless it is clear that ablated expression of eEF1A2 causes motor neuron degeneration from the phenotype homozygous mice present. The mass spectrometry experiment, although hindered by the absence of tandem digestion, identified a range of proteins in the spinal cords some of which were determined to be differentially expressed between the eEF1A2 nulls and wildtypes. However the overall proteome profiles remained unchanged with the two groups being undiscernible from one another. The significantly differentially expressed proteins were revealed to be associated with functions and pathways upon GO analysis, that when dysregulated have been implicated in MND as well as the non-canonical roles of eEF1A2. From these Fgf12, Dync1i2, Kif5c and Dnajc5 demonstrated significant differential expression, functional roles that were associated to eEF1A2s non-canonical roles and also emerged as being involved in the pathology or neurodegeneration and MND. The data analysis conducted on the microarray data demonstrated far greater differences in the gene expression profiles between *wasted* and wildtype than was observed in the proteomics experiment. They were however, enriched for similar functions and pathways that were represented by the significantly differentially expressed proteins as well. Although the majority proteins of promise as biomarkers were not shown to be as significantly different between the *wasted* and wildtype. The functions and pathways that they are part of include stress response in cells. Kif5C demonstrated the greatest potential as a novel biomarker for MND. It was

differentially expressed at both a protein and transcriptomic level as well as being heavily implicated in the pathogenesis of MND.

Although there are promising biomarkers being developed, they as of yet still fail to be translated into practise. It is always of benefit to search further and investigate other possibilities and this investigation has identified potential markers that, upon further examination of their validity, may be useful as prognostic tools in the development of therapies. The exploration of biomarkers has also hinted at the disease's enigmatic mechanisms and pathogenesis which in turn may lend itself towards further possible targets for novel therapies.

References:

1. Abbas W, Kumar A, Herbein G. (2015). The eEF1A proteins: at the crossroads of oncogenesis, apoptosis, and viral infections. Front Oncol.; 5: 75.
2. Abbott C. M. Helen J., Newbery H. Squires C.E, Brownstein D., Griffiths, L.A., Soares, D.C. (2009). eEF1A2 and neuronal degeneration. Biochemical Society Transactions; 37(6)1293-1297.
3. Aizawa H, Sekine Y, Takemura R, Zhang Z, Nangaku M, Hirokawa N. (1992). Kinesin family in murine central nervous system. J Cell Biol 119:1287–1296.
4. Akerfelt, M., Morimoto, R.I., Sistone L. (2010). Heat shock factors: integrators of cell stress, development and lifespan. Nature Reviews Molecular Cell Biology 11, 545-555.
5. Alberts, B. et al. (2008). Molecular Biology of the Cell, 5th ed. New York: Garland Science.
6. Alfarez, D. N., Joëls, M. and Krugers, H. J. (2003), Chronic unpredictable stress impairs long-term potentiation in rat hippocampal CA1 area and dentate gyrus *in vitro*. European Journal of Neuroscience, 17: 1928–1934.
7. Al-Mehmadi, S., Splitt, M., For DDD Study group, Ramesh, V., DeBrosse, S., Dessoffy, K., *et al.* (2016). FHF1 (FGF12) epileptic encephalopathy. Neurology: Genetics, 2(6), e115.
8. Amaral, E., Guatimosim, S., Guatimosim, C. (2011). Using the fluorescent styryl dye FM1-43 to visualize synaptic vesicles exocytosis and endocytosis in motor nerve terminals. Methods Mol Biol.; 689:137-48.
9. America, A. H. P. and Cordewener, J. H. G. (2008), Comparative LC-MS: A landscape of peaks and valleys. Proteomics, 8: 731–749.
10. Analysis Tool Web Services from the EMBL-EBI. (2013 July) Nucleic acids research 41 (Web Server issue):W597-600.
11. Angelastro JM, Klimaschewski L, Tang S, Vitolo OV, Weissman TA, et al. (2000) Identification of diverse nerve growth factor-regulated genes by serial analysis of gene expression (SAGE) profiling. Proc Natl Acad Sci U S A 97: 10424–10429.
12. Atalay M., Oksala N., Lappalainen J., Laaksonen D.E., Sen C.K., Roy S. (2009). Heat shock proteins in diabetes and wound healing. Curr. Protein Pept. Sci. ;10:85–95.
13. Babu, G. N., Kumar, A., Chandra R. *et al*., (2008). Oxidant-antioxidant imbalance in the erythrocytes of sporadic amyotrophic lateral sclerosis patients correlates with the progression of disease. Neurochemistry International, vol. 52, no. 6, 1284–1289.
14. Bantscheff, M., Lemeer, S., Savitski, M.M. et al. (2012). Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem. 404: 939.
15. Beckhauser, T. F., Francis-Oliveira, J., & De Pasquale, R. (2016). Reactive Oxygen Species: Physiological and Physiopathological Effects on Synaptic Plasticity. *Journal of Experimental Neuroscience*, *10*(Suppl 1), 23–48.
16. Bommel, H., Xie, G., Rossoll, W., Wiese, S., Jablonka, S., Boehm, T., Sendtner, M., 2002. Missense mutation in the tubulin-specific chaperone E (Tbce) gene in the mouse mutant progressive motor neuronopathy, a model of human motoneuron disease. J. Cell Biol. 159, 563 – 569.
17. Bottley A, Kondrashov A. (2013). Aberrant translation of proteins implicated in Alzheimer's disease pathology. OA Genetics 01;1(1):5.
18. Boylan, K. (2015). Familial ALS. Neurologic Clinics, 33(4), 807–830. Dadon-Nachum, M., Melamed, E., Offen, D. (2011). The "Dying-Back" phenomenon of motor neurons in ALS. J Mol Neurosci Vol 43; 3. 470–477
19. Bunai, F., Ando, K., Ueno, H., and Numata, O. (2006). Tetrahymena eukaryotic translation elongation factor 1A (eEF1A) bundles filamentous actin through dimer formation. J Biochem 140, 393–399.
20. Burda JE, Sofroniew MV. (2014). Reactive gliosis and the multicellular response to CNS damage and disease. Neuron.; 81:229–248.

21. Burgoyne R. D., Morgan A. (2015). Cysteine string protein (CSP) and its role in preventing neurodegeneration. Semin. Cell Dev. Biol. 40, 153–159.
22. Byrne, A. B., McWhirter, R. D., Sekine, Y., Strittmatter, S. M., Miller, D. M., Hammarlund, M. (2016). Inhibiting poly(ADP-ribosylation) improves axon regeneration. eLife, 5, e12734.
23. Cadieux-Dion M., Andermann E., Lachance-Touchette P., Ansorge O., Meloche C., Barnabe A. (2013). Recurrent mutations in DNAJC5 cause autosomal dominant Kufs disease. Clin Genet.; 83:571–575.
24. Cao, S., Smith, L.L., Padilla-Lopez, S.R. et al. (2017). Homozygous *EEF1A2* mutation causes dilated cardiomyopathy, failure to thrive, global developmental delay, epilepsy and early death. *Hum Mol Genet*. ddx239.
25. Chakravarty S, Reddy BR, Sudhakar SR, Saxena S, Das T, Meghah V, et al. (2013) Chronic Unpredictable Stress (CUS)-Induced Anxiety and Related Mood Disorders in a Zebrafish Model: Altered Brain Proteome Profile Implicates Mitochondrial Dysfunction. PLoS ONE 8(5): e63302.
26. Chambers DM, Peters J, Abbott CM. (1998).The lethal mutation of the mouse wasted (wst) is a deletion that abolishes expression of a tissue-specific isoform of translation elongation factor 1α, encoded by the Eef1a2 gene. Proc Natl Acad Sci; 95:4463–4468.
27. Chambers, M., Cleveland, W.S., Kleiner, B., Tukey, P.A. (1983). Graphical Methods for Data Analysis. The Wadsworth statistics / probability series. Wadsworth and Brooks/Cole, Pacific Grove, CA.
28. Chandra S., Gallardo G., Fernández-Chacón R., Schlüter O. M., Südhof T. C. (2005). α-synuclein cooperates with CSPα in preventing neurodegeneration. Cell 123, 383–396.
29. Chang, R., and Wang, E. (2007). Mouse translation elongation factor eEF1A-2 interacts with Prdx-I to protect cells against apoptotic death induced by oxidative stress. J Cell Biochem 100, 267–278.
30. Chen, E., Proestou, G., Bourbeau, D., and Wang, E. (2000). Rapid up-regulation of peptide elongation factor EF-1a protein levels is an immediate early event during oxidative stress-induced apoptosis.Exp. Cell Res., 259, 140–148
31. Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., Chan, A.P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE 7(10): e46688.
32. Cingolani, L.A. and Goda, Y. (2008). Actin in action: the interplay between the actin cytoskeleton and synaptic efficacy. Nature Reviews Neuroscience 9, 344-356.
33. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, Colantuoni EA, Elkahloun AG, Herman MM, Weinberger DR, Kleinman JE. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature; 478(7370):519-23.
34. Condeelis, J. (1995). Elongation factor 1a,translation and the cytoskeleton.TIBS,20, 169–170.
35. Conner, J. M., Franks, K. M., Titterness, A. K., Russell, K., Merrill, D. A., Christie, B. R., Tuszynski, M. H. (2009). NGF is Essential for Hippocampal Plasticity and Learning. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 29(35), 10883–10889.
36. Court, F.A., Coleman, M.P. (2012). Mitochondria as a central sensor for axonal degenerative stimuli. Vol. 35, Issue 6, 364-372.
37. Cova, E., Bongioanni, P., Cereda, C. *et al*., (2010). Time course of oxidant markers and antioxidant defenses in subgroups of amyotrophic lateral sclerosis patients, Neurochemistry International, vol. 56, no. 5. 687–693.
38. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ . Molecular & Cellular Proteomics : MCP, 13(9), 2513–2526.
39. Dalman, M.R. Deeter, A., Nimishakavi, G., Duan, Z. (2012). Fold change and p-value cutoffs significantly alter microarray interpretations. BMC Bioinformatics 2012**13(Suppl 2)**:S11
40. Davies, F.C.J.; Hope, J.E.; McLachlan, F.; Nunez, F.; Doig, J.; Bengani, H.; Smith, C.; Abbott, C.M. (2017). Biallelic mutations in the gene encoding eEF1A2 cause seizures and sudden death in F0 mice. Scientific Reports 7, Article number: 46019.
41. de Kovel, C. G.F., Brilstra, E. H., van Kempen, M. J.A., van't Slot, R., Nijman, I. J., Afawi, Z., De Jonghe, P., Djémié, T., Guerrini, R., Hardies, K., Helbig, I., Hendrickx, R., Kanaan, M., Kramer, U., Lehesjoki, A.-E. E., Lemke, J. R., Marini, C., Mei, D., Møller, R. S., Pendziwiat, M.,

Stamberger, H., Suls, A., Weckhuysen, S., EuroEPINOMICS RES Consortium and Koeleman, B. P.C. (2016). Targeted sequencing of 351 candidate genes for epileptic encephalopathy in a large cohort of patients. Mol Genet Genomic Med, 4: 568–580.
42. de Ligt J. ,Willemsen M.H., van Bon B.W., Kleefstra T., Yntema H.G., Kroes T., Vulto-van Silfhout A.T., Koolen D.A., de Vries P., Gilissen C. et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl J. Med ., 367, 1921–1929.
43. Debaisieux S, Rayne F, Yezid H, Beaumelle B (2012). "The ins and outs of HIV-1 Tat.". Traffic. 13 (3): 355–63.
44. Doig J, Griffiths LA, Peberdy D, Dharmasaroja P, Vera M, Davies FJ, Newbery HJ, Brownstein D, Abbott CM. (2013). In vivo characterization of the role of tissue-specific translation elongation factor 1A2 in protein synthesis reveals insights into muscle atrophy. FEBS J.; 280(24) 6528-6540.
45. Duttaroy A., Bourbeau D., Wang X. L., Wang E. (1998). Apoptosis Rate Can Be Accelerated or Decelerated by Overexpression or Reduction of the Level of Elongation Factor-1α. Exp. Cell Res. 238, 168–176.
46. Fernandez-Chacon R, Wolfel M, Nishimune H, Tabares L, Schmitz F, Castellano-Munoz M, Rosenmund C, Montesinos ML, Sanes JR, Schneggenburger R, Sudhof TC. (2004). The synaptic vesicle protein CSP alpha prevents presynaptic degeneration. Neuron. ;42:237–251.
47. Filipowicz, W. and Hohn, T. (1996). Post-Transcriptional Control of Gene Expression in Plants. 1st ed. Netherlands: Kluwer academic publishers.
48. Firth, H.V. et al (2009).DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources. Am.J.Hum.Genet 84, 524-533 Iossifov et al. (2014). The contribution of *de novo* coding mutations to autism spectrum disorder. Nature.; 515(7526):216-21.
49. Fischer, L.R., Culver, D.G., Tennant, P. et al. (2004). Amyotrophic lateral sclerosis is a distal axonopathy: evidance in mice and man. Exp Neurol 185(2):232–240.
50. Gaiottino, J., *et al*. (2013). Increased neurofilament light chain blood levels in neurodegenerative neurological diseases. PLoS ONE;8:e75091.
51. Gandhi S., Abramov A. Y. (2012). Mechanism of oxidative stress in neurodegeneration. Oxid. Med. Cell. Longev. 2012:428010.
52. Garcia-Esparcia, P., Hernández-Ortega, K., Koneti, A., Gil, L. et al. (2015). Altered machinery of protein synthesis is region- and stage-dependent and is associated with α-synuclein oligomers in Parkinson's disease. Acta Neuropathologica Communications; 3:76
53. García-Junco-Clemente P, Cantero G, Gómez-Sánchez L, Linares-Clemente P, Martínez-López JA, Luján R, Fernández-Chacón R. (2010). Cysteine string protein-alpha prevents activity-dependent degeneration in GABAergic synapses. J Neurosci.; 30(21):7377-91.
54. Garman, R.H. (2010). Histology of the Central Nervous System. Toxicologic Pathology Vol: 39 issue: 1: 22-35.
55. Ge, M. and Pan, X.M. (2009). The contribution of proline residues to protein stability is associated with isomerization equilibrium in both unfolded and folded states. Extremophiles; 13:481.
56. Giansanti, Tsiatsiani, L., Low, T.Y., Heck, A.J. (2016). Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. Nature Protocols 11, 993–1006.
57. Goldfarb, M. (2005). Fibroblast growth factor homologous factors: evolution, structure and function. Cytokine & Growth Factor Reviews, 16(2), 215–220.
58. Gorenberg, E. L., & Chandra, S. S. (2017). The Role of Co-chaperones in Synaptic Proteostasis and Neurodegenerative Disease. Frontiers in Neuroscience, *11*, 248.
59. Gray, T. M., Arnoys, E. J., Blankespoor, S., Born, T., Jagar, R., Everman, R., Plowman, D., Stair, A. and Zhang, D. (1996). Destabilizing effect of proline substitutions in two helical regions of T4 lysozyme: Leucine 66 to proline and leucine 91 to proline. Protein Science, 5: 742–751.
60. Griffiths L. A., Doig J., Churchhouse A. M., Davies F. C., Squires C. E., Newbery H. J., et al. (2012). Haploinsufficiency for translation elongation factor eEF1A2 in aged mouse muscle and neurons is compatible with normal function. PLoS One 7:e41917.

61. Griffiths, L.A., Doig, J., Churchhouse, A.M.D., Davies, F.C.J., Squires, C.E., Newbery, H.J., et al. (2012). Haploinsufficiency for Translation Elongation Factor eEF1A2 in Aged Mouse Muscle and Neurons Is Compatible with Normal Function. PLoS ONE 7(7): e41917.
62. Gross S.R. and Kinzy, T.G. (2005). Translation elongation factor 1A is essential for regulation of the actin cytoskeleton and cell morphology. Nature Structural & Molecular Biology 12, 772 – 778.
63. Guella, I., Huh, L., Mckenzie, M., et al. (2016). De novo FGF12 mutation in 2 patients with neonatal-onset epilepsy. Neurology Genetics. 2. e120..
64. Guella, I., Huh, L., Mckenzie, M., Toyota, E. *et al.*. (2016). De novo FGF12 mutation in 2 patients with neonatal-onset epilepsy. Neurology Genetics. 2. e120.
65. Gupta N, Pevzner PA. (2009). False discovery rates of protein identifications: a strike against the two-peptide rule. J. Proteome Res.;8(9):4173–4181.
66. Hafezparast M, Klocke R, Ruhrberg C, Marquardt A, Ahmad-Annuar A, et al. (2003) Mutations in dynein link motor neuron degeneration to defects in retrograde transport. Science 300: 808–812.
67. Hamacher, M. *et al*. (2011) Data Mining in Proteomics: From Standards to Applications, Methods in Molecular Biology, vol. 696, Ch.1, pp.18. Springer Science + Business Media, LLC.
68. Hensel, N., Ratzka, A., Brinkmann, H., Klimaschewski, L., Grothe, C., Claus, P. (2012). Analysis of the Fibroblast Growth Factor System Reveals Alterations in a Mouse Model of Spinal Muscular Atrophy. PLoS ONE 7(2): e31202.
69. Hildebrand MS, Dahl HM, Damiano JA, et al. (2013). Recent advances in the molecular genetics of epilepsy. Journal of Medical Genetics; 50:271-279.
70. Hirakawa, A., Sato, Y., Sozu, T., Hamada, C., & Yoshimura, I. (2007). Estimating the False Discovery Rate Using Mixed Normal Distribution for Identifying Differentially Expressed Genes in Microarray Data Analysis. *Cancer Informatics*, *3*, 140–148.
71. Hoeffer, C. A., Sanchez, E., Hagerman, R. J., Mu, Y., Nguyen, D. V., Wong, H., Whelan, A. M., Zukin, R. S., Klann, E. and Tassone, F. (2012), Altered mTOR signaling and enhanced CYFIP2 expression levels in subjects with fragile X syndrome. Genes, Brain and Behavior, 11: 332–341
72. Hug, N., Longman, D., & Cáceres, J. F. (2016). Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Research, 44(4), 1483–1495.
73. Hughes SM, Vaughan KT, Herskovits JS, Vallee RB (1995). Molecular analysis of a cytoplasmic dynein light intermediate chain reveals homology to a family of ATPases. J Cell Sci 108: 17–24.
74. Ince, P.G., Lowe, J., Shaw, P.J. (1998) Amyotrophic lateral sclerosis: current issues in classification, pathogenesis and molecular pathology. Neuropathol. Appl. Neurobiol., 24:104-117.
75. Inestrosa, N.C. and Toledo, M.E. (2008). The role of *Wnt* signaling in neuronal dysfunction in Alzheimer's disease. Molecular Neurodegeneration **3**:9.
76. International League against Epilepsy Consortium on Complex Epilepsies epilepsy. (2014). Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. The Lancet. Neurology, 13(9), 893–903.
77. Jackson, R.J., Hellen, C.U., Pestova, T.V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11, 113–127 (2010).
78. Jackson, R.S. II, Cho, Y.J., Stein, S. & Liang, P. (2007). CYFIP2, a direct p53 target, is leptomycin-B sensitive.Cell Cycle 6, 95–103.
79. Jeanmougin M, de Reynies A, Marisa L, Paccard C, Nuel G, Guedj M. (2010). Should We Abandon the *t*-Test in the Analysis of Gene Expression Microarray Data: A Comparison of Variance Modeling Strategies. PLoS ONE 5(9): e12336.
80. Kahns S, Lund A, Kristensen P, Knudsen CR, Clark BF, et al. (1998). The elongation factor 1 A-2 isoform from rabbit: cloning of the cDNA and characterization of the protein. Nucleic Acids Res 26: 1884–1890.
81. Kanai, Y., Okada, Y., Tanaka, Y., Harada, A., Terada, S., Hirokawa, N. (2000). KIF5C, a novel neuronal kinesin enriched in motor neurons. J. Neurosci. 20:6374–6384.
82. Karpievitch, Y. V., Dabney, A. R., & Smith, R. D. (2012). Normalization and missing value imputation for label-free LC-MS analysis. BMC Bioinformatics, 13(Suppl 16), S5.

83. Kaufman, L., Ayub, M., & Vincent, J. B. (2010). The genetic basis of non-syndromic intellectual disability: a review. Journal of Neurodevelopmental Disorders, 2(4), 182–209.
84. Kawaguchi, Y.,Bruni,R., Roizman, B. (1997). Interaction of herpes simplex virus 1 alpha regulatory protein ICP0 with elongation factor 1delta: ICP0 affects translational machinery. J Virol.; 71(2): 1019–1024.
85. Khacho, M., Mekhail, K., Pilon-Larose, K., Pause, A., Cote, J., Lee, S. (2008). eEF1A Is a Novel Component of the Mammalian Nuclear Protein Export Machinery. Molecular Biology of the Cell Vol. 19, 5296–5308.
86. Khalyfa A, Bourbeau D, Chen E, Petroulakis E, Pan J, Xu S, et al. (2001). Characterization of elongation factor-1A (eEF1A-1) and eEF1A-2/S1 protein expression in normal and wasted mice. J Biol Chem.;276: 22915–22922.
87. Kim GH, Kim JE, Rhie SJ, Yoon S. (2015). The role of oxidative stress in neurodegenerative diseases. Exp Neurobiol.; 24(4):325–40.
88. Kohan S. A., Pescatori M., Brecha N. C., Mastrogiacomo A., Umbach J. A., Gundersen C. B. (1995). Cysteine string protein immunoreactivity in the nervous system and adrenal gland of rat. J. Neurosci.15, 6230–6238.
89. Kokić, A. N., Stević, Z., Stojanović, S. *et al*. (2005). Biotransformation of nitric oxide in the cerebrospinal fluid of amyotrophic lateral sclerosis patients. Redox Report, vol. 10, no. 5, 265–270.
90. Kumar, A. and Ratan, R.R. (2016). Oxidative Stress and Huntington ’s disease: The Good, The Bad, and The Ugly. J Huntingtons Dis.; 5(3): 217–237.
91. Lai, Y.-C., Baker, J. S., Donti, T., Graham, B. H., Craigen, W. J., Anderson, A. E. (2017). Mitochondrial Dysfunction Mediated by Poly(ADP-Ribose) Polymerase-1 Activation Contributes to Hippocampal Neuronal Damage Following Status Epilepticus. International Journal of Molecular Sciences, 18(7), 1502.
92. Lam, W. W.K., Millichap, J. J., Soares, D. C., Chin, R., McLellan, A., FitzPatrick, D. R., Elmslie, F., Lees, M. M., Schaefer, G. B., DDD study and Abbott, C. M. (2016), Novel de novo EEF1A2 missense mutations causing epilepsy and intellectual disability. Mol Genet Genomic Med, 4: 465–474.
93. Latosinska, A., Vougas, K., Makridakis, M., Klein, J., Mullen, W., Abbas, M. Jankowski, V. (2015). Comparative Analysis of Label-Free and 8-Plex iTRAQ Approach for Quantitative Tissue Proteomic Analysis. *PLoS ONE, 10*(9), e0137048.
94. Lazar, C., Gatto, L., Ferro, M., Bruley, C., Burger, T. (2016). Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. J. Proteome Res., 15, 1116−1125.
95. Le Masson G, Przedborski S, Abbott LF. (2014**).** A computational model of motor neuron degeneration**.** Neuron*. 20;83(4):975-88*.
96. Lee S., Wolfraim L.A., Wang E. (1993). Differential expression of S1 and elongation factor-1 alpha during rat development. J Biol Chem.;268(32):24453-9.
97. Levine, T.D., Bowser, R., Hank, N., Saperstein, D.. (2010). A pilot trial of memantine and riluzole in ALS: Correlation to CSF biomarkers. Amyotrophic Lateral Sclerosis Vol. 11 , Iss. 6,2010.
98. Levy, M., Faas, G.C., Saggau, P., Craigen, W.J., Sweatt, J.D. (2003). Mitochondrial regulation of synaptic plasticity in the hippocampus. J. Biol. Chem. 278:17727–17734.
99. Li D, Wei T, Abbott CM, Harrich D. (2013). The unexpected roles of eukaryotic translation elongation factors in RNA virus replication and pathogenesis. Microbiology and Molecular Biology Reviews; 77: 253–266.
100. Li D., Wei T., Abbott C. M., Harrich D. (2013). The unexpected roles of eukaryotic translation elongation factors in RNA virus replication and pathogenesis. *Microbiol. Mol. Biol. Rev.* 77 253–266.
101. Lopes, F., Barbosa, M., Ameur, A., Soares, G., de Sá, J., Dias, A.I., Oliveira, G., Cabral, P., Temudo, T., Calado, E., et al. (2016). Identification of novel genetic causes of Rett syndrome-like phenotypes. J Med Genet. 53(3), 190-199

102. Lund A, Knudsen SM, Vissing H, et al. (1996). Assignment of human elongation factor 1alpha genes: EEF1A maps to chromosome 6q14 and EEF1A2 to 20q13.3. Genomics;36:359–61 5
103. M.-C. Boll, M.C., Alcaraz-Zubeldia, M., Montes, S., Murillo-Bonilla, L., Rios, C. (2003). Raised nitrate concentration and low SOD activity in the CSF of sporadic ALS patients. Neurochemical Research, vol. 28, no. 5, 699–703.
104. Macleod, S., and Appleton, R. E. (2007). Neurological disorders presenting mainly in adolescence. Archives of Disease in Childhood, 92(2), 170–175.
105. Marco, E., Martin-Santamaria, S., Cuevas, C., Gago, F. (2004). Structural Basis for the Binding of Didemnins to Human Elongation Factor eEF1A and Rationale for the Potent Antitumor Activity of These Marine Natural Products. 4439J. Med. Chem. 2004, 47, 4439-4452.
106. Marraffini L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature 526, 55–61.
107. Mateyak, M.K., Kinzy, T.G. (2010). eEF1A: Thinking Outside the Ribosome. Journal of Biological Chemistry 285, 21209-21213.
108. Mi, H., Muruganujan, A., Casagrande, J.T.,Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nature Protocols 8, 1551 – 1566.
109. Miecznikowski, J. C., Damodaran, S., Sellers, K. F., & Rabin, R. A. (2010). A comparison of imputation procedures and statistical tests for the analysis of two-dimensional electrophoresis data. Proteome Science, 8, 66.
110. Miki H, Setou M, Kaneshiro K, Hirokawa N. (2001). All kinesin superfamily protein, KIF, genes in mouse and human. Proc Natl Acad Sci ;98:7004–11.
111. Motor neuron disease. NHS. (Last reviewed: January, 2015). http://www.nhs.uk/Conditions/Motor-neurone-disease/Pages/Introduction.aspx.
112. Munch C, Sedlmeier R, Meyer T, Homberg V, Sperfeld AD, et al. (2004) Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS. Neurology 63: 724–726.
113. Murie, C., Woody, O., Lee, A. Y., & Nadon, R. (2009). Comparison of small n statistical tests of differential expression applied to microarrays. BMC Bioinformatics, 10, 45.
114. Murray, L.M., Thomson, D., Conklin, A., Wishart, T.M., Gillingwater, T.H. (2008) Loss of translation elongation factor (eEF1A2) expression in vivo differentiates between Wallerian degeneration and dying-back neuronal pathology. J Anat 213:633–45.
115. Myers CT, Mefford HC. Advancing epilepsy genetics in the genomic era. Genome Med. 2015;7: 91.
116. Nakajima J, Okamoto N, Tohyama J, Kato M, Arai H, Funahashi O, *et al*. (2015). De novo eEF1A2 mutations in patients with characteristic facial features, intellectual disability, autistic behaviors and epilepsy. Clin Genet 87(4):356–61.
117. NCBI BioProject, (23rd,July, 2017). EEF1A1 eukaryotic translation elongation factor 1 alpha 1 [*Homo sapiens* (human)]. Retrieved from: https://www.ncbi.nlm.nih.gov/gene/1915/?report=expression.
118. NCBI BioProject, (6th August, 2017 Dnajc5 DnaJ heat shock protein family (Hsp40) member C5 [*Mus musculus* (house mouse)]. Retrieved from: https://www.ncbi.nlm.nih.gov/gene/13002/?report=expression.
119. NCBI BioProject, (9th July, 2017). EEF1A2 eukaryotic translation elongation factor 1 alpha 2 [*Homo sapiens* (human)]. Retrieved from: https://www.ncbi.nlm.nih.gov/gene/1917/?report=expression&bioproject=PRJNA280600.
120. Nesvizhskii, A. and Aebersold, R. (2005). Interpretation of Shotgun Proteomic Data; The Protein Inference Problem. MCP 4, 1419-1440.
121. Newbery H., Gillingwater T., Dharmasaroja P., Peters J., Wharton S., Thomson D., Ribchester R., Abbott C. (2005). Progressive Loss of Motor Neuron Function in Wasted Mice: Effects of a Spontaneous Null Mutation in the Gene for the eEF1A2 Translation Factor. J Neuropathol Exp Neurol. vol: 64 (4) pp: 295-303.
122. Newbery H., Gillingwater T., Dharmasaroja P., Peters J., Wharton S., Thomson D., Ribchester R., Abbott C. (2005). Progressive Loss of Motor Neuron Function in Wasted Mice: Effects of a Spontaneous Null Mutation in the Gene for the eEF1A2 Translation Factor. J Neuropathol Exp Neurol. vol: 64 (4) pp: 295-303

123. Newbery H.; Loh D.; O'Donoghue J.; Tomlinson V.; Chau Y.; Boyd J.; Bergmann J.; Brownstein D.; Abbott C. (2007). Translation elongation factor eEF1A2 is essential for post-weaning survival in mice. Journal of Biological Chemistry. 282.
124. Noble WS, MacCoss MJ (2012) Computational and Statistical Analysis of Protein Mass Spectrometry Data. PLoS Comput Biol 8(1): e1002296.
125. Novosylna O, Doyle A, Vlasenko D, Murphy M, Negrutskii B, El'skaya A. (2017). Comparison of the ability of mammalian eEF1A1 and its oncogenic variant eEF1A2 to interact with actin and calmodulin. Biol Chem. 1;398(1):113-124.
126. O'Callaghan, J.P. and Sriram, K. (2005).Glial fibrillary acidic protein and related glial proteins as biomarkers of neurotoxicity. Expert Opin. Drug Saf. 4(3):433-442.
127. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D,*et al.*. (2016**).** Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation**.** Nucleic Acids Res.2016 Jan 4;44(D1):D733-45.
128. Ott D.E., Coren L.V., Johnson D.G., Kane B.P., Sowder R.C. 2nd, Kim Y.D., Fisher .R.J, Zhou X.Z., Lu K.P., Henderson L.E. (2000). Actin-binding cellular proteins inside human immunodeficiency virus type 1. Virology; 266(1):42-51.
129. Pan J., Ruest L., Xu S., Wang E. (2004) Immuno-characterization of the switch of peptide elongation factors eEF1A-1/EF-1α and eEF1A-2/S1 in the central nervous system during mouse development. Developmental Brain Research. vol: 149 (1) pp: 1-8.
130. Panayiotis Tsokas, P., Grace, E.A., Chan, P., Ma, T., Sealfon, S.C., Iyengar, R., Landau, E.M., Blitzer R.D. (2005). Local Protein Synthesis Mediates a Rapid Increase in Dendritic Elongation Factor 1A after Induction of Late Long-Term Potentiation. Journal of Neuroscience, 25 (24) 5833-5843.
131. Pasinelli, P. and R. H. Brown (2006). Molecular biology of amyotrophic lateral sclerosis: insights from genetics. Nat Rev Neurosci **7**(9): 710-23.
132. Patel P., Prescott G. R., Burgoyne R. D., Lian L.-Y., Morgan A. (2016). Phosphorylation of cysteine string protein triggers a major conformational switch. Structure 24, 1380–1386.
133. Perrin, S. (2014). Preclinical research: Make mouse studies work. Nature 507, 423–425.
134. Pfister, K.K., Shah, P.R., Hummerich, H., Russ, A., Cotton, J., Annuar, A.A., *et al*. (2006) Genetic Analysis of the Cytoplasmic Dynein Subunit Families. PLoS Genet 2(1): e1.
135. Pilling, A.D., Horiuchi, D., Lively, C.M., Saxton, W.M. (2006). Kinesin-1 and Dynein are the primary motors for fast transport of mitochondria in *Drosophila* motor axons. Mol. Biol. Cell. 17:2057–2068.
136. Puls I, Jonnakuty C, LaMonte BH, Holzbaur EL, Tokito M, et al. (2003) Mutant dynactin in motor neuron disease. Nat Genet 33: 455–456.
137. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F. (2013). Genome engineering using the CRISPR-Cas9system. Nature Protocols: 8, 2281–2308.
138. Raven, P.H, Johnson, G.B, Mason, K.A., Losos, J.B., Singer, S.R., (2014). Biology. 10th ed. New York, NY 10020: McGraw Hill.
139. Resing KA, Ahn NG, Katheryn A. Resing, Natalie G. Ahn. (2005). Proteomics strategies for protein identification. FEBS Letters**,** vol. 579:4, pp: 885–889.
140. Reynaud, E. (2010). Protein Misfolding and Degenerative Diseases. Nature Education 3(9):28.
141. Reynet, C., and Kahn, C. R. (2001). Unbalanced expression of the different subunits of elongation factor 1 in diabetic skeletal muscle. Proc. Natl. Acad. Sci. USA 98, 3422-3427.
142. Riluzole. MND Association (February, 2017). www.mndassociation.org/publications.
143. Rotimi, C.N., Chen, G., Adeyemo, A.A. *et al.* (2004). A Genome-Wide Search for Type 2 Diabetes Susceptibility Genes in West Africans. Diabetes, 53 (3) 838-841.
144. Rozas, José Luis *et al.*(2012). Motorneurons Require Cysteine String Protein-α to Maintain the Readily Releasable Vesicular Pool and Synaptic Vesicle Recycling. Neuron, Volume 74, Issue 1, 151 – 165.
145. Ruest, L. B., Marcotte, R., Wang, E. (2002). Peptide elongation factor eEF1A-2/S1 expression in cultured differentiated myotubes and its protective effect against caspase-3-mediated apoptosis. J Biol Chem 277(7): 5418-25.

146. Ruiz R., Biea I.A., Tabares L. (2014). alpha-Synuclein A30P decreases neurodegeneration and increases synaptic vesicle release probability in CSPalpha-null mice. Neuropharmacology.;76 Pt A:106–117.
147. Samocha, K. E., Robinson, E. B., Sanders, S. J., Stevens, C., Sabo, A., McGrath, L. M., Kosmicki, J. A., Rehnström, et al..(2014). A framework for the interpretation of de novo mutation in human disease. Nature Genetics 46(9), 944–950.
148. Sandra L. Taylor, L. Renee Ruhaak, Karen Kelly, Robert H. Weiss, Kyoungmi Kim. (2017). Effects of imputation on correlation: implications for analysis of mass spectrometry data from multiple biological matrices, *Briefings in Bioinformatics*, Vol. 18, Issue 2; 312–320.
149. Sasikumar AN, Perez WB, Kinzy TG. (2012). The many roles of the eukaryotic elongation factor 1 complex. Wiley Interdiscip. Rev RNA; 3: 543–55.
150. Scheper, G.C., Knaap M.S., Proud, C.G. (2007). Translation matters: protein synthesis defects in inherited disease. Nature Reviews Genetics 8, 711-723.
151. Schiavo, G., Greensmith, L., Hafezparast, M., Fisher, E.M.C. (2013). Cytoplasmic dynein heavy chain: the servant of many masters. Vol: 36, Issue 11; 641-651.
152. Scitable by Nature. (Accessed at 08, 2017). Definition: Microarray. https://www.nature.com/scitable/.
153. Sellers, K. F. and Miecznikowski, J.C. (2010). Feature Detection Techniques for Preprocessing Proteomic Data. International Journal of Biomedical Imaging, vol. 2010, Article ID 896718, 9.
154. Shamovsky I, Ivannikov M, Kandel ES, Gershon D, Nudler E. (2006). RNA-mediated response to heat shock in mammalian cells. Nature 440:556–560.
155. Shen, X., Banga, S., Liu, Y., Xu, L., Gao, P., Shamovsky, I., Nudler, E. and Luo, Z.-Q. (2009), Targeting eEF1A by a *Legionella pneumophila* effector leads to inhibition of protein synthesis and induction of host stress response. Cellular Microbiology, 11: 911–926.
156. Sheng, Z. (2014). Mitochondrial trafficking and anchoring in neurons: New insights and implications. J. Cell Biol; 204(7) 1087-1098.
157. Shin-ichiro Ejiri. (2002). Moonlighting Functions of Polypeptide Elongation Factor 1: From Actin Bundling to Zinc Finger Protein R1-Associated Nuclear Localization, Bioscience, Biotechnology, and Biochemistry, 66:1, 1-21.
158. Siekierska, A., Isrie, M., Liu, Y., Scheldeman, C., Vanthillo, N., Lagae, L., Buyse, G. M. (2016). Gain-of-function FHF1 mutation causes early-onset epileptic encephalopathy with cerebellar atrophy. Neurology, 86(23), 2162–2170.
159. Skaper, S.D. (2003). Poly(ADP-Ribose) polymerase-1 in acute neuronal death and inflammation: a strategy for neuroprotection. Ann N Y Acad Sci.; 993:217-28; discussion 287-8.
160. Soares DC, Barlow PN, Newbery HJ, Porteous DJ, Abbott CM (2009) Structural Models of Human eEF1A1 and eEF1A2 Reveal Two Distinct Surface Clusters of Sequence Variation and Potential Differences in Phosphorylation. PLoS ONE 4(7).
161. Soares, D. and Abbott, C. (2013). Highly homologous eEF1A1 and eEF1A2 exhibit differential post-translational modification with significant enrichment around localised sites of sequence variation. Biology Direct 8:29.
162. Stanley, S., Domingos, A.I., Kelly, L., Garfield, A., Damanpour, S., Heisler L., Friedman, J. (2013). Profiling of Glucose-Sensing Neurons Reveals that GHRH Neurons Are Activated by Hypoglycemia. Vol. 18;4, 586-607.
163. Stefanis, L. (2012). α-Synuclein in Parkinson's Disease. Cold Spring Harbor Perspectives in Medicine, 2(2), a009399.
164. Suzuki, Y., Oishi, Y., Nakano, H., Nagayama, T. (1987). A strong correlation between the increase in number of proline residues and the rise in thermostability of five Bacillus oligo-1, 6-glucosidase. Appl Microbiol Biotechnol 26:546–551.
165. Tanaka Y, Kanai Y, Okada Y, Nonaka S, Takeda S, Harada A, Hirokawa N. (1998). Targeted disruption of mouse conventional kinesin heavy chain, kif5B, results in abnormal perinuclear clustering of mitochondria. Cell; 93:1147–1158
166. Ting, L., Cowley, M. J., Hoon, S. L., Guilhaus, M., Raftery, M. J., & Cavicchioli, R. (2009). Normalization and Statistical Analysis of Quantitative Proteomics Data Generated by Metabolic Labeling. Molecular & Cellular Proteomics : MCP, 8(10), 2227–2242.

167. Ting, L., Cowley, M. J., Hoon, S. L., Guilhaus, M., Raftery, M. J., & Cavicchioli, R. (2009). Normalization and Statistical Analysis of Quantitative Proteomics Data Generated by Metabolic Labeling. Molecular & Cellular Proteomics : MCP, 8(10), 2227–2242.
168. Tuncel, D., Aydin, N., Kocatürk, P.A., Kavas, G.O., Sarikaya, S. (2006). Red cell superoxide dismutase activity in sporadic amyotrophic lateral sclerosis. Journal of Clinical Neuroscience, vol. 13, no. 10, 991–994.
169. Vaughan, K.T., Vallee, R.B. (1995). Cytoplasmic dynein binds dynactin through a direct interaction between the intermediate chains and p150Glued. J Cell Biol 131: 1507–1516.
170. Veeramah, K.R., Johnstone, L., Karafet, T.M., Wolf, D., Sprissler, R., Salogiannis, J., Barth-Maron, A., Greenberg, M.E., Stuhlmann, T., Weinert, S., et al. (2013). Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. Epilepsia 54, 1270–1281.
171. Vera M, Pani B, Griffiths LA, Muchardt C, Abbott CM, Singer RH et al. (2014). The translation elongation factor eEF1A1 couples transcription to translation during heat shock response. Elife; 3: e03164.
172. Vivar, C., Potter, M. C., & van Praag, H. (2013). All About Running: Synaptic Plasticity, Growth Factors and Adult Hippocampal Neurogenesis. *Current Topics in Behavioral Neurosciences*, *15*, 189–210.
173. Wang H, Alvarez S, Hicks LM (2012). Comprehensive Comparison of iTRAQ and Label-free LC-Based Quantitative Proteomics Approaches Using Two *Chlamydomonas reinhardtii* Strains of Interest for Biofuels Engineering. J Proteome Res 11(1):487–501.
174. Webb-Robertson, B.-J. M., Wiberg, H. K., Matzke, M. M., Brown, J. N., Wang, J., McDermott, J. E., Waters, K. M. (2015). Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics. Journal of Proteome Research, 14(5), 1993–2001.
175. Whittemore E.R., Loo D.T., Watt J.A., Cotman C.W. (1995). A detailed analysis of hydrogen peroxide-induced cell death in primary neuronal culture. Neuroscience;67(4):921–932.
176. Willemsen MH, Ba W, Wissink-Lindhout WM, *et al*. (2014). Involvement of the kinesin family members KIF4A and KIF5C in intellectual disability and synaptic function. Journal of Medical Genetics ;51:487-494.
177. Witze, E.S., Old, W.M., Resing, K. A., Ahn, N.G. (2007). Mapping protein post-translational modifications with mass spectrometry. Nature Methods 4, 798 – 806.
178. Wu, B., Eliscovich,C., Yoon, Y.J. Singer, R.H. (2016). Translation dynamics of single mRNAs in live cells and neurons. Science; 352(6292): 1430–1435.
179. Xiao, H., Neuveut, C., Benkirane, M., and Jeang, K.-T. (1998). Interaction of the second coding exon of Tat with human EF-1 delta delineates a mechanism for HIV-1-mediated shut-off of host mRNA translation. Biochem. Biophys. Res. Commun., 244, 384–389.
180. Xiao, H., Neuveut, C., Benkirane, M., Jeang, K. (1998). Interaction of the Second Coding Exon of Tat with Human EF-1δ Delineates a Mechanism for HIV-1-Mediated Shut-Off of Host mRNA Translation. Biochemical and Biophysical Research Communications, Volume 244, Issue 2, 384-389
181. Yudkoff, M., Daikhin, Y., Nissim, I., Horyn, O., Luhovyy, B., Lazarow, A., Nissim, I. (2005). Brain amino acid requirements and toxicity: the example of leucine. J Nutr.;135(6 Suppl):1531S-8S.
182. Zhang Y.-Q., Henderson M. X., Colangelo C. M., Ginsberg S. D., Bruce C., Wu T., *et al*. . (2012). Identification of CSPα clients reveals a role in dynamin 1 regulation. Neuron 74, 136–150.
183. Zhang, S. (2007). A comprehensive evaluation of SAM, the SAM R-package and a simple modification to improve its performance. *BMC Bioinformatics*, *8*, 230.
184. Zinsmaier K. E., Eberle K. K., Buchner E., Walter N., Benzer S. (1994). Paralysis and early death in cysteine string protein mutants of Drosophila. Science 263, 977–980.
185. Zinsmaier K. E., Hofbauer A., Heimbeck G., Pflugfelder G. O., Buchner S., Buchner E. (1990). A cysteine-string protein is expressed in retina and brain of Drosophila. J. Neurogenet. 7, 15–29.

Supplementary Information:

CLUSTAL O(1.2.4) multiple sequence alignment of EEF1A2.

```
Danio           ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Homo            ------------------------------------------------------------
Mus             ------------------------------------------------------------
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          CCACCCCTCCATTTCTCTCATCCCTCCATCTTCCTCCCATCTCCTTACCTCCATCTCTTG

Danio           ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Homo            ------------------------------------------------------------
Mus             ------------------------------------------------------------
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          TTTCTCTTTCCTCTTTTCCTCTTATTGTTCAATTTCTCCCTCCCTCCATCTTCCTCCCTT

Danio           ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Homo            ------------------------------------------------------------
Mus             ------------------------------------------------------------
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          CTCTTTACCTCCCATCCTTCTGTCTGCATCCTCCTCTCTCCATCTCTCTTCCATTTCTCC

Danio           ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Homo            ------------------------------------------------------------
Mus             ------------------------------------------------------------
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          TAACCTTCCACTCTTCCTAACCCTCTATTTATCTCTCCCTCCATTTCTCCCATCCCTTTA

Danio           ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Homo            ------------------------------------------------------------
Mus             ------------------------------------------------------------
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          TCCCCCATCTCTTCCCATCCATCCATCTCCCACCCTTCCACCTCTCCCACCCATCTGTCT

Danio           ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Homo            ------------------------------------------------------------
Mus             ------------------------------------------------------------
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
```

```
Gallus          ACTGCTCCTCCATCTCTCCTCCCTTCCTCCCATCCTTCTATTCCCCCTATCCCCCTATTC

Danio           ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Homo            --------------GCCCC-------------------------GC--------------
Mus             ------------------------------------------GT--------------
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          CTCCCATCCCTTTACCCCTGGTCTCTACATCTCCCACTACTCCAGCTTTCATCCATTTCT

Danio           ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Homo            ------------CCCCGCCCGCGGCGCGTTTCTCCCCCGCCTCCCGCGTCCGTCTTTGCA
Mus             ------------TCTCGCTCA----CTGGTTCTCTCCCTC-GCTCCGGTGCATCATTGCA
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          CCCATCCCTCCATCTTCCTCCCAT-CCCTTTATCCCCC--ATCTCCTCCCCATCTCCCAC

Danio           ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Homo            GCCCGCGCCTCCCGCATCGC-----------CTCGCGTCCCCGTG------GCGCCCGCC
Mus             GCTGCGTCCTCTCGGATCCTCATTACGCCGGCCCGCGTCCGTGGGTGCGCGGCCCCTGCG
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          TCTTCCACCTCTTCCATCTACCTTCCATCCCCAGTTCTCCCC--CTCC---GCCTCTCCC

Danio           ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Homo            -----CGCG-----------------------------------CGCGTCCGCGCCCCGCC
Mus             -----TCCACGCATCTTTCGCATCCCA--------TCTGCCAGCCGCTCGGCGCC-CG-C
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          TGCTCTCTCCTCCTCTTTCTCTTCCTCCTCCTCCTCCTCCCCCCGGCT---CGCCTCCTC

Danio           ----------------------------------------GAGTTCATCACACACACAC-
Oryctolagus     ------------------------------------------------------------
Homo            CCCTCCCGCGCGGTTCCGCATTGGCGTGCTGCAGG-GCG-CGGTGCACTGCGCCGCCACC
Mus             CTCCCCCCTCCGGTACCGCATTGCCGTACTGCAGGGGCG-CAGTGCATTGCGCCGGCACC
Bos             ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          CTCCCCCGCCCGGTACCGCATTGCCGTAGTGCGGGGGGGCCGCTGCATTGCGCTGCCGCC

Danio           ----AGACCTGGACT-CTTCCGCCAGCTTTGAA------------GGATTTCATAGTGTC
Oryctolagus     ------------------------------------------------------------
Homo            GTCAATAGGTGGACCCCCTCCCGGA--GATAAAACCGCCGGCGCCGGCGCCGCCAGTCCC
Mus             GTCAATAGGTGGACCCCCTCCTGGAGAGATAAAACCGCCGGCGCCGGCGCCACCAGTCCT
Bos             ------------------------------------------------------CCC
Xenopus         ------------------------------------------------------------
Gallus          GTCAATAGGTGGGCC-CCTCCCGGGGAGATAAAGCCGCCGGAGCCCAAGCTGGCAGCCTC

Danio           TGTTTGCTGAGCGGAAAGCAGGGTCTGC--------------CTGTCTGCTGCAGAC-AG
Oryctolagus     ------------------------------------------------------------
Homo            TCT-GGCTGAG----ACCTCGGCTCCGGAATCACTGCAGCCC-----CCCT-CGCCCTGA
```

```
Mus             TCT-GACTGAG----TCCTCGGCTCTGGAGTTCCTGCCCAGCATATACCCTCAACCCCAA
Bos             TCT-GGCTGAG----ACCTCGGCTCTGGACTCACTGCTCAGCTTC--CCCT-CACCCTGA
Xenopus         ----------------------------------------------------------C-TT
Gallus          TGC-CGCCCCG----ACCGCCGCTCCGCCTCT--------CCATAAACGCAGCTGCG-TC


Danio           AAGAAAGCACCT--CTTACTGTTCTCTCTTGCC--GCCCCGAGTCAACATGGGGAAAGAG
Oryctolagus     ----------------------CTCGGCTCCGG-AGCCCCCGGCAGAATGGGCAAGGAG
Homo            GCCAGAGCACCCCGGGTCCCGCCAGCCCCTCAC---ACTCCCAGCAAAATGGGCAAGGAG
Mus             ACCAGAGCCCCCA---CAGTGCCAGCCCCTCCCT--CACCCAGGCAGAATGGGCAAGGAG
Bos             GCCAGAGCACCCCAGGTCGTGCCAGCCCCTCCCCACGCCCCAGGCAGAATGGGCAAGGAG
Xenopus         CCT---------------TGCACTCTCGCTAGTGTCACCCCAGGAGGATGGGGAAAGAG
Gallus          CCCAGCGCTCCC--CCCACTGCCCGCCCTCCACGCGTAGCCCACCAGCATGGGGAAGGAG
                                        *            *     *  ***** ** ***

Danio           AAGATCCACATCAACATTGTGGTGATCGGCCATGTTGATTCTGGGAAATCAACCACCACT
Oryctolagus     AAGACGCACATCAACATCGTGGTCATCGGCCATGTGGACTCGGGCAAGTCCACCACCACC
Homo            AAGACCCACATCAACATCGTGGTCATCGGCCACGTGGACTCCGGAAAGTCCACCACCACG
Mus             AAGACACACATCAACATTGTGGTCATTGGCCACGTGGACTCAGGCAAGTCCACCACGACA
Bos             AAGACCCACATCAACATAGTGGTCATCGGCCACGTGGACTCAGGCAAGTCCACCACGACT
Xenopus         AAGACACACATCAACATCGTGGTCATTGGCCACGTGGACTCTGGCAAGTCCACAACCACC
Gallus          AAGACGCACATCAACATTGTCGTCATCGGGCATGTGGACTCTGGGAAATCCACCACCACC
                ****  *********** ** ** ** ** ** ** ** ** ** ** ** ** ** **

Danio           GGGCATCTCATCTACAAATGTGGAGGAATTGATAAAAGAACCATTGAGAAGTTTGAGAAA
Oryctolagus     GGCCACCTCATCTACAAGTGCGGGGGCATCGACAAGAGGACCATCGAGAAGTTTGAAAAG
Homo            GGCCACCTCATCTACAAATGCGGAGGTATTGACAAAAGGACCATTGAGAAGTTCGAGAAG
Mus             GGCCACCTCATCTACAAGTGTGGTGGCATCGACAAGCGGACCATCGAGAAGTTTGAGAAG
Bos             GGCCACCTCATCTACAAATGCGGGGGCATCGACAAGAGGACCATCGAGAAGTTTGAGAAG
Xenopus         GGCCACCTGATCTACAAGTGCGGGGGCATCGACAAAAGGACGATAGAGAAGTTTGAGAAG
Gallus          GGGCACCTCATCTACAAATGCGGGGGCATTGACAAAAGGACCATTGAGAAATTCGAGAAG
                ** ** ** ******** ** ** ** ** ** **  * ** ** ***** ** ** **

Danio           GAGGCAGCTGAGATGGGAAAAGGTTCTTTTAAGTATGCCTGGGTCCTGGATAAGTTGAAG
Oryctolagus     GAGGCGGCCGAGATGGGGAAAGGCTCCTTCAAGTACGCCTGGGTGCTGGACAAGCTGAAG
Homo            GAGGCGGCTGAGATGGGGAAGGGATCCTTCAAGTATGCCTGGGTGCTGGACAAGCTGAAG
Mus             GAGGCAGCAGAGATGGGGAAGGGCTCTTTTAAATATGCCTGGGTGCTGGACAAGCTGAAG
Bos             GAGGCGGCCGAGATGGGAAAGGGCTCCTTCAAGTATGCCTGGGTACTGGACAAGCTGAAG
Xenopus         GAGGCGGCCGAGATGGGGAAAGGTTCCTTCAAGTACGCTTGGGTTTTGGACAAGCTGAAG
Gallus          GAGGCTGCCGAGATGGGGAAGGGGTCCTTCAAATACGCCTGGGTGCTGGACAAGCTGAAG
                ***** ** ******** ** ** ** ** ** ** ** *****  **** *** *****

Danio           GCTGAGAGGGAGAGAGGCATCACCATAGACATCTCACTCTGGAAGTTTGAGACCACTAAA
Oryctolagus     GCCGAGCGGGAGCGCGGCATCACCATCGACATCTCCCTCTGGAAGTTCGAGACCACCAAG
Homo            GCGGAGCGTGAGCGCGGCATCACCATCGACATCTCCCTCTGGAAGTTCGAGACCACCAAG
Mus             GCCGAGCGGGAACGAGGCATCACCATCGACATCTCCCTCTGGAAGTTTGAGACCACCAAG
Bos             GCAGAGCGGGAACGCGGCATCACCATCGACATCTCCCTCTGGAAATTTGAGACCACCAAA
Xenopus         GCTGAGAGGGAGCGAGGAATCACCATTGATATCTCCCTTTGGAAGTTTGAGACAAACAAA
Gallus          GCCGAGCGTGAGCGTGGCATCACCATTGACATCTCACTGTGGAAATTTGAAACAAGCAAG
                ** *** * **  * ** ******** ** ***** ** ***** ** ** ** *  **

Danio           TACTACATAACCATAATAGATGCTCCAGGACATAGAGACTTTATCAAAAACATGATCACT
Oryctolagus     TACTACATCACCATCATCGACGCGCCCGGCCACCGCGACTTCATCAAGAACATGATCACG
Homo            TACTACATCACCATCATCGATGCCCCCGGCCACCGCGACTTCATCAAGAACATGATCACG
Mus             TACTACATCACCATCATCGATGCTCCAGGACACCGAGACTTCATCAAGAATATGATTACA
Bos             TACTACATCACCATCATCGACGCCCCAGGCCACCGCGACTTCATTAAGAACATGATCACA
Xenopus         TATTACATCACCATCATTGATGCCCCCGGACATCGAGACTTCATCAAGAATATGATCACT
Gallus          TACTACGTCACCATCATCGATGCGCCCGGCCACAGGGACTTCATCAAGAACATGATCACT
                ** *** * ***** ** ** ** ** ** **  * ***** ** ** ** ***** **
```

```
Danio            GGGACATCTCAGGCGGATTGTGCTGTCTTAATTGTAGCGGCTGGAGTGGGTGAATTTGAG
Oryctolagus      GGCACGTCCCAGGCGGACTGCGCGGTGCTCATCGTGGCCGCGGGCGTGGGCGAGTTTGAG
Homo             GGTACATCCCAGGCGGACTGCGCAGTGCTGATCGTGGCGGCGGGCGTGGGCGAGTTCGAG
Mus              GGCACATCCCAGGCGGACTGCGCAGTGCTGATCGTGGCAGCCGGTGTGGGCGAGTTTGAG
Bos              GGCACATCCCAGGCGGACTGCGCCGTGCTGATTGTGGCCGCAGGAGTGGGTGAGTTCGAG
Xenopus          GGAACCTCTCAGGCAGACTGTGCAGTGCTGATAGTGGCAGCCGGAGTGGGTGAGTTTGAA
Gallus           GGCACATCCCAGGCTGACTGCGCCGTACTGATTGTTGCTGCCGGTGTCGGTGAGTTTGAA
                 ** ** ** ***** ** ** ** **  * ** ** ** ** ** ** ** ** ** **

Danio            GCAGGCATTTCCAAAAATGGCCAAACAAGGGAACACGCCCTGCTGGCCTACACACTTGGT
Oryctolagus      GCCGGCATCTCCAAGAACGGGCAGACGCGGGAGCACGCGCTGCTGGCCTACACGCTGGGC
Homo             GCGGGCATCTCCAAGAATGGGCAGACGCGGGAGCATGCCCTGCTGGCCTACACGCTGGGT
Mus              GCGGGCATCTCCAAGAACGGGCAAACCCGGGAACACGCACTCCTGGCCTACACTCTGGGT
Bos              GCAGGCATCTCCAAGAATGGGCAGACCCGGGAGCACGCGCTGCTGGCCTACACGCTGGGC
Xenopus          GCTGGCATCTCCAAGAATGGACAGACCCGTGAACATGCCCTCCTGGCTTACACCCTCGGA
Gallus           GCCGGCATCTCCAAGAACGGGCAGACCCGTGAGCACGCCCTGCTGGCTTACACCCTGGGG
                 ** ***** ***** ** ** ** **  * ** ** ** ** ***** ***** ** **

Danio            GTCAAGCAACTGATCGTAGCCGTCAACAAGATGGACTCCACCGAGCCTTCCTACAGTGAG
Oryctolagus      GTGAAGCAGCTCATCGTGGGCGTCAACAAGATGGACTCCACAGAGCCGGCCTACAGCGAG
Homo             GTGAAGCAGCTCATCGTGGGCGTGAACAAAATGGACTCCACAGAGCCGGCCTACAGCGAG
Mus              GTGAAGCAGCTCATTGTGGGTGTCAACAAGATGGACTCCACGGAACCAGCCTACAGCGAG
Bos              GTGAAGCAGCTCATCGTGGGGGTGAACAAGATGGACTCCACGGAGCCCGCCTACAGCGAG
Xenopus          GTCAAACAGCTCATCGTGGGAATCAATAAAATGGACTCCACCGAGCCTCCCTACAGCGAG
Gallus           GTGAAGCAGCTCATTGTGGGCATCAACAAGATGGATTCCACGGAGCCTGCATACAGCGAG
                 ** ** ** ** ** ** *   * ** ** ***** ***** ** **  * ***** ***

Danio            AAACGCTATGATGAGATCGTCAAAGAAGTGAGCGCCTACATCAAAAAGATTGGTTATAGT
Oryctolagus      AAGCGCTACGACGAGATCGTCAAGGAGGTCAGCGCCTACATCAAGAAGATCGGCTACAAC
Homo             AAGCGCTACGACGAGATCGTCAAGGAAGTCAGCGCCTACATCAAGAAGATCGGCTACAAC
Mus              AAGCGCTATGATGAGATTGTTAAGGAGGTCAGCGCCTACATCAAGAAGATCGGCTACAAC
Bos              AAGCGCTATGATGAAATTGTCAAGGAGGTCAGCGCCTACATCAAGAAGATCGGCTACAAC
Xenopus          AAACGTTATGATGAGATCGTCAAGGAGGTCAGCGCCTACATCAAGAAGATTGGCTACAAC
Gallus           AAACGCTATGATGAGATCGTCAAGGAGGTCAGCGCCTACATCAAGAAGATCGGCTACAAC
                 ** ** ** ** ** ** ** ** ** ** ************** ***** ** ** *

Danio            CCAGCTTCCGTACCCTTTGTCCCTATTTCAGGCTGGCATGGCGACAACATGCTGGAACCG
Oryctolagus      CCGGCCACCGTGCCCTTCGTGCCCATCTCGGGCTGGCACGGGGACAACATGCTGGAGCCC
Homo             CCGGCCACCGTGCCCTTTGTGCCCATCTCCGGCTGGCACGGTGACAACATGCTGGAGCCC
Mus              CCAGCCACGGTGCCCTTCGTGCCCATCTCGGGCTGGCATGGTGACAACATGCTGGAGCCT
Bos              CCCGCCACTGTACCCTTCGTGCCCATCTCAGGCTGGCACGGCGACAACATGCTGGAGCCC
Xenopus          CCAGCTACAGTTCCCTTCGTACCCATTTCTGGCTGGCATGGAGATAACATGTTGGAGCCC
Gallus           CCAGCCACGGTTCCCTTCGTGCCCATCTCGGGCTGGCACGGGGACAACATGCTGGAGCCA
                 ** **  * ** ***** ** ** ** ** ******** ** ** ****** **** **

Danio            TCTTCCAATATGCCATGGTTTAAAGGCTGGAAGCTGGACAGGAAGGAGCACCATGCCGGT
Oryctolagus      TCCCCCAACATGCCGTGGTTCAAGGGCTGGAAAGTGGAGCGGAAGGAAGGCAATGCCAGC
Homo             TCCCCCAACATGCCGTGGTTCAAGGGCTGGAAGGTGGAGCGTAAGGAGGGCAACGCAAGC
Mus              TCACCTAATATGCCATGGTTCAAGGGCTGGAAAGTAGAGCGTAAGGAAGGAAATGCAAGC
Bos              TCACCCAACATGCCCTGGTTCAAGGGCTGGAAAGTGGAGAGGAAGGAAGGGAACGCCAGT
Xenopus          TCTCCCAATATGCCATGGTTCAAGGGATGGAAAGTGGAGAGGAAGGAAGGCAATGCCAAT
Gallus           TCCCCCAATATGCCTTGGTTCAAGGGCTGGAAGGTGGAGCGCAAGGAAGGCAACGCAAGC
                 **  * ** ***** ***** ** ** *****  * **  * *****     * **

Danio            GGTGTTACTCTATTGGAAGCTCTTGATACCATCATGCCTCCAACACGGCCCACTGATAAA
Oryctolagus      GGCGTGTCCCTGCTCGAGGCGCTGGACACCATCCTGCCCCCCCACGCGCCCCCACAGACAAG
Homo             GGCGTGTCCCTGCTGGAGGCCCTGGACACCATCCTGCCCCCCCACGCGCCCCCACGGACAAG
Mus              GGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCCCACCCGCCCCCACTGACAAG
Bos              GGCGTGTCCCTCCTGGAGGCCCTGGACACCATCCTGCCCCCCCACACGCCCCCACAGACAAG
Xenopus          GGAGTTTCCCTGCTTGAGGCTTTGGACACCATCCTTCCCCCAACTCGCCCAACAGACAAA
```

```
Gallus          GGGGTGTCCCTCCTGGAGGCCCTGGACACCATCCTGCCCCCCACCCGCCCCACAGACAAA
                ** **  * **  * ** **  * ** ** *** * ** ** ** ** ** ** ** **

Danio           CCCTTACGTCTTCCACTACAAGATGTCTACAAGATTGGAGGAATCGGGACTGTGCCAGTG
Oryctolagus     CCGCTGCGCCTGCCCCTGCAGGACGTGTACAAGATTGGCGGCATCGGCACGGTGCCCGTG
Homo            CCCCTGCGCCTGCCGCTGCAGGACGTGTACAAGATTGGCGGCATTGGCACGGTGCCCGTG
Mus             CCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGCATTGGGACCGTGCCTGTG
Bos             CCCCTGCGTCTGCCACTGCAGGATGTGTACAAGATTGGTGGCATTGGCACTGTGCCCGTG
Xenopus         CCTCTGCGTCTTCCCCTGCAAGATGTCTATAAAATTGGAGGAATCGGCACAGTTCCAGTG
Gallus          CCCCTGCGCCTGCCCCTGCAGGATGTCTACAAAATTGGAGGAATTGGCACAGTTCCCGTG
                **  * ** ** ** ** ** ** ** ** ** ***** ** ** ** ** ** ** ***

Danio           GGCAGGGTAGAGACGGGTGTTCTCCGGCCCAGTATGGTGGTGACCTTTGCCCCAGTCAAC
Oryctolagus     GGCCGCGTGGAGACCGGCATCCTGCGGCCCGGCATGGTGGTGACCTTTGCCCCCGTGAAC
Homo            GGCCGGGTGGAGACCGGCATCCTGCGGCCGGGCATGGTGGTGACCTTTGCGCCAGTGAAC
Mus             GGCCGAGTGGAGACCGGTATCCTCCGGCCTGGTATGGTGGTGACCTTTGCGCCAGTCAAC
Bos             GGCCGAGTGGAGACAGGGATCCTGCGGCCTGGCATGGTGGTGACCTTCGCGCCCGTGAAC
Xenopus         GGTCGTGTAGAGACTGGCATTCTAAAGCCGGGCATGGTGGTGACCTTTGCTCCAGTCAAT
Gallus          GGCCGAGTGGAGACTGGCATCCTGCGACCCGGCATGGTGGTCACCTTTGCGCCTGTGAAT
                **  * ** ***** **  * **    **  * ******** ***** ** ** ** **

Danio           ATCACTACAGAAGTGAAGTCCGTGGAGATGCATCACGAGTCTCTAAGTGAAGCTCTTCCA
Oryctolagus     ATCACCACGGAGGTGAAGTCGGTGGAGATGCACCATGAGGCGCTGAGCGAGGCGCTGCCC
Homo            ATCACCACTGAGGTGAAGTCAGTGGAGATGCACCACGAGGCTCTGAGCGAAGCTCTGCCC
Mus             ATCACCACAGAGGTGAAGTCTGTGGAAATGCACCATGAGGCACTTAGCGAGGCCCTGCCT
Bos             ATCACCACGGAGGTGAAGTCGGTGGAGATGCACCACGAGGCTCTGAGTGAGGCCCTTCCT
Xenopus         ATCACAACTGAGGTCAAGTCCGTTGAGATGCACCATGAGGCTCTGAGCGAGGCTCTGCCT
Gallus          ATCACCACTGAGGTGAAGTCAGTGGAGATGCACCACGAGGCGCTGAGCGAGGCCCTGCCT
                ***** ** ** ** ***** ** ** ***** ** *** * ** ** ** ** ** **

Danio           GGAGACAATGTGGGCTTTAATGTGAAGAACGTGTCCGTAAAAGACATTCGAAGAGGTAAC
Oryctolagus     GGGGACAACGTGGGCTTCAACGTCAAGAACGTGTCCGTGAAGGACATCCGGCGGGGCAAC
Homo            GGCGACAACGTCGGCTTCAATGTGAAGAACGTGTCGGTGAAGGACATCCGGCGGGGCAAC
Mus             GGTGACAATGTCGGGTTCAATGTGAAGAATGTGTCCGTTAAGGATATTCGCCGGGGCAAT
Bos             GGGGACAATGTTGGCTTCAACGTGAAGAACGTGTCAGTCAAGGACATCCGCCGGGGCAAC
Xenopus         GGGGACAATGTTGGCTTCAATGTCAAGAACGTGTCAGTAAAGGACATTCGCCGAGGCAAC
Gallus          GGAGACAATGTTGGCTTCAACGTGAAGAATGTCTCCGTCAAGGACATCCGCCGTGGGAAT
                ** ***** ** ** ** ** ** ***** ** ** ** ** ** ** **  * ** **

Danio           GTTTGTGGAGACAGTAAGTCCGACCCGCCTCAGGAAGCATCAGGGTTTACAGCACAGGTC
Oryctolagus     GTGTGTGGGGACAGCAAGTCCGACCCGCCGCAGGAGGCCGCGCAGTTCACCTCCCAGGTC
Homo            GTGTGTGGGGACAGCAAGTCTGACCCGCCGCAGGAGGCTGCTCAGTTCACCTCCCAGGTC
Mus             GTCTGCGGGGACAGCAAAGCTGACCCGCCTCAGGAGGCTGCCCAGTTCACCTCTCAGGTT
Bos             GTGTGTGGGGACAGCAAGTCCGACCCACCCCAGGAAGCCGCCCAGTTCACGTCCCAGGTC
Xenopus         GTTTGTGGGGACAGCAAGAGTGACCCACCCCAGGAAGCTGCTGGTTTCACTTCTCAGGTG
Gallus          GTCTGTGGGGACAGCAAGTCAGACCCGCCGCAGGAGGCAGCACAGTTCACGTCTCAGGTG
                ** ** ** ***** **    ***** ** ***** **  *    ** **  * *****

Danio           ATCATTTTGAATCACCCAGGACAGATCAGTTCAGGTTACTCTCCTGTCATAGACTGTCAC
Oryctolagus     ATCATCCTGAACCACCCCGGCCAGATCAGCGCCGGCTACTCGCCGGTCATCGACTGCCAC
Homo            ATCATCCTGAACCACCCGGGGCAGATTAGCGCCGGCTACTCCCCGGTCATCGACTGCCAC
Mus             ATCATCCTGAACCACCCTGGGCAAATCAGCGCTGGCTACTCGCCAGTCATCGACTGTCAC
Bos             ATCATTCTGAACCACCCTGGGCAGATCAGCGCTGGCTACTCACCAGTCATTGACTGCCAC
Xenopus         ATCATCTTAAACCACCCTGGTCAGATCAGTGCTGGATATTCCCCAGTCATTGACTGCCAC
Gallus          ATCATCCTGAACCACCCTGGGCAGATCAGCGCCGGATACTCACCTGTCATCGACTGCCAC
                *****  * ** ***** ** ** ** **  * ** ** ** ** ***** ***** ***

Danio           ACTGCTCATATCGCCTGCAAGTTTGCTGAACTCAAGGAGAAGATTGATCGCCGCTCAGGC
Oryctolagus     ACGGCCCACATCGCCTGCAAGTTCGCCGAGCTCAAGGAGAAGATCGACCGGCGCTCGGGC
Homo            ACAGCCCACATCGCCTGCAAGTTTGCGGAGCTGAAGGAGAAGATTGACCGGCGCTCTGGC
```

```
Mus             ACGGCCCACATTGCCTGCAAGTTTGCCGAGCTAAAGGAGAAGATTGACCGTCGTTCTGGC
Bos             ACAGCCCACATCGCCTGCAAGTTTGCTGAGCTGAAGGAGAAGATTGACCGGCGCTCTGGC
Xenopus         ACTGCCCACATCGCCTGTAAGTTTGCAGAGCTGAAAGAGAAGATCGATCGCCGGTCCGGC
Gallus          ACTGCTCACATCGCCTGCAAGTTTGCTGAGCTGAAGGAGAAGATCGACCGACGCTCTGGC
                ** ** ** ** ***** ***** ** ** ** ** ******** ** ** ** ** ***

Danio           AAGAAGCTAGAAGACAATCCCAAAAGCCTGAAGTCTGGAGATGCCGCCATAGTGGACATG
Oryctolagus     AAGAAGCTGGAGGACAACCCCAAGTCCCTCAAGTCCGGGGACGCGGCCATCGTGGAGATG
Homo            AAGAAGCTGGAGGACAACCCCAAGTCCCTGAAGTCTGGAGACGCGGCCATCGTGGAGATG
Mus             AAGAAGCTGGAGGATAACCCCAAGTCCCTGAAGTCTGGTGATGCAGCCATTGTTGAGATG
Bos             AAGAAGTTGGAGGACAACCCCAAGTCCCTGAAGTCCGGTGATGCAGCCATTGTGGAGATG
Xenopus         AAGAAGCTTGAGGACAACCCCAAGTCCCTGAAATCTGGAGACGCGGCTATTGTGGAGATG
Gallus          AAGAAGCTGGAGGACAACCCCAAATCCCTGAAATCGGGTGATGCGGCCATCGTGGAGATG
                ****** * ** ** ** *****   *** ** ** ** ** ** ** ** ** ** ***

Danio           ATCCCAGGAAAACCAATGTGTGTGGAGAGCTTCTCTCAGTATCCTCCACTGGGACGCTTT
Oryctolagus     GTGCCCGGGAAGCCCATGTGTGTGGAGAGCTTCTCCCAGTACCCGCCCCTCGGCCGCTTC
Homo            GTGCCGGGAAAGCCCATGTGTGTGGAGAGCTTCTCCCAGTACCCGCCTCTCGGCCGCTTC
Mus             GTCCCTGGAAAACCCATGTGTGTGGAGAGCTTCTCACAGTACCCACCTCTCGGCCGCTTC
Bos             GTCCCGGGGAAGCCTATGTGTGTGGAGAGCTTCTCCCAGTACCCACCTCTCGGCCGCTTC
Xenopus         ATCCCTGGGAAGCCTATGTGTGTGTAGAGAGCTTCTCCCAGTACCCACCTCTTGGGCGCTTT
Gallus          ATTCCTGGCAAGCCGATGTGTGTGGAGAGCTTCTCCCAGTACCCACCCCTTGGCCGCTTT
                 * ** ** ** ** ******** ********** ***** ** ** ** ** *****

Danio           GCTGTCCGAGATATGAGACAGACCGTTGCAGTCGGTGTGATCAAAAATGTGGAGAAGAAG
Oryctolagus     GCCGTGCGCGACATGCGGCAGACGGTGGCCGTGGGCGTCATCAAGAACGTGGAGAAGAAG
Homo            GCCGTGCGCGACATGAGGCAGACGGTGGCCGTAGGCGTCATCAAGAACGTGGAGAAGAAG
Mus             GCCGTGCGCGACATGCGGCAGACTGTGGCCGTGGGCGTCATCAAGAACGTGGAGAAGAAG
Bos             GCCGTGCGCGACATGCGGCAGACAGTGGCCGTGGGCGTCATCAAGAACGTGGAGAAGAAG
Xenopus         GCAGTGAGAGACATGAGGCAGACTGTGGCCGTGGGAGTCATTAAGAACGTGGAGAAGAAA
Gallus          GCTGTCCGTGACATGCGGCAGACCGTGGCCGTGGGCGTCATCAAGAACGTGGAGAAGAAG
                ** **  * ** *** * ***** ** ** ** ** ** ** ** ** ***********

Danio           ATTGGCGGCAGCGGGAGAGTGACCAAATCAGCTCAGAAAGCTCAAAAATCTAGCAAATGA
Oryctolagus     AGCGGCGGCGCCGGCAAGGTCACCAAGTCGGCGCAGAAGGCGCAGAAGGCCGGCAAGTGA
Homo            AGCGGCGGCGCCGGCAAGGTCACCAAGTCGGCGCAGAAGGCGCAGAAGGCGGGCAAGTGA
Mus             AGCGGCGGCGCAGGCAAGGTCACCAAGTCCGCACAGAAGGCTCAGAAAGCGGGCAAGTGA
Bos             AGCGGCGGCGCCGGCAAGGTCACCAAGTCGGCGCAGAAGGCACAGAAGGCGGGCAAGTGA
Xenopus         AGCGGAGGAGCCGGCAAGGTGACCAAGTCCGCACAGAAAGCCCAGAAGGCTGGCAAATGA
Gallus          AGCGGCGGGGCCGGCAAAGTCACCAAGTCTGCCCAGAAGGCCCAGAAGGCTGGCAAATGA
                *  ** **    ** *  ** ***** ** ** ***** ** ** **  *  **** ***

Danio           ATCTGAATCTCCAAGACAGTCACCTTA------GGCCCTGTCCCA-GCTTACATGCCTCT
Oryctolagus     AGCGCGGGCGCCCGCGGCGCGGCCCCCGCCGGCGGCGCCGCGCCCCCGCCCCCGCCCCCG
Homo            AGCGCGGGCGCCCGCGGCGCGACCCTCCCCGGCGGCGCCGCGCTCCGAACCCCGGCCCG-
Mus             AGCGCGGGCGCCCGCGGCGTGGCCCTCCCCAACGGCGCCGCGCCGCGCCCCCAGCCCCG-
Bos             AGCGCGGGCGCCCGCGGCGCGACCCTCCCCGGCGGCGCCGCGCCCCGCCCCCGCCCCCA-
Xenopus         ATTGCCGGTTCCCTCCGTCTGGCACAC--------------------AAGCCCTGCCCC-
Gallus          ATCGTGGGCTCCCAGTGCGTAGCGCAG--------------------AAACCATCCCTG-
                *         **          *                           *   **

Danio           CTCATTTAGGCATGCTCAGTCAGTTCCTTCCCTGTGTGCTTGAAATATATACTCGAACCA
Oryctolagus     C----CGCGGC-CGCGCGCCGCGCCCCGCCCCCCG--------------GCCCCGCCCCG
Homo            ------------------------------------G--------------CCCCC------
Mus             ----------------------------CCCCCGG--------------CACTG------
Bos             ---------------------------CC-------------------------------
Xenopus         ----------------------------T--CTGG--------------GAA--------
Gallus          ----------------------------ACACCAG--------------GAC--------
```

```
Danio           ACTGGA-GTT-TGATAGACTGAAGGAAAATATTGAAAAACTAGCATTATCACATTTTGAC
Oryctolagus     GCCGCGCGCCCCGCC---CCGCCC--------CCAGACCCCGGCCCT-GCCCC------G
Homo            -------GCCCCGCC----------------------------CCCG-CCCCG------C
Mus             -------GCCCCGCC----------------------------CCCG-CCCCA------G
Bos             ---------CCCGGC----------------------------CCCG-TCCCA------G
Xenopus         -------GGTCCGTC--------------------------------------------
Gallus          -------GCT--GCC--------------------------------------------
                           *

Danio           TGGTTGCACTGTATATTCACTTTTAATAGCAGACCGGTACACACGTTGCATGGATGTTTG
Oryctolagus     GCGCGGCCCCG------------------GCGCCGCGCGCCCCCGCCAGGCGCACGTCTG
Homo            GCGCCGCTCCG------------------GCGCCCCGCACCCCCGCCAGGCGCATGTCTG
Mus             GCGCGGCCCCT------------------CTGCCCCGACCCCCTGCCAGGCGCATGTCTG
Bos             GCGCGGCCCCG------------------GCGCCCCGCCCCCCCGCCAGGCGCATGTCTG
Xenopus         ------------------------------------CCCTTCTCTTTGAGGTGCACGCCAG
Gallus          ------------------------------------ACCGTCTCCCCCCGGCGCATGTGTG
                                                                 * * *   *

Danio           CACGACATCTGTTAAATGAATGTAGCTGGTATTGCTGTGTGTGTGCGTGCGTGCGTGCGT
Oryctolagus     CACCTCCGCTTGTCGGCGGCTGTC-----------------GGTCAGCGACTGG------
Homo            CACCTCCGCTTGCCAGAGGCCCTC-----------------GGTCAGCGACTGG------
Mus             CACCTCCGCTTGTAAGAGGCTCTA-----------------CGTCAGCGACTGG------
Bos             CACCTCCGCTTGTAAGAGGCTCTC-----------------CGTCAGCGACTGG------
Xenopus         CCCATCTGCTTGTAAAAGCCTGTA-----------------TGTCAACGACTGG------
Gallus          CACATCAGCTTGTAAGAGTTTATA-----------------TGTCAACGACTGG------
                * *  *  **       *    *                  **    *  **

Danio           GCGTGCGTGTGTGCGCGCATGTTGAAAGAAGAAACTGCACTTTGACCAAACTGGAAAGCA
Oryctolagus     ----------ATGCTCGC--CATCAAGGT--CCAGTGGAAGTTCTTCAAGCGGAAAG-CG
Homo            ----------ATGCTCGC--CATCAAGGT--CCAGTGGAAGTTCTTCAAGAGGAAAGGCG
Mus             ----------ATGCTCGC--CATCAAAGT--CCAGTGGAAATTCTTCAAGAGGAAAAGCG
Bos             ----------ATGCTCGC--CATCAAGGT--CCAGTGGAAGTTCTTTAAGAGGAAAAGCG
Xenopus         ----------ATGCTCAC--CATTAAAGT--CCAGTGGAAGTTCTTTAAGAGGAAAAGCA
Gallus          ----------ATGCTCAC--CATTAAGGT--CCAGTGGAAATTCTTTAAAAGGAAAAGCA
                           *** * *    * ** *     * ** *  **    **   * **  *

Danio           AAC-TGCTGATGATAATTTTGTATGATTTATAAATGAGCACTGATGATGAAAGGCTATTC
Oryctolagus     ---CCGCCGCC-----------------------------------------CCGGCTTCGC
Homo            CCCCCGCC--C-----------------------------------------CAGGCTTCCG
Mus             CCCCCGCC--C-----------------------------------------CAGGCTTCCG
Bos             TCCCCGCCGCC-----------------------------------------CCGGCTTCCG
Xenopus         TGC-------------------------------------------------------------
Gallus          TGT-------------------------------------------------------------


Danio           GATCCAAAGCACCATTTGTCTCCTTTCGGACCCTTGCTGC---TCAAA--------CCTG
Oryctolagus     GCCC-----GCG-----CCCCCGCCCCGTGCCCGTGTTTCCAATAAACCGAGC-CCCCG-
Homo            CGCC-----CAGCGCTCGCCACGCTCAGTGCCCGTTTTACCAATAAACTGAGC-GACCCC
Mus             CG-C-----CAGCGCTCACCACGCTCAGTGCCCGTTTTCCCAATAAACTGAGC-GACCCC
Bos             CGTC-----CAGCCTTTGTCACGCTCAGTGCCCGTTTTACCAATAAACTGAGC-GACCCC
Xenopus         --TC-----CTGC-TCTGTAATCTTTAGTGTCCATTTTACCAATTAAACTGGTTCAACAT
Gallus          --TC-----CAGCGTTTGTGAGGCTTCATGTTAATTTTACCAATAAAACTGGTACAACAT
                   *                             *  * *   * **          *

Danio           TGACAAATAAAG---------A-----AACT-GTTGT---CCA-----AAAAA-AAA-AA
Oryctolagus     -------------------------------------------------------------
Homo            AGAAAAAAAAAAAAAAAAAA---------------------------------------
Mus             AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA-----A-AAA-AAA-AA
Bos             CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
Xenopus         TCAAAAAAAA--------AAAAAAA--AAAA-AA------A-------A-A---A-----
```

```
Gallus          CCACA-------------------------------------------------------


Danio           A-AA-AAA-A-AA-AAAAA
Oryctolagus     --------------------
Homo            --------------------
Mus             A-AA-AAA-A-AA-AAAAA
Bos             AAAAAAAAAAAAAAAAAAA
Xenopus         --------------------
Gallus          --------------------
```

CLUSTAL O(1.2.4) multiple sequence alignment of EEF1A1.

```
Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             CTGGTGCCTGGTGGAGGCGGCGCGGGGTAATCTGGGAAAGTGGTGTCGTGTGCTGGCTCC
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------


Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          -------------------------------------------------ACGCCGTGCGGGT
Bos             GCCCTTTTCCCCGAGGGTGGGGGGAGAACCGTATATAAGTGCCGTAGTCTCCGTGAACGT
Mus             -------TTCCCGAGGGTGG-GGGAGAACGGTATATAAGTGCGGCAGTCGCCTTGGACGT
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------


Danio           ----------------------------GAGTGATCTCTC--------AATCTTGAAA---
Xenopus         --------GCGGCGAGTT------------TTAAGTGTC--------CACCGCCAAACAT
Gallus          GTCGTTTCTC-----TTTG---GCCGGAAGAAAG---------------AAGCTAAAG--
Bos             TCTTTTTCGCAACGGGTTTGCCGCCGGGACACAGGTGTCGTGAAAACCACCGTTAAAC--
Mus             TCTTTTTCGCAACGGGTTTGCCGTCAGAACGCAGGTGTTGTGAAAACCACCGCTAATT--
Homo            -CTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTGTCGTGAAAACTACCCCTAA----
Oryctolagus     --------------------------GGCACGAGCTCGTGCTGAAAACCACCGCTAAAT--
                                                                      **


Danio           ---CTTATCAATCATGGGAAAGGAAAAGACCCACATTAACATCGTGGTTATTGGCCACGT
Xenopus         CTAACAATCCACAATGGGAAAGGAAAAGACTCACATCAACATCGTCGTCATTGGACACGT
Gallus          ---ACCATCCGAAATGGGAAAGGAGAAGACCCACATCAACATCGTCGTCATCGGCCACGT
Bos             ---CTAAGCCAAAATGGGAAAGGAGAAGACCCACATCAACATCGTTGTCATTGGGCACGT
Mus             ---CAAAGCAAAAATGGGAAAGGAAAAGACTCACATCAACATCGTCGTAATCGGACACGT
Homo            -----AAGCCAAAATGGGAAAGGAAAAGACTCATATCAACATTGTCGTCATTGGACACGT
Oryctolagus     ---CAAAGCCAAAATGGGAAAGGAAAAGACTCACATCAACATCGTCGTCATTGGCCACGT
                      * *    *********** ***** ** ** ***** ** ** ** ** *****


Danio           CGACTCCGGAAAGTCCACCACCACCGGCCATCTGATCTACAAATGCGGTGGAATCGACAA
Xenopus         AGATTCTGGAAAGTCCACAACAACTGGACATCTTATCTACAAATGTGGTGGTATCGACAA
Gallus          CGATTCCGGCAAGTCCACCACCACCGGGCACCTCATCTACAAATGTGGTGGCATCGACAA
Bos             AGATTCAGGGAAGTCTACCACGACTGGCCATCTGATCTATAAATGTGGCGGGATCGACAA
Mus             AGATTCCGGCAAGTCCACCACAACCGGCCACCTGATCTACAAATGTGGTGGAATCGACAA
Homo            AGATTCGGGCAAGTCCACCACTACTGGCCATCTGATCTATAAATGCGGTGGCATCGACAA
Oryctolagus     AGATTCGGGCAAGTCCACCACCACTGGCCATCTGATCTACAAATGTGGTGGCATCGACAA
                 ** ** ** ***** ** ** ** ** ** ** ***** ***** ** ** ********
```

```
Danio            GAGAACCATCGAGAAGTTCGAGAAGGAAGCCGCTGAGATGGGCAAGGGCTCCTTCAAGTA
Xenopus          GAGAACCATCGAAAAGTTCGAGAAGGAAGCTGCTGAGATGGGCAAGGGCTCCTTCAAGTA
Gallus           GAGGACCATCGAGAAGTTCGAGAAGGAGGCGGCCGAGATGGGCAAAGGTTCCTTCAAATA
Bos              GAGAACAATTGAAAAGTTCGAGAAGGAGGCTGCCGAGATGGGAAAGGGCTCCTTCAAATA
Mus              GCGAACCATCGAAAAGTTTGAGAAGGAGGCTGCTGAGATGGGAAAGGGCTCCTTCAAGTA
Homo             AAGAACCATTGAAAAATTTGAGAAGGAGGCTGCTGAGATGGGAAAGGGCTCCTTCAAGTA
Oryctolagus      AAGAACCATTGAAAAATTTGAGAAGGAGGCTGCCGAGATGGGAAAGGGCTCCTTCAAGTA
                   * ** ** ** ** ** ******** ** ** ******** ** ** ******** **

Danio            CGCCTGGGTGTTGGACAAACTGAAGGCCGAGCGTGAGCGTGGTATCACCATTGACATTGC
Xenopus          CGCCTGGGTCTTGGACAAACTGAAGGCCGAGCGTGAACGTGGTATCACCATTGACATCTC
Gallus           TGCCTGGGTCTTGGACAAGCTCAAGGCTGAGCGTGAGCGTGGTATCACTATCGATATTTC
Bos              TGCCTGGGTCTTGGACAAACTTAAAGCTGAACGTGAGCGTGGTATCACCATTGATATCTC
Mus              CGCCTGGGTCTTAGACAAACTGAAAGCTGAGCGTGAGCGTGGTATCACTATTGACATCTC
Homo             TGCCTGGGTCTTGGATAAACTGAAAGCTGAGCGTGAACGTGGTATCACCATTGATATCTC
Oryctolagus      TGCCTGGGTCTTGGATAAACTGAAAGCCGAGCGTGAGCGTGGTATCACCATCGACATCTC
                  ******** ** ** ** ** ** ** ** ***** ********** ** ** **  *

Danio            TCTCTGGAAATTCGAGACCAGCAAATACTACGTCACCATCATTGATGCCCCTGGACACAG
Xenopus          CCTGTGGAAATTTGAGACCAGCAAATATTATGTTACTATCATTGATGCTCCAGGACACAG
Gallus           CCTGTGGAAATTTGAAACAAGCAAGTACTACGTCACCATCATCGATGCTCCTGGGCACAG
Bos              CCTGTGGAAATTTGAGACCAGCAAGTACTATGTTACCATCATTGATGCCCCAGGACACAG
Mus              CCTGTGGAAATTCGAGACCAGCAAATACTATGTGACCATCATTGATGCCCCAGGACACAG
Homo             CTTGTGGAAATTTGAGACCAGCAAGTACTATGTGACTATCATTGATGCCCCAGGACACAG
Oryctolagus      CCTGTGGAAATTTGAGACCAGCAAGTATTACGTGACTATCATTGATGCCCCAGGACACAG
                   * ******** ** ** ***** ** ** ** ** ***** ***** ** ** *****

Danio            AGACTTCATCAAGAACATGATCACTGGTACTTCTCAGGCTGACTGTGCTGTGCTGATTGT
Xenopus          AGACTTCATCAAGAACATGATCACTGGTACCTCTCAGGCTGACTGTGCTGTCCTGATTGT
Gallus           AGACTTCATTAAGAACATGATTACTGGAACTTCTCAGGCTGATTGTGCTGTCCTGATTGT
Bos              AGACTTCATCAAAAACATGATTACAGGCACATCCCAGGCTGACTGTGCTGTCCTGATCGT
Mus              AGACTTCATCAAAAACATGATTACAGGCACATCCCAGGCTGACTGTGCTGTCCTGATTGT
Homo             AGACTTTATCAAAAACATGATTACAGGGACATCTCAGGCTGACTGTGCTGTCCTGATTGT
Oryctolagus      AGACTTCATCAAAAACATGATTACAGGCACATCTCAGGCTGACTGTGCCGTCCTGATTGT
                 ****** ** ** ******** ** ** ** ** ******** ***** ** ***** **

Danio            TGCTGGTGGTGTCGGTGAGTTTGAGGCTGGTATCTCCAAGAACGGACAGACCCGTGAGCA
Xenopus          TGCTGCTGGTGTTGGTGAATTTGAAGCTGGTATCTCAAAGAACGGACAAACTCGTGAGCA
Gallus           TGCTGCTGGTGTTGGTGAGTTCGAGGCCGGTATTTCTAAGAACGGGCAGACCCGTGAGCA
Bos              TGCTGCTGGTGTTGGTGAATTTGAAGCCGGTATCTCCAAGAACGGGCAGACCCGTGAGCA
Mus              TGCTGCTGGTGTTGGTGAATTTGAAGCTGGTATCTCCAAGAACGGGCAGACCCGCGAGCA
Homo             TGCTGCTGGTGTTGGTGAATTTGAAGCTGGTATCTCCAAGAATGGGCAGACCCGAGAGCA
Oryctolagus      TGCTGCTGGTGTCGGGGAATTTGAAGCTGGTATCTCCAAGAACGGGCAGACCCGTGAGCA
                 ***** ****** ** ** ** ** ** ***** ** ***** ** ** ** ** *****

Danio            CGCCCTCCTGGCTTTCACCCTGGGAGTGAAACAGCTGATCGTTGGAGTCAACAAGATGGA
Xenopus          TGCCCTCCTTGCCTACACTCTGGGAGTAAAGCAACTGATCGTTGGTGTTAACAAAATGGA
Gallus           CGCTCTTCTGGCCTACACCCTGGGTGTGAAACAGCTGATCGTTGGTGTTAACAAGATGGA
Bos              TGCCCTTTTGGCTTACACCCTGGGTGTGAAACAACTAATTGTTGGCGTTAACAAAATGGA
Mus              TGCTCTTCTGGCTTACACCCTGGGTGTGAAACAGCTGATTGTTGGTGTCAACAAAATGGA
Homo             TGCCCTTCTGGCTTACACACTGGGTGTGAAACAACTAATTGTCGGTGTTAACAAAATGGA
Oryctolagus      TGCCCTTCTGGCTTACACGCTGGGTGTGAAACAGCTAATTGTTGGTGTTAACAAGATGGA
                  ** **  * ** * *** ***** ** ** ** ** ** ** ** ** ***** *****

Danio            CTCCACTGAGCCCCCTTACAGCCAGGCTCGTTTTGAGGAAATCACCAAGGAAGTCAGCGC
Xenopus          TTCAACTGAACCCCCATACAGCCAGAAAAGATATGAGGAAATCGTTAAGGAAGTCAGCAC
Gallus           TTCCACTGAGCCACCTTACAGCCAGAAGAGATACGAAGAGATCGTCAAGGAAGTCAGCAC
Bos              TTCCACTGAGCCACCCTATAGCCAGAAGAGATACGAAGAAATTGTTAAGGAAGTCAGCAC
Mus              TTCCACCGAGCCACCATACAGTCAGAAGAGATACGAGGAAATCGTTAAGGAAGTCAGCAC
Homo             TTCCACTGAGCCACCCTACAGCCAGAAGAGATATGAGGAAATTGTTAAGGAAGTCAGCAC
```

```
Oryctolagus      TTCCACTGAGCCACCCTACAGCCAGAAGAGATACGAGGAAATCGTTAAGGAAGTCAGCAC
                  ** ** ** ** ** ** ** ***    * *  ** ** **    ************ *

Danio            ATACATCAAGAAGATCGGCTACAACCCTGCCAGTGTTGCCTTCGTCCCAATTTCAGGATG
Xenopus          ATACATCAAGAAGATTGGTTACAACCCTGATACTGTTGCCTTTGTACCTATTTCTGGATG
Gallus           TTACATCAAGAAAATTGGCTACAACCCAGACACTGTAGCTTTTGTGCCAATCTCTGGTTG
Bos              CTATATTAAGAAAATTGGCTACAACCCCGACACAGTAGCATTTGTGCCAATTTCTGGCTG
Mus              CTACATTAAGAAAATTGGCTACAACCCTGACACAGTAGCATTTGTGCCAATTTCTGGTTG
Homo             TTACATTAAGAAAATTGGCTACAACCCCGACACAGTAGCATTTGTGCCAATTTCTGGTTG
Oryctolagus      CTACATTAAGAAAATTGGCTACAACCCTGACACAGTAGCATTTGTGCCAATTTCTGGTTG
                  ** ** ***** ** ** ******** *  *  ** ** ** ** ** ** ** ** **

Danio            GCACGGTGACAACATGCTGGAGGCCAGCTCAAACATGGGCTGGTTCAAGGGATGGAAGAT
Xenopus          GAACGGTGACAACATGCTTGAGCCCAGCGCCAATATGCCTTGGTTTAAGGGGTGGAAAAT
Gallus           GAACGGGGACAACATGCTGGAGCCTAGCTCTAACATGCCCTGGTTCAAGGGATGGAAGGT
Bos              GAATGGTGACAACATGCTAGAACCAAGTGCTAATATGCCATGGTTCAAGGGATGGAAAGT
Mus              GAATGGTGACAACATGCTGGAGCCAAGTGCTAATATGCCTTGGTTCAAGGGATGGAAAGT
Homo             GAATGGTGACAACATGCTGGAGCCAAGTGCTAACATGCCTTGGTTCAAGGGATGGAAAGT
Oryctolagus      GAACGGTGACAACATGCTGGAGCCAAGTGCTAATATGCCGTGGTTCAAGGGATGGAAAGT
                 * * ** ************ **  * **  * ** ***   ***** ***** *****  *

Danio            TGAGCGCAAGGAGGGTAATGCTAGCGGTACTACTCTTCTTGATGCCCTTGATGCCATTCT
Xenopus          CTCACGTAAAGAGGGATCTGGCAGCGGAACTACCCTGCTGGAAGCTCTTGACTGCATTTT
Gallus           TACCCGGAAAGATGGCAATGCCAGTGGAACCACCCTCCTGGAAGCCTTGGACTGCATCCT
Bos              CACCCGTAAGGACGGCAATGCCAGTGGAACCACCCTGCTTGAAGCTCTGGATTGCATTCT
Mus              CACCCGCAAAGATGGCAGTGCCAGTGGCACCACGCTGCTGGAAGCTTTGGATTGTATCCT
Homo             CACCCGTAAGGATGGCAATGCCAGTGGAACCACGCTGCTTGAGGCTCTGGACTGCATCCT
Oryctolagus      CACCCGCAAAGATGGCAATGCCAGTGGAACCACACTGCTTGAAGCCCTGGACTGCATCCT
                     ** ** ** **   **  ** ** ** ** ** ** ** **  * **    **  *

Danio            GCCCCCTAGCCGTCCCACCGACAAGCCCCTCCGTCTGCCACTTCAGGATGTGTACAAAAT
Xenopus          GCCACCAAGTCGCCCAACTGATAAGCCTCTGCGTCTGCCTCTGCAGGATGTCTACAAAAT
Gallus           GCCTCCAACTCGTCCAACTGACAAACCTCTGCGTCTGCCTCTTCAAGATGTCTACAAAAT
Bos              GCCACCAACTCGCCCAACTGACAAACCCTTGCGTTTGCCTCTCCAGGATGTCTATAAAAT
Mus              ACCACCAACTCGTCCAACTGACAAGCCCCTGCGACTGCCCCTCCAGGATGTCTATAAAAT
Homo             ACCACCAACTCGTCCAACTGACAAGCCCTTGCGCCTGCCTCTCCAGGATGTCTACAAAAT
Oryctolagus      TCCACCAACTAGACCAACTGACAAGCCTCTGCGTCTGCCCCTACAGGATGTCTACAAAAT
                  ** ** *   * ** ** ** ** **  * **  **** ** ** ***** ** *****

Danio            TGGAGGTATTGGAACTGTACCTGTGGGTCGTGTGGAGACTGGTGTCCTCAAGCCTGGTAT
Xenopus          TGGCGGTATTGGTACTGTACCAGTTGGTCGTGTGGAGACTGGTGTCCTTAAGCCAGGCAT
Gallus           TGGTGGCATTGGTACTGTACCAGTTGGCCGTGTGGAAACTGGTGTCCTGAAGCCAGGTAT
Bos              TGGTGGTATTGGTACTGTCCCTGTGGGTCGTGTGGAGACTGGTGTTCTCAAACCTGGCAT
Mus              TGGAGGCATTGGCACTGTCCCTGTGGGCCGAGTGGAGACTGGTGTTCTCAAGCCTGGCAT
Homo             TGGTGGTATTGGTACTGTTCCTGTTGGCCGAGTGGAGACTGGTGTTCTCAAACCCGGTAT
Oryctolagus      TGGTGGTATTGGCACTGTCCCTGTGGGCCGAGTGGAGACTGGTGTTCTCAAACCTGGCAT
                 *** ** ***** ***** ** ** ** ** ***** ******** ** ** ** ** **

Danio            GGTTGTGACCTTCGCCCCTGCCAATGTAACCACTGAGGTCAAGTCTGTTGAGATGCACCA
Xenopus          GGTGGTTACTTTTGCCCCTGTTAATGTAACAACTGAAGTTAAATCTGTTGAAATGCACCA
Gallus           GGTGGTCACATTTGCCCCCGTCAACGTTACAACTGAAGTAAAGTCTGTTGAGATGCACCA
Bos              GGTGGTCACCTTTGCTCCAGTCAATGTAACAACTGAAGTGAAGTCTGTAGAAATGCACCA
Mus              GGTGGTTACCTTTGCTCCAGTCAATGTAACAACTGAAGTCAAGTCTGTTGAAATGCACCA
Homo             GGTGGTCACCTTTGCTCCAGTCAACGTTACAACGGAAGTAAAATCTGTCGAAATGCACCA
Oryctolagus      GGTGGTAACTTTTGCTCCAGTCAATGTCACAACTGAAGTCAAGTCCGTCGAAATGCACCA
                 *** ** ** ** ** ** *  ** ** ** ** ** ** ** ** ** ** ********

Danio            CGAGTCTCTGACTGAGGCCACTCCTGGTGACAACGTTGGCTTCAACGTTAAGAACGTGTC
Xenopus          TGAAGCCCTTAGCGAGGCCATGCCCGGTGACAATGTTGGCTTTAACGTGAAAAACGTTTC
Gallus           TGAAGCCCTTAGCGAAGCTCTGCCTGGTGATAACGTTGGCTTCAACGTCAAGAACGTGTC
```

```
Bos             TGAAGCATTGAGTGAAGCCCTTCCTGGGGACAATGTGGGCTTTAATGTCAAAAACGTGTC
Mus             TGAAGCTTTGAGTGAAGCTCTTCCTGGGGACAATGTGGGCTTCAATGTAAAGAACGTGTC
Homo            TGAAGCTTTGAGTGAAGCTCTTCCTGGGGACAATGTGGGCTTCAATGTCAAGAATGTGTC
Oryctolagus     TGAAGCTTTGAGTGAAGCTCTTCCTGGGGACAATGTGGGCTTCAATGTCAAGAACGTGTC
                 **  *  * *  ** **    ** ** ** ** ** ***** ** ** ** ** ** **

Danio           AGTCAAGGACATCCGTCGTGGTAATGTGGCTGGAGACAGCAAGAACGACCCACCCATGGA
Xenopus         TGTGAAGGACGTCCGTCGTGGCAACGTTGCTGGTGACAGCAAGAATGACCCACCAATGGA
Gallus          TGTGAAAGATGTCCGCCGTGGTAACGTTGCTGGTGACAGCAAGAATGATCCTCCAATGGA
Bos             TGTCAAAGATGTCCGTCGTGGCAATGTGGCTGGTGACAGCAAAAATGATCCACCCATGGA
Mus             GGTCAAAGATGTTAGACGAGGCAATGTTGCTGGTGACAGCAAAAACGACCCACCAATGGA
Homo            TGTCAAGGATGTTCGTCGTGGCAACGTTGCTGGTGACAGCAAAAATGACCCACCAATGGA
Oryctolagus     TGTCAAAGATGTTCGTCGTGGCAACGTTGCTGGTGACAGCAAAAATGACCCACCAATGGA
                 ** ** **  *  * ** ** ** ** ***** ******** ** ** ** ** *****

Danio           GGCTGCCAACTTCAACGCTCAGGTCATCATCCTGAACCACCCTGGTCAGATCTCTCAGGG
Xenopus         AGCTGGTACCTTTACAGCACAGGTTATCATCCTGAACCACCCAGGCCAGATTGGTGCTGG
Gallus          AGCTGCTGGCTTTACTGCGCAGGTTATTATCCTGAACCACCCTGGCCAAATCAGTGCTGG
Bos             AGCTGCTGGCTTCACAGCTCAGGTGATTATTTTGAACCATCCAGGCCAAATCAGTGCTGG
Mus             AGCAGCTGGCTTCACTGCTCAGGTGATTATCCTGAACCATCCAGGCCAAATCAGTGCTGG
Homo            AGCAGCTGGCTTCACTGCTCAGGTGATTATCCTGAACCATCCAGGCCAAATAAGCGCCGG
Oryctolagus     AGCAGCTGGCTTCACTGCTCAGGTGATCATCCTGAACCATCCAGGTCAGATCAGTGCTGG
                 ** *    *** *  ** ***** ** **  ******* ** ** ** **       **

Danio           TTACGCCCCAGTGCTGGATTGCCACACTGCTCACATCGCCTGCAAGTTTGCTGAGCTCAA
Xenopus         ATATGCCCCTGTGTTGGATTGCCACACAGCTCACATTGCTTGCAAGTTTGCTGAGCTGAA
Gallus          TTATGCCCCTGTGCTGGATTGCCACACTGCTCACATTGCCTGCAAGTTTGCTGAGCTCAA
Bos             ATATGCACCTGTGCTGGATTGTCACACAGCTCACATTGCTTGCAAGTTTGCTGAGCTGAA
Mus             CTACGCTCCTGTTCTGGATTGTCACACAGCCCACATAGCATGCAAGTTTGCTGAGCTTAA
Homo            CTATGCCCCTGTATTGGATTGCCACACGGCTCACATTGCATGCAAGTTTGCTGAGCTGAA
Oryctolagus     GTATGCCCCTGTACTGGATTGTCACACAGCTCACATTGCTTGCAAGTTTGCTGAGCTGAA
                 ** ** ** **  ******* ***** ** ***** ** ***************** **

Danio           GGAGAAGATCGACCGTCGTTCTGGCAAGAAGCTTGAAGACAACCCCAAGGCTCTCAAATC
Xenopus         GGAAAAGATTGATCGCCGTTCTGGTAAGAAACTGGAAGACAATCCCAAGTTCCTGAAGTC
Gallus          GGAGAAGATTGATCGTCGTTCCGGCAAGAAGCTGGAGGATGGCCCGAAGTTCCTGAAATC
Bos             GGAGAAGATTGATCGTCGTTCTGGGAAAAAGCTGGAAGATGGCCCTAAATTCTTGAAATC
Mus             AGAAAAGATCGATCGTCGTTCTGGTAAGAAGCTGGAAGATGGCCCCAAGTTCCTGAAGTC
Homo            GGAAAAGATTGATCGCCGTTCTGGTAAAAAGCTGGAAGATGGCCCTAAATTCTTGAAGTC
Oryctolagus     GGAAAAGATTGATCGTCGTTCTGGAAAGAAGTTGGAAGATGGCCCTAAATTCTTGAAATC
                 ** ***** ** ** ***** ** ** **  * ** **    ** **     * ** **

Danio           CGGAGATGCCGCCATTGTTGAGATGGTCCCTGGCAAGCCCATGTGTGTGGAGAGCTTCTC
Xenopus         TGGTGATGCTGCCATTGTTGACATGATCCCAGGAAAGCCTATGTGCGTGGAGAGCTTCTC
Gallus          TGGAGACGCTGCCATTGTTGATATGATCCCTGGCAAACCCATGTGTGTTGAGAGCTTCTC
Bos             TGGTGACGCTGCCATCGTTGATATGGTTCCTGGCAAGCCCATGTGTGTCGAGAGCTTCTC
Mus             TGGCGATGCTGCCATTGTTGATATGGTCCCTGGCAAGCCCATGTGTGTTGAGAGCTTCTC
Homo            TGGTGATGCTGCCATTGTTGATATGGTTCCTGGCAAGCCCATGTGTGTTGAGAGCTTCTC
Oryctolagus     TGGTGATGCTGCCATTGTTGATATGGTTCCCGGCAAGCCCATGTGTGTTGAGAGCTTCTC
                 ** ** ** ***** ***** *** * ** ** ** ** ***** ** ***********

Danio           TACCTACCCTCCTCTTGGTCGCTTTGCTGTGCGTGACATGAGGCAGACCGTTGCTGTCGG
Xenopus         TGACTACCCCCCTCTTGGTCGTTTTGCTGTCCGTGACATGAGGCAGACTGTTGCTGTAGG
Gallus          TGATTATCCTCCTCTGGGTCGTTTCGCTGTGCGTGACATGAGACAGACGGTTGCTGTTGG
Bos             TGATTATCCTCCCCTGGGCCGTTTTGCTGTGCGTGACATGAGACAGACAGTCGCTGTGGG
Mus             TGACTACCCTCCACTTGGTCGCTTTGCTGTTCGTGACATGAGGCAGACAGTTGCTGTGGG
Homo            AGACTATCCACCTTTGGGTCGCTTTGCTGTTCGTGATATGAGACAGACAGTTGCGGTGGG
Oryctolagus     TGACTATCCTCCTCTGGGTCGTTTCGCTGTCCGTGATATGAGACAGACGGTTGCTGTGGG
                    ** ** **  * ** ** ** ***** ***** ***** ***** ** ** ** **
```

```
Danio           CGTCATCAAGAGCGTTGAGAAGAAAATCGGTGGTGCTGGCAAGGTCACAAAGTCTGCACA
Xenopus         AGTCATCAAGGCGGTCGATAAGAAGGCTGCTGGAGCTGGCAAAGTCACAAAGTCTGCTCA
Gallus          TGTCATCAAGGCCGTCGACAAGAAGGCTGGTGGAGCCGGCAAGGTCACAAAGTCTGCTCA
Bos             TGTCATCAAAGCAGTGGACAAGAAGGCAGCTGGAGCTGGCAAGGTCACCAAGTCTGCCCA
Mus             TGTCATCAAAGCTGTGGACAAGAAGGCTGCTGGAGCTGGCAAAGTCACCAAGTCTGCCCA
Homo            TGTCATCAAAGCAGTGGACAAGAAGGCTGCTGGAGCTGGCAAGGTCACCAAGTCTGCCCA
Oryctolagus     TGTGATCAAAGCAGTGGACAAGAAGGCTGCTGGAGCTGGCAAGGTCACCAAGTCTGCCCA
                 ** *****    ** ** *****    * *** ** ***** ***** ******** **

Danio           GAAGGCTGCCAAGACCAAGTGAATTTCC-CTCAATC-------ACACCGTTC-----CAA
Xenopus         GAAAGCACAGAAAGGCAAATGAATATT---TCCAGCATCTCTCACCTCAGTCATAACCAG
Gallus          GAAGGCCCAGAAGGCTAAATGAAAATTCTGTACATCAGCTGCCACCTCAGTCGTAATCAG
Bos             GAAAGCTCAGAAGGCTAAATGAATATTATCCCCAATACCTGCCACCCCAGTCTTAATCAG
Mus             GAAAGCTCAGAAGGCTAAATGAATATTACCCCTAACACCTGCCACCCCAGTCTTAATCAG
Homo            GAAAGCTCAGAAGGCTAAATGAATATTATCCCTAATACCTGCCACCCCACTCTTAATCAG
Oryctolagus     GAAAGCTCAGAAGGCTAAATGAATATTACCCCTAATACCTGCCACCCCAGTCTTAATCAG
                *** **    **    ** ****  *       *         **  *  **     **

Danio           AGGTTGCGGCGTGTTCTTCCCAACCTCTTGGAATTTCTCTAAACCTGGGCAC--------
Xenopus         TGGTGGAGGATTGGTCTC-A-GAACTCT------TTCTATCAA-TTGGCCATCAGAGTTT
Gallus          TGGTGGAAGAACGGTCTC-A-GAACTGT------TTGTCTCAA-TTGGCCATTTAAGTTT
Bos             TGGTGGAAGAACGGTCTC-A-GAACTGT------TTGTCTCAA-TTGGCCATTTAAGTTT
Mus             TGGTGGAAGAACGGTCTC-A-GAACTGT------TTGTCTCAA-TTGGCCATTTAAGTTT
Homo            TGGTGGAAGAACGGTCTC-A-GAACTGT------TTGTTTCAA-TTGGCCATTTAAGTTT
Oryctolagus     TGGTGGAAGAACGGTCTC-A-GAACTGT------TTGTCTCAA-TTGGCCATTTAAGTTT
                 *** *  *   * ***     * ** *      ** * * **  *** **

Danio           TCTACTTAAGGACTGGATAATGCTGATTAAAACCCATCGGAAAAATTTTCGCAGGAAAGG
Xenopus         AATAGTCAAAGACTGGTTAAT---GATTACAATGCATCGCAAAAGCTTCAGAAGGAAAAA
Gallus          AATAGTAAAAGACTGGTTAAT---GATTACAATGCATCGTAAAAGCTTCAGAAGGAAAGG
Bos             AATAGTAAAAGACTGGTTAAT---GATAACAATGCATCGTAAAACCTTCAGAAGGAAAGG
Mus             AATAGTAAAAGACTGGTTAAT---GATAACAATGCATCGTAAAACCTTCAGAAGGAAAGA
Homo            AGTAGTAAAAGACTGGTTAAT---GATAACAATGCATCGTAAAACCTTCAGAAGGAAAGG
Oryctolagus     AATAGTAAAAGACTGGTTAAT---GATAACAATGCATCGTAAAACCTTCAGAAGGAAAGG
                  ** * ** ****** ****   *** * **  ***** ****  **  * ******

Danio           AAAACAACT-TGGATTTAAGTGTGGCTCCATTTATTGACTGATAGTGCCT--CTTTCAGT
Xenopus         A--ATG--CTCGTGGACACAT-------------TT----GTTTGTGGCAGTTTTTAAGT
Gallus          A--ATG--TTTGTGGACCATTT---------GTTT--------CGTGGCAGTT--TAAGT
Bos             AGAATGTTTTTGTGGACCATAT---------GTTTT----GTGTGTGGCAGTT--TAAGT
Mus             A---T--G-TTGTGGACCATTT---------TTTTT----GTGTGTGGCAGTT-TTAAGT
Homo            AGAATGTT-TTGTGGACCACTTTGGTTTTCTTTTTT----GCGTGTGGCAGTT-TTAAGT
Oryctolagus     AGAATGTT-TTGTGGACCATTT--------TTTTTT----GTGCGTGGCAGTT-TTAAGT
                *          *                       *         *** *      * ***

Danio           TATTAAATTTGTG------------TTTTGATGGTTTAGAACTGCACCT----GTTGCCA
Xenopus         TATTAGTGTTTTAAATCCAGTAATTTCTAAATGGA-AGCAACTTGACCAA-AATCTGTCA
Gallus          TATTAGTCTTTAAAATCGATTAATTT-TTTAAAAT-GGAAACTTGACCAAAAATCTGTCA
Bos             TATTAGTTTTTAAAATCAGT-ACTTT-TTAATGAA-AACAACTTGACCAAAAATCTGTCA
Mus             TATTAGTTTTCAAAATCAGT-ACTTT-TTAATGGA-AACAACTTGACCAAAAATCTGTCA
Homo            TATTAGTTTTTAAAATCAGT-ACTTT-TTAATGGA-AACAACTTGACCAAAAATTTGTCA
Oryctolagus     TATTAGTTTTTAAAATCAGT-ACTTT-TTAATGGA-AACAACTTGACCAAAAATCTGTCA
                *****   **               * *  *        ****  ***       ** **

Danio           CAGTACAATTTGGAAACGC-TGATGAATAAACTAAT----AAAG----GTAT--------
Xenopus         CC---AAA--TTGAGAC-CATTAAAAAAAAGTTAAAAGAAAAAAAAAAA-----------
Gallus          ----------------------------------------------------------
Bos             CA---GAAT-TTGAGACCCATTAAAA-AAAGTTTAATGAGAAACCTGTGTCTTCCTTTTG
Mus             CA---GAATTTTGAGACCA-TTAAAA-CAAGTTTAATGAGAAA-----------------
Homo            CA---GAATTTTGAGACCCATTAAA--AAAGTTAAATGAGAAACCTGTGTGTTCCTT-TG
```

```
Oryctolagus      CA---GAATTTTGAGACCCATTAAAACAAAGTTTAATGAG--------------------

Danio            -TAAAAATTGAAAAAA--------AAA-----------A----------------------
Xenopus          ---AAAAAAAAAAAA---AAA--AAA--AAAAAAAAAA---------AA----------
Gallus           ------------------------------------------------------------
Bos              GTCAACACTGTAACTCCCTACAGTACTACTTTGGTAAGAGTTGCTCATAAGCTATTTCTG
Mus              ------------------------------------------------------------
Homo             GTCAACACCGAGACAT-------------TTAGGTGAAAGA-------CATCTAATTCTG
Oryctolagus      ------------------------------------------------------------

Danio            ------------------------------------------------------------
Xenopus          ------------------------------------------------------------
Gallus           ------------------------------------------------------------
Bos              GTAAAACAATTT-----------TC-------AAATTATGGGTTTGTATTTCTAGGGTGG
Mus              ------------------------------------------------------------
Homo             GTTTTACGAATCTGGAAACTTCTTGAAAATGTAATTCTTGAGTTAACACTTCT-GGGTGG
Oryctolagus      ------------------------------------------------------------

Danio            ------------------------------------------------------------
Xenopus          ------------------------------------------------------------
Gallus           ------------------------------------------------------------
Bos              AGCTTCAGGTTTGTTAACCTTG---TGTTGAGAACTCATCTGTTTTAATAACATCTTAAG
Mus              ------------------------------------------------------------
Homo             AGAATAGGGTTGTTTTCCCCCCACATAATTGGAA----GGGGAAGGAATATCATTTAAAG
Oryctolagus      ------------------------------------------------------------

Danio            ------------------------------------------------------------
Xenopus          ------------------------------------------------------------
Gallus           ------------------------------------------------------------
Bos              CATTGGTGCAACACTTTTCTAGATTAGGACAGATGAACA-AAGTAACTATGTTTTATATG
Mus              ------------------------------------------------------------
Homo             CTATGGG---AGGGTTGCTTTGATTACAACACTGGAGAGAAATGCAGCATGTTGCTGAT-
Oryctolagus      ------------------------------------------------------------

Danio            ------------------------------------------------------------
Xenopus          ------------------------------------------------------------
Gallus           ------------------------------------------------------------
Bos              TAAGCTAGTTTGTA----AGGTCAGATTCTAGAGTATAGAAGCTCCTGTTGCATATAAAA
Mus              ------------------------------------------------------------
Homo             --TGCCTGTCACTAAAACAGGCCAAAA-------ACTGAGTCCTTGTGTTGCATAGAAAG
Oryctolagus      ------------------------------------------------------------

Danio            ------------------------------------------------------------
Xenopus          ------------------------------------------------------------
Gallus           ------------------------------------------------------------
Bos              TTTTGTATGGTTACATAAATG---AATAAATCTATGTCAT---TTAGTTTGCCAGGTATG
Mus              ------------------------------------------------------------
Homo             CTTCATGTTGCTAAACCAATGTTAAGTGAATCTTTGGAAACAAAATGTTTCCAAATTACT
Oryctolagus      ------------------------------------------------------------

Danio            ------------------------------------------------------------
Xenopus          ------------------------------------------------------------
Gallus           ------------------------------------------------------------
```

```
Bos             AGGATGTGCATTCACATTTATAATAAGTAGTTAATCTAAAATGTGAGACTTAATAGTATT
Mus             ------------------------------------------------------------
Homo            GGGATGTGCATGTTG----------------------AAACGTGG--GTTAAAATGACT
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             GAGTACTGCCTTGCTAGAGTTAATTGTATAC-AGGTTCTAGGAAAAATAAAAGATGCTGG
Mus             ------------------------------------------------------------
Homo            GGGCA-------GTGAAAGTTGACTATTTGCCATGACATAAG--AAATAA----GTGTAG
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             AGGCTAATTGTGTTTCCTGTCACTTAAATAGAAATGAACTTTATAGGAATTCTCATCAAA
Mus             ------------------------------------------------------------
Homo            TGGCTAG-TGTACACCCTATGA--------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             TGGGCAAGGTTGGGTGAATGAGAACATTGCTACATTGGTGAGGAGGTAAGGCCCATTTGG
Mus             ------------------------------------------------------------
Homo            -------------GT-----------------------------GGAAGGGTCCATTTTG
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             GTCTGAGTGTAGG----TTCAT-CT-AGTTAATGTTTTCAAA-G-----TTAACTGCCAT
Mus             ------------------------------------------------------------
Homo            AAGTCAGTGGAGTAAGCTTTATGCCAGTTTGATGGTTTCACAAGTTCTATTGAGTGCTAT
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             TTTAAAAACTAGAAAGGA-----AACTGGAACTTGTCATTTGCAGGTTGCTATAACTTGA
Mus             ------------------------------------------------------------
Homo            TCAGAATAGGAACAAGGTTCTAATAGAAAAAGATGGCAATTTGAAGTAGCTATAAAATTA
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             CAGTTGAATGAAAATAACCCTTAACTCTAGAG-----GAATTACTTTCATAACCTGTCAA
Mus             ------------------------------------------------------------
Homo            GACT--AATCTACATTG--CTTTTCTCCTGCAGAGTCTAATACCTTTTATGCTTTGATAA
Oryctolagus     ------------------------------------------------------------
```

```
Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            TTGTAGGGCTTGTTCACAAATGGGAAAAACTAGGTGGTCAGTTGATAATTGATCTCTGGT
Mus            ------------------------------------------------------------
Homo           TTA-------------------------GCAGTTTGT------CTACTTGGTCACT---
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            ATAAGCAATGTAATTCTAAGTTAACCTTGGTTTTGATAGTCTTACACATTTGCAGAAAC-
Mus            ------------------------------------------------------------
Homo           ---AGGAATGAAACTACATG------------GTAATAGGCTTAA-CAGGTGTAATAGCC
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            ------------------TGAGTAATTCTGTATCTGATAACTAGGCTTTTATAATAGGAA
Mus            ------------------------------------------------------------
Homo           CACTTACTCCTGAATCTTTAAGCATTTGTGCATTTGAAAAAT--GCTTTT----------
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            AATTAATCCAGCTGAAACTGGTGAAATCACACC----AGATACCATTCTAGAAGCCTTTT
Mus            ------------------------------------------------------------
Homo           -------CGCGATCTTCCTGCTGGGATTACAGGCATGAGCCACTGTGCCTGAC--CT---
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            TTATGAATAGAAAGCATCCTGTGAGCTGTAGACAGATGGATTAATTGCAGTTTTTCCTAG
Mus            ------------------------------------------------------------
Homo           --------------------CCCATATGTAAAA--GTG-TCTAAAGGTTTTTTTTTGGTT
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            ATAAAA--ATAATTTATATGGCCATTGGTAGGGACTCTAAAATGACAAAGGATTTGGTCC
Mus            ------------------------------------------------------------
Homo           ATAAAAGGAAAATTTTTGCTTAAGTTTGAAGGATAGGTAAAATT--AAAGGACATGCTTT
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            CTAGCTTGTTCCTTAATGTTTTGATAGATATGGTGTTAACTGATGAAAGGAGATATGAAT
Mus            ------------------------------------------------------------
Homo           CTGT-TTG---TGT----------------------------------------------
```

```
Oryctolagus     ------------------------------------------------------------


Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             AAAGATTGCTAGCTAGTAAATAGCTGGTGAAGACTTGTCTTCACAAGGAAGGTTTCATTG
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------GATGGTT-----T
Oryctolagus     ------------------------------------------------------------


Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             CAGAAAATAGCTTTTAAGTGGAAGTTTAAGGAGAGCAGGACAGTTGGCATGCTTGGAGCC
Mus             ------------------------------------------------------------
Homo            TTAAAAATTTTTTTTAAGATGGAGTTCTTGTTGCCCAGGCTAGAATGCAAT---GG---C
Oryctolagus     ------------------------------------------------------------


Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             ATAGATTTGTCTGTACAGAAGTGCTGTGAGCGCCACCTACTGG-----------------
Mus             ------------------------------------------------------------
Homo            A-AAATCT-------------CACTGCAATCTCCTCCTCCTGGGTTCAAGCAATTCTCCT
Oryctolagus     ------------------------------------------------------------


Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             ------------------TGGTCACTACAGACTAGTTTAAAACATGAATTTCTAGATTC
Mus             ------------------------------------------------------------
Homo            ACTTCAGCCTCCCAAGTAGCTGGGATTACAGGCATGTGCTA--------------------
Oryctolagus     ------------------------------------------------------------


Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             AGGATGAATGTTTTGCTCAGTTTTATACAGTTGC--TTTCTGAATA-TGAACTGACTGAA
Mus             ------------------------------------------------------------
Homo            ----------ATTTGGTGTTTTTAATAGAGATGAGGTTTTTCCATGTTGGTCAGGCTGG-
Oryctolagus     ------------------------------------------------------------


Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             AAGATTTATAAAACAAACTGGTGCTTGACCTTAAACCCCATTTTTAAAATTGCCTAGC--
Mus             ------------------------------------------------------------
Homo            ------TC----TCAAACT----CCTGACCTTAGG-----------TGATCGCCTCGGCC
Oryctolagus     ------------------------------------------------------------


Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
```

```
Bos             --------TGTTGAAA-TGGAATCTTGTTGAAA----CTTGGCCCCAGGAATTCTTCTGT
Mus             ------------------------------------------------------------
Homo            TCCTAAAGTGCTGGAATTACAGGCATGAGCCACCATGCCTGG--CCAGGACATG---TGT
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             GGTTTCAGAATATGAGAGGCAGTATG-A---TTTCT---GAAAAAATGATTCTTAACCCT
Mus             ------------------------------------------------------------
Homo            T-CTTAAGGACATGCTAAGCAGGAGTTAAAGCAGCCCAAGAGATAAGGCCTCTTAAAGT-
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             ATTCCCAGGATAGGCAGTATAGTTGAGTTTTGGCCAGATTAACTATGCTTCAGGGGTTTT
Mus             ------------------------------------------------------------
Homo            --------GACTGGCAAT------GTGTATTGCTC---------AAGATTCAAAGGTACT
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             TTCAAGA-GCATGAGCAATTCA-TGTGGTTGGGTTCTGGTCTGCCCAAAAG------CAA
Mus             ------------------------------------------------------------
Homo            TGAATTGGCCATAGACAAGTCTGTAATGAAGTGTTATCGTTTTCCCTCATCTGAGTCTGA
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             ATTGCATATACTGAAGAGCTTGATTATTAGTAATTTGCAGAATGGGGTATATAAATACCA
Mus             ------------------------------------------------------------
Homo            ATTAGATAAA-----ATGCCTTCCCATCAGCCAGTG---CTCTGAGGTATC--AAGTC-T
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             GGACAGGACTATAGATATTTTTATTAAAATGTATTTACCTTGGTTACCTTTTGTTAGGTT
Mus             ------------------------------------------------------------
Homo            AAATTGAACTAGAGATTTTTGTCCTTA------GTTT-CTTTGCTATCTAATGTTTAC--
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             GAGCACTGAGCTCATGCCAAAAACAGGCTGTTTCTGGCTCTGTATACAATTCTTGAAATG
Mus             ------------------------------------------------------------
Homo            -----------ACAAG---TAAATAGTCTAAGATTTGCT--GGA--------------TG
Oryctolagus     ------------------------------------------------------------
```

```
Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            GGCAGCTCTAGAAAAGGGATGAAAACTTGTGTGTCAAGATGTCCTTTTATGTTTTCTGTT
Mus            ------------------------------------------------------------
Homo           -------ACAGAAAA--------AACAGGT--------AAGGCCTTTA---------ATA
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            GATAGGTGTTAGATGTCCTTGTGAAATGCTGCAAGATGATGTATAGTCACTTGGGAATTA
Mus            ------------------------------------------------------------
Homo           GATGGCCAATAGATGCCCTGATA-----ATGAAAGTTGA--------CACCTGTAAG-AT
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            TTAAAGGTATATTAGTTATTCCCTTGTATCTTAAAAGGAATGAGCAGCATGAGTTCTCAA
Mus            ------------------------------------------------------------
Homo           TTACCAGTAGAGA-ATTCTTGA-------CATGCAAGGA------AGCAAGATTTAA---
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            TCTTGTGATGAGAAGAGCAAGCTCTTCCATATGAGAAAGAAGGTGGAATGAAATGCATCT
Mus            ------------------------------------------------------------
Homo           -------CTGAA------AAATTGTTCCCACTG-----GAAGCAGGAATGAG-------T
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            TCCTTTATTTGCAACGTTCCTGAAACTTTTCCTGATCTGGTTTTGGTTTTCATGACTTCT
Mus            ------------------------------------------------------------
Homo           CAGTTTACTTGCATA--TACTGAGA-----------------TTGAGATT---AACT---
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            GTCCTTTTTTCTTCCAAACTTAGAGGTTCTCTCACCCAGGTGGCATCAGAGTCCTCTGTT
Mus            ------------------------------------------------------------
Homo           --------TCCTGTGAAACCCAGTG--TCTTAGACAACTGTGGCTTGAGCACCACCTGCT
Oryctolagus    ------------------------------------------------------------

Danio          ------------------------------------------------------------
Xenopus        ------------------------------------------------------------
Gallus         ------------------------------------------------------------
Bos            TGCAGGCCTCCGGGAACTTCCTAGACTAGTTAAAATCACAATTGAAGTTGCATTGCTTGC
Mus            ------------------------------------------------------------
Homo           GGTATTCATTA-CAAACTTGCTC-ACT-------ACAATAAATGAATTTTAA--------
```

```
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             TCAGGTGATAGTGATTTAGAAAAAACTTTGGAAACTGCCACATGGCTTCTGAGAGGTCAT
Mus             ------------------------------------------------------------
Homo            ------------GCTTTAAAAAAAAAAAAAAAAAA------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             TGACCTCTGATCCAATGGAAAACTTGGAAAGGATCATATGGGTTTCTTTTCTAAATAGGT
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             TATGCTACATAGTTCTGCTGTTAATAGAAATGAGTGAGATTAGTGAGTTTGTCATATACC
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             TTGGACAACCTTCAGCAGGTCCATGGATCACCAAGCCTTGAGTGCTCAGTGACAACAAAG
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             AGCATGTTAATATTTCAAGTAAGCAATAATAGAAAAGTTAGCAGAATATAAAGTTCTTAC
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             AACTTTTTTCATTAAAATTTACCCTGTAAAAAGCATTTTGCAAGTTTTGATAAAGTTTAC
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
```

```
Bos             CAAAAAACAGTTTATTTACCACAATCATGGTATAGAATATTTCCGTCACTTTTCCCCTTC
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             CCAAATTTCCCACGTCCCTGTATGGTCAATCACTCTTCTCTACCTCTTATTCTTGGTAAC
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             TACTGACCTATATTCTTTCCCCGTTTCGATTTTTAGCAGTGTCACATACCTGAGTTGATA
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             CCATATGTAGCCTTTGAATCTGGTTTCTTTACTTAGCATGATGCCTTTGAGATGATCTGT
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             TTTGCATGTATCAGTAATCTGTTCCTTTTTAATAACTATAAGCTTGTTTACAGCTGAGGG
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Gallus          ------------------------------------------------------------
Bos             ACATTGGTTGTTTTGGATAATTACAAATAAACCACTAAACATTCACATATAAAAAAAAAA
Mus             ------------------------------------------------------------
Homo            ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------

Danio           --------
Xenopus         --------
Gallus          --------
Bos             AAAAAAAA
Mus             --------
Homo            --------
Oryctolagus     --------
```

Comparison table of currently commercially available mouse models for ALS as offered by the Jackson Laboratory. Outlining mutation and associated phenotypes with average survival rates. (http://mousemutant.jax.org/).

Strain Name	Common Name	Molecular Mutation	Phenotype	Survival
B6SJL-Tg(SOD1*G93A)1Gur/J 002726	SOD1-G93A	Human SOD1 with glycine to alanine transition at position 93	Decreased grip strength, impaired coordination, motor neuron degeneration, severe muscle weakness beyond 3 months old, hind limb tremors at 14 weeks old, become paralyzed in one or more limbs	50% survival at 128.9+/-9.1 days
B6.Cg-Tg(SOD1*G93A)1Gur/J 004435	B6 SOD1-G93A	Human SOD1 with glycine to alanine transition at position 93	Decreased grip strength, impaired coordination, motor neuron degeneration, decreased muscle size, hind limb tremors at 14 weeks of age, become paralyzed in one or more limbs	50% survival at 157.1+/-9.3 days
B6.Cg-Tg(Prnp-TARDBP*A315T)95Balo/J 010700	Prp-TDP43-A315T	Mouse prion promoter driving a modified human TAR DNA binding protein with an A315T amino acid substitution	Progressive and fatal neurodegenerative disease, frontotemporal lobar degeneration with ubiquitin aggregates	Average survival: 97+/-11 days, females live longer than males
B6SJL-Tg(SOD1)2Gur/J 002297	WT SOD1	Normal human SOD1	Normal	Normal

Sequences highlighting various alleles and annotated from D252H mice.

EXON 5
PAM D252H Mutation Insertion Deletion

#10 Alleles

```
#1      10.13R      ACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
        10.1R       ACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
        10.4        ACCCGCCCCACTGGCAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
        10.2R       ACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
        10.6R       ACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
        10.11       ACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
        10.9        ACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
        10.8        ACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
        10.7        ACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
        10.10       ACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
        10.5        ACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT


#2      10.12R      CCCCCCACCCGC-----TCCAAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGG
```

#11 Alleles

```
 #1         11.4R       CCCACCCGCCCCACTGACAAGCCCCTTCGTC-GCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.6R       CCCACCCGCCCCACTGACAAGCCCCTTCGTC-GCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.3        CCCACCCGCCCCACTGACAAGCCCCTTCGTC-GCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.13       CCCACCCGCCCCACTGACAAGCCCCTTCGTC-GCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.11       CCCACCCGCCCCACTGACAAGCCCCTTCGTC-GCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.7        CCCACCCGCCCCACTGACAAGCCCCTTCGTC-GCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.1R       CCCACCCGCCCCACTGACAAGCCCCTTCGTC-GCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCT


  #2        11.2        CCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.12R      CCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.8        CCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.9        CCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.10       CCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCT
            11.5        CCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCT
```

 # 12 Alleles

```
  #1        12.13       CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC


  #2        12.4        CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTG----------ACAGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC
            12.8        CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTG----------ACAGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC
            12.7        CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTG----------ACAGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC
            12.5        CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTG----------ACAGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC
            12.1        CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTG----------ACAGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC
```

```
          12.3        CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTG----------ACAGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC
          12.10R      CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTG----------ACAGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC
          12.6R       CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTG----------ACAGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC
          12.12R      CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTG----------ACAGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC
          12.2R       CCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTG----------ACAGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTC
```

 # 14 Alleles

```
#1        14.11R      CCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTC
          14.6R       CCCCCCACCCGCCCCGCTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTC
          14.7        CCCCCCACCAGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTC
          14.9        CCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTC
          14.8        CCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTC
          14.1R       CCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTC

#2        14.10       GACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTCA----------AGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTG
          14.13R      GACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTCA----------AGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTG

#3        14.5R       CTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTG
          14.3        CTGGACACAATCCCGCCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTG
```

#15 Alleles

```
#1        15.2        CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGSATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTCCAAA
          15.4        CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAAGCGCTTGCCTACTCTGCTCCAAA
          15.8        CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTCCAAA
          15.13       CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTCCAAA
          15.7        CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCGGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTCCAAA
          15.6        CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTCCAAA
          15.11R      CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTCCAAA
          15.1        CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTCCAAA

#2        15.3R       CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTCCAAA
          15.10R      CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTCCAAA
          15.9R       CCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCTGCTCCAAA
```

#16 Alleles

```
#1        16.13       CTGCCCCCCATCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
          16.11R      CTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
          16.4        CTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
          16.8R       CTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
          16.9        CTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT

#2        16.6        CTGCCCCCCACCCGCCCCACTGACCAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
```

```
#3        16.7         CTGCCCCCCACCCGCCCTACTGACAAGCCCCTTCGTCTGCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
          16.12R       CTGCCCCCCACCCGCCCTACTGACAAGCCCCTTCGTCTGCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
          16.2R        CTGCCCCCCACCCGCCCTACTGACAGGCCCCTTCGTCTGCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
          16.5R        CTGCCCCCCACCCGCCCTACTGACAGGCCCCTTCGTCTGCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
          16.10R       CTGCCCCCCACCCGCCCTACTGACAAGCCCCTTCGTCTGCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
          16.1         CTGCCCCCCACCCGCCCTACTGACAAGCCCCTTCGTCTGCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
          16.3R        CTGCCCCCCACCCGCCCTACTGACAAGCCCCTTCGTCTGCCTCTGCAGCATGTGTACAAGATTGGCGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCAGGCGCTTGCCTACTCT
```

#17 Alleles

```
#1        17.12        GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTG

#2        17.4         GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACT----------------------------------------GTATTGGGGGTGAGTGAGGGTTCAGTG
          17.5         GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACT----------------------------------------GTATTGGGGGTGAGTGAGGGTTCAGTG
          17.6         GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACT----------------------------------------GTATTGGGGGTGAGTGAGGGTTCAGTG
          17.8         GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACT----------------------------------------GTATTGGGGGTGAGTGAGGGTTCAGTG
          17.13R       GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACT----------------------------------------GTATTGGGGGTGAGTGAGGGTTCAGTG
          17.10R       GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACT----------------------------------------GTATTGGGGGTGAGTGAGGGTTCAGTG
          17.1R        GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACT----------------------------------------GTATTGGGGGTGAGTGAGGGTTCAGTG
          17.11R       GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACT----------------------------------------GTATTGGGGGTGAGTGAGGGTTCAGTG
          17.2R        GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACT----------------------------------------GTATTGGGGGTGAGTGAGGGTTCAGTG
          17.3R        GCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACT----------------------------------------GTATTGGGGGTGAGTGAGGGTTCAGTG
```

#18 Alleles

```
#1        18.4           GGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGG
          18.8           GGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGG
          18.14          GGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGG

#2        18.3           GGACACAATCCTGCCCCCCACCCGCCCTACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGG

          18.3-Continued         AAACCCTGACTAGGATATCCTGAGTGGATATCTGCAGGATGTGCACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACT
          18.3-Continued         GCGGCAGGCGCTTGCCTACTCTGCTCCAAACCCTGACTAGGANATCAACAAGGAATCTTG


#3        18.6           GGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGT------CTGCAGGACGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGG

#4        18.15          GGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGT------CTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGG
```

#21 Alleles

```
#1        21.2R        CCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCGGTGCTGGGGCTCACTGCGGCA
          21.12R       CCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCA
```

#23 Alleles

```
#1          23.13R    AAGCGGCGTGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTC
```

#24 Alleles

```
#1       24.6        AAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAG-------------------TGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCA
         24.4        AAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAG-------------------TGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCA
         24.3        AAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAG-------------------TGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCA
         24.11       AAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAG-------------------TGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCA
         24.13R      AAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAG-------------------TGGATGTGTACAAGACTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCCCACTGCGGCA
         24.8R       AAGACCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAG-------------------TGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGGCTCACTGCGGCA

#2       24.5R       AAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCACAAGCCCCTTCGTCTGCCTCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTT
```

#25 Alleles

```
#1         25.6          TGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCG------TCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGG
           25.10         TGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCG------TCTGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTCAGTGCTGGGG

#2         25.13         TGTCCCTGCTGGAAGCCCTGGACACAATCCTGCCCCCCACCCGCCCCACTGACAAGCCCCTTCGTCTGCAGGATGTGTACAAGCAGGATGTGTACAAGATTGGGGGTGAGTGAGGGTTC
```

Proteins with p-value <0.05.

Down regulated in eEF1A2 null
Up regulated in eEF1A2 null

Gene.name	p.value
Fgf12	0.001518
Ttn	0.045856
Gas7	0.039288
Cbs	0.004292
Aimp1	0.003389
Ckap4	0.038731
Syne1	0.015807
S100a16	0.006661
Ndrg4	0.023392
Gga1	0.027117
Tomm20	0.045197
Iglon5	0.038257
Cox7c	0.021239
Caprin1	0.010657
Ppp1r14a	0.01775
Kif5c	0.02016
Ghitm	0.050441
Grpel1	0.048353
Tagln2	0.032055
Glra1	0.036788
Rbm8a	0.025845
Sugt1	0.040627
Cyp51a1	0.040718
Hsph1	0.034543
Dnajc5	0.047138
Prpf8	0.019353
Acin1	0.030751
Gcn1l1	0.016245
Napb	0.005158
Ndufa2	0.023928
Flna	0.022916
Naprt	0.04933
Fkbp4	0.019564
Ndufb11	0.039723

Gene.name	p.value
Uso1	0.049705
Pfdn2	0.040341
Txnrd2	0.037786
Txnl1	0.002896
Sec22b	0.01954
Slc25a25	0.003476
Uqcrfs1	0.037489
Psmd6	0.018114
Rap1gds1	0.0518
Cfl2	0.013105
Srgap3	0.035698
Fam49b	0.042124
Acy1	0.02001
Napa	0.049048
Safb	0.021779
Dync1i2	0.050162
Scfd1	0.003106
Tspan7	0.037685
Dnm1l	0.012052
Esyt1	0.026192
Erc1	0.034952
Vbp1	0.018681
Ndufs5	0.012351
Ppp5c	0.033012
Atp5j2	0.036039
Slc30a3	0.049244
Oxr1	0.016948
Atp1a3	0.015822
Cox7b	0.042208
Ube2d3	0.044727
Ndufa7	0.017251
Cul1	0.025685
Ogdh	0.042284
Cnp	0.030698

Gene.name	p.value
Acsl6	0.047726
Slc25a3	0.019052
Slc1a2	0.012341
Eif2a	0.009392
Psmd11	0.019979
Guk1	0.039977
Rbm39	0.020238
Hapln4	0.043426
Gdi1	0.045631
Qdpr	0.036639
Nedd4	0.045003
Ccbl2	0.020516
Dpysl2	0.021393
Eef1a1	0.009621
Maoa	0.021419
Aldh2	0.015313
Anxa2	0.027545
Yars	0.032789
Zadh2	0.043666
Chmp6	0.026847
Hp1bp3	0.016429
Ykt6	0.017213
Tnr	0.016393
Acsbg1	0.010376
Add1	0.043025
Ddt;Gm20441	0.041602
Camk2b	0.030519
Eif4a1	0.024624
Ddx1	0.036658
Tardbp	0.029109
Cyfip2	0.015926
Hccs	0.05452
Slc25a22	0.048158
Ppm1b	0.006912
Pabpc1	0.035859
Sbf1	0.001248
Pcmt1	0.0179
Prrt2	0.01962
Pgm1;Pgm2	0.047961

Gene.name	p.value
Kctd12	0.011375
Plcl1	0.053561
Ddx17	0.013856
Acta1;Actc1;Actg2;Acta2	0.005828
Kpna3	0.053606
Sec23a	0.03418
Rpl7a	0.049824
Gnao1	0.023497
Adap1	0.0086
Cnrip1	0.00492
Stt3b	0.039038
Necap1	0.023091
FAM120A	0.028652
Cav1	0.054152
Srm	0.052242
Gja1	0.030767
Idh3g	0.01602
Gstm5	0.021886
Igsf8	0.003151
Impa1	0.030378
Wdr37	0.053356
Ap3d1	0.010559
Fech	0.042371
Wars	0.038933
Rpl35a	0.036247
Rpl19	0.048655
Lxn	0.015364
Mtnd4	0.01699
Mblac2	0.00865
Wdr47	0.013506
Arf1;Arf3;Arf2	0.053493
Gpx4	0.015325
Prpsap2	0.046007
Sdhd	0.03899
Tcp11l1	0.033077
C1qa	0.023543
Dhrs4	0.033835
Psmd4	0.040312
Rpl14	0.008582

Gene.name	p.value
Far1	0.047808
Rpl18	0.044418
Rpl12	0.034373
Rims1	0.048656

Gene.name	p.value
Acads	0.045078
Stub1	0.048602
Nap1l5	0.030117
Mtfr1l	0.046472

SAM 500 genes.

Down regulated in *wasted*
Up regulated in *wasted*

Gene.name	q.value	log2.fc
Eef1a2	0.0327	-7.02808
Eef1a2	0.0327	-6.8819
Eef1a2	0.0549	-4.99321
Chrna4	0.0414	-1.83851
Ahnak2	0.0483	-1.1011
Kcng4	0.0404	-1.03422
Fbxo9	0.0327	-0.89958
Igfbp2	0.0494	-0.85266
Igfbp2	0.0523	-0.80323
Kcng4	0.0532	-0.76408
Pla2g3	0.0553	-0.76024
Mrvi1	0.0416	-0.62258
Card10	0.0788	-0.60361
Eml1	0.0799	-0.56785
Dhcr24	0.0756	-0.54255
Prph	0.053	-0.53
LOC100041194	0.0813	-0.52565
Tmprss5	0.0416	-0.52511
Etv4	0.0637	-0.52338
Dhcr24	0.0549	-0.51004
E130012A19Rik	0.0637	-0.50453
6330503H08Rik	0.0743	-0.49859
Unc13c	0.0549	-0.49703
Dync1i1	0.0523	-0.49652
Dysf	0.0532	-0.48723
Gpd1	0.0494	-0.48561

Gene.name	q.value	log2.fc
Itih3	0.0862	-0.48045
Unc13c	0.0523	-0.46791
Plekho2	0.0523	-0.46727
Mvd	0.045	-0.46285
Itih3	0.0615	-0.45993
Acy3	0.081	-0.45057
Aacs	0.0826	-0.444
Txnl4a	0.0587	-0.44098
Prph	0.0646	-0.43986
Gprasp1	0.0577	-0.43313
Whrn	0.0735	-0.42905
Whrn	0.059	-0.42407
Fgfrl1	0.0579	-0.41937
Calr3	0.0539	-0.4189
Gprasp1	0.045	-0.41612
Elmo2	0.0735	-0.40706
Stk32a	0.0584	-0.39763
Pcyt1b	0.075	-0.39723
2310047M10Rik	0.0742	-0.39577
Acly	0.059	-0.38436
Gls2	0.0607	-0.384
Parva	0.0606	-0.3826
Rgs11	0.075	-0.3826
E2f1	0.0742	-0.37747
2310047M10Rik	0.0556	-0.37466
Megf11	0.0615	-0.36257

Gene.name	q.value	log2.fc
Gjc2	0.0749	-0.36106
1500031L02Rik	0.0681	-0.35785
Tubb2c	0.076	-0.35664
Arhgef4	0.045	-0.35326
BC067047	0.0735	-0.3493
2310058D17Rik	0.0494	-0.34712
Fmr1	0.0679	-0.3443
Tlcd1	0.0646	-0.34146
Amhr2	0.0532	-0.3358
Rgs7bp	0.0569	-0.33376
Blvrb	0.0682	-0.33183
Sqle	0.0523	-0.3318
Ankmy2	0.0509	-0.33027
6330442E10Rik	0.0561	-0.32585
Pstpip1	0.0605	-0.3236
LOC100045304	0.0494	-0.3231
2610019E17Rik	0.075	-0.32266
Cryba2	0.0603	-0.32084
Lgr6	0.0416	-0.31931
Serpine2	0.0743	-0.31509
Nat8l	0.0735	-0.31474
Bcan	0.0759	-0.31458
Frmd8	0.0855	-0.31235
Dos	0.081	-0.3084
Gls2	0.0584	-0.30809
Copg	0.0615	-0.30499
Gpld1	0.0742	-0.30454
Hrh3	0.0404	-0.30293
Spata13	0.076	-0.30195
Hebp2	0.0742	-0.30174
Esrrb	0.0549	-0.3003
6330442E10Rik	0.059	-0.29985
Kcnip3	0.0561	-0.29798
9330175B01Rik	0.0749	-0.29745
Necab3	0.0587	-0.2973
Gpld1	0.076	-0.29634
Asb1	0.0657	-0.29577
Elovl5	0.0764	-0.29493
Sh3bp5l	0.0436	-0.29049

Gene.name	q.value	log2.fc
Pgrmc2	0.0742	-0.29039
Sc5d	0.0494	-0.28738
Acaa1b	0.045	-0.28408
Ctnnbip1	0.0523	-0.28406
Rdh11	0.075	-0.28245
Stk16	0.075	-0.27354
C130057N11Rik	0.0735	-0.27067
Ppp1r3f	0.0549	-0.27062
Mid1ip1	0.045	-0.26963
Eif2ak1	0.0742	-0.26529
Tmem2	0.0799	-0.2636
Eif5	0.0532	-0.26169
1500012F01Rik	0.0549	-0.25865
2810017I02Rik	0.0735	-0.25843
Mtap	0.0549	-0.25834
LOC100047651	0.0786	-0.25692
Spna2	0.0735	-0.25452
Rasl10b	0.0608	-0.25296
Snn	0.0416	-0.25208
Mtap	0.075	-0.25022
Tnfrsf22	0.0862	-0.24991
Slc29a4	0.0735	-0.24909
Rassf4	0.0808	-0.24864
Cyp27a1	0.0555	-0.24735
Tubb2b	0.0564	-0.24665
Rassf4	0.0603	-0.24579
Nat8l	0.0826	-0.24226
Plxnb3	0.0697	-0.24064
Smox	0.0799	-0.23802
Mar-02	0.0657	-0.238
LOC100046996	0.0826	-0.23761
Nefm	0.0756	-0.23577
Eif4a2	0.0756	-0.23267
BC014795	0.0782	-0.23135
Zswim6	0.0615	-0.22877
Muted	0.0523	-0.22871
Tpm1	0.0549	-0.22794
9030024J15Rik	0.0561	-0.22663
Def8	0.0837	-0.22617

Gene.name	q.value	log2.fc
Ttf2	0.0657	-0.22542
Oaz2	0.0816	-0.22541
Snhg10	0.0786	-0.22439
Snx27	0.0561	-0.22367
Mag	0.073	-0.22294
Phf21b	0.0742	-0.22275
LOC100043671	0.0742	-0.22274
Spg7	0.0839	-0.22262
Rasgef1a	0.0787	-0.22078
Cand1	0.0799	-0.21619
Tmem38b	0.0602	-0.21516
Slc39a13	0.0726	-0.21324
Usp7	0.075	-0.20747
Efcab2	0.0618	-0.2067
Tmem98	0.0844	-0.2064
Sord	0.059	-0.20612
Lrig1	0.0757	-0.20567
Mtap7d1	0.0799	-0.20455
Ufsp2	0.075	-0.20394
Tnip1	0.0584	-0.20301
Rap2a	0.0739	-0.20286
Mfsd2	0.0681	-0.20244
Thsd7a	0.0754	-0.20243
Stk3	0.0523	-0.20233
Pcdh12	0.0681	-0.20231
Nsf	0.076	-0.20185
Atmin	0.0504	-0.20156
Fsd2	0.0539	-0.20015
Dennd2a	0.0561	-0.19848
Vegfb	0.0523	-0.19824
Igbp1	0.076	-0.19379
2310028O11Rik	0.0615	-0.19257
Eif5	0.0549	-0.18955
Sap30l	0.075	-0.18827
Tmem141	0.0816	-0.18808
Sfrp1	0.0638	-0.18806
Farp2	0.0756	-0.1879
Serpine2	0.075	-0.18747
Shisa2	0.0556	-0.18725

Gene.name	q.value	log2.fc
Zfp251	0.0726	-0.18678
Paip2b	0.0638	-0.18634
4932415G12Rik	0.0735	-0.18525
LOC100044190	0.0735	-0.18499
Rab39	0.0556	-0.18475
Adora1	0.075	-0.18341
Nutf2	0.0787	-0.18272
LOC100045697	0.0816	-0.18146
Sorl1	0.0681	-0.18118
Nsmaf	0.0696	-0.18031
1110029I05Rik	0.075	-0.17979
Lrsam1	0.0816	-0.17874
1110038D17Rik	0.0638	-0.17848
Pmvk	0.0657	-0.17802
Lcmt1	0.0555	-0.17703
Rpl22	0.045	-0.17488
Col22a1	0.045	-0.17425
H2afj	0.0813	-0.17391
4930572J05Rik	0.0801	-0.1732
Impa2	0.0816	-0.17045
Ppa1	0.0483	-0.16995
Slc27a1	0.0754	-0.1695
Syn2	0.0799	-0.16919
Dscr1	0.0754	-0.16909
Ank3	0.0608	-0.16719
Gnptab	0.0799	-0.16565
4930563E22Rik	0.045	-0.16038
Gng11	0.081	-0.15902
LOC100046953	0.075	-0.15887
Pgs1	0.0651	-0.1585
Haghl	0.075	-0.15844
Pcyt2	0.0845	-0.15743
A830080H07Rik	0.0658	-0.15728
Etl4	0.0754	-0.15691
Zdhhc3	0.053	-0.15675
L3mbtl3	0.0549	-0.15599
Dlgap3	0.0688	-0.15551
Zyx	0.0556	-0.15325
scl0002275.1_1	0.061	-0.15314

Gene.name	q.value	log2.fc
Depdc5	0.0726	-0.15068
9530064J02	0.0799	-0.14855
Pipox	0.0816	-0.148
Rnpepl1	0.0862	-0.14799
LOC385156	0.0556	-0.14757
Phf20	0.075	-0.14639
Rhox12	0.0549	-0.14481
Pfkm	0.076	-0.14244
Rapgef6	0.0826	-0.14112
Zkscan6	0.0556	-0.13749
Ankrd6	0.0695	-0.1351
LOC100041343	0.0735	-0.13405
Slc6a9	0.0826	-0.13165
Slc4a2	0.0626	-0.13148
EG626549	0.0813	-0.12942
Tceb1	0.0681	-0.1269
Lrrfip2	0.0872	-0.12215
2010317E24Rik	0.0556	-0.12126
C130090G16Rik	0.075	-0.11593
Aak1	0.0839	-0.11567
LOC672339	0.0839	-0.11469
Harbi1	0.0754	-0.11464
C230098K08Rik	0.0799	-0.11421
Rit1	0.0532	-0.1139
2610110G12Rik	0.0801	-0.11352
Tmem121	0.0756	-0.11339
H2-T18	0.0523	-0.11302
Ctage5	0.0561	-0.1116
Slc28a1	0.0556	-0.11119
Akap11	0.0754	-0.11067
4833411B01Rik	0.0523	-0.11043
A130076G11Rik	0.045	-0.11027
Rhoc	0.0848	-0.10735
Calu	0.0743	-0.10719
Rabep1	0.0756	-0.10641
Pcdha7	0.075	-0.10206
Pi4ka	0.0657	-0.10081
Col5a2	0.081	-0.09981
Xpc	0.0532	-0.09446

Gene.name	q.value	log2.fc
E330036I19Rik	0.0681	-0.09095
Plekhm3	0.0681	-0.09026
Zbtb8os	0.075	-0.08416
Add1	0.0811	-0.08323
Tlr12	0.0608	-0.08113
AA536749	0.053	-0.07966
Ugcgl2	0.045	-0.07729
Tbcel	0.0638	-0.063
3830408C21Rik	0.075	-0.06219
Arhgef1	0.0754	-0.06187
Synj1	0.0799	-0.05242
1700010M22Rik	0.0845	0.043094
4933415I03Rik	0.076	0.066763
Krt2-1	0.0742	0.071064
Chst7	0.0615	0.076695
2610024G14Rik	0.0579	0.082937
LOC385985	0.076	0.086921
Tptf-pending	0.0556	0.087338
Ptgfrn	0.075	0.092881
D430036M17Rik	0.0416	0.100094
Klhl23	0.075	0.109857
LOC272714	0.0742	0.111978
Stxbp4	0.0646	0.113948
9430022P05Rik	0.0404	0.114923
1810026J23Rik	0.0743	0.115461
Sap30bp	0.0799	0.119045
2310003F16Rik	0.0555	0.121071
Ube2h	0.0839	0.121233
Tardbp	0.075	0.122426
9530048O09Rik	0.0845	0.1289
LOC385111	0.0735	0.128979
LOC100044812	0.0579	0.130231
Pomp	0.0756	0.132681
Syngr3	0.085	0.133064
Slc25a19	0.0579	0.135193
Eif3i	0.0598	0.135825
Pelo	0.0523	0.136615
Erlin1	0.045	0.136714
Srgap3	0.0782	0.138793

Gene.name	q.value	log2.fc
Mapbpip-pending	0.0764	0.139446
Skiv2l2	0.0807	0.140494
Chchd1	0.078	0.142138
Sorcs3	0.0657	0.143041
Hsf1	0.0754	0.146896
Htf9c	0.0766	0.147031
Ubl5	0.0764	0.14722
Ppp2r5c	0.0839	0.147956
Lhx9	0.0798	0.14814
Idh3a	0.0523	0.151136
Tomm7	0.045	0.154455
Blcap	0.0615	0.155789
2310061F22Rik	0.0587	0.156537
Trim26	0.0681	0.156835
Gnptg	0.0799	0.157897
Ergic1	0.0603	0.157975
2210015D19Rik	0.0549	0.158015
Rab40c	0.0742	0.158067
Ndufs8	0.0556	0.158472
Erp29	0.0825	0.160188
Lsm4	0.0584	0.162286
Vti1a	0.0494	0.163822
Cbln2	0.0598	0.168021
Gabbr2	0.0765	0.16866
Gpr19	0.0523	0.169462
Wdr74	0.0638	0.169641
Ndufa9	0.0555	0.170207
Rars	0.076	0.170789
Pdcd2	0.0726	0.174155
Maged1	0.075	0.17445
Slc25a38	0.076	0.174801
Sar1b	0.0726	0.178197
Ttc19	0.0826	0.178894
D10Ertd641e	0.0672	0.179399
Uck2	0.075	0.180541
Ppapdc1	0.0743	0.180922
Ndufa13	0.0532	0.18222
C330023M02Rik	0.0756	0.183026

Gene.name	q.value	log2.fc
Nudt19	0.0804	0.184195
Eif1ad	0.0799	0.184327
B230210A04Rik	0.0584	0.185029
2700062C07Rik	0.0826	0.18609
Shmt2	0.0735	0.186753
Sf3b5	0.0556	0.187823
2700062C07Rik	0.0799	0.188424
A830006F12Rik	0.0681	0.189706
Pcdhac2	0.0632	0.189731
Chic2	0.0598	0.189993
Mkln1	0.0549	0.191883
Gria3	0.0681	0.191964
Kctd1	0.0404	0.197811
Ypel3	0.0743	0.197938
EG666387	0.0799	0.198215
Mettl1	0.081	0.198442
Rag1ap1	0.0799	0.198749
Ndufa2	0.0735	0.198927
Nxph1	0.0553	0.198976
Rnf185	0.0606	0.203597
Tmem120a	0.0807	0.20386
Ndufb6	0.0681	0.204013
2310066E14Rik	0.0631	0.20409
Hint1	0.0743	0.204559
6720458F09Rik	0.069	0.204846
2310014G06Rik	0.0523	0.205697
LOC100041835	0.0775	0.206471
9030025P20Rik	0.0638	0.20743
Srpk1	0.0756	0.207591
Klhl26	0.0606	0.210569
4933411K20Rik	0.076	0.212715
Map3k12	0.0821	0.212811
Cnpy2	0.0826	0.214376
1110002E23Rik	0.0735	0.215243
Dhodh	0.0735	0.217043
Ndufa7	0.0695	0.217345
EG574403	0.0799	0.21788
Btbd11	0.0416	0.22226
Ddx52	0.059	0.225269

Gene.name	q.value	log2.fc
Angptl6	0.0804	0.226248
Tmsb10	0.0759	0.226995
Ogfr	0.045	0.229161
2010003O18Rik	0.0579	0.229409
Ndufs6	0.0549	0.233281
Ppp2r5e	0.0549	0.233817
Dusp4	0.0698	0.234588
Acp2	0.0404	0.237202
Mrpl49	0.0681	0.239236
Lypd1	0.0756	0.240299
BC029169	0.0799	0.24238
Tro	0.0587	0.243054
Gcap26	0.0764	0.244058
Nme1	0.053	0.244166
Nt5c	0.0549	0.244727
Golga3	0.0556	0.246693
Brp16	0.076	0.247201
Ppih	0.0799	0.247883
Pqlc3	0.0549	0.249532
Pfdn4	0.0532	0.249651
Ssr4	0.0555	0.251496
Slc19a1	0.0679	0.251817
Eme2	0.0754	0.25349
Cebpg	0.0656	0.263251
Rpl24	0.0638	0.26468
Rbm42	0.0799	0.265469
Vamp2	0.0681	0.267441
2310004N11Rik	0.0726	0.268146
Nol1	0.0735	0.269544
AI593442	0.0816	0.272164
Hspa9	0.0742	0.272165
Mrps33	0.0416	0.273369
Cited2	0.083	0.273823
Sco1	0.0742	0.275207
2310045N01Rik	0.0626	0.275701
Resp18	0.059	0.279951
Kcnq2	0.0532	0.28061
B930075F07	0.0555	0.280965
0610009O20Rik	0.0799	0.283195

Gene.name	q.value	log2.fc
Atg12	0.0494	0.28384
Cited2	0.0532	0.288039
Aars	0.053	0.288954
Prodh	0.0799	0.290377
Itfg2	0.0749	0.291127
Dph2	0.0556	0.292175
B930076A02	0.0813	0.292318
Txndc13	0.0549	0.292341
Spry3	0.0566	0.293702
Lrrc24	0.0811	0.293813
Gars	0.073	0.294331
Mrpl34	0.075	0.297914
Lars	0.0737	0.299517
Creb3	0.0735	0.299579
Psenen	0.0799	0.300067
Ppp1r11	0.0404	0.300205
Ttc9c	0.0534	0.300392
Prkrip1	0.0726	0.30044
Scg2	0.0615	0.30054
4932417H02Rik	0.0556	0.302188
A630084D02Rik	0.0799	0.304275
Herpud1	0.0572	0.307249
2310001H12Rik	0.059	0.312295
Crcp	0.0799	0.312586
Phf5a	0.076	0.31422
Lor	0.0603	0.315525
Sobp	0.0783	0.316202
Slc1a4	0.0327	0.316264
2700007P21Rik	0.0483	0.319815
Lars	0.0742	0.31988
Sobp	0.0816	0.321579
Dph2	0.0637	0.322526
Ppfia1	0.0523	0.32476
LOC219106	0.0756	0.328118
Armc6	0.045	0.332464
Ccl3	0.0735	0.338499
Samd14	0.0523	0.342155
LOC100042179	0.0756	0.34237
Bdnf	0.0681	0.342407

Gene.name	q.value	log2.fc
Cebpb	0.0749	0.34316
Fbxw7	0.0799	0.344797
Silg111	0.0549	0.3499
Polrmt	0.0807	0.355044
Hn1	0.0816	0.357982
Tmsb10	0.0799	0.359909
Hspa9	0.0549	0.361298
Traf3	0.0555	0.36274
Cda	0.078	0.362949
Ppp2r5d	0.0799	0.366861
Krtcap2	0.0523	0.369508
Rpp21	0.0773	0.371909
Kif5c	0.0523	0.372773
Otub2	0.0822	0.373273
Asns	0.0523	0.377895
Adprh	0.0756	0.379301
Traf3	0.0822	0.38194
2210418O10Rik	0.0726	0.388602
Atf4	0.0756	0.388824
Lonp1	0.045	0.388884
Aldh18a1	0.0613	0.39186
Cox10	0.0523	0.391958
Cox6a2	0.0598	0.397847
Inpp5e	0.0788	0.406217
Herpud1	0.0742	0.41637
Trib3	0.0556	0.424718
Spin2	0.045	0.431905
Asns	0.0832	0.433114
LOC100044736	0.0549	0.437268
Gm129	0.0579	0.440883
Fbxw7	0.045	0.442991
Otub2	0.0735	0.446521
Spin2	0.0589	0.447936
Plekhm2	0.0416	0.450606
Nol5a	0.0799	0.457729
1810005K13Rik	0.078	0.462703
Otub2	0.0436	0.471952
Ppfia1	0.0556	0.47242
B230206N24Rik	0.0735	0.491784

Gene.name	q.value	log2.fc
Alkbh7	0.0607	0.493248
Bdnf	0.045	0.52505
Kif5a	0.0539	0.542309
Ppfia3	0.0603	0.544676
1110008P14Rik	0.075	0.550547
Crhbp	0.0549	0.55113
Serpinf1	0.0404	0.556089
Osbpl6	0.0532	0.566368
Stbd1	0.075	0.649952
Kif5a	0.0416	0.660935
Fos	0.0688	0.667382
Kif5a	0.0532	0.718964
Tnfrsf12a	0.0523	0.742107
Mthfd2	0.0726	0.767083
D330014H01Rik	0.0756	0.840207
Ccl4	0.0579	0.857928
Vgf	0.0416	0.928688
Ddit3	0.0549	1.211976
Ddit3	0.0735	1.364746
Chac1	0.045	1.855671

mRNA expression of eEF1A2 from Brain Cloud database.