University of Nottingham

Biological Functions of Cap Adjacent RNA Methylation in Brain

By

Eleanor Bellows

September 2019

Thesis submitted to the University of Nottingham for Degree of Doctor of Philosophy

Acknowledgements

First I would like to express my great appreciation to Prof Rupert Fray and Dr Helen Miranda Knight for the continuous support during my PhD studies and related research. I would like to thank them for all their guidance, ideas and enthusiasm, without them this project would not have been possible.

I would like to thank Dr Zsuzsa Bodi for all her help and assistance in and out of the lab and for teaching me how to perform TLCs, which have become a backbone of this project. I would like to thank Dr Dan Domininssini and Dr Efrat Glick Saar for their advice on performing protein pull downs and for carrying out the mass spectrometry and proteomic analysis of my samples. I would like to thank Dr Claire Gachon and Dr Marie-Mathilde Perrineau for kindly growing the *Ectocarpus siliculosus.* I would like to thank Cameron Grundy for his work advising on the trees of similarity used in my thesis and Nathan Archer for running some of the trees of similarity on his computer. Frank Baron for synthesising the RNA oligonucleotides used in the SCARLET protocol and designing the modified SCARLET protocol and Nick Rielly for extracting the human brain RNA with me. I express my appreciation for all the people who kindly donated animal tissue to this project, without it the work would not have been possible.

I would like to thank past and present members of the Fray and Knight labs for making the labs such an enjoyable place to work, or their help, and for the many discussions of problems and their solutions.

Last but not least, I would like to thank my parents, Ian and Stephanie, and my partner, Stuart for all their support and encouragement over the last four years.

Declaration

I declare that all work presented in this thesis has been produced by myself and has not been submitted for any other degree. Every effort has been made to acknowledge collaborations and the sources of previous research have been cited throughout.

Table of contents

Acknowledgements I
Declaration II
List of Figures VIII
List of Tables X
List of Abbreviations XII
Abstract 1
1.0 Introduction
1.1 The mammalian brain 4
1.1.1 Synaptic plasticity 8
1.1.2 Neurodegenerative disease 9
1.2 Ribonucleic acid species 13
1.2.1 The epitranscriptome 13
1.2.2 Co-transcriptional RNA processing 17
1.3 History of RNA modifications 20
1.4 Readers, writers and erasers 21
1.4.1 m^6A readers, writers and erasers 21
1.4.2 Cap1 readers, writers and erasers 22
1.4.3 m^6Am readers, writers and erasers 22
1.5 Function of RNA modifications 26
1.5.1 m^6A 26
1.5.2 Am 27
1.5.3 m^6Am 27
1.6 Epitranscriptome in the brain 28
1.7 Thesis aims 29
2.0 Materials and Methods
2.1 Materials 33
2.1.1 Human tissue 33
2.1.2 Animal tissue 35
2.1.3 Algae 35
2.1.4 Bacterial strains 36
2.1.5 Plasmids 36
2.1.6 Reagents 37
2.1.7 Enzymes 42
2.1.8 Radioactive nucleotides 44

2.1.9 Buffers, media and stock solutions 44
2.1.9.1 Buffers 44
2.1.9.2 Media 46
2.1.9.3 Stock solutions 46
2.1.10 Kits 49
2.1.11 Primers 50
2.2 Methods 52
2.2.1 RNA extraction 52
2.2.2 Phenol chloroform purification 53
2.2.3 mRNA enrichment 54
2.2.3.1 Terminator exonuclease digestion 54
2.2.3.2 Poly-A purification 54
2.2.3.3 RNase H digestion 55
2.2.4 RNA measurements 56
2.2.5 Thin Layer Chromatography 56
2.2.5.1 Cap adjacent TLC 56
2.2.5.2 Internal TLC 57
2.2.5.3 Running a TLC 57
2.2.5.4 Imaging a TLC screen 58
2.2.6 SCARLET 58
2.2.6.1 Oligonucleotide SCARLET 58
2.2.6.2 U1 snRNA SCARLET 60
2.2.7 Polymerase Chain Reaction 62
2.2.8 Protein cloning 62
2.2.8.1 Plasmid assembly 62
2.2.8.2 Transformation 63
2.2.8.3 Plasmid extraction 63
2.2.8.4 Glycerol stocks 63
2.2.9 Sequencing 64
2.2.10 Pilot experiments 64
2.2.11 Protein gel 64
2.2.12 Protein production and purification 65
2.2.13 DCP2 treatment of poly A RNA 66
2.2.14 Protein pull down 67
2.2.14.1 RNA oligonucleotide transcription and capping 67
2.2.14.2 First protein pull down method 69

2.2.14.3 Second protein pull down method 69
2.2.14.4 Protein digestion 70
2.2.14.5 Mass spectrometry and proteomics 70
2.2.15 Bioinformatic analysis 71
2.2.15.1 Searching for m^{6}Am in m^{6}A-seq data 71
2.2.15.2 Trees of Similarity 73
3.0 Evolutionary Conservation of the m^{6}Am Modification
3.1 Preface 77
3.2 Results 78
3.2.1 Optimisation of rRNA removal 78
3.2.2 Thin layer Chromatography to detect m^{6}Am 82
3.2.3 Variation in biological replicates 83
3.2.4 Do *D. melanogaster* snRNAs contain m^{6}Am? 87
3.2.5 Identifying orthologues of m^{6}Am associated proteins by evolutionary conservation of protein sequences 93
3.2.6 Searching for m^{6}Am in *E. siliculosus* 101
3.3 Discussion 108
3.3.1 Analysis of rRNA removal 108
3.3.2 TLC as a method for modified nucleotide detection 109
3.3.3 Evolution of m^{6}Am associated proteins 110
3.3.4 Proteins required for the m^{6}Am modification 111
3.3.5 The role of m^{6}Am 113
3.3.6 Potential role of modified nucleotides in the brain 114
4.0 m^{6}Am in the Human Brain
4.1 Preface 116
4.2 Results 117
4.2.1 Detection of m^{6}Am in brain 117
4.2.2 Detection of m^{6}Am from m^{6}A sequencing 122
4.2.2.1 *De novo* motif analysis 130
4.2.2.2 GO analysis of potential m^{6}Am transcripts 133
4.3 Discussion 142
4.3.1 Human brain function and disease 142
4.3.2 m^{6}Am in neurodegenerative diseases 143
4.3.3 Detection of m^{6}Am from sequencing data 145
4.3.4 The function of m^{6}Am in the human brain 147
4.3.5 Future work 147

5.0 Identifying m^{6}Am Mofications by DCP2 Enzymatic Treatment
5.1. Preface 150
5.2. Results 151
5.2.1. Expressing soluble DCP2 protein 151
5.2.2. Using DCP2 to decap *R. norvegicus* brain RNA 156
5.3. Discussion 158
5.3.1. Sequencing transcripts with m^{6}Am 158
5.3.2. Technical differences with published findings 159
5.3.3. DCP2 *in vivo* 160
5.3.4. Regulation of translation by DCP2 160
6.0 Identification of m^{6}Am binding proteins
6.1 Preface 163
6.2 Results 164
6.2.1 Optimisation of substrate synthesis 164
6.2.2 Proteomic analysis of protein pull downs 169
6.2.3 Comparison of proteomic analysis 181
6.2.4 Gene ontology of modification binding proteins 184
6.3 Discussion 192
6.3.1 Transcription of oligonucleotides 192
6.3.2 Protein pull downs 193
6.3.3 Writers and erasers 193
6.3.4 m^{6}Am reader proteins 194
6.3.5 Future studies 195
7.0 Genera Discussionl
7.1 Global changes in cap adjacent methylation 197
7.2 Molecular function of cap adjacent methylation 198
7.2.1 Am binding proteins indicate the molecular effect of Am methylation ... 199
7.2.2 m^{6}Am binding proteins indicate the molecular effect of m^{6}Am methylation 200
7.2.3 Transcripts modified by m^{6}Am suggest cellular functions 201
7.3 Cap adjacent methylation and neurodegenerative disease 202
7.4 Cap adjacent methylation and memory formation 204
7.5 Future directions of research into cap adjacent methylation research 204
7.5.1 Function of m^{6}Am in healthy and diseased cells 205
7.6 Concluding remarks 206
References 208

Appendix 226
Accession numbers for proteins in evolutionary trees 226
Protein Alignments 238
Ensembl IDs for putative m^6Am binding proteins 312

List of Figures

Chapter 1

Figure 1.1 Schematic representation of the human brain……7

Figure 1.2 A diagrammatic illustration of the structure of the hippocampus and dentate gyrus……7

Figure 1.3 Healthy synapse and a synapse from a patient with Parkinson's disease……12

Figure 1.4 Structure of cap0, cap1 and m^6Am first nucleotides……16

Figure 1.5 Co-transcriptional processing of mRNA……19

Figure 1.6 Readers, writers and erasers of m^6A……..24

Figure 1.7 Readers, writers and erasers of the Am and m^6Am modifications……25

Chapter 3

Figure 3.1 Agilent graphs showing the quality of total and rRNA depleted RNA……79

Figure 3.2 TLCs showing the first nucleotide of *R. norvegicus* RNA enriched for different RNA species……...81

Figure 3.3 Representative TLCs showing the first nucleotide of RNA from brain tissue from different species……85

Figure 3.4 Biological variation in the different organisms……..86

Figure 3.5 TLC showing the first nucleotide from poly(A) RNA from *G. gallus* hypothalamus……..86

Figure 3.6 Optimisation of the SCARLET protocol and analysis of U2 snRNA……..90

Figure 3.7 Diagrammatic Illustration of the modified SCARLET method……..91

Figure 3.8 FTO phylogenetic tree of ALKB family members from animals and algae protein sequences……95

Figure 3.9 Phylogenetic tree of CMTR1 and CMTR2 protein orthologues……..99

Figure 3.10 Phylogenetic tree of PCIF1 protein orthologues……..100

Figure 3.11 Representative TLCs detecting modified nucleotides in *E. siliculosus.*103

Figure 3.12 Phylogenetic tree of METTL3 protein orthologues……..104

Figure 3.13 Phylogenetic tree of METTL14 protein orthologues……105

Figure 3.14 Phylogenetic tree of WTAP protein orthologues……..106

Chapter 4

Figure 4.1 Representative TLCs examining cap methylation in RNA from non-disease control, Lewy body dementia and Parkinson's disease hippocampal brain samples……..118

Figure 4.2 Work flow of bioinformatic analysis to detect m^6Am sites……..124

Figure 4.3 Peak annotations and starting nucleotides of cDNA transcripts of the genes in the total dataset, the total peaks and TSS peaks in the foetal brain, white matter and grey matter m^6A sequencing…….128

Figure 4.4 *De novo* motif analysis of potential m^6Am peaks……...132

Chapter 5

Figure 5.1. Map of the pMAL-c5x-hDCP2 construct……..153

Figure 5.2 Protein gels from the pilot experiments optimising soluble DCP2 protein production……..154

Figure 5.3 Protein gel showing DCP2 protein purification elutions……155

Figure 5.4 TLCs showing the first nucleotide of control and DCP2 treated RNA….157

Chapter 6

Figure 6.1 Optimisation of capping and 2′-O-methylation of the transcribed cap1 oligonucleotide……167

Figure 6.2 TLCs showing the first nucleotide of the three transcribed oligonucleotides……...171

List of Tables

Chapter 2

Table 2.1 Patient details from which the hippocampal tissue was donated……...34

Table 2.2 Bacterial strains used in this study……...36

Table 2.3 Table of reagents……37

Table 2.4 Table of enzymes……....42

Table 2.5 Kits used for specific assays……49

Table 2.6 Primers and oligonucleotides used in this study……50

Table 2.7 Pilot experiment conditions……64

Chapter 3

Table 3.1 Table showing the presence of modified nucleotides and the associated methylases and demethylases……89

Chapter 4

Table 4.1 Number of peaks and unique gene IDs at the different stages of the m^6Am sequencing analysis work flow……...125

Table 4.2 GO terms associated with transcripts that putatively have the m^6Am modification in foetal brain……136

Table 4.3 Reactome analysis for the transcripts that putatively have the m^6Am modification in foetal brain……137

Table 4.4 GO terms associated with transcripts that putatively have the m^6Am modification in grey matter……138

Table 4.5 Reactome analysis for the transcripts that putatively have the m^6Am modification in grey matter……139

Table 4.6 GO terms associated with transcripts that putatively have the m^6Am modification in white matter……...140

Table 4.7 Reactome analysis for the transcripts that putatively have the m^6Am modification in white matter……...141

Chapter 6

Table 6.1 Table of proteins which were found to be bound to the three RNA oligonucleotides……...172

Table 6.2 Proteins which were found to be bound to the cap0 RNA oligonucleotide (m^7GpppA)……175

Table 6.3 Proteins that were found to be bound to the cap1 RNA oligonucleotide (m^7GpppAm)……177

Table 6.4 Proteins that were found to be bound to the m^6Am RNA oligonucleotide (m^7Gpppm^{6}Am)……...179

Table 6.5 Proteins that were found to bind to the RNA cap modifications were also found to bind to RNA in the database searches…….……182

Table 6.6 GO terms associated with the proteins that bound to the cap0 oligonucleotide……186

Table 6.7 Reactome analysis for the proteins that bound to the cap0 oligonucleotide……187

Table 6.8 GO terms associated with the proteins that bound to the cap1 oligonucleotide……188

Table 6.9 Reactome analysis for the proteins that bound to the cap1 oligonucleotide……....189

Table 6.10 GO terms associated with the proteins that bound to the m^6Am oligo……...190

Table 6.11 Reactome analysis for the proteins that bound to the m^6Am oligonucleotide……191

List of Abbreviations

A	Adenosine
ALKBH1	AlkB Homolog 1
ALKBH2	AlkB Homolog 2
ALKBH3	AlkB Homolog 3
ALKBH4	AlkB Homolog 4
ALKBH5	AlkB Homolog 5
ALKBH6	AlkB Homolog 6
ALKBH7	AlkB Homolog 7
ALKBH8	AlkB Homolog 8
Am	2′O-ribose methyladenosine
ANOVA	Analysis of variance
BDNF	Brain-derived neurotrophic factor
C	Cytosine
CA1	Cornu Ammon 1
CA2	Cornu Ammon 2
CA3	Cornu Ammon 3
cap0	m^7GpppN structure
cap1	m^7GpppNm structure
cap2	m^7GpppNmNm structure
cDNA	complementary DNA
CFIIM	Cleavage factor 2
CFIM	Cleavage factor 1
Cm	2′O-ribose methylcytosine
CMTR1	Cap Methyltransferase 1
CMTR2	Cap Methyltransferase 2
CNS	Central nervous system
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulation factor
D2R	Dopamine receptor 2
D3R	Dopamine receptor 3
DAT	Dopamine transporter
DCP1	Decapping MRNA 1
DCP2	Decapping MRNA 2
DNA	Deoxyribonucleic acid
DXO	Decapping Exoribonuclease
EDC1	Enhancer Of MRNA Decapping 1
EDC3	Enhancer Of MRNA Decapping 3
EDC4	Enhancer Of MRNA Decapping 4
eRIC	RNA interactome capture
FTO	Fat mass and obesity-associated protein
G	Guanine
Gm	2′O-ribose methylguanine
GO	Gene ontology
GRIK	Glutamate Ionotropic Receptor Kainate Type Subunit 1

HNRNPA2B1	Heterogeneous Nuclear Ribonucleoprotein A2/B1
HNRNPC	Heterogeneous Nuclear Ribonucleoprotein C
Homer	Hypergeometric Optimization of Motif EnRichment
IFIT1	Interferon Induced Protein With Tetratricopeptide Repeats 1
ILF2	Interleukin Enhancer Binding Factor 2
IPTG	Isopropyl-β-D-thiogalactoside
kDa	KiloDaltons
KIAA1429	Vir Like M^6A Methyltransferase Associated
LB	Lysogeny broth
LBD	Lewy body dementia
lncRNA	Long no coding RNA
LTD	Long term depression
LTP	Long term potentiation
m^6A	N^6-methyladenosine
m^6Am	m^7Gpppm^6Am structure
m^6Am	N^6-2′O-dimethyladenosine
me-RIP-seq	m^6A methylation RNA immunoprecipitation chip sequencing
METTL14	Methyltransferase Like 14
METTL3	Methyltransferase Like 3
mi-CLIP	m^6A individual-nucleotide-resolution cross-linking and immunoprecipitation
miRNA	Micro RNA
mRNA	Messenger RNA
MTA	N6-adenosine-methyltransferase MT-A70-like
MTB	N6-adenosine-methyltransferase non-catalytic subunit MTB
NCL	Nucleolin
NSPC	Neural stem and progenitors cells
Nudt16	Nudix Hydrolase 16
Nudt3	Nudix Hydrolase 3
OOPS	Orthogonal organic phase separation
PAGE	Polyacrylamide gel electrophoresis
PAP	Polyadenylate polymerase
PARP1	Poly(ADP-Ribose) Polymerase 1
PB	Parkinson's disease
PCIF1	PDX1 C-Terminal Inhibiting Factor 1
PCR	Polymerase chain reaction
PNK	Poly nucleotide kinase
pTEX	Phenol Toluol extraction
RBM15	RNA Binding Motif Protein 15
RIG1	Retinoic acid inducible 1
RIP-seq	RNA immunoprecipitation chip sequencing
RNA	Ribonucleic acid
RNA pol II	RNA polymerase II
RNA pol II CTD	RNA polymerase II C terminal binding domain
RppH	RNA 5' Pyrophosphohydrolase
rRNA	Ribosomal RNA

SAM	S-adenosyl-L-methionine
SCARLET	Single-nucleotide resolution, termed as "site-specific cleavage and radioactive-labelling followed by ligation-assisted extraction and thin-layer chromatography
SEM	Standard error of the mean
SNARE complex	SNAP Receptor complex
snRNA	Small nuclear RNA
SRSF10	Serine And Arginine Rich Splicing Factor 10
SRSF3	Serine And Arginine Rich Splicing Factor 3
ssDNA	Single stranded DNA
ssRNA	Single stranded RNA
TLC	Thin layer chromatography
tRNA	Transfer RNA
TSS	Transcription start site
U	Uridine
Um	2′O-ribose methyluridine
WTAP	Wilms Tumour 1-Associating Protein
XRN1	5′-3′ Exoribonuclease 1
YTHDC1	YTH Domain Containing 1
YTHDC2	YTH Domain Containing 2
YTHDF1	YTH N6-Methyladenosine RNA Binding Protein 1
YTHDF2	YTH N6-Methyladenosine RNA Binding Protein 2
YTHDF3	YTH N6-Methyladenosine RNA Binding Protein 3
ZC3H13	Zinc Finger CCCH-Type Containing 13
ZEB1	Zinc finger E-box binding homeobox 1
ZFP217	Zinc Finger Protein 217

Abstract

In eukaryotes, mRNA transcripts that are transcribed by RNA polymerase II (RNA pol II) and then subsequently co-transcriptionally modified. One of the co-transcriptional modifications is the addition of the 7-methyl guanine cap to the 5′ end of the RNA. If no further modifications are added to the first two nucleotides, this cap structure is referred to as cap0, and is found in plants and yeast. Additional modifications take place in higher eukaryotes. For example the ribose sugar of the first nucleotide can be methylated at the 2′-O position to give cap1. If this cap1 methylation occurs on an adenosine, the nitrogenous base can be further methylated generating N^6,2′-O-dimethyladenosine (m^6Am). The function of the m^6Am modification is currently unknown. We propose that the m^6Am modification may regulate translational processes which underlie neuronal and synaptic function. To understand this, transcripts that have the modification, proteins that bind the modification and evolutionary profile of m^6Am will be investigated by preforming Thin Layer Chromatography (TLCs), protein pull downs using RNA oligonucleotides and RNA sequencing and gene orthologue analysis.

Within the evolutionary tree, m^6Am was found in vertebrates and preliminary evidence suggests its presence in the marine algae *E. siliculosus.* The organisms which contain m^6Am methylation all had CMTR1, (the Am methylase), PCIF1, (the m^6Am methylase) and FTO, (the m^6Am demethylase). However, the invertebrate species where m^6Am was absent, also contained PCIF1 and CMTR1 enzymes.

In this study, proteins binding to the Am and m^6Am modification were found to be different indicating that Am has a different role than m^6Am. Am-binding proteins were found to be associated with RNA processing such as splicing, removal of RNA secondary structures and translation. In contrast m^6Am binding proteins were shown

to be associated with microtubules and extracellular vesicles suggesting a function in RNA transport throughout the cell and between cells.

Previously DCP2 was reported to not decap transcripts that started with m^6Am, it was hypothesised that if this could be confirmed it could be used in a novel enzymatic enrichment sequencing method for m^6Am. Contrary to this published data DCP2 was found to preferentially decap Am and m^6Am transcripts. This information combined with the activity of other known decapping enzymes highlights a potential mechanism of regulating different sub-populations of RNA denoted by the cap structure.

Putative m^6Am methylated transcripts were identified using a bioinformatic approach from m^6A sequencing of human foetal brain, and parahippocampus brain- grey matter and white matter. Cell type specific transcripts were indicated to be methylated suggesting that m^6Am may increase translation or stability of these transcripts. An increase in global Am and m^6Am abundance in Parkinson's disease and Lewy body dementia-affected hippocampus tissue specimens was observed. Further research is needed to provide insight into which specific transcripts show changes in Am and m^6Am methylation as the diseases progress.

Taken together, these results suggest that the function of the m^6Am modification is context dependant. At the molecular level, m^6Am may instigate the same effect across all cells but it is m^6Am binding proteins and the individual transcripts that are methylated that provide the unique cell specific regulation of mRNA fate. Observations made in this thesis provide important questions for further research into the role of m^6Am in brain function and neurodegenerative diseases.

Chapter 1: Introduction

1.1 The mammalian brain

The brain is the control centre of the human body. It regulates the autonomic nervous system which controls breathing, swallowing and circadian rhythm as well as many other function that occur without conscious control (Svorc, 2018). The brain regulates the somatic nervous system which control conscious movements like the movement of limbs (Cuevas, 2007). The brain is also the centre of higher level functions such as cognition, the process of acquiring knowledge and understanding through thought, and experience, as well as the formation and recall of memories.

In the adult human brain, such cognitive processes are performed by approximately 86 billion neuronal cells and 84 billion non-neuronal cells (Azevedo et al., 2009). These cells are arranged into networks and regions which perform distinct functions. The brain can be divided up into the hind brain which is made up of the cerebellum, pons and medulla oblongata; the mid brain; and the forebrain which is made up of the thalamus, hypothalamus and cerebrum (Figure 1.1). In the hindbrain the medulla controls the autonomic nervous system such as breathing and blood flow, the pons affects activities such as sleeping, waking and dreaming; and the cerebellum controls balance and coordinated movement. The mid brain's main function is in processing and relaying visual and auditory stimuli and is also involved in the control of body movement (Mack et al., 2013).

The cerebrum is the largest part of the brain and performs higher function processes. The cerebrum can be divided into smaller structures called the frontal, occipital, parietal and temporal lobes. The frontal lobe carries out processes such as thinking, decision making and planning. The occipital lobe processes visual information from the eyes. The parietal lobe processes information to do with taste, temperature and touch. The temporal lobe process sounds, interprets language and effectuates emotional and contextual responses. Within the temporal lobe is the limbic system which is responsible for processing emotion and memory. In this system, one structure,

the hippocampus is predominantly where learning and episodic memories are performed and it is also associated with emotional processing through its connections with the amygdala (Mack et al., 2013). Within the limbic system surrounding the hippocampus is the parahippocampal cortex, this structure is also involved in visuospatial processing and episodic memory (Aminoff et al., 2013). There is much debate on how the small curved structure of the hippocampus is involved in the formation of memories (Maguire et al., 2016).

The hippocampal formation includes the dentate gyrus an area where adult neurogenesis occurs, and three main neuronal populations are known as Cornu Ammon 1 (CA1), Cornu Ammon 2 (CA2) and Cornu Ammon 3 (CA3) (van Dijk et al., 2016) (Figure 1.2). The Cornu Ammon (CA) pyramidal cells in the hippocampus are unique in that, they form a unidirectional circuit termed the performant pathway. In a simplified view, input comes from the dentate gyrus to the CA3 region neurons, through to the CA2 and CA1 cells. In partnership, the activity in all these regions with the integration of information from other regions of the brain orchestrates memory formation (Basu and Siegelbaum, 2015). It is well known that atrophy of the hippocampus occurs with old age (Golomb et al., 1993). However in neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease and Dementia with Lewy bodies, atrophy is more extreme owing to abnormal cell aggregates which disrupt cell integrity (Irvine et al., 2008).

Brain tissue can be divided into two types by cell composition. Grey matter is predominantly composed of neuronal cells, neuronal cell bodies and branching dendrites. In contrast, white matter is comprised of glial cells which support neuronal cell function, e.g. oligodendrocytes, astrocytes, ependymal cells and microglia (Jäkel and Dimou, 2017) as well as myelinated axons. Oligodendrocytes are the myelinating cells of the central nervous system (CNS). They insulate the cells, so that the electrical impulses that travel along the axons require less energy, 5000 times less energy than

if the cells were unmyelinated (Ritchie Joseph, 1982). In the cortex, oligodendrocytes wrap around the axons creating up to 60 lamellae. They form between 30 and 80 internodes allowing the action potential to jump along the axon (Matthews and Duncan, 1971, Chong et al., 2012). Astrocytes are a heterogeneous population of cells that support a wide range of function such as the formation and maturation of synapses (Diniz et al., 2014) as well as maintenance, and the pruning of synapses (Chung et al., 2015, Chung et al., 2016). In the hippocampus, pyramidal cells make up most of the grey matter while the stratum oriens and stratum radiatum/lacunosum/moleculare are classified as white matter.

The cells in the brain communicate through action potentials, i.e. electric currents created by ions travelling through voltage gated ion channels in the membrane changing the polarity of the cell. Action potentials propagate from one neuron to another through the process of synaptic transmission. This is when the presynapse neuron becomes depolarised by an action potential, resulting in voltage sensitive Ca^{2+} channels opening and exocytosis of vesicles containing neurotransmitter. The released neurotransmitter molecules interacts with receptors in the postsynaptic neuron membrane resulting in local depolarisation of the post synaptic cell which can lead to an action potential to propagate along the postsynaptic neuron (Holz RW, 1999). If synapses are co-activated frequently, communication between the pre and post synaptic cells strengthen (Murthy, 1998).

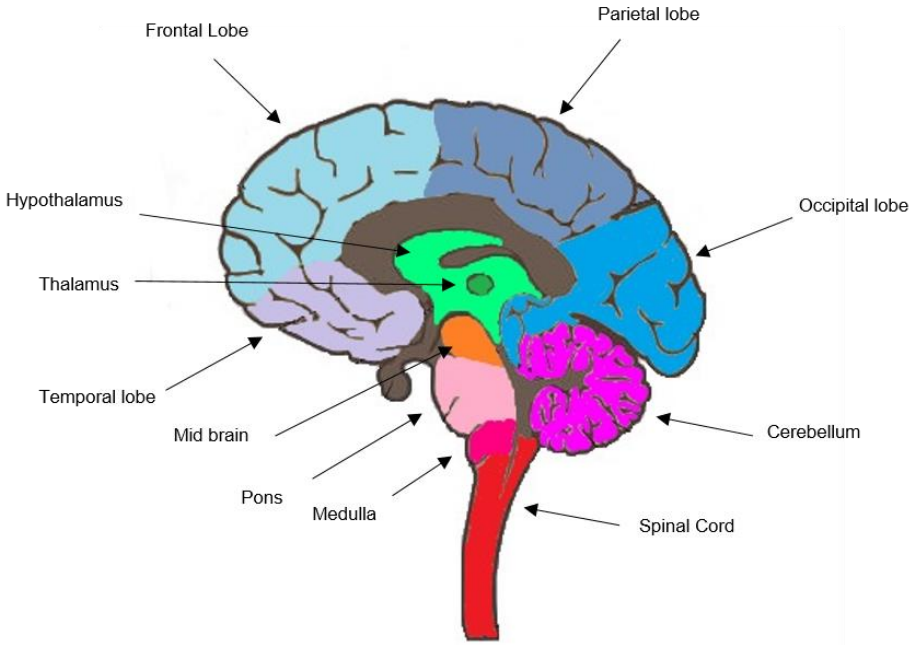


Figure 1.1. Schematic representation of the human brain. The brain can be divided up into three main areas: the hind brain made of the pons (pale pink), medulla (magenta) and cerebellum (purple); the mid brain (orange); and the forebrain made of the thalamus (dark green), hypothalamus (light green) and cerebrum (blue). The cerebrum can be further divided into the temporal, frontal, occipital and parietal lobe regions. Image adapted from https://www.thoughtco.com/mesencephalon-anatomy-373223

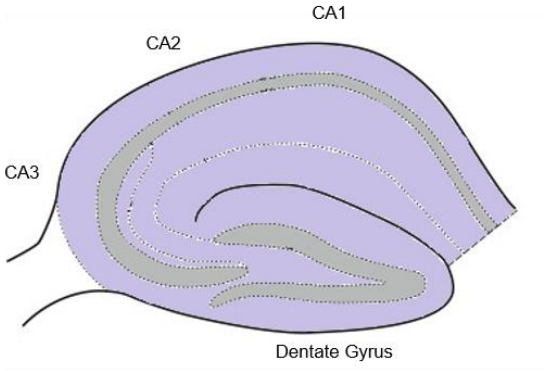


Figure 1.2. A diagrammatic illustration of the structure of the hippocampus and dentate gyrus. The hippocampus is located within the temporal lobe and is composed of three cell populations CA1, CA2 and CA3. Input signals come from the dentate gyrus, through to the CA3 neurons and then through the CA2 and CA1 neurons in a unidirectional circuit. Adapted from image in (Sheppard et al., 2019).

1.1.1 Synaptic plasticity

Synaptic plasticity is the term given to describe the ability of synapses to strengthen or weaken over time in response to neuronal activity. This process is protein synthesis dependant and underpins the process of memory formation (Pfeiffer and Huber, 2006). The first person to propose that protein synthesis was required for memory formation was Monné in 1948. He wrote that "… *it is possible that synthesis of specific proteins is the essential physical phenomenon paralleling memory, fantasy, and intuition. This hypothesis is supported by the fact that protein synthesis occurs in strongly stimulated neurons and that cells are able to 'learn' to synthesize new specific proteins*…" (Monne, 1948). As research has progressed this idea has developed and become central to the synaptic plasticity and memory hypothesis. Asserting that activity dependant synaptic plasticity happens at the relevant synapses during memory formation and is both necessary and sufficient to the encoding and storage of memory in that area of the brain (Takeuchi et al., 2014).

A molecular model of synaptic plasticity can be measured by changes in Long Term Potentiation (LTP), a persistent increase in synaptic strength following high frequency stimulation of a chemical synapse, and Long Term Depression (LTD), a persistent decrease in synaptic strength following long term stimulation of a synapse. For LTP and LTD to occur and memories to be formed, translation of RNAs to produce proteins which are involved in synaptic processes is required (Costa-Mattioli et al., 2009). The first evidence of this was shown in 1963 when translational inhibitors were injected into the frontal cortex of mice which stopped the formation of memories of where an electric shock was given in a Y maze (Flexner et al., 1965). Later through a series of experiments on rat hippocampal sections, Kang et al., showed that Brain-Derived Neurotrophic Factor (BDNF) stimulates local protein synthesis and is required for synaptic enhancement and LTP (Kang and Schuman, 1996).

At dendrites there are over 2500 mRNAs (Cajigas et al., 2012) and they need to be regulated so that the appropriate transcripts get translated in a location specific, time specific and signal specific manner. At synapses in the hippocampus and dentate gyrus there is on average only 8 polyribosomes (Steward and Reeves, 1988, Ostroff et al., 2002). This suggests there may be layers of regulatory control so that when synaptic activity occurs, the correct RNA transcripts are translated quickly. How this is achieved is not fully understood, Schuman et al., proposed two hypotheses. One, that synaptic signalling controls protein translation; two, that synaptic signalling controls the mRNAs that are being chosen for translation (Schuman et al., 2006).

Defects in translational control has been linked to cognitive and neurodegenerative disorders (Buffington et al., 2014, Richter and Coller, 2015). There is increasing evidence that subtle disruptions in translational regulation can have dramatic consequences for neurons, including dysregulation of EIF2 (Kapur et al., 2017). Alzheimer's disease has been linked to reduced translation, in the neurons less RNA is associated with the polysomes (Langstrom et al., 1989). There is also increased levels of phosphorylated EIF2α in the cortex of patients with Alzheimer's disease (Hoozemans et al., 2009) and in the substantia nigra in patients with Parkinson's disease (Hoozemans et al., 2007). Phosphorylated EIF2α also has a role in inhibiting translation (Chesnokova et al., 2017). There are many other perturbations of translations in neurodegenerative diseases, there is evidence that translation can increase as well as decrease (Kapur et al., 2017).

1.1.2 Neurodegenerative disease

Neurodegenerative disease is an umbrella term for diseases which progressively affect neurons commonly leading to impairment of cognition and movement. Neurodegenerative diseases are often age related and incurable. As there are putatively only two regions of the brain which undergo adult neurogenesis e.g.

integration of new neurones during adulthood, the brain cannot reverse the effect of cell death if large areas of tissue matter become atrophied. Two examples of neurodegenerative disease which cause neuronal atrophy are Lewy body dementia and Parkinson's disease. Both diseases are characterised by the accumulation of protein aggregates called Lewy bodies which are deposits of α-synuclein. These aggregates progressively spread from localised regions throughout the brain causing loss of neurones and disruption of grey and white matter integrity (Goedert, 2015, Braak et al., 2003).

In Parkinson's disease, cell loss starts in the substantia nigra in the mid brain and in Lewy body dementia loss is localised to the cerebral cortex and limbic regions (Power et al., 2017, Naoi and Maruyama, 1999). Disease progression in Parkinson's disease is slower and longer than Lewy body dementia. In Parkinson's disease after the initial loss of cells in the mid brain there is a slow progression of the Lewy bodies throughout the brain. Dementia like symptoms occur when the Lewy bodies reach the neocortex. Lewy body dementia has a shorter disease course with substantial Lewy body and Alzheimer's disease pathology spreading throughout the brain (Halliday and McCann, 2010). As the pathologies spread both diseases often result in hippocampal atrophy and a loss of memory formation (Camicioli et al., 2003, Adamowicz et al., 2017).

In the early prodromal stages of Parkinson's disease before motor defects become apparent there are already synaptic and axonal abnormalities (Dijkstra et al., 2015). In this stage of the disease there is an increase in α-synuclein which accumulates as oligomers and protofibrils at the presynaptic terminal (Figure 1.3b) (Ingelsson, 2016, Li et al., 2001). The increase in toxic α-synuclein forms reduce the synaptic vesicle pool, impairs their trafficking and can misregulate members of the SNARE complex impairing tethering, docking, priming and fusion of synaptic vesicles (Murphy et al., 2000, Larsen et al., 2006, Choi et al., 2013). α-synuclein aggregates are also found where synaptic vesicles should be fusing, replacing the proteins that should be present

(Martin et al., 2006). Dopaminergic neurons are specifically vulnerable because α-synuclein negatively modulates the proteins that are responsible for dopamine synthesis and also inhibits the enzymes that regulate the uptake of dopamine such as VMAT2 and DAT (Lotharius and Brundin, 2002, Vaughan and Foster, 2013). Reduced dopamine and impaired synaptic communication, resulting cell in death causes motor symptoms, such a as bradykinesia, rigidity and resting tremor (Bridi and Hirth, 2018). Figure 1.3a depicts a healthy synapse and a synapse from a patient with Parkinson's disease showing the effects of the α-synuclein.

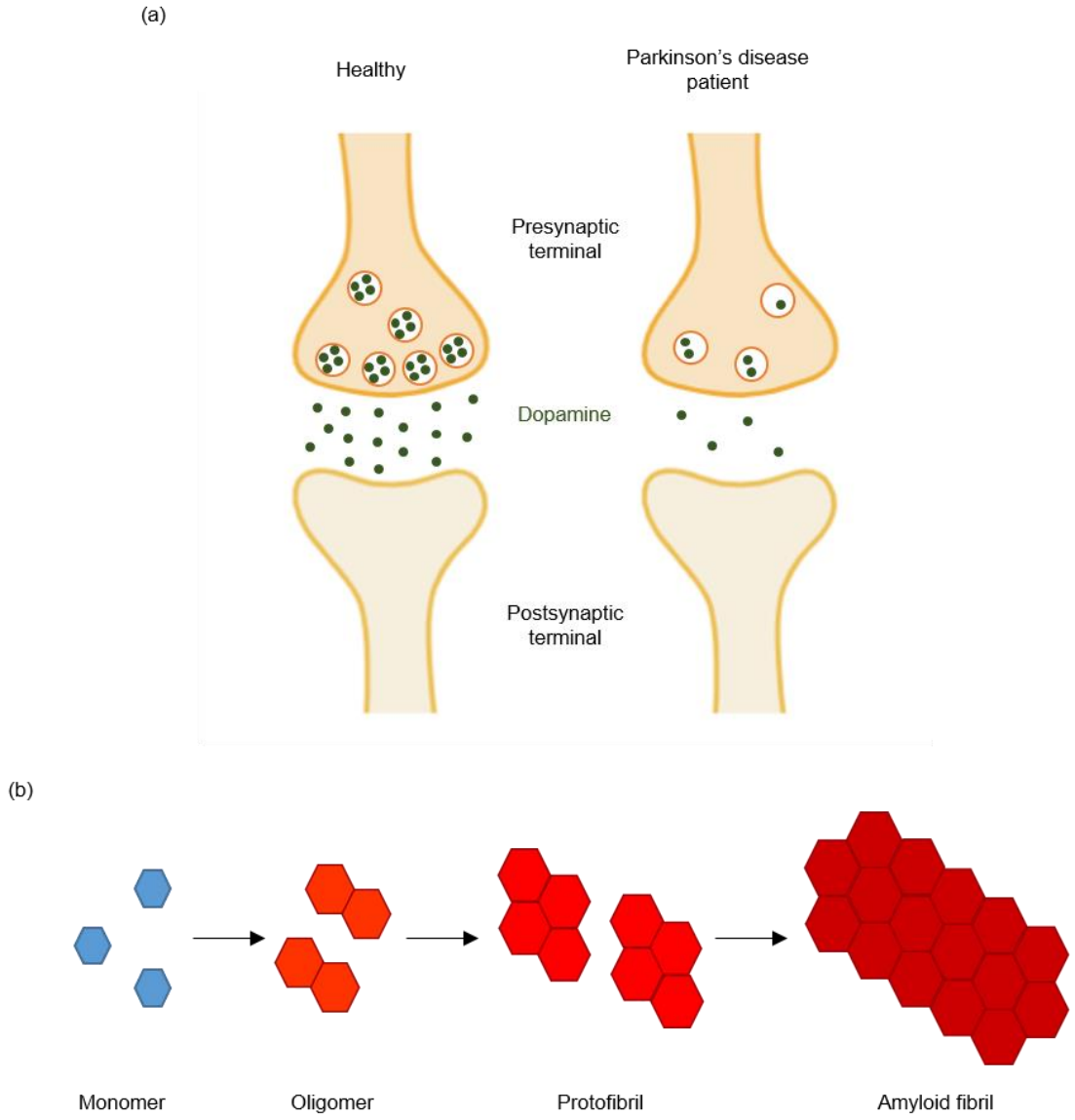


Figure 1.3. A healthy synapse and a synapse from a patient with Parkinson's disease. a) In a healthy synapse in a dopaminergic neuron there are many vesicles waiting to fuse to the membrane in the event of an action potential and the vesicles are full of dopamine. At a synapse in Parkinson's disease patients there is a reduced number of vesicles and the vesicles contain less dopamine. b) In healthy cells α-synuclein acts as a monomer involved in dopamine production and regulation of vesicle binding so dopamine can be released into the synaptic cleft. In Parkinson's disease the monomers form oligomers and protofibrils in early stages of the disease which are toxic to cell and synapse function. Later in the disease amyloid fibrils aggregate together to form Lewy bodies causing further disruption. This illustration was partially created using Biorender.

1.2 Ribonucleic acid species

Ribonucleic acid (RNA) is a molecule made of 4 bases adenine, cytosine, guanine and uracil. Each base is made of a ribose sugar, phosphate molecule and a nitrogenous base. They are assembled by three RNA polymerases to form macromolecules which perform many functions in eukaryotic cells. In rapidly growing cells, ~80% of the RNA is ribosomal RNA (rRNA), a further ~15% is transfer RNA (tRNA). ~1% of the RNA is messenger RNA (mRNA) and the remaining 4% is made up of non-coding RNAs such as microRNAs (miRNAs), long noncoding RNAs (lncRNAs) and small nuclear RNAs (snRNA) .

mRNA carries genetic information transcribed from the DNA out into the cytoplasm where it will be translated into protein. mRNA is transcribed by RNA polymerase II and during this process it is modified by the addition of the 5′ guanine cap, a 3′ poly adenylated tail and the removal of introns. This results in a stable mRNA that can leave the nucleus and go to the ribosomes in the cytoplasm. mRNA is highly regulated through co-transcriptional processing, proteins that bind the RNA and RNA modifications that are added to it, so that the correct proteins are translated at the correct time.

1.2.1 The epitranscriptome

The prefix ‘epi’ is from Greek and means above or upon. Epigenetics relates to the addition of modifications on DNA and histones that regulate gene expression without changing the nucleic acid sequence. The recent term epitranscriptomics, refers to all covalent modifications of the transcriptome. Any modification added to ribonucleotides irrespective of function or heritability is considered to be part of the epitranscriptome (2017). Currently 172 post transcriptional modifications have been identified on RNA molecules. These 172 modifications are found on non-coding RNAs such as rRNAs

and tRNAs as well as coding mRNAs, playing important roles in their functions. (http://modomics.genesilico.pl/, accessed 13/09/19) (Boccaletto et al., 2017).

rRNAs are heavily modified on the human 80S ribosomal RNA and it has been suggested that one of 14 posttranscriptional modifications can be found on 228 of the nucleotides (Taoka et al., 2018). These modification stabilise the structure of the ribosome increasing protein production (Sloan et al., 2017). 93 RNA modifications can be found on tRNAs (Boccaletto et al., 2017). Some of these are universally conserved while others are species specific (Lyons et al., 2018). One modification found universally is a pseudouridine at position 55 which facilitates proper tertiary structure, the T-arm and D-arm of the tRNA interact through this pseudouridine 55 (Robertus et al., 1974).

The most common internal modification of mRNA is N^6-methyladenosine (m^6A) which is the addition of a methyl group to the 6^{th} nitrogen of an adenosine nucleotide in any internal position along the RNA transcript (Meyer et al., 2012). It is typically found at levels equivalent to 1.8 nucleotides per 1000 bases (Dubin and Taylor, 1975). There are also modifications which occur next to the 5′ guanine cap on the mRNA. The first two nucleotides of the mRNA transcript can be methylated on the ribose sugar, irrespective of the nucleotide forming 2′-O-ribose methyladenosine, 2′-O-ribose methylcytosine, 2′-O-ribose methylguanine, 2′-O-ribose methyluridine (Am, Cm, Gm, Um), this is known as cap1 on the first nucleotide and cap2 on the second nucleotide. If the cap1 nucleotide is adenosine, the nitrogenous base can be further modified to N^6,2′-O-dimethyladenosine (m^6Am). In HeLa mRNA between 20-30% of RNA transcripts have the m^6Am modification (Wei et al., 1975).

Adenine methylation is thought to originate from prokaryote restriction modification systems. This system discriminates between “self” DNA and the invasive “non-self” DNA so that the “non-self” DNA can be degraded (Kobayashi, 2001, Rao et al., 2013). Eukaryotes have acquired adenine methylases in 13 independent events and these

methylases methylate both RNA and DNA (Iyer et al., 2016). mRNA modifications are not found in all eukaryotes. The internal m^6A modification is found in Eukaryotes; animals (Meyer et al., 2012, Dominissini et al., 2012, Dubin and Taylor, 1975), plants (Zhong et al., 2008), and yeast (Clancy et al., 2002). The cap1 modification is found in animals (Furuichi, 2015) and m^6Am is found in vertebrates (Sendinc et al., 2019, Furuichi, 2015). Viruses with RNA genomes also have these modifications. This is so the virus can use the cell machinery to replicate and to avoid the host's immune system (Manners et al., 2019, Dang et al., 2019). Both m^6A and m^6Am have been found on different virus strains on both genomic RNA and mRNAs (Wei et al., 1975, Tirumuru and Wu, 2019, Manners et al., 2019).

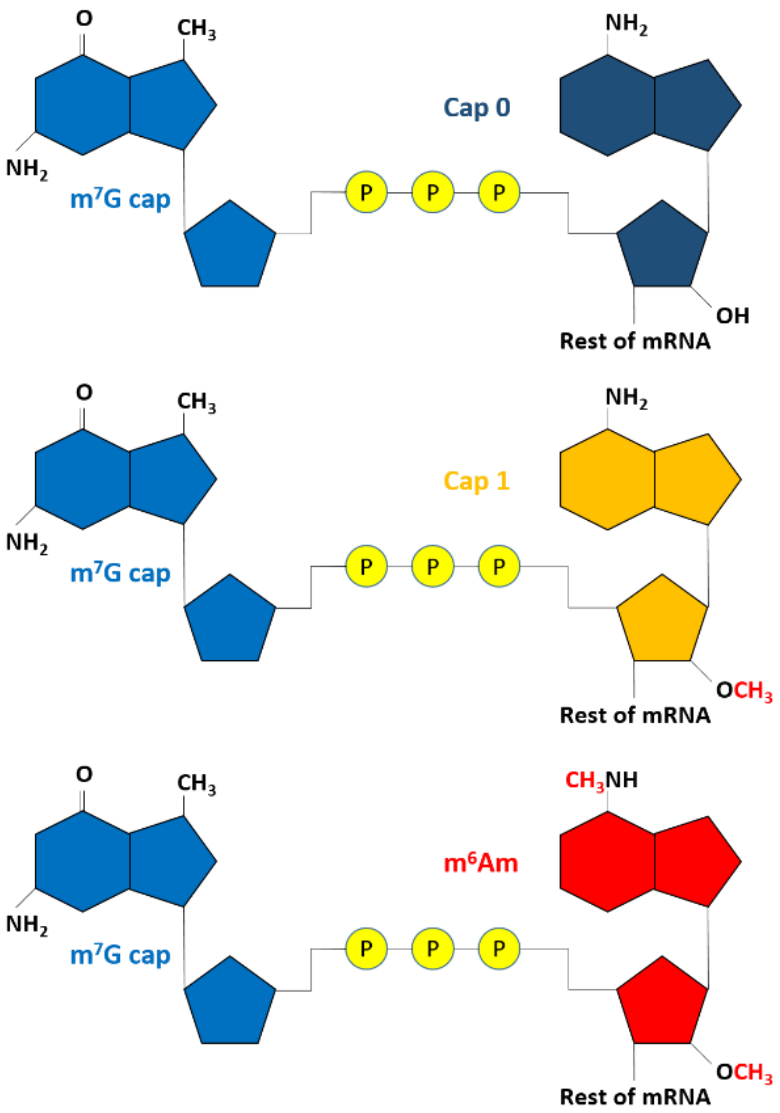


Figure 1.4. Structure of cap0, cap1 and m^{6}Am first nucleotides. The guanine cap (light blue) is attached to the first nucleotide through a 5′-5′ linkage. Cap0 has no modifications on the first nucleotide, in this diagram the first nucleotide is adenosine (dark blue); cap1 has a methylation on the ribose sugar of the first nucleotide, in this diagram the first nucleotide is Am (yellow); and m^{6}Am has two methylations, one on the ribose sugar and the other on the nitrogenous base and can only occur on adenosine (red). Methyl groups on the first nucleotide are highlighted in red.

1.2.2 Co-transcriptional RNA processing

RNA Polymerase II (RNA Pol II) transcribes mRNAs and some non-coding RNAs.Nascent RNAs are processed by the addition of a 5′ guanine cap, 3′ poly-A tail and the removal of introns. This processing increases the stability of the RNA, allows for nuclear export and the correct coding sequence to be translated within the cytoplasm (Ramanathan et al., 2016). RNA modifications may be added to the nascent transcript co-transcriptionally at the same time as these other processing events.. The proteins that carry out the processing bind to the C-terminal domain (CTD) of RNA pol II which is made up of a series of heptad repeats (Y1-S2-P3-T4-S5-P6-S7) (Corden, 1990). The recruitment of these proteins is regulated by the phosphorylation of serine 2 and serine 5 of this heptad repeat (Figure 1.5) (Hsin and Manley, 2012).

Serine 5 of the CTD of RNA pol II is phosphorylated shortly after transcription has started, this is when the guanine cap is added. Usually the guanine cap is added to the nascent RNA when it is between 25-30 nucleotide long (Jove and Manley, 1984, Rasmussen and Lis, 1993). There are three enzymes involved in adding the m^7guanine cap RNA triphosphatase, RNA guanylyltransferase, and RNA guanine-N^7 methyltransferase (Shuman, 2000).

After the capping process the first epitranscriptomic modifications are added. CMTR1 methylates the ribose sugar of the first nucleotide producing the cap1 structure (Belanger et al., 2010). The second nucleotide can also be methylated by CMTR2 forming the cap2 structure (Werner et al., 2011). This process happens after the guanine cap is added and occurs co-transcriptionally (Haline-Vaz et al., 2008). Potentially before the serine 5 of the pol II CTD is dephosphorylated a second RNA modification can be added. If the cap 1 is added to an adenosine the nucleotide can also be methylated at the N^6 position producing m^6Am (Akichika et al., 2018). PCIF1 stops the dephosphorylation of serine 5, so m^6Am methylation may occur when PCIF1 is bound to the CTD of pol II (Yeo et al., 2003). The methylation is known to occur after

the guanine cap is methylated as the m^7G cap is required for protein recognition (Sendinc et al., 2019).The structures of the modified first nucleotides attached to the guanine cap can be seen in figure 1.4.

Splicing begins to occur when both serine 5 and serine 2 of RNA pol II CTD are phosphorylated. Between 92%–94% of human transcripts are subject to alternative splicing (Wang et al., 2008). Splicing is the processes where the spliceosome and snRNAs removes introns from pre-mRNA and joins the exons together. The last processing step which occurs co-transcriptionally is the addition of the poly-A tail which occurs when only serine 2 of RNA pol II CTD is phosphorylated. 10-30 nucleotides downstream of the AAUAAA conserved sequence endonucleolytic cleavage occurs, followed by 3′ poly adenylylation (Colgan and Manley, 1997). The cleavage is carried out by CPSF, CstF and two cleavage factors CFIm and CFIIm, the polyadenylation is carried out by PAP (Neve et al., 2017).

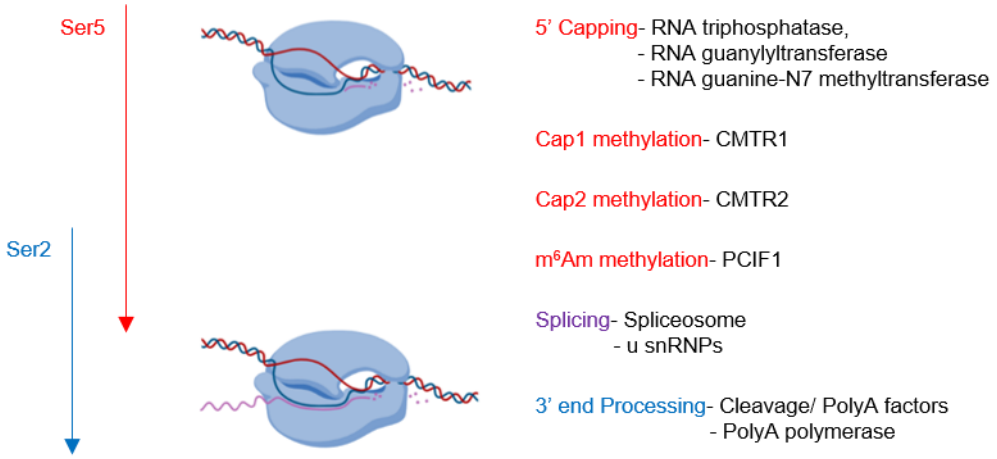


Figure 1.5. Co-transcriptional processing of mRNA. The changing phosphorylation of RNA Pol II CTD influences the proteins that bind controlling the stages of co transcriptional processing. When serine 5 is phosphorylated 5′ capping occurs along with cap1 methylation, cap2 methylation and potentially methylation of N^6 adenosine on the pre-mRNA. When serine 2 and 5 are methylated splicing of the pre-mRNA occurs. When serine 2 is methylated 3′ end processing and poly-adenylylation occurs. This illustration was partially created using Biorender.

1.3 History of RNA modifications

In the 1970's two pioneering groups led by Desrosier and Perry added [H^3]-methionine to Novikoff heptoma cells and mouse L cells respectively which was incorporated into S-adenosylmethionine (SAM) the enzymatic cofactor involved in DNA and RNA methylation (Desrosie et al., 1974, Perry and Kelley, 1974). Through enzyme treatments and chromatography, radioactive groups were found on poly(A) RNA providing evidence that nucleotides were methylated. The first nucleotides to be found methylated were the guanine cap and the first two nucleotides of the transcript which were N^6,2′-O-methyladenosine and 2′-O-methyl nucleotides (Perry and Kelley, 1974, Desrosie et al., 1974). Ensuing research identified that the guanine cap was methylated at the N7 position (Dubin and Taylor, 1975) and that Am, Cm, Gm and Um could all be found at the cap1 and cap2 positions but m^6Am was only found at the cap1 (Wei et al., 1976, Wei et al., 1975). Furthermore it was found that at the cap1 position, m^6Am was more common (28%) than Am (14%) (Wei et al., 1976, Heckle et al., 1977) and the methylation of the cap structure made up 50% of the RNA methylation with the remaining being m^6A and m^5C (40% and 10% respectively) in both Novikoff heptoma cells and cultured hamster cells (Dubin and Taylor, 1975, Desrosie et al., 1974). Investigations into m^6A indicated that in HeLa cells this modification was preceeded by an adenosine or guanine and always followed by a cytosine (Wei et al., 1976) and that m^6A could be found 4-5 times a transcript or 1.8 nucleotides per 1000 base pairs (Dubin and Taylor, 1975).

Interest in methylated mRNA was renewed in 2012 with the development of two next generation sequencing techniques termed m^6A-seq (Dominissini et al., 2012) and MeRIP-Seq (Meyer et al., 2012). These methods use an m^6A antibody approach to immunoprecipitate fragmented RNA that has the modified nucleotide which enables m^6A modified regions to be detected at a resolution of 100 base pairs. These methods

paired with thin layer chromatography allowed for whole transcriptome changes in RNA methylation to be studied at the transcript level.

Interest in m^6Am cap modification is currently increasing. In 2016 it was shown that fat mass and obesity-associated protein (FTO) the m^6A demethylase could demethylate m^6Am (Mauer et al., 2016). However it was not until late 2018 when the protein PCIF1 was found to be a m^6Am specific methylase (Akichika et al., 2018). Moreover, since the 1970s little work has focused on the cap1 modification and currently there is no sequencing method available to detect which transcripts are cap1 modified.

1.4 Readers, writers and erasers

Proteins which bind to specific RNA modifications can be classified as three types known as writers, readers and erasers. Writers are the methylases that transfer the modification on to RNA. Readers are the proteins that interact with the modifications which putatively provide a signal initiating a cellular activity. Erasers are demethylases that remove the modification from the RNA and provide a mechanism where the modification is reversible. m^6A, m^6Am and Am all have distinct proteins which carry out these writer, reader and eraser processes.

1.4.1 m^6A readers, writers and erasers

The m^6A methylase complex is made up of three core proteins METTL3, METTL4 and WTAP. METTL3 is the catalytic unit, METTL14 acts as the RNA binding platform and WTAP modulates levels of methylation on transcripts (Bokar et al., 1997, Liu et al., 2013a, Ping et al., 2014). The methylase complex contains a number of other supporting proteins including KIAA1429 (Schwartz et al., 2014), RBM15 (Patil et al., 2016) and ZC3H13 (Wen et al., 2018). The m^6A modification is read by nuclear readers YTHDC1 and HNRNPA2B1, and cytoplasmic readers, YTHDF1-3 and YTHDC2 (Xu et al., 2014, Wu et al., 2018, Hsu et al., 2017, Shi et al., 2018, Du et al., 2016, Shi et al.,

2017). These proteins detect the modification and are thought to be involved in many steps in the mRNA lifecycle from splicing to translation efficiency to degradation. There are two m^6A erasers FTO and ALKBH5 (Jia et al., 2008, Zheng et al., 2013). Figure 1.6 shows a diagram of the reader, writer and eraser proteins.

1.4.2 Cap1 readers, writers and erasers

The cap1 and cap2 methylases are known to be Cap Methyltransferase 1 (CMTR1) and Cap Methyltransferase 2 (CMTR2) respectively (Belanger et al., 2010, Werner et al., 2011). Methylation occurs co-transcriptionally when CMTR1 binds to the 7 methyl guanine cap (Haline-Vaz et al., 2008). However, it has been shown that *in vitro* CMTR1 does not require the 7′N methylation of the guanine cap to methylate the ribose of the first nucleotide (Belanger et al., 2010). There are a few known proteins that interact with the cap1 modification like the IFIT proteins in the immune response (Abbas et al., 2017). It has been reported that 137 proteins interact with CMTR1 and many of these are associated with pre-mRNA processing. A proportion of these protein interactions are RNA dependant, suggesting that PARP1, ILF2, XRCC6, eIF2α, and NCL may be potential cap1 readers (Simabuco et al., 2019). Currently there is not a reported eraser for the cap1 or cap2 modifications. Figure 1.7 shows the proteins that are known to be associated with the Am modification.

1.4.3 m^6Am readers, writers and erasers

Only two proteins have been suggested to interact with the m^6Am modification, these are FTO (Mauer et al., 2016) a suspected demethylase and PCIF1 a suspected methylase (Sendinc et al., 2019, Akichika et al., 2018, Sun et al., 2018, Boulias et al., 2019). PCIF1 is a nuclear protein that has been shown to bind to RNA polymerase II CTD domain (Sendinc et al., 2019, Fan et al., 2003, Hirose et al., 2008). FTO was first shown to be a demethylase of the internal m^6A (Jia et al., 2011, Zheng et al., 2013). Later it was reported that the preferred substrate of FTO was m^6Am rather than m^6A

(Mauer and Jaffrey, 2018). Recently it has been suggested that the cap binding complex may be bound to the cap and consequently m^6Am and this could potentially be stopping FTO from dephosphorylating m^6Am in the nucleus. As a result in the nucleus there would be more m^6A that FTO could access and demethylate even if it was not the optimal and preferred substrate (Shi et al., 2019). As of yet there are no reader protein published to interact with m^6Am. Figure 1.7 shows the proteins that are known to be associated with the m^6Am modification.

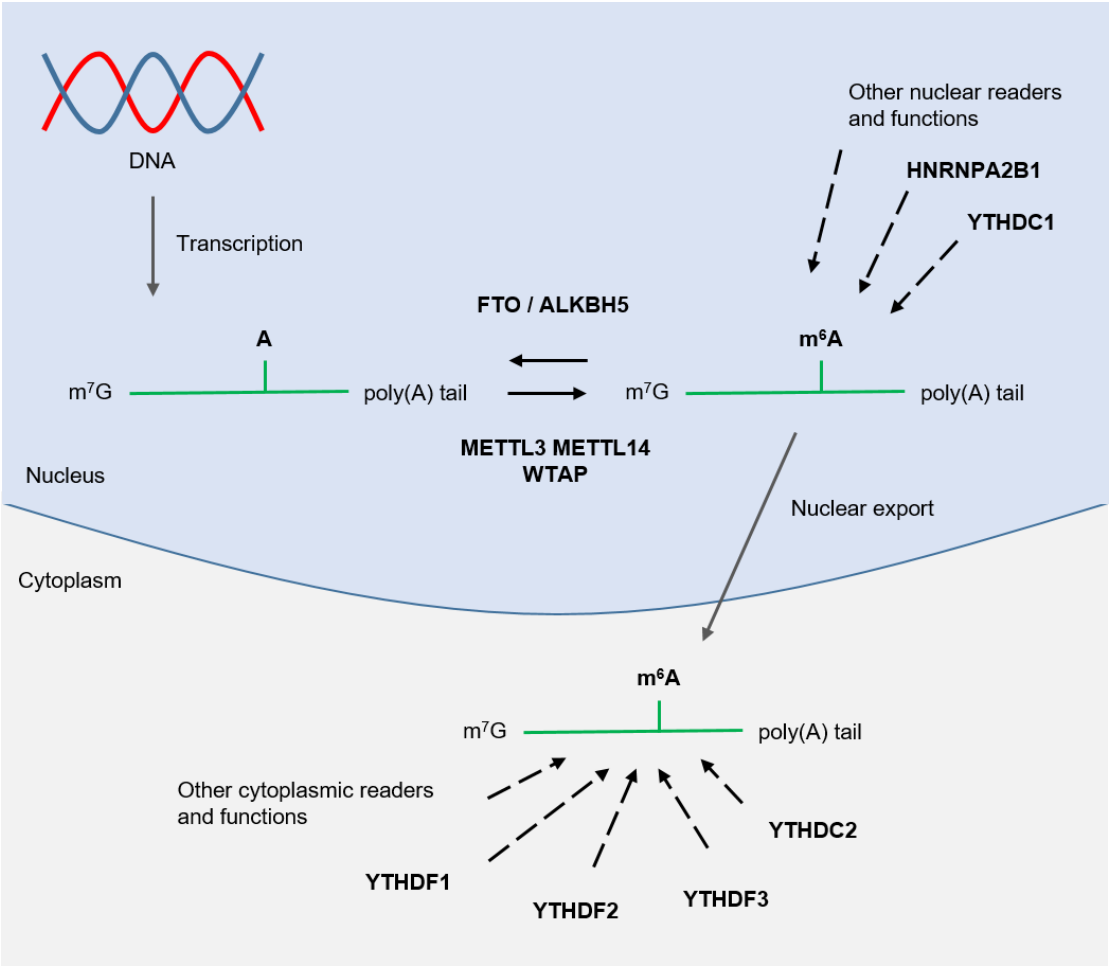


Figure 1.6. Readers, writers and erasers of m^{6}A. m^{6}A is the methylation of the N^6 position of internal adenosine nucleotides. The methylation is added to the RNA by a core complex of METTL3, METTL14 and WTAP. The modification can be removed by ALKBH5 and FTO. There are nuclear and cytoplasmic reader proteins that bind to the methylation and change the fate of the RNA. A selection of nuclear readers shown in the figure are HNRNPA2B1 and YTHDC1 and cytoplasmic readers YTHDF1, YTHDF2, YTHDF3, YTHDC2.

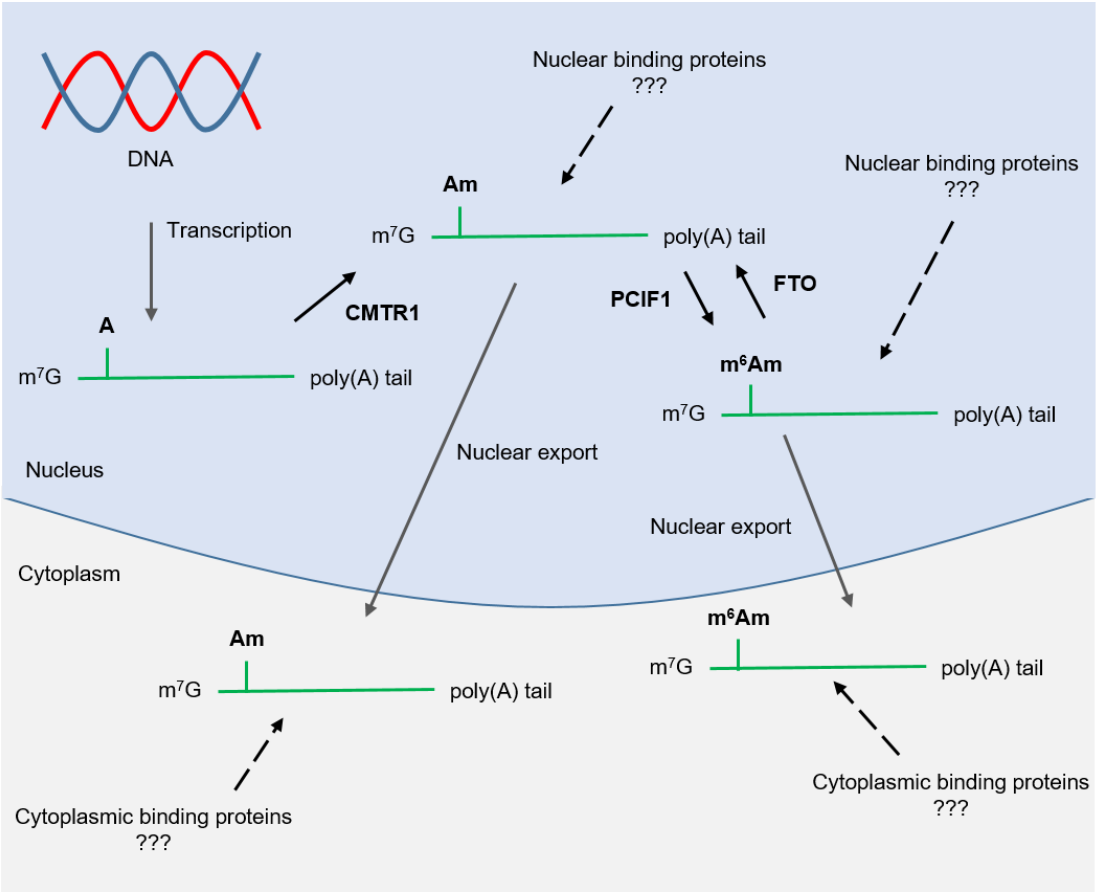


Figure 1.7. Readers, writers and erasers of the Am and m^{6}Am modifications. The cap1 2′-O-methylation of the ribose sugar is added to the first nucleotide of the transcript by CMTR1. If the first nucleotide is adenosine, it can be methylated again at the N^6 position of the nitrogenous base by PCIF1 to form m^{6}Am. FTO has been shown to be the demethylase of m^{6}Am. Currently there are no known cytoplasmic or nuclear readers for either modification.

1.5 Function of RNA modifications

1.5.1 m^6A

m^6A modification has been found to have an effect at the molecular, physiological and organism level. At the molecular level, m^6A modification has an effect on splicing, nuclear export, stability and translational efficiency (Wang et al., 2015, Wang et al., 2014, Edens et al., 2019, Zhao et al., 2014). The m^6A reader protein YTHDC1 has been shown to be involved in the regulation of mRNA splicing. YTHDC1 is reported to interact with SRSF3 and SRSF10 to promotes exon inclusion or facilitate exon skipping respectively (Xiao et al., 2016). Reader proteins can increase and decrease the stability of mRNA. When YTHDF2 binds to the m^6A modification on mRNA transcripts, the complex is then taken to P bodies for degradation (Wang et al., 2014). However when YTHDF1 binds to the same mark on nascent mRNA it increases the efficiency of ribosome binding and translation (Wang et al., 2015).

At the pathway or cellular level, m^6A modifications have roles in regulating meiosis in *S. Cerevisiae* (Schwartz et al., 2013), differentiation of stem cells (Heck and Wilusz, 2019), and local proteins translation at synapses (Braulio Martinez unpublished). RNA methylation can destabilise pluripotency transcripts resulting in differentiation. An example of this is the methylation of NOTCH1 mRNA results in its degradation via YTHDF2 resulting in endothelial-hematopoietic transition (Zhang et al., 2017, Lv et al., 2018). At synapses, our lab has shown that YTHDF1-2 and the eraser protein ALKBH5 are involved in the processes of synaptic plasticity at local active ribosomes (Braulio Martinez unpublished).

At the physiological level, m^6A mutants do not develop normally and indicate developmental defects as shown in *Arabidopsis* (Bodi et al., 2012), *Drosophila* (Hongay and Orr-Weaver, 2011) and Zebrafish (Ping et al., 2014). In *Arabidopsis* homozygous knockout of N^6-adenosine-methyltransferase MT-A70-like (MTA), a

METTL3 writer orthologue, is embryonic lethal (Zhong et al., 2008). Furthermore, even heterozygous knockout plants die at the globular stage of embryo development (Zhong et al., 2008). In Zebrafish m^6A modifications have been shown to be required for tissue differentiation. In Zebrafish WTAP and METTL3 are maternally expressed from the 4 cell stage into embryogenesis and enriched expression is noted in the brain at 36 hours post fertilisation (hpf). When *Wtap* is knocked out using morpholinos, there are developmental defects seen from 24hpf, with the embryos have smaller heads, eyes and brain ventricles and curved notochords (Ping et al., 2014). At the molecular level YTHDF2 is involved in degradation of mRNA with m^6A modification. At a physiological level if *Ythdf2* is knocked out in mice there is a significant decrease in proliferation and differentiation capabilities of neural stem/progenitor cell (NSPC). This results in the neurones being unable to form neurites as there is a delayed degradation of key neural developmental mRNAs (Li et al., 2018).

1.5.2 Am

The cap structure and potentially the cap1 modification has a role in splicing and translation through the proteins that bind including the cap binding complex and CMTR1 (Inesta-Vaquera and Cowling, 2017, Simabuco et al., 2019). Cap1 and cap2 methylation are used to mark RNA as “self” so the RNA does not trigger the innate immune response through the IFIT1 and RIG1 proteins (Züst et al., 2011). Both RIG1 and IFIT1 binding is blocked by the cap1 methylation, and if there are very high levels of IFIT1 cap2 methylation provides more resistance to these proteins (Abbas et al., 2017, Schuberth-Wagner et al., 2015).

1.5.3 m^6Am

It is not fully understood what might be the function of the m^6Am modification might be. There are conflicting reports on its effect on mRNA. For example, initial work suggested that m^6Am increased the stability of mRNA and might confer resistance to

DCP2 a decapping enzyme (Mauer et al., 2016). The Jaffrey group also showed that transcripts which contained a m^6Am modification have increased half-lives by ~2.5 hours (Mauer et al., 2016). Later re-analysis of this data showed that the increased stability was down to internal m^6A and not m^6Am (Wei et al., 2018). Two more labs independently provided evidence that m^6Am does not have a role in transcript stability (Sendinc et al., 2019, Akichika et al., 2018). Akichika et al., 2019 found that capped transcripts had increased translation in wild type HEK293T cells compared to *PCIF1* knock out HEK293T cells suggesting a role for m^6Am in translation regulation. This is contested as Boulias et al., found there was no effect on translation using knockout of *PCIF1* in HEK293T and HeLa cells, and Sendinc et al., 2019 found that m^6Am supresses translation in MEL624 cells.

Mauer et al., 2016 reported that m^6Am conferred resistance to the decapping enzyme DCP2. DCP2 decapping is the first step in 5′-3′ degradation which takes place in the cytoplasm, DCP2 works in a complex with DCP1, EDC4, EDC1 and EDC3. EDC4 acts as a scaffold and DCP1, EDC1 and EDC3 activate DCP2 changing its conformation to the active form (Valkov et al., 2016, Mugridge et al., 2018, Chang et al., 2014). After decapping, XRN1 an exoribonuclease digests the RNA into single nucleotides (Nagarajan et al., 2013). This prevention of RNA degradation through this pathway enables spatial temporal control over translation.

1.6 Epitranscriptome in the brain

m^6A, cap1 and m^6Am are all enriched in the brain (Dominissini et al., 2012, Kruse et al., 2011, Ven Murthy, 1982). Recent research has been performed showing m^6A, has a role in many physiological processes including influencing synaptic function (Merkurjev et al., 2018).

The m^6A/m demethylase FTO has been implicated in dopaminergic neuronal function in the mid brain. FTO conditional knockout impairs GRIK-Dopamine receptor 2 (D2R)-

Dopamine receptor 3 (D3R) mediated signalling, a dopamine signalling pathway found in the midbrain that controls neuronal activity and behavioural responses. *FTO* knockout mice are reported to have 5000 new m^6A modification peaks as shown by MeRIP-seq analysis when compared to wild type mice. GO analysis indicated that these transcripts were disproportionately enriched for transcripts relating to synaptic transmission and neuronal cell signalling. This enrichment in neuronal transcripts suggests that FTO physiologically regulates m^6A in mRNA and that FTO does not globally target all m^6A modified transcripts but a specific subsection (Hess et al., 2013).

The m^6A modification has recently been linked directly to memory formation in rodent studies. It has been shown that there is an upregulation in m^6A in the medial prefrontal cortex in fear conditioned mice. As expected the transcripts that were methylated were involved in dendritic and post synaptic regulation, synaptic transmission, and transmembrane transport regulation. They also found that mice had increased fear memory 24hrs after training as a result of decreased levels of FTO in the medial prefrontal cortex (Widagdo et al., 2016). In the hippocampus, during contextual-dependant fear conditioning tasks, FTO transiently decreases in CA1 neurones with the largest decrease found at the synapse (Walters et al., 2017).

Very little research has been carried out into the role of the cap modification and brain function. It has been shown the cap0 is not found on polysome bound mRNA in rodent brain (Ven Murthy, 1982). In *Drosophila*, the cap2 methyltransferase adrift has a role in tracheal nerve path finding in the CNS linking the cap modification to essential nervous system processes (Pereanu et al., 2007, Englund et al., 1999).

1.7 Thesis aims

Relatively little is still known about the role of m^6Am in mammalian processes and physiology and no work had been carried out on the role of m^6Am in brain function.

There are four aims to this thesis. One, to investigate where in evolution m^6Am could be found to aid in the identification of candidate methylase proteins. Two, to assess if there was a perturbation in global m^6Am levels in human neurodegenerative disease. Three, to characterise which transcripts are m^6Am methylated in human tissue. Four, to identify proteins which bind to the m^6Am modification.

The specific objectives are as follows:

Chapter 3- Evolutionary conservation of m^6Am

The aim of the work described in this chapter was to investigate the evolutionary conservation of m^6Am to help in the identification of candidate methylase proteins. This was carried out by assaying for the presence, and determining global levels of m^6Am on capped RNA species from different species. After initiating this study, genes identified as encoding the m^6Am writer and eraser were reported by other groups, and a further aim became to test the correlation of these genes with the presence of m^6Am containing RNA.

Chapter 4- m^6Am in the human brain

The aim of work described in this chapter was to test if there was evidence for misregulation or altered m^6Am methylation levels in neurodegenerative disease. This was performed by comparing global m^6Am methylation levels on capped RNA transcripts between unaffected control hippocampal samples and Parkinson's disease and Lewy body dementia samples. The second aim of this chapter was to analyse m^6A sequencing data from human foetal brain, white matter and grey matter to identify m^6Am peaks and the transcripts they were located on.

Chapter 5- Identifying m^6Am modifications by DCP2 enzymatic treatment

The aim of the work described in chapter 5 was to test if m^6Am was resistant to the DCP2 decapping enzyme, and potentially create a novel sequencing method for

identifying transcripts with m^6Am. DCP2 could then be used to decap transcripts without m^6Am, as these could be digested away with an exonuclease and the remaining transcripts could be made into a sequencing library.

Chapter 6- Identification of m^6Am binding proteins

The aim of work described in this chapter was to identify the m^6Am methylase and m^6Am binding proteins through protein pull downs using modified RNA oligonucleotides that started with Am and m^6Am after the guanine cap.

Thesis Hypothesis

This will test the hypothesis that m^6Am plays a key role in brain function through the transcripts that are methylated and the proteins that interact with the modification. The m^6Am modification may have a function in synaptic plasticity by selecting for transcripts that are important for localised translation.

Chapter 2: Methods and Materials

2.1 Materials

2.1.1 Human tissue

Human brain tissue from individuals with unaffected with brain disease, or who had Parkinson's disease or Lewy body dementia were acquired from the Parkinson's UK Brain Bank, Imperial College. The use of this tissue is covered by the ethics approval that the Parkinson's UK Brain Bank at Imperial College London has as a Research Tissue Bank by the Wales Research Ethics Committee (Ref. No. 08/MRE09/31+5). The Material Transfer Agreement (MTA) was between A. C. Shone (Imperial Brain Bank) and Dr Helen Miranda Knight (University of Nottingham). HTA licence number is 12265. RNA was extracted (2.2.1) from fresh frozen hippocampus tissue from the unaffected control, Parkinson's and Lewy body dementia samples. Nick Rielly helped in the extraction of the RNA. Details of the human samples are provided in Table 2.1.

	Age at death (years)	Sex	Clinical diagnosis	Duration of pathology (years)	Reported cause of death
Control					
1	79	Male	N/A	N/A	Brainstem stroke, bronchopneumonia
2	94	Female	N/A	N/A	Respiratory failure, bronchopneumonia, fractured sternum die to fall
3	87	Male	N/A	N/A	Congestive cardiac failure, atrial fibrillation, mitral regurgitation, chronic kidney disease stage 5

Parkinson's disease					
1	63	Male	Parkinson's with Lewy Bodies	9	bronchopneumonia
2	83	Female	Tremulous Parkinsonism	3	Not reported
3	77	Male	Akinetic-rigid syndrome	3	Aspiration pneumonia and Lewy body disease
Lewy Body Dementia					
1	86	Male	Brainstem Lewy body disease	19	End stage Parkinson's disease
2	75	Female	Brainstem Lewy body disease	16	Parkinson's disease
3	84	Male	Brainstem Lewy body disease	Not available	Multiple system atrophy

Table 2.1. Patient details from which the hippocampal tissue was donated. Tissue from unaffected control, Parkinson's disease and Lewy body dementia patients was used in this study. Information about the patients was taken from the pathology reports and include age at time of death, gender, clinical diagnosis, duration of pathology and reported cause of death.

2.1.2 Animal tissue

All work involving vertebrate animals conducted by the University of Nottingham is reviewed by the University's Animal Welfare and Ethical Review Body (AWERB). *Rattus norvegicus* brains were obtained from four adult rats (gender unknown) donated by Dr Michael Garle (University Of Nottingham, UK). *Ambystoma mexicanum* brains were obtained from five adult male axolotls donated by Professor Andrew Johnson (University Of Nottingham). *Danio rerio* brains were obtained from three zebrafish (age and gender unknown) donated by Dr Martin Gering (University Of Nottingham). *Drosophila melanogaster* heads were obtained from three feeding pots of *D. melanogaster* donated by Dr Brenda Ivonne Canales Coutiño (University Of Nottingham). The *D. melanogaster* used in the SCARLET method were donated by Dr Matthias Stoller (University of Birmingham). *Octopus vulgaris* brains were obtained from three octopi purchased from Alix Harvey (Marine Biological Association). *Apis mellifera* heads were donated by Dr Reinhard Stoger (University of Nottingham). *Gallus Gallus* hypothalamus RNA was donated by Dr Nathan Archer (University of Nottingham).

2.1.3 Algae

Three biological replicates of *Ectocarpus siliculosus* (Dillwyn) Lyngbye (1819) (CCAP 1310/4) were grown by Dr Marie-Mathilde Perrineau (Scottish Association for Marine Science).

2.1.4 Bacterial strains

Bacterial Strain	Company (Headquarters)	Use
DH5α	New England Biolabs Biolabs (Hitchin, England)	Standard Cloning
Rosetta-gami	Merck (Feltham, England)	Optimisation of Protein Production
SoluBL21	Genlantis (San Diego, California, USA)	Optimisation of protein production

Table 2.2. Bacterial strains used in this study. The table includes the strain, company they were purchased from, and use. The Rosetta-gami genotype is F– ompT hsdSB (rB– mB–) gal dcm lacY1 ahpC (DE3) gor522::Tn10 trxB pRARE (CamR, KanR, TetR).

2.1.5 Plasmids

pMAL-c5x was purchased from New England Biolabs (Hitchin, England) and was used in the Gibson assembly process. The plasmid is ampicillin resistant and contains maltose binding protein so amylose purification can be performed on expressed protein. pet28a-hDCP2 was purchased from Addgene (Teddington, England), plasmid number #72214. The plasmid is kanamycin resistant and contains the human DCP2 coding sequence. pMAL-c5x hDCP2 was constructed by myself from the DCP2 coding sequence from pet28a-hDCP2 and the pMAL-c5x plasmid. The plasmid is ampicillin resistant. The expressed DCP2 protein can be purified using both maltose binding protein and His tag.

2.1.6 Reagents

Table 2.3 provides details of reagents of analytical grade or higher which were used in this study. LB and solid agar plates for bacterial growth were supplied by Central Services (Plant Sciences).

Reagent	Company (Headquarters)
10mM ATP	New England Biolabs (Hitchin, England)
10x TBE	Roche (Welwyn Garden City, England)
2-(N-morpholine)-ethanesulfonic acid	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
AE phenol	Sigma Aldrich (Merck, Feltham, England)
Agarose	Sigma Aldrich (Merck, Feltham, England)
Agilent chips	Agilent Technologies (Stockport, England)
Ampicillin	Sigma Aldrich (Merck, Feltham, England)
Biotin-16-UTP	Sigma Aldrich (Merck, Feltham, England)
Bromophenol blue	BDH Laboratory Supplies (Poole, England)
$CaCl_2$	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
Cellulose TLC plates	Merck (Feltham, England)
Cellulose-F TLC plates	Merck (Feltham, England)
Chloramphenicol	Sigma Aldrich (Merck, Feltham, England)
Chloroform	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
Chloroform isoamyl alcohol	Sigma Aldrich (Merck, Feltham, England)
Complete protease inhibitors	Roche (Welwyn Garden City, England)

DMSO	Sigma Aldrich (Merck, Feltham, England)
dNTPs	VWR (Lutterworth, England)
Dithiothreitol (DTT)	Sigma Aldrich (Merck, Feltham, England)
Econo-Pac® Chromatography Columns	Biorad Laboratories (Watford, England)
Ethylenediaminetetraacetic acid (EDTA)	Sigma Aldrich (Merck, Feltham, England)
Ethanol	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
Ethidium bromide	Sigma Aldrich (Merck, Feltham, England)
Gel Loading Dye, Purple (6X)	New England Biolabs (Hitchin, England)
Glacial acetic acid	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
Glucose	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
Glycerol	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
Glycine	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
Glycogen, RNA grade	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
10mM GTP	New England Biolabs (Hitchin, England)
HCL	Merck (Feltham, England)
Hepes	Sigma Aldrich (Merck, Feltham, England)
HisPur™ Ni-NTA Resin	Thermo Scientific (Thermo Fisher Scientific, Altrincham, England)
Hydrophilic Streptavidin Magnetic Beads	New England Biolabs (Hitchin, England)
HyperLadder™ 1kb	Bioline (Memphis, Tennessee, USA)
Indole-3-acetic acid (IAA)	Sigma Aldrich (Merck, Feltham, England)
Imidazole	Fluka (Honeywell, Bucharest, Romania)

Imperial Protein Stain	Thermo Scientific (Thermo Fisher Scientific, Altrincham, England)
Isopropyl-β-D-thiogalactoside (IPTG)	Sigma Aldrich (Merck, Feltham, England)
Isobutyric acid	Sigma Aldrich (Merck, Feltham, England)
Isopropanol	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
Kanamycin	Sigma Aldrich (Merck, Feltham, England)
KCl	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
KH_2PO_4	Acros Organics (Thermo Fisher Scientific, Altrincham, England)
KOAc	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
LiCl	Fluka (Honeywell, Bucharest, Romania)
$MgCl_2$	Sigma Aldrich (Merck, Feltham, England)
$MgSO_4$	Sigma Aldrich (Merck, Feltham, England)
$MnCl_2$	Sigma Aldrich (Merck, Feltham, England)
100mM N^6-Methyl-ATP	Jena Biosciences (Ely, England)
Na_2HPO_4	Sigma Aldrich (Merck, Feltham, England)
NaCl	Sigma Aldrich (Merck, Feltham, England)
NaOAc	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
NH_4OH	Sigma Aldrich (Merck, Feltham, England)
NHCl	Sigma Aldrich (Merck, Feltham, England)
Np-40	Sigma Aldrich (Merck, Feltham, England)
Oligo d(T)25 Magnetic Beads	New England Biolabs (Hitchin, England)
Page blue stain	Fermentas (Thermo Fisher Scientific, Altrincham, England)

Phosphate buffered saline, tablet	Sigma Aldrich (Merck, Feltham, England)
Phenol Chloroform Isoamyl-alcohol	Sigma Aldrich (Merck, Feltham, England)
Mini-PROTEAN TGX Gels	Biorad Laboratories (Watford, England)
Color-coded Prestained Protein Marker, Broad Range (11-250 kDa)	Cell signaling Technologies (London, England)
Ribolock	Thermo Scientific (Thermo Fisher Scientific, Altrincham, England)
32mM S-adenosylmethionine (SAM)	New England Biolabs (Hitchin, England)
Sarkosyl	Sigma Aldrich (Merck, Feltham, England)
Sodium Dodecylsulfate (SDS)	Applichem GmbH (Darmstadt, Germany)
Soc media	New England Biolabs (Hitchin, England)
Tetracycline	Sigma Aldrich (Merck, Feltham, England)
Thermopol buffer	New England Biolabs (Hitchin, England)
Trifluoractic acid	Sigma Aldrich (Merck, Feltham, England)
Trizma base	Sigma Aldrich (Merck, Feltham, England)
Tryptone	OXOID (Thermo Fisher Scientific, Altrincham, England)
Tween-20	Sigma Aldrich (Merck, Feltham, England)
Urea	Fisher Scientific (Thermo Fisher Scientific, Altrincham, England)
Yeast extract	OXOID (Thermo Fisher Scientific, Altrincham, England)
Yeast tRNA	Invitrogen (Thermo Fisher Scientific, Altrincham, England)
β-mercaptoethanol	Sigma Aldrich (Merck, Feltham, England)

Table 2.3. Table of reagents. The table contain the reagents used with the company and address of the headquarters.

2.1.7 Enzymes

Table 2.4 provides details of the enzymes used.

Enzyme	Company (Headquarters)
2′-O Methyltransferase	New England Biolabs (Hitchin, England)
Calf intestinal Phosphatase (CIP)	New England Biolabs (Hitchin, England)
FastAP Thermosensitive Alkaline Phosphatase	Thermo Scientific (Thermo Fisher Scientific, Altrincham, England)
P1 nuclease	Sigma Aldrich (Merck, Feltham, England)
Phusion® High-Fidelity DNA Polymerase	New England Biolabs (Hitchin, England)
Poly Nucleotide Kinase (PNK)	Thermo Scientific (Thermo Fisher Scientific, Altrincham, England)
RppH (RNA 5′ pyrophosphohydrolase)	New England Biolabs (Hitchin, England)
T1 RNase	Thermo Scientific (Thermo Fisher Scientific, Altrincham, England)
T4 DNA ligase	New England Biolabs (Hitchin, England)
T4 RNA ligase	New England Biolabs (Hitchin, England)
Taq DNA Polymerases	New England Biolabs (Hitchin, England)
Terminator Exonuclease	Epicenter (Cambridge Biosciences, Cambridge, England)
Trypsin	Promega (Southampton, England)
TURBO DNase	Invitrogen (Thermo Fisher Scientific, Altrincham, England)
Vaccinia Virus Capping Enzyme	New England Biolabs (Hitchin, England)
Xmn1	New England Biolabs (Hitchin, England)

Table 2.4. Table of enzymes. The table contains the enzymes used and the company they were purchased from and the address of the headquarters.

2.1.8 Radioactive nucleotides

ATP, [γ-32P]- 3000Ci/mmol 10mCi/ml EASYTIDE, 250 µCi, Perkin Elmer (Waltham, Massachusetts, United States), was used to label exposed nucleotides for TLC analysis

2.1.9 Buffers, media and stock solutions

2.1.9.1 Buffers

Buffers used are detailed below.

AE buffer- 50mM NaOAc pH, 5.3; 10mM EDTA

Algae extraction buffer- 100mM Tris-HCL, pH 9.5; 150mM NaCl; 1% Sarkosyl; 5mM DTT

5x DCP2 decapping buffer- 50mM Tris-HCL, pH7.5; 500mM KCL; 10mM $MgCl_2$; 10mM DTT, 0.5mM $MnCl_2$

2x Lamelli buffer- 62.5mM Tris-HCL, pH 6.8; 25% glycerol; 0.01% (w/v) bromophenol blue; 2% SDS; 5% β-mercaptoethanol

10x SDS running buffer- 250mM Tris base; 1.92mM glycine, 1% SDS

P1 buffer- 30mM NaOAc

Nickle column protein purification buffers

Ni column Elution buffer- 20mM Na_3PO_4; 300mM NaCl; 250mM Imidazole; adjust to pH 7.4, do not autoclave imidazole

Ni column equilibration buffer- 20mM Na_3PO_4; 300mM sodium chloride; 10mM Imidazole; adjust to pH 7.4, do not autoclave imidazole

Ni column MES buffer- 20mM 2-(N-morpholine)-ethanesuulfonic acid; 100mM NaCl, adjust to pH 5

Ni column wash buffer- 20mM Na_3PO_4; 300mM NaCl; 25mM Imidazole; adjust to pH 7.4, do not autoclave imidazole

Poly A buffers

RNA binding buffer- 1M LiCl; 40mM Tris-HCL, pH 7.5; 2mM EDTA; 0.1% NP40

RNA wash buffer- 150mM LiCl; 20mM Tris-HCL, pH 7.5; 1mM EDTA; 0.01% NP40

Oligo d(T) storage and equilibrium buffer- PBS; 0.1% Tween 20

Protein pull down 1 buffers

IP Lysis Buffer- 25mM Tris-HCL, pH 7.4; 150mM NaCl; 1% NP40; 1mM EDTA; 5% glycerol

RNA Capture Buffer- 20mM Tris, pH 7.5; 1M NaCl; 1mM EDTA

Protein RNA binding buffer (10x)- 200mM Tris, pH 7.4; 500mM NaCl; 20mM $MgCl_2$; 1%Tween 20

Wash Buffer- 20mM Tris, pH 7.4; 10mM NaCl; 0.1% Tween 20

Protein pull down 2 buffers

RNA binding Buffer- 50mM Hepes-HCL, pH 7.5; 150mM NaCl; 0.5% (v/v) Np-40; 10mM $MgCl_2$

RNA Wash buffer- 50mM Hepes-HCL, pH 7.5; 250mM NaCl; 0.5% (v/v) Np-40; 10mM $MgCl_2$

Protein incubation buffer- 10mM Tris-HCL, Ph 7.5; 150mM KCl; 1.5mM $MgCl_2$; 0.1% (v/v) Np-40; 0.5mM DTT, Complete Protease Inhibitor

Elution Buffer- 50mM Tris-HCL pH 8.5; 2M Urea; 10mM DTT

TLC Running buffers

First dimension buffer- 5:3, Isobutyric acid:0.5M NH4OH

Second dimension buffer- 70:15:15, Isopropanol:HCL:Water

2.1.9.2 Media

Media made is detailed below.

M9 minimal media- M9 salts (3g/L KH_2PO_4; 6g/L Na_2HPO_4; 0.5g NaCl; 1g/L NH_4Cl); 0.1mM $CaCl_2$; 1mM $MgS0_4$; 0.3% glycerol or 0.4% glucose

Rich growth media- per litre; 10g tryptone, 5g yeast extract, 5g NaCl, 2g glucose

2.1.9.3 Stock solutions

Stock solution used to make up buffers are detailed below.

50mM Ammonium Hydrogen carbonate (ABC) buffer- Dissolve 1.58g in 400ml sterile distilled water.

Ampicillin (100 mg/mL)- Dissolve 1g Ampicillin in 10ml water, filter sterilise and store in the freezer.

Chloramphenicol (25 mg/mL)- Dissolve 0.26g in 10ml Ethanol, filter sterilise and store in the freezer.

100mM $CaCl_2$- Dissolve 1.11g of $CaCl_2$ into 100ml of water. Autoclave to sterilise.

1M DTT- Dissolve 1.542g DTT in 10ml of 50mM ABC Buffer, filter sterilise and store in the freezer.

100mM EDTA pH8- Dissolve 29.224g EDTA in 400ml, adjust to pH8 with NaOH pellets. Make up to 500ml, autoclave to sterilise.

500mM EDTA pH8- Dissolve 93.05g EDTA in 400ml, adjust to pH8with NaOH pellets. Make up to 500ml, autoclave to sterilise.

50% glycerol- Mix 20ml of glycerol and 20ml of water, place on rollers till mixed. Filter sterilise.

0.5M Hepes-HCL pH 7.5- Dissolve 23.83g of Hepes in 60ml of distilled water, adjust pH to 7.5 with NaOH. Make up to 100ml with distilled water. Filter sterilise and store in the freezer.

0.5M IAA- Dissolve 0.925g IAA in 10 ml of 50mM ABC buffer, filter sterilise and store in the freezer.

0.1M IPTG- Dissolve 23.83g of IPTG in 100ml of water. Filter sterilise and store in the freezer.

Kanamycin (50mg/ml)- Dissolve 0.5g of Kanamycin in 10ml of water. Filter sterilise and store in freezer

1M KCl- Dissolve 7.45g of KCl in 100ml of distilled water. Autoclave to sterilise

3M KOAc pH4.8- Dissolve 9.81g of KOAc in 60ml of water, adject pH using Glacial Acetic Acid to pH 4.8. Autoclave to sterilise.

2M LiCl- Dissolve 8.46g of LiCl in 100ml of distilled water. Autoclave to sterilise.

200mM $MgCl_2$- Dissolve 1.9g of $MgCl_2$ in 100ml of distilled water. Autoclave to sterilise.

1M $MgS0_4$- Dissolve 12.03g of MgSO4 in 100ml of distilled water. Autoclave to sterilise.

20mM $MnCl_2$- Dissolve 2.51g of $MnCl_2$ in 100ml of distilled water. Autoclave to sterilise.

2M NaCl- Dissolve 11.68 of NaCl in 100ml of distilled water. Autoclave to sterilise.

5M NaCl- Dissolve 29.22g of NaCl in 100ml of distilled water. Autoclave to sterilise.

3M NaOAc pH5.3- Dissolve 24.61g of NaOAc in 60ml of water, adjust pH using glacial Acetic Acid to pH 4.8. Autoclave to sterilise.

100mM NaOAc pH5.3- Dissolve 0.82g of NaOAc in 60ml of water, adjust pH using glacial Acetic Acid to pH5.3. Autoclave to sterilise.

0.1N NaOH- Dissolve 1g NaOH in 250ml.Filter sterilise.

10% NP40- Mix 1ml of NP40 with 9ml of water. Mix until dissolved, filter sterilise.

1x PBS- Put 1 tablet in 500ml water. Autoclave to sterilise.

10% Sarkosyl solution- Dissolve 1g in 10ml of distilled water. Filter sterilise.

10% SDS- Mix 10ml of SDS into 90ml distilled water. Filter sterilise.

1x TBE- Mix 100ml of 10x TBE into 900ml distilled water

Tetracycline (12.5mg/ml)- Dissolve 0.125g tetracycline in 10ml distilled water. Filter sterilise.

1M Tris-HCL- Dissolve 60.5g of Tris base in 400ml of distilled water. Adjust the PH accordingly (ph7.5, 8.5, 9.5) with HCL. Autoclave to sterilise.

.

2.1.10 Kits

Kits purchased to perform specific assays are presented in table 2.5.

Method	Kit Name	Manufacturer (Headquaters)
RNA extraction	mirVana™ miRNA Isolation Kit, with phenol	Invitrogen (Thermo Fisher Scientific, Altrincham, England)
RNA and DNA extraction	AllPrep DNA/RNA Mini Kit	Qiagen (Manchester, England)
Nucleotide and enzyme removal	QIAquick nucleotide removal columns	Qiagen (Manchester, England)
Removal of fragmented RNA and clean-up of synthesised RNA oligonucleotides	RNeasy micro kit	Qiagen (Manchester, England)
Plasmid extraction	GeneJET Plasmid Miniprep Kit	Thermo Scientific (Thermo Fisher Scientific, Altrincham, England)
T7 invitro transcription	RiboMAX™ Large Scale RNA Production Systems– SP6 and T7	Promega (Southampton, England)
Detection of protein concentration	Peirce BCA protein assay kit	Thermo Scientific (Thermo Fisher Scientific, Altrincham, England)
Ligating gene of interest into plasmid backbone	Gibson Assembly kit	NEB (Hitchin, England)

Table 2.5. Kits used for specific assays. The table includes the kit used, the method it was used for and the manufacture which produced it.

2.1.11 Primers

Primers were ordered from Sigma Aldrich and were synthesis at a scale of 0.025 µmole, purification desalt. The oligonucleotides used in the SCARLET were made at a synthesis scale of 0.025 µmole, HPLC purification. Details of the primers are presented in table 2.6.

Purpose	Target	Name	Sequence	Notes
Oligonucleotide generation	-	T7sigma2.5	TAATACGACTCACTATTAC	T7promoter sequence to bind to T7sigma2.5polyA Oligonucleotide sequence and T7 promoter
Oligonucleotide generation	-	T7sigma2.5polyA	AAAAAAAAAAAAGGCCAGACCGAGACCCACACGAGTAATAGT GAGTCGTATTA	
Amplification of DCP2	DCP2	DCP2_F	ATGGAGACCAAACGGGT	
Amplification of DCP2	DCP2	DCP2_R	TCAAAGGTCCAAGATTTTCATTAT	
Adding His tag to DCP2	DCP2	DCP2_long_F	ATGGAGACCAAACGGGTGGAGATTCC	
Adding His tag to DCP2	DCP2	DCP2_His_R	TCAGTGATGGTGATGGTGATGAAGGTCCAAGATTTTC	
Adding Gibson Assembly overhangs	DCP2	DCP2_gibson_F	ACCTCGGGATCGAGGGAAGGATGGAGACCAAACGGGTG	
Adding Gibson Assembly overhangs	DCP2	DCP2_Gibson_R	CCCATGGACATATGTGAAATTCAGTGATGGTGATGGTG	
Sequencing plasmid insert	pMAL-c5x	pMAL_F	GGTCGTCAGACTGTCGATGAAGCC	
Sequencing Plasmid insert	pMAL-c5x	pMAL_R	TGTCCTACTCAGGAGAGCGTTCAC	
Oligonucleotide used in optimising Scarlet method	-	A oligonucleotide	AACUGAUCAUCAUCACCACAGAGCAGGU(biot)	Oligonucleotide synthesised by Frank

Oligonucleotide used in optimising Scarlet method	-	Am oligonucleotide	AmACUGAUCAUCAUCACCACAGAGCAGGU(biot)	Oligonucleotide synthesised by Frank
Oligonucleotide used in optimising Scarlet method	-	m^6Am oligonucleotide	m^6AmACUGAUCAUCAUCACCACAGAGCAGGU(biot)	Oligonucleotide synthesised by Frank
SCARLET Splint		Oligonucleotide Splint	TGTGGTGATGATGATCAGTTCCGATGGGTGAGTCCGTAAG	Hybridisation oligonucleotide for Frank's Oligonucleotide
SCARLET ssDNA		ssDNA for Scarlet	GGACTTGACGACTTACGGACCTTACGGACTCACCCATCGG	ssDNA used in optimisation of SCARLET methods
SCARLET ssRNA		ssRNA for Scarlet	GGACUUGACGACUUACGGACCUUACGGACUCACCCAUCGG	ssRNA used in optimisation of SCARLET methods
Drosophila Splint		Drosophila Splint	CCTCTACGCCAGGTAAGTATCCGATGGGTGAGTCCGTAAG	Hybridisation oligonucleotide for U2 snRNA drosophila

Table 2.6. Primers and oligonucleotides used in this study. The table contains the sequence, name or the oligo, what it was designed for, the target sequence and any relevant notes.

2.2 Methods

2.2.1 RNA extraction

The RNA used in Thin Layer Chromatography (TLC) experiments to investigate the evolutionary conservation of m^6Am presented in chapter 3, was extracted using the mirVana™ miRNA Isolation Kit, with phenol. Tissue samples were snap frozen in liquid nitrogen and stored at -80°C until the RNA was extracted. Rat brain was sliced using a cryostat (Leica, CM3050) before vortexing in lysis buffer to disrupt the cells. Other brain tissues were manually ground using a plastic pestle in lysis buffer before vortexing to further disrupt the tissue. The RNA isolation procedure was followed according to the kit's manual.

RNA was extracted from *Ectocarpus siliclosus* using a method described by Greco et al., (Greco et al., 2014). 0.1g of tissue was used for each extraction. Two changes were made to the extraction protocol. One, to disrupt the cells the tissue was ground in a pestle and mortar after being frozen in liquid nitrogen. Two, turbo DNase was used to remove the DNA from the total nucleic acid extracted. 10µg of total nucleic acid was diluted up to 44µl of distilled water and 5µl of Turbo DNase buffer was added with 1µl of Turbo DNase. The reaction was left for 30minutes at 37°C. To end the reaction phenol chloroform was used according to the paper's instructions.

Human brain RNA used for the TLCs presented in chapter 4, was extracted using AllPrep DNA/RNA Mini Kit which extracts RNA and DNA simultaneously. The DNA and total RNA isolation procedure was followed according to the kit's manual.

Acid phenol extraction was used to extract rat brain RNA which was used in the DCP2 decapping experiments presented in chapter 5. The protocol was modified from a yeast RNA extraction protocol (Schmitt et al., 1990). Sliced rat brain tissue was vortexed with

400µl of AE buffer and 40µl of 10% SDS before an equal volume of AE saturated phenol was added. This was heated to 65°C for 4 minutes before being dipped into liquid nitrogen until phenol crystals appeared. Tubes were then centrifuged for 5 minutes at a maximum speed of 13,000 rpm. The aqueous layer was then mixed with an equal volume of phenol chloroform before being centrifuged for 3 minutes on maximum speed. The RNA in the aqueous layer was precipitated overnight at -20°C with 1/10 volume of 3M NaOAc pH5 and 2.5 volumes absolute ethanol. RNA was pelleted by centrifugation at room temperature for 10 minutes 13,000 rpm. The pellet was then washed in 80% ethanol and dried in a vacuum drier for 2 minutes. The RNA was resuspended in 50µl of distilled water.

2.2.2 Phenol chloroform purification

After enzymatic reactions, phenol chloroform was used to remove the enzyme and buffer. The RNA was then precipitated using sodium acetate and ethanol. The enzyme reaction was made up to 200µl with sterile distilled water and 1 volume of phenol-chloroform was added, vortexed for 1 minute and centrifuged at 13,000rpm for 5 minutes. The aqueous phase was transferred into a new tube and 1 volume of chloroform was added, vortexed for 1 minute and centrifuged for 3 minutes at 13,000rpm. The aqueous phase was transferred to a new tube and 2.5 volumes of absolute ethanol and 1/10 volume 3M sodium acetate pH 5, was added before the RNA was precipitated over night at -20°C. If a small amount of RNA was being precipitated or the RNA length was short, 1µl of glycogen was added to the precipitation mix.

The following day, precipitated RNA was centrifuged for 10 minutes at maximum speed 13,000 rpm. The pellet was washed in 500µl of 80% ethanol. The supernatant was

removed and the pellet vacuum dried for 2 minutes before being resuspened in sterile distilled water. The amount of water varied depending on the next reaction and the amount of RNA in the pellet.

2.2.3 mRNA enrichment

2.2.3.1 Terminator exonuclease digestion

Total RNA was cold labelled using T4 Polynucleotide Kinase (PNK). This adds a 5′monophosphate onto the rRNA. Reaction mixture consisted of 4µl 10mM ATP, 4µl 10x PNK A buffer, 4µl PNK enzyme, 4µg total RNA. The reaction mixture was incubated at 37°C for 2 hours, followed by phenol chloroform purification and ethanol precipitation. Terminator exonuclease was used to remove the RNA which had a 5′ monophosphate, including rRNA, other uncapped RNA species and fragmented RNA. The RNA pellet was resuspended in 24µl of water, 3µl 10x reaction buffer A, 3µl terminator exonuclease. This was incubated for 3 hours at 30°C, followed by phenol chloroform purification and ethanol precipitation over night at -20°C.

For the second and third rat, drosophila, axolotl and zebrafish samples the terminator treated RNA was put though Qiagen RNeasy micro kit column, following the clean-up and concentration protocol to remove degraded RNA fragments.

2.2.3.2 Poly-A purification

Poly-A purification was based on the NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB). The oligo d(T) beads were bought separately and the buffers made up in the lab according to the NEBNext® manufacturer's instructions. 5µg of total RNA was diluted up to 50µl with sterile distilled water. 20µl of beads were aliquoted into a 1.5ml Eppendorf and left to stand on the magnetic rack for 2 minutes. The supernatant

was then discarded and the beads re-suspended in 100µl RNA binding buffer. The tube was placed back on the rack and the supernatant removed. This was repeated once more. The beads were re-suspended in 50µl of RNA binding buffer and the 50µl of RNA was added. The tube was placed in a heat block at 65°C for 5 minutes, and cooled on ice. The beads were pipetted up and down 6 times before being put on rollers at room temperature for 20 minutes. The tube was put back on the magnetic rack for two minutes. The supernatant was removed which was the –poly(A) RNA. The beads were washed twice with 200µl wash buffer. 50µl of pre heated 80°C water was added and the beads were resuspended and incubated at 80°C for two minutes. The tube was cooled at room temperature. 50µl of RNA binding buffer was added and the beads were pipetted up and down six times and incubated on the rollers for 15 minutes at room temperature. The supernatant was discarded and the beads were washed with 200µl of wash buffer. The beads were resuspended in 10µl of pre warmed 80°C water, incubated for 2 minutes at 80°C, and placed on to the magnetic rack. The supernatant which was the poly(A) RNA was transferred into a clean tube.

2.2.3.2.1 Bead regeneration

The oligo d(T) beads were regenerated so they could be used again. Beads were washed in 200µl 0.1N NaOH on the rollers for 5 minutes. The beads were washed with distilled water until a neutral PH was achieved; and then equilibrated and stored in PBS Tween buffer.

2.2.3.3 RNase H digestion

RNase H digestion was performed using NEBNext® rRNA Depletion Kit. The manufacturer’s instructions were followed.

2.2.4 RNA measurements

For all methods, the concentration of the RNA was measured using Nanodrop (NanoDrop1000, Thermofisher). Some samples were also run on Agilent pico chips by Dr Zsuzsa Bodi to generate a RNA Integrity Number (RIN) a value which is an indicator of quality.

2.2.5 Thin Layer Chromatography

2.2.5.1 Cap adjacent TLC

The TLC method for cap adjacent modifications was first described (Kruse et al., 2011). The changes made to this method during the project are detailed below.

rRNA free RNA was decapped using 5′pyrophosphohydrolase (RppH) which leaves a 5′ monophosphate. The reaction mixture consisted of 500ng of poly A RNA or RNA from terminator removal of rRNA, 5µl Thermopol reaction buffer, 5µl RppH, sterile distilled water up to 50µl. This was incubated for 1hr at 37°C before Calf Intestinal Alkaline Phosphatase (CIP) was added to remove 5′ phosphate. 1.5µl CIP was added and incubated for 20 minutes at 37°C, followed by phenol chloroform purification and ethanol precipitation over night at -20°C.

The RNA was pelleted by centrifugation at room temperature for 10 minutes 14,000 rpm. The pellet was then washed in 80% ethanol and dried in a vacuum drier for 2 minutes. The decapped RNA was suspended in 5µl of water.

PNK was used to add a radioactive 5′ phosphate to the first nucleotide of the RNA. Reaction mixture was composed of, 2.5µl decapped RNA, 1µl PNK A buffer, 5µl water, 0.5µl ^{32}pATP, 1µl PNK. This was incubated for 1 hour at 37°C. To remove free nucleotides the reaction was put through QIAquick nucleotide removal columns,

following the instructions for radioactive samples. The elution of the RNA was changed and the RNA was eluted in 20µl twice. These were combined and precipitated over night at -20°C in 100µl of absolute ethanol and 40µl 3M NaOAc pH5.

Labelled RNA was precipitated by centrifugation at 13,000rpm for 20 minutes. The supernatant was removed and the pellet was dried for 5 minutes in a 37°C hot block. RNA was digested using P1 enzyme in a reaction mixture consisting of the RNA pellet, 1µl 10x P1 buffer, 0.5µl P1, 8.5µl water and incubated for 2 hours at 37°C.

2.2.5.2 Internal TLC

m^6A was detected in the consensus sequence GAC. The T1 enzyme used in the TLC experiments cuts RNA after G exposing the next nucleotide including m^6A modified adenosine. The reaction mixture used consisted of 25ng of RNA, 1µl PNK buffer A, 1µl T1 and water up to 8.5µl and was incubated for 15 minutes at 37°C. 1µl PNK and 0.5µl 32pATP was then added to label the exposed nucleotides. The PNK reaction was incubated for 1hr at 37°C. The reaction mixture was subsequently ethanol precipitated overnight at -20°C with the addition of 1µl 3M NaOAc ph5 and 40µl ethanol. Labelled RNA was precipitated by centrifugation at 13,000rpm for 20 minutes. The supernatant was removed and the pellet was dried for 5 minutes in a 37°C hot block. The RNA was digested using P1 enzyme in a reaction mixture consisting of the RNA pellet, 1µl 10x P1 buffer, 0.5µl P1, 8.5µl water and this was incubated for 2 hours at 37°C.

2.2.5.3 Running a TLC

1µl of the labelled nucleotides were added to cellulose TLC plates and placed in the TLC first buffer overnight and left to dry before the plate was turned and placed into the second buffer overnight. The plates were dried and placed on a phospho-screen

and imaged using a FX scanner. The TLCs were run on cellulose F and cellulose TLC plates, the change was a result of a supplier shortage of cellulose F TLC plates. The nucleotides run in slightly different positions on the two different types of TLC plates but all the nucleotides can be detected.

2.2.5.4 Imaging a TLC screen

The TLC screens were imaged using a FX scanner (Bio-Rad). Quantity One software (Bio-Rad) was used to analyse the TLC images. The software detects the saturation of the pixels in a given area. A ratio was created between the saturation of the A spot and the Am spot and the Am spot and the m^6Am spot. To analyse the significance between the samples Prism was used to perform Kruskal Wallis statistical tests and create the graphs.

2.2.6 SCARLET

A modified version of a published site-specific cleavage and radioactive-labelling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET) protocol reported by Lui et al., (Liu et al., 2013b) was performed so that the first nucleotide of transcripts could be identified. The new method was optimised using RNA oligonucleotides before being performed on wild type and *Cmtr1*$^{-/-}$/*Cmtr2*$^{-/-}$ total Drosophila RNA samples.

2.2.6.1 Oligonucleotide SCARLET

To test the changes that were made to the published SCARLET method RNA oligonucleotides were used. These were used as the first nucleotide of the oligonucleotides was known. Three oligonucleotides starting with A, Am or m^6Am were synthesised by Frank Baron (University of Nottingham). The m^6A and m^6Am

phosphoramidites were synthesised according to Kruse et al., 2012. The oligonucleotides were synthesised using an Applied Biosystems 394 DNA Synthesizer (School of Chemistry, University of Nottingham) with the standard 0.2 µM RNA synthesis cycle.

To act as a standard, and understand where the labelled nucleotides would migrate to, the oligonucleotides were labelled individually and digested to single nucleotides. Two labelling controls were also performed. For the first control, the three oligonucleotides were labelled individually and then equal volumes were mixed and P1 treated. For the second control, equal volumes of the three oligonucleotides were mixed together, radioactively labelled and then P1 treated.

The oligonucleotides were generated with a 5′ OH so they could be labelled without dephosphorylation. The labelling mixture consisted of 3pmol oligonucleotide, 1µl p^{32}ATP, 1µl PNK, 1µl PNK buffer A and sterile distilled water to 10µl. The reaction was incubated for 1hour at 37°C before heat inactivation at 75°C. The reaction was put through Qiagen nucleotide removal columns to remove unincorporated ATP and eluted in 20µl sterile distilled water. At this point both the controls and standards were P1 digested. Reaction mixture consisted of 4µl oligonucleotide, 4µl water, 1µl P1, 1µl 10x P1 buffer and incubated for 10 minutes at 60°C and then 37°C for 1.5 hours.

For the SCARLET method itself, the three oligonucleotides were mixed together in equal amounts and then annealed to a single stranded RNA or DNA oligonucleotide using a splint. The reaction mixture consisted of 0.2pmol (4µl of labelled oligonucleotide), 1µl PNK buffer A, 5µl water, 1µl (4pmol) splint, 1µl (5pmol) ssRNA/DNA and was incubated for 3 minutes at 75°C and 10 minutes at room temperature. The single stranded RNA oligonucleotide and mix of A, Am, and m^6Am

oligonucleotides were ligated. The reaction mixture consisted of 3µl sterile distilled water, 2µl DMSO, 1µl PNK buffer A, 1µl 10mM ATP, 1µl T4 RNA ligase 2 or T4 DNA ligase. The ligation was incubated at 37°C for 3.5 hours. The reaction was then dephosphorylated to remove the radioactive phosphate from any oligonucleotides that had not ligated to the single stranded RNA/DNA. 1µl of FastAP Thermosensitive Alkaline Phosphatase was added. This mixture was incubated at 37°C for 30 minutes followed by heat inactivation at 75°C for 5 minutes.

The oligonucleotides had a biotinylated UTP at the 3′ end to clean up the reaction they were bound to hydrophilic streptavidin beads and washed. 5µl of hydrophilic streptavidin beads were put into new 1.5ml Eppendorf tubes and washed twice with 20mM Tris-HCL pH 7.5. To wash the beads were resuspended in 100µl Tris-HCL, pipetted up and down 6 times and then placed on the magnetic rack and the supernatant removed. To bind the RNA onto the beads, 20µl RNA and 20µl of RNA capture buffer were incubated with the beads on rollers at room temperature for 20 minutes. The beads were then washed twice in 1x PBS before P1 digestion. P1 digestion mixture consisted of the RNA bound beads, 8µl of sterile distilled water, 1µl P1, 1µl P1 buffer. This was incubated at 60°C for 10 minutes followed by 37°C 1.5 hours. 1µl of each P1 digestion was subsequently placed on a TLC plate and ran in the first dimension only. The buffer was the same as for the standard TLC protocol.

2.2.6.2 U1 snRNA SCARLET

The RNA single stranded oligonucleotide ligated with a greater efficiency to the A, Am, and m^6Am RNA oligonucleotides so this was used in the method aimed at detecting the first nucleotide of U1 snRNA. First the snRNAs needed to be enriched for. To do this –poly(A), the RNA that did not bind to the oligo d(T) beads, was treated with

terminator exonuclease to remove uncapped RNAs. –poly(A) RNA from wild type drosophila and *Cmtr1*$^{-/-}$/*Cmtr2*$^{-/-}$ knockout drosophila was first treated with PNK to add a 5′ monophosphate. The reaction mixture consisted of 2µg –poly(A) RNA, 3µl PNK buffer A, 3 µl 10mM ATP, 2µl PNK, 2µl sterile distilled water. This was incubated for 2 hours at 37°C. To end the reaction phenol chloroform purification was performed followed by ethanol precipitation. Next RNA was treated with terminator exonuclease, the reaction mixture consisted of RNA pellet, 1µl Terminator exonuclease, 2µl buffer A, 16µl sterile distilled water. This was incubated for 3 hours at 30°C. To end the reaction phenol chloroform purification was performed followed by ethanol precipitation.

RNA was decapped and dephosphorylated so that the first nucleotide could be labelled with a radioactive 5′ phosphate. Two tubes were prepared for wild type and knock out RNA, the reaction mixture consisted of 500ng RNA, 5µl Thermopol buffer, 5µl RppH, sterile distilled water to 50µl. This was incubated for 1 hour at 37°C. 1.5µl of CIP was added and this was incubated for a further 20 minutes at 37°C. To end the reaction phenol chloroform purification was performed followed by ethanol precipitation.

~1µg of RNA enriched for snRNAs was input into the SCARLET assay. The decapped RNA was labelled, using a reaction mixture consisting of 1µg RNA, 7µl sterile distilled water, 1µl PNK, 1µl PNK buffer A and 1µl p^{32}ATP. This was incubated for 37°C for 1 hour and then heat inactivated for 5 minutes at 75°C. To anneal the ssRNA, 1µl (4pmol) splint DNA, 1µl (5pmol) ssRNA was added to the reaction. This was incubated for 3 minutes at 75°C and 10 minutes at room temperature. To ligate the ssRNA to the U2 RNA RNA T4 ligase 2 was used. Reaction mixture consisted of annealing reaction, 3µl sterile distilled water, 2µl DMSO, 1µl PNK buffer A, 1µl 10mM ATP, 1µl T4 RNA ligase 2 was incubated at 37°C for 3.5 hours. The reaction was then dephosphorylated to

remove the radioactive phosphate from the RNA that had not ligated to the ssRNA, 3µl of CIP was added to the reaction and was incubated at 37°C for 40 minutes. To end the reaction and remove radioactive phosphate, the reaction was put through a Qiagen Qiaquick nucleotide removal column and eluted in 100µl. This was then ethanol precipitated overnight. The SCARLET reaction was P1 digested before being run on a 2D TLC. Reaction mixture consisting of the RNA pellet, 1µl P1, 1µl P1 buffer, 8µl water was incubated at 60°C for 15 minutes followed by 37°C 1 hours 45 minutes.

2.2.7 Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) was used amplify the human *Dcp2* coding sequence from the pet28a-*hDcp2* plasmid and to add on the His tag and add the over hangs for Gibson assembly cloning. For these PCR reactions the phusion high fidelity DNA polymerase was used as it has a proof reading ability. Colony PCR was used to assess if the DCP2 coding sequence had been cloned into the plasmid successfully. For this reaction Taq DNA polymerase was used. For both enzymes, standard reaction volumes and thermocycler conditions were used according to manufacturer's instructions. The PCR products were run on 1% agarose Tris/Borate/EDTA (TBE) gels, with 5µl ethidium bromide solution in 50ml of agarose. 1kb ladder was used to identify the band size. Between 5 and 20µl of PCR product was loaded with 6x loading dye. The gels were run at 100 volts for 45 minutes, and imaged using a biorad imager.

2.2.8 Protein cloning

2.2.8.1 Plasmid assembly

The Gibson assembly method was used to make the pMAL-c5X *hDcp2* plasmid. An online calculator (http://nebuilder.neb.com/) designed the overhangs which were

added to the Dcp2 coding sequence. The pMAL-c5X was cut using Xmn1 in cut smart buffer for 30 minutes and heat inactivated for 20 minutes at 65°C. The Dcp2 PCR product (Dcp2 coding sequence, with HIS tag and overhangs added) and cut plasmid were ligated using the Gibson assembly master mix according to manufactures instructions.

2.2.8.2 Transformation

The assembled pMAL-c5X-*hDcp2* plasmid was transformed into DH5-α competent *E. coli* using heat shock method according to the protocol given with the Gibson Assembly kit. To transform the plasmid into the Rosetta-gami cells and the SoluBL21 cells, a standard lab transformation protocol was followed. 25ng of plasmid was added to the cells and this was left on ice for 30 minutes before heat shock of 30 seconds at 42°C. After heat shock the cells were put back on ice for two minutes and then 700µl of SOC media was added. The cells were incubated for one hour at 37°C. 200µl of cells were plated on agar plates with the relevant antibiotics and grown at 37°C overnight.

2.2.8.3 Plasmid extraction

GeneJET Plasmid Miniprep Kit was used to extract the plasmid according to manufacturer's instructions.

2.2.8.4 Glycerol stocks

Glycerol stocks of all the transformed *E. coli* were generated by mixing 500µl of overnight culture with 500µl of 50% glycerol. The stocks were frozen in liquid nitrogen before being transferred to the -80°C freezer for storage.

2.2.9 Sequencing

To check that the Dcp2 sequence was accurate and that the Dcp2 PCR product was ligated into the plasmids correctly, the Eurofins premixed sequencing service was used to perform Sanger sequencing.

2.2.10 Pilot experiments

To check that the expressed protein was soluble and to further optimise soluble protein production pilot experiments were performed according to the protocol from the pMAL protein fusions and purification system manual (NEB). Details are provided in table 2.7. For all the pilot experiments *E. coli* was grown at 37°C until the cells reached an OD A_{600} of 0.5 and then protein production was induced with IPTG.

Cells	Media	IPTG	Other additions	Temp	Induction time
DH5α	Rich broth	0.3mM		37°C	1,2,3 hours
Rosetta-gami	Rich broth	0.3mM		37°C	2 hours
Rosetta-gami	Rich broth	0.15mM		37°C	2 hours
SoluBL21	LB	0.4mM	2% ethanol	18°C	20 hours
SoluBL21	M9 + glucose	0.4mM		18°C	20 hours
SoluBL21	M9 + glycerol	1mM		18°C	20 hours

Table 2.7. Pilot experiment conditions. The table contains details of the cell types and conditions that were tested to obtain soluble DCP2 protein.

2.2.11 Protein gel

Protein gels were run to detect if the DCP2 protein was successfully produced. 10µl of sample was mixed with the same volume of lamelli buffer and heated to 90°C for 5 minutes. The samples were loaded on to Biorad TGX precast gels along with 7µl of

protein ladder. The gels were run for 1 hour at 110 volts. After the gels were washed 3 times in distilled water for 10 minutes and, stained in page blue or imperial stain for 1hour on a rocker. The stain was removed and the gel was washed for 60 minutes in distilled water, the water was changed every 10 minutes. The gel was left in distilled water overnight before being imaged using the Biorad GS800 scanner.

2.2.12 Protein production and purification

To grow *E. coli* for protein production 1 Litre of LB media was inoculated with 10ml culture and 1ml ampicillin overnight. The culture was grown at 37°C until OD600=0.5 and then induced with 2% ethanol and 0.4mM IPTG and grown for 20 hours at 18°C. The cells were pelleted by centrifugation at 4000rpm for 20 minutes. The pellets were resuspended in equilibrium buffer and frozen.

DCP2 was purified using the gravity flow column method in the HisPUr Ni-NTA resin manual. 1ml of Ni-NTA resin was put into a gravity flow column and storage buffer was allowed to drain out. The resin was washed with 2 bed volumes of equilibration buffer. The protein equilibrium buffer mix was put through the column. The column was washed with two bed volumes of wash buffer, and repeated until the absorbance of the flow through at 280nm approached baseline. The first and last wash were collected to run on a protein gel later. The protein was eluted using two volumes of elution buffer, repeated twice. The first 0.5ml was collected separately, this was the last of the column buffer. From each elution a 10µl aliquot was taken to run on a protein gel. The rest was mixed with 50% glycerol and an aliquot of 50µl was removed to detect the protein concentration. The remainder of the elution was frozen in liquid nitrogen. The concentration was detected using Pierce™ BCA Protein Assay Kit, according to manufacturer's instructions.

To regenerate the column, the column bed was washed with 10 resin-bed volumes of MES buffer; followed by 10 bed volumes of distilled water. The resin was then stored in 50% slurry of 20% ethanol.

2.2.13 DCP2 treatment of poly A RNA

Poly(A) RNA was decapped using DCP2 in a method similar to a method reported by the Jaffrey's lab (Mauer et al., 2016). 20mM of DCP2 was mixed with 400ng of poly(A) RNA in decapping buffer 37°C for 30 minutes. The reaction was stopped by phenol chloroform and precipitated using ethanol and sodium acetate.

After DCP2 treatment the RNA was treated to isolate different populations of decapped and capped transcripts. Between each of the reactions, phenol chloroform purification was performed and the RNA precipitated.

To assess capped transcripts, the DCP2 treated RNA was then treated with PNK, Terminator, RppH, CIP and input in to the labelling step of the TLC protocol. The PNK reaction consisted of the RNA pellet, 1µl PNK, 2µl PNK buffer A, 2µl 10mM ATP, sterile distilled water up to 20µl and was incubated for 1hour at 37°C. Terminator exonuclease reaction consisted of the RNA pellet, 1µl terminator, 2µl terminator buffer A, sterile distilled water up to 20ul and was incubated for 3 hours at 30°C. RppH and CIP reaction consisted of the RNA pellet, 5µl thermopol buffer, 4µl RppH, sterile distilled water up to 50µl. This was incubated for 1 hour at 37°C before 1.5µl CIP was added and incubated for a further 20 minutes.

To look at the 5′OH decapped transcripts the DCP2 treated RNA was input into the TLC protocol at the labelling step. To investigate the 5′ P and 5′OH decapped transcripts the DCP2 treated RNA was treated with CIP before being input into the TLC

at the labelling step. CIP reaction consisted of the RNA pellet, 5µl thermopol buffer, 1.5µl CIP, sterile distilled water up to 50µl and was incubated for 30 minutes at 37°C

2.2.14 Protein pull down

2.2.14.1 RNA oligonucleotide transcription and capping

To generate the RNA oligonucleotides for protein pull down experiment, the RiboMAX Large Scale RNA Production System was used. The RiboMAX Large Scale RNA production System was chosen as the dNTPs are added individually. Three oligonucleotides were generated, Cap0 (GpppA), Cap1 (GpppAm) and m^6Am (Gpppm^6Am). 1µl of each template primer (100mM) was used (T7sigma2.5, T7sigma2.5polyA). The template was heated for 5 minutes at 65°C and cooled at room temperature before being added to the reaction mixture. The rest of the mix was made up according to the kit's guidelines and left to incubate for 4 hours. To generate the A and Am oligonucleotide ATP was added to the reaction. To generate the m^6Am oligonucleotide N^6-methyl-ATP was added. Biotinylated UTP was added according to manufacturer's instructions.

To cap and generate the 2′-O-methyladenosine the NEB *Vaccinia* capping system and cap 2′-O methyltransferase enzyme was used. The standard reaction, 1ng enzyme to 37.5ng RNA, was suggested for capping standard length RNAs not short oligonucleotides. Therefore a number of changes were made to the given protocol to optimise the reactions for short oligonucleotides. The amount of RNA was decreased to 1:18.74 and 1:12.5 ng enzyme to RNA ratio. 2mM $MgCl_2$ and 0.4mM SAM were also added to reactions to increase the efficiency of the reaction.

To compare the reactions, the capping efficiency was calculated by calculating the input RNA and the RNA after everything without a cap was digested. This value will be

skewed because of the RNA lost when the reactions were cleaned up using phenol chloroform and the RNA precipitated. The methyltransferase efficiency was calculated using the Quantity One software (Bio-Rad) and is the ratio of A:Am. The final capping and methylation reactions were performed using 2µl of each enzyme and 4500ng of RNA. The protocol that was supplied with the enzymes was followed. The capping reaction for the Cap0 oligonucleotide consisted of 4.5µg oligonucleotide in 15µl of water, 2µl 10x capping buffer, 1µl 10mM GTP, 1µl 2mM SAM, 2µl Vaccinica capping enzyme. The capping and methylation reaction for Cap1 and m^6Am oligonucleotide consisted of 4.5µg RNA in 14µl of water, 2µl 10x capping buffer, 1µl 10mM GTP, 1µl 2mM Sam, 2µl *Vaccinia* capping enzyme, 2µl mRNA cap 2-O-methyltransferase. The reactions were incubated for 3 hours at 37°C.

To remove the uncapped transcripts, oligonucleotides were dephosphorylated using CIP, the reaction consisted of 2µl CIP, 5µl cut smart buffer, 500ng RNA and was incubated for 1 hour at 37°C. Phenol chloroform purification was used to end the reaction and the RNA was ethanol precipitated. The RNA was then phosphorylated by PNK to ensure there was a 5′ monophosphate with a reaction consisting of 100ng of RNA, 1µl 10mM ATP, 1µl PNK buffer A, 1µl PNK. This was incubated for 1hour at 37°C. Following this reaction phenol chloroform purification and ethanol precipitation was performed. The RNA was then treated wither terminator exonuclease to remove the uncapped oligonucleotides which started with a monophosphate, the reaction mixture consisted of 1µg oligonucleotides, 1µl terminator exonuclease, 3µl buffer A, sterile distilled water up to 30µl. This was incubated for 3 hours at 30°C. The RNA was put through the Qiagen RNA clean up kit and eluted in 50µl.

2.2.14.2 First protein pull down method

The pull down method from the Peirce Magnetic RNA-Protein Pull-Down kit protocol (Thermofisher) was followed. All the buffers from the kit and the IP lysis buffer were made in the lab. Two further changes to the protocol were made. First, to make the protein lysate. 75µl of IP lysis buffer was mixed with 7µl protease inhibitor and 0.1g of rat brain tissue was vortexed for 1 minute and centrifuged maximum speed, 13,000 rpm for 10 minutes. The supernatant was removed and used in the protein pull downs. Second, the protein pull down protocol was followed using 25µl hydrophilic streptavidin beads, 10pmol of RNA oligonucleotides, and 20µl of protein lysate. The protein pull downs were performed on three biological replicates with four treatments, Cap0 oligonucleotides, Cap1 oligonucleotides, m^6Am oligonucleotides and bead only. After the protein binding, the beads were washed three times with wash buffer and 3 times with wash buffer without the tween 20.

2.2.14.3 Second protein pull down method

For the second protein pull down, a different method supplied by the Dominissini lab who performed the proteomic analysis was performed on 4 biological replicates. The protein crude extract was made by passing sliced rat brain through a syringe needle with 500µl of whole cell extract buffer. This was incubated for 90 minutes at 4°C on a rotating wheel. This was then centrifuged for 30 minutes at 14000rpm at 4°C. The soluble cell extract was snap frozen in liquid nitrogen until it was needed.

50µl of hydrophilic streptavidin magnetic beads were washed twice in 1ml of RNA binding buffer. To inactivate any RNases 100µl of RNA binding buffer with 2.5µl of Ribolock, was added to the beads and incubated on ice for 30 minutes. The tubes were put on the magnetic rack to remove the buffer. The beads were pre-blocked with yeast

tRNA (50µg/ml) in 500µl of RNA binding buffer overnight at 4°C on a rotating wheel. The pre blocked beads were washed twice with 1ml of RNA binding buffer. 250ng of biotinylated RNA probe diluted to 600µl with RNA binding buffer was incubated for 30 minutes at 4°C on a rotating wheel. The beads were washed once with 1ml RNA wash buffer and twice with 1ml protein incubation buffer. 1000µg of protein extract, 30µg yeast tRNA and Ribolock was diluted to a total volume of 600µl protein binding buffer and incubated for 30 minutes at room temperature on a rotating wheel and then 90 minutes at 4°C. The beads were washed 3 times in 1ml protein incubation buffer and a further three times in ice cold PBS to remove detergent from the beads.

2.2.14.4 Protein digestion

Protein digestion was performed on both protein pull downs. The beads were re-suspended in 100µl of elution buffer and incubated on a rotating wheel for 20 minutes. Indole-3-acetic acid (IAA) was added to a final concertation of 55mM and incubated for a further 10 minutes in the dark. The digestion was performed by adding 250ng of trypsin and incubated for 2 hours at room temperature. After incubation, the supernatant was collected in a separate tube. 50 µl of elution buffer was added to the beads for 5 minutes at room temperature. The second elution was added to the first and 100ng of fresh trypsin was added to the pooled elutions. The elutions were left to digest overnight at room temperature. 1µl of Trifluoroacetic acid was added and then the tubes were vortexed. The samples were frozen in liquid nitrogen and stored at -80°C.

2.2.14.5 Mass spectrometry and proteomics

The mass spectrometry and proteomics were completed by Dr Efrat GlickSaar at the University of Tel Aviv. The tryptic peptides were desalted using C18 tips prior to Mass

spectrometry analysis. They were then separated using an Easy-nLC 1200 (Thermo) connected to a Q Exactive HF mass spectrometer (Thermo). Samples were separated using acetonitrile gradient followed by washes for a total of 140 minutes of data acquisition. For measurements on the Q Exactive HF, the top 10 most abundant peptides were fragmented for every full scan. Data processing was performed using the MaxQuant software. Normalized ratios from MaxQuant output tables were used to identify significant interactors.

2.2.15 Bioinformatic analysis

2.2.15.1 Searching for m^6Am in m^6A-seq data

m^6A RNA sequencing was performed as part of a project funded by the Stoneygate Trust. Preprocessing of reads was performed using FASTX. The reads were aligned to the human genome build GRCh38 using Tophat2 using default settings (Kim et al., 2013). The grch38 gtf file was downloaded from ensemble.com. The m^6A-tagged regions were identified using exomePeak peak calling algorithm with the input library as the background (Zhang et al., 2008). m^6A RNA sequencing data was provided pre-processed in CSV files containing the Ensembl ID, peak location strand and score. Three data sets were analysed, human foetal brain RNA, white matter RNA and grey matter RNA for the parahippocampus.

2.2.15.1.1 Annotating peaks

The Homer programme was used to annotate the sequencing peaks (Heinz et al., 2010). The chromosome, start of the peak, end of the peak and strand were added to a text file and Homer annotated the peaks with the following options, TSS, 5′UTR, exon, intron, 3′UTR, TTS, non-coding, intergenic and NA. The code is as follows.

```
annotatePeaks.pl input_bed_file.txt hg38 > output_file.txt
```

2.2.15.1.2 Starting nucleotide

The m^6Am motif only happens when the starting nucleotide is an adenosine. The sequence of the peaks annotated TSS were extracted using Homer using the following command.

homerTools extract input_coordinates.txt genome location >output file.txt

In excel the peaks starting with A were taken, these are the potential m^6Am peaks. As a control the Ensembl IDs of the peaks from the total data set were put into Ensembl Biomart. The cDNA sequence for the Ensembl ID's was extracted from the Hg38 aligned genome. This control was examined to assess if the first nucleotide of the TSS peaks were biased by the sequencing process and so were compared to the starting nucleotide of the total peak sequences and the starting nucleotide of the cDNA transcripts for the total data set.

2.2.15.1.3 Motif analysis

De novo motif enrichment was conducted using Homer. *De novo* motif analysis was performed looking for motifs five nucleotides long and using the RNA parameters. The analysis was performed using sites characterised as potential m^6Am peaks, the TSS peaks and the whole set of sequencing peaks compared to the random GC matched background sequences provided by Homer. The potential m^6Am peaks were also compared to the total set of peaks from the sequencing data. The code used is as follows.

findMotifsGenome.pl input_Bed_file.txt hg38 output_folder/ -len 5 -rna

findMotifsGenome.pl input_Bed_file.txt hg38 output_folder/ -len 5 -rna -bg background_bed_file.txt

2.2.15.1.4 Gene ontology and reactome analysis

DAVID was used to perform the Gene Ontology (GO) and reactome analysis for the potential m^6Am peaks and for the Cap0, Cap1, and m^6Am binding proteins. A p-value cut off of 0.05 was used for all analysis and the top 20 were put into tables found in the results sections. For the potential m^6Am peaks the total data set of sequencing peaks was used as the background. For the m^6Am binding proteins *Rattus norvegicus* genome was used as the background.

2.2.15.2 Trees of Similarity

The amino acid sequence of PCIF1, CMTR1, CMTR2 and FTO were obtained from the NCBI protein database for the species of interest i.e. *Homo sapiens*, Rattus norvegicus, *Xenopus tropicalis*, *Danio rerio*, *Drosophila melanogaster, Octopus vulgaris, Apis mellifera, Gallus gallus.* As the Xenopus genome is better annotated, the Xenopus genome was used instead of axolotl. Human FTO was compared against marine algae proteomes annotated in the JGI genome and Ensembl protist databases. In the species which contained a FTO orthologue, human PCIF1, CMTR1 and CMRT2 were also compared.

The amino acid sequence for METTL3, METTL14 and WTAP were procured from the NCBI protein database for the species of interest. The human and *Arabidopsis* sequences for the three proteins were compared against the proteomes of the marine algae used in the previous trees of similarity using the JGI genome and Ensembl protist databases.

The proteins alignments were performed using MUSCLE3.8.31 (Edgar, 2004b, Edgar, 2004a). MUSCLE performs a pairwise alignment, the default settings were used. MUSCLE performs 16 iterations of the alignment or until convergence is reached. The

code used is as follows, a phylip output format was required for RAxML to build the trees of similarity and a multiple sequence file was required for Genedoc to visualise the protein alignments.

./muscle3.8.31_i86linux64 -in <folder/inputfile.fasta> -phyiout <outputfile>

./muscle3.8.31_i86linux64 -in <folder/inputfile.fasta> -msf -out <outputfile>

The protein alignments were visualised in Genedoc2.7 (Nicholas et al., 1997) using the default settings.

The trees of similarity were built using RAxML 8.0.0 (Randomized Axelerated Maximum Likelihood) (Stamatakis, 2014). The trees were made using a maximum likelihood algorithm, a GAMMA model of heterogeneity and most likelihood estimate of alpha-parameter was used. The Blosum62 amino acid substitution matrix was used to identify distant evolutionary relationships. Blosum62 is a 'deep' scoring matrix and was used because the protein sequences were full length and the alignments shared between 20-30% identities. The Trees were bootstrapped, using the data from 100 trees.

To create a tree without bootstrapping the code is as follows.

./raxmlHPC-SSE3 -p 12345 -s <folder/muscleoutput.phy>
-m PROTGAMMABLOSUM62 -n <folder/outputfile>

-p = Parsimony random seed

-s = name of alignment file

-m = model of binary amino acid substitutions

-n = name of output file

To create a tree with bootstrapping the code is as follows.

./raxmlHPC-SSE3 –f a -p 12345 –x 12345 –N100 -s <muscleoutput.phy>
-m PROTGAMMABLOSUM62 -n <outputfile>

-f a = rapid bootstrap analysis and search for best-scoring ML tree in one program run

-p = Parsimony random seed

-x = random see

-N = number of trees that will be made

-s = name of alignment file

-m = model of binary amino acid substitutions

-n = name of output file

The trees of similarity were visualised in Dendroscope3 (Huson et al., 2007, Huson and Scornavacca, 2012). The trees were viewed as an unrooted rectangular phylogram with a scale bar of amino acid substitutions per site.

Chapter 3: Evolutionary Conservation of the m^6Am Modification

3.1 Preface

At the time the work described in this chapter commenced, the protein which methylates cap1 adenosine to form m^6Am, i.e. the m^6Am methylase, was unknown. To assist in the identification of this protein, the presence or absence of the m^6Am modification was assessed in brain tissue from different species, i.e. in mammals, birds, amphibians, insects and cephalopods. Detection of the first nucleotide was performed by thin layer chromatography (TLC), which has previously been shown to be a reliable, reproducible, and quantifiable method (Kruse et al., 2011).

In 2011 CMTR1, the enzyme that methylates the first nucleotide at the 2′ ribose position to give cap1, the substrate on which the m^6Am methylase acts, was discovered (Belanger et al., 2010). During this study investigating the evolutionary conservation of m^6Am two further proteins associated with m^6Am were identified; FTO, a demethylase that acts on m^6Am (as well as internal m^6A) in 2016 (Mauer et al., 2016) and the m^6Am methylase, PCIF1 which was discovered in 2018 (Akichika et al., 2018, Boulias et al., 2019, Sendinc et al., 2019). The correlation between these three proteins and m^6Am modification was analysed. On searching the literature, it was apparent that FTO orthologues are present in marine algae genomes (Fredriksson et al., 2008, Robbens et al., 2008) but were otherwise missing from the plant and fungal kingdoms. This was explored further to examine if there was a correlation between the presence of FTO in the proteome and the presence of m^6Am.

3.2 Results

3.2.1 Optimisation of rRNA removal

This study investigated cap adjacent modifications found on capped RNA. To achieve this mRNA was enriched by removing ribosomal RNA which constitutes approximately 90% of total RNA in a cell (Lodish et al., 2000). There are three common methods that enrich for mRNA and capped species. i) poly(A) purification uses magnetic beads with oligonucleotides that bind the poly(A) tail of the RNA. ii) rRNA depletion, which uses an RNase H method to digest away the ribosomal RNA which has been bound by DNA oligonucleotides homologous to the rRNA. iii) Terminator exonuclease digests RNA with a monophosphate at the 5′ end so will digest the rRNAs, it will also digest any fragmented RNA if PNK treatment is first preformed to add the required 5′ monophosphate.

The quality of the RNA was measured using RNA Integrity number (RIN) which is the ratio between the 18s and 28s rRNA peaks and is used to detect degradation. As shown in figure 3.1, plots produced by the Agilent software for total RNA and treated RNA indicate the quality of the total RNA and the effectiveness of each mRNA purification method. The total RNA graph for SH-SY5Y cell RNA shows there is some degradation, as the 28s peak is shorter than the 18S peak. There are no publications describing what the *O. vulgaris* transcriptome profile should look like on an Agilent chip and therefore interpretation about quality of RNA for this species might be limited. rRNA depletion and poly(A) purification were the most successful methods; on the Agilent graphs the 18s and 28s ribosomal peaks cannot be detected (Figure 3.1b,e). The results of using the terminator method was improved after increasing the incubation time from two hours to three hours (Figure 3.1c, f). Across all rRNA depleted samples the RIN is low, as the ribosomal ratio is required for the calculation.

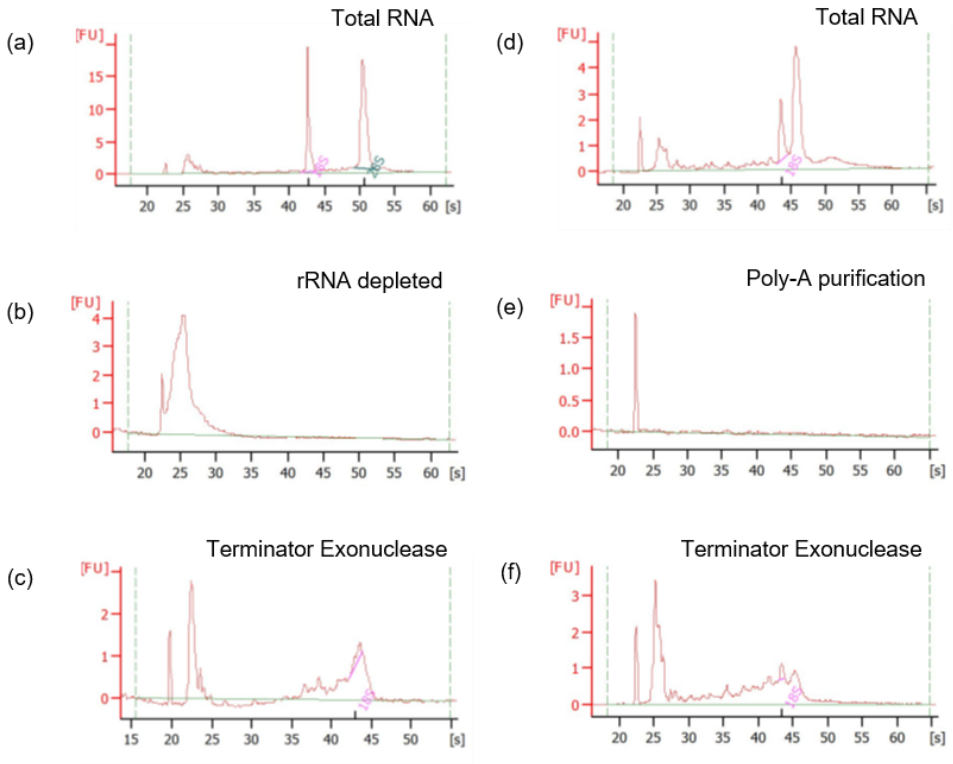
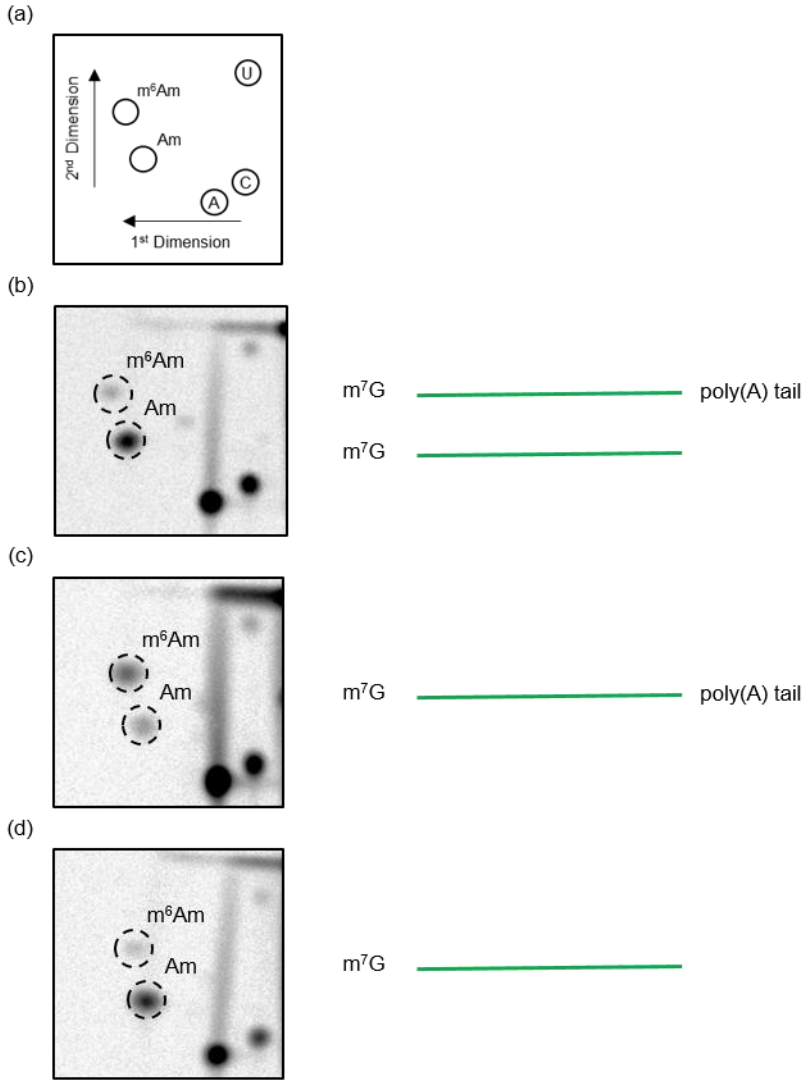


Figure 3.1. Agilent graphs showing the quality of total and rRNA depleted RNA. The first peak on the graph is the marker. The following peaks are in size order from smallest to largest. The 18s and 28s rRNAs are marked in pink and blue text respectively on the graphs by the software. a) Total RNA extracted from SH-SY5Y cells, RIN=9. b) NEBNext rRNA depletion of SH-SY5Y RNA, RIN=2.6. c) Terminator exonuclease treated SH-SY5Y RNA for 2hours, RIN=3.2. d) Total RNA extracted from *O. vulgaris* brain, RIN=3.9. e) poly(A) purified *O. vulgaris* RNA, RIN=NA. f) Terminator exonuclease treated *O. vulgaris* brain RNA for 3hours, RIN=2.6.

It is well known that m^6Am is found adjacent to the cap on mRNA but it was not reported to be cap associated in other species until recently, when it was reported on snRNAs (Mauer et al., 2019). To broadly investigate which classes of RNA contained the m^6Am modification, TLCs on RNA treated in three ways were performed, to detect m^6Am on RNA that had been enriched for different RNA species. The three treatments were PNK and terminator exonuclease (which enriches for capped RNA species); poly(A), PNK and terminator exonuclease treated RNA (which enriches for capped species with a poly(A) tail); and poly(A) minus, PNK and terminator exonuclease (which enriches for capped RNA species without poly(A) tails).

The TLC showing the first nucleotide of transcripts enriched for mRNA through poly(A) purification has the highest percentage of m^6Am at 81.3% relative to Am (Figure 3.2b). The TLC showing transcripts enriched for capped RNAs but without a poly(A) tail which will include snRNAs, snoRNAs, pre-miRNAs and pre-tRNAs, (Ohira and Suzuki, 2016, Abdelhamid et al., 2014) contained less m^6Am, but it is still present, 3.2% m^6Am relative to Am (Figure 3.2c). To gain the best overall picture of where m^6Am is located and to give the best chance of detecting it on RNA in the brain samples being tested, RNA enriched for all capped species was chosen for the initial investigation. RNA enriched for capped transcripts had 7.5% m^6Am relative to Am (Figure 3.2a). To enrich for capped RNA species terminator exonuclease digestion was chosen as the rRNA removal technique for this investigation.



Sample	A:Am	Am:m6Am
b	83.6 : 16.4	92.5 : 7.5
c	99.8 : 0.2	18.7 : 81.3
d	79.4 : 20.6	96.8 : 3.2

Figure 3.2. TLCs showing the first nucleotide of *R. norvegicus* RNA enriched for different RNA species. a) Schematic of a TLC labelled with A, C, U, Am and m^{6}Am. b) PNK and terminator exonuclease treated RNA enriching for RNA species with 5′cap, no selection on the 3′end. c) Poly(A) purified RNA treated with PNK and terminator exonuclease enriching for RNA species that have a 5′cap and a poly(A) tail. d) poly(A) minus RNA (RNA that did not bind to beads in poly(A) purification) treated with PNK and terminator enriching for RNA species with a 5′cap and no poly(A) tail at the 3′end. TLCs performed in one replicate.

3.2.2 Thin layer Chromatography to detect m^6Am

Detection of m^6Am using TLC was first published by the Fray lab in 2011 (Kruse et al., 2011). Two dimensional TLCs were performed to detect Am and m^6Am methylation in *H. sapiens* hippocampus tissue, *R. norvegicus*, *A. mexicanum*, *D. rerio*, *O. vulgaris* whole brain, *G. gallus* hypothalamus and *D. melanogaster* and *A. mellifera* heads. TLCs separate out nucleotides into an identifiable pattern, the first nucleotide of the transcript can be detected because it is radioactively labelled at a 5′ position. Figure 3.3 contains an annotated representation of the pattern expected to be observed for m^6Am and Am modified nucleotides and A, C and U nucleotides. TLC experiments were performed in biological replicates of three. A ratio of A:Am and Am:m^6Am were calculated using the Quantity One software (Bio-Rad). The ratio was calculated, to remove loading variation and differences in the strength of the radiation which changes the intensity of the dots on the imaging screen.

The TLC experiments showed the Am nucleotide to be present in the brains of all organisms tested (Figure 3.3, 3.5). The levels of Am were found to vary between the different species, and the difference was statistically significant (kruskal wallis, T=13.4, N=21, p=0.037). The *O. vulgaris* RNA had the highest levels of Am of all the organisms with a mean of 32% Am relative to A (Figure 3.3). Octopus extensively modify their transcriptome through A to I, and, C to U editing (Yablonovitch et al., 2017) to diversify their proteome (Albertin et al., 2015). It is interesting that cap-adjacent modifications are also high in this species as they should not change the encoded protein. However, it may regulate the amount of protein translated.

The m^6Am nucleotide was detected in *Homo sapiens* hippocampus tissue, and *R. norvegicus*, *A. mexicanum*, *D. rerio*, whole brain and *G. gallus* hypothalamus; all of which are vertebrates (Figure 3.3, 3.5). m^6Am levels between these organisms were

observed to vary but not significantly (kruskal wallis, T= 5.869, N=12, p=0.112). *D. rerio* brain contained the highest level with a mean of 13% (Figure 3.3). There may be biological and technical reasons for *D. rerio* having the highest level of m^6Am. Some of these reasons are discussed later in this chapter. m^6Am was not found in the *O. vulgaris brain* and the *D. melanogaster* and *A. mellifera heads* (Figure 3.3).

The *G. gallus* TLC is separate in figure 3.5 as it is on poly(A) RNA rather than all capped transcripts the presence of the nucleotides can be compared to the rest of the work but the percentage of nucleotides cannot. The RNA contained both Am and m^6Am, the ratio of A:Am was 98.85:1.15 and the ratio of Am:m^6Am was 24.19:75.81.

3.2.3 Variation in biological replicates

For each organism three biological replicates were performed to examine the biological variation between different individuals. Looking at the percentage of Am relative to A, there is large variation between the biological replicates for both *R. norvegicus* and *O. vulgaris* samples. In contrast, the difference in biological replicate variation, is much smaller in *A. mexicanum*, *D. rerio*, *A. mellifera* and *D. melanogaster* (Figure 3.4.). In addition the relative amount of m^6Am to Am showed variation between the individuals of each species examined, with *D. rerio* having the greatest variation of the five organisms that contained m^6Am (Figure 3.4).

There are technical and biological reasons for the variation between animals of the same species. One, the quality of the RNA may vary between the different samples this will affect the results. Two, the levels of rRNA contamination will change between the samples and these modifications are potentially seen on the TLCs. Three, in biological terms, all the animals maybe at slightly different developmental stages. Four,

the animals may have responded differently to environmental stresses, and hence will be producing slightly different RNA transcripts with different levels of modifications.

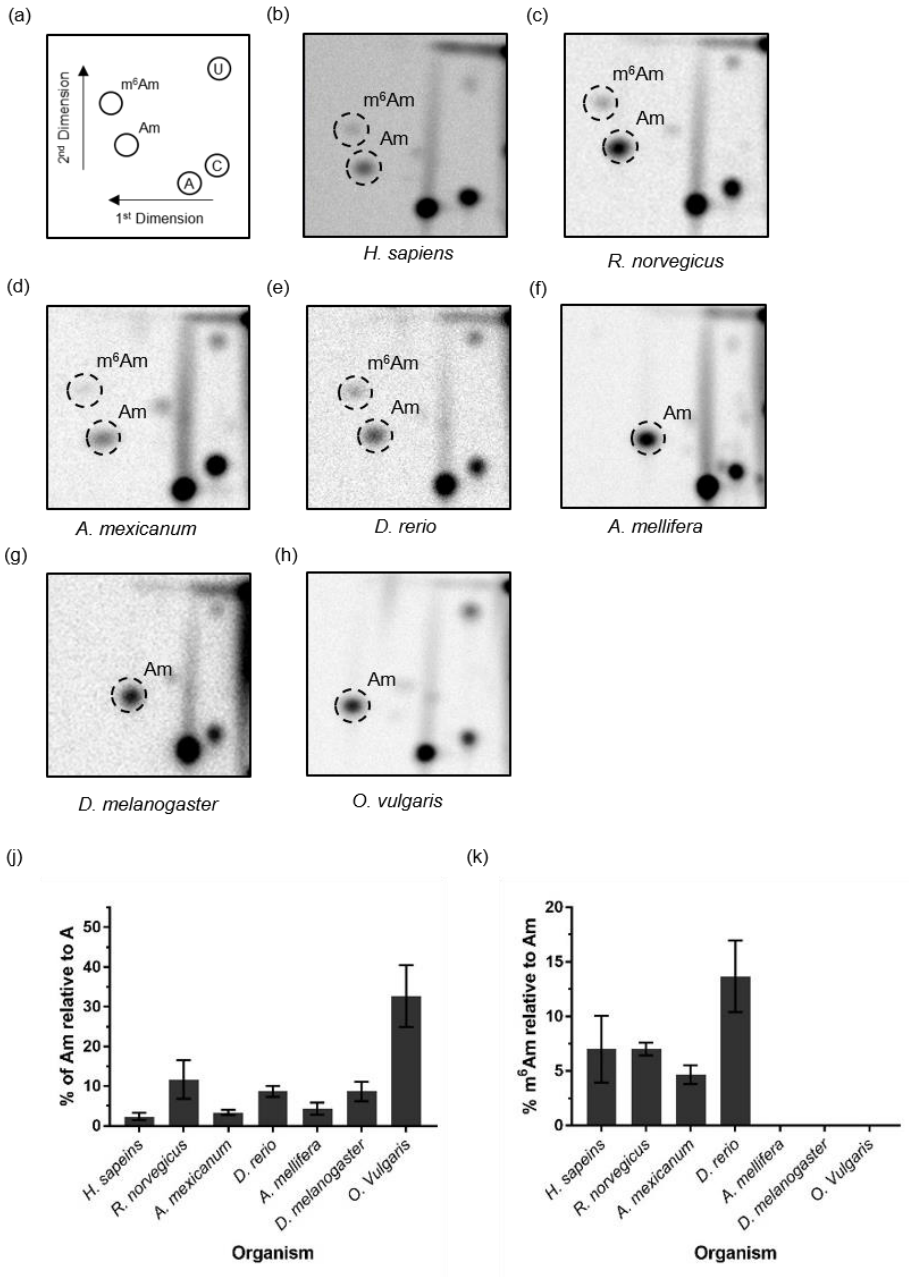


Figure 3.3. Representative TLCs showing the first nucleotide of RNA from brain tissue from different species. TLCs were done intriplicate. a) TLC with annotated with A, C and U nucleotides and Am and m^{6}Am modified nucleotides. b) TLC on RNA from the *H. sapiens* hippocampus. c) TLC on *R. norvegicus* brain RNA d) TLC on *A. mexicanum* brain RNA. e) TLC on *D. rerio* brain RNA. f) TLC on *A. mellifera* head RNA. g) TLC on D*. melanogaster* head RNA. h) TLC on *O. vulgaris* brain RNA. j) Graph showing the average ratio of Am relative to A showing mean with SEM. k) Graph showing the average ratio of m^{6}Am relative to Am showing mean with SEM.

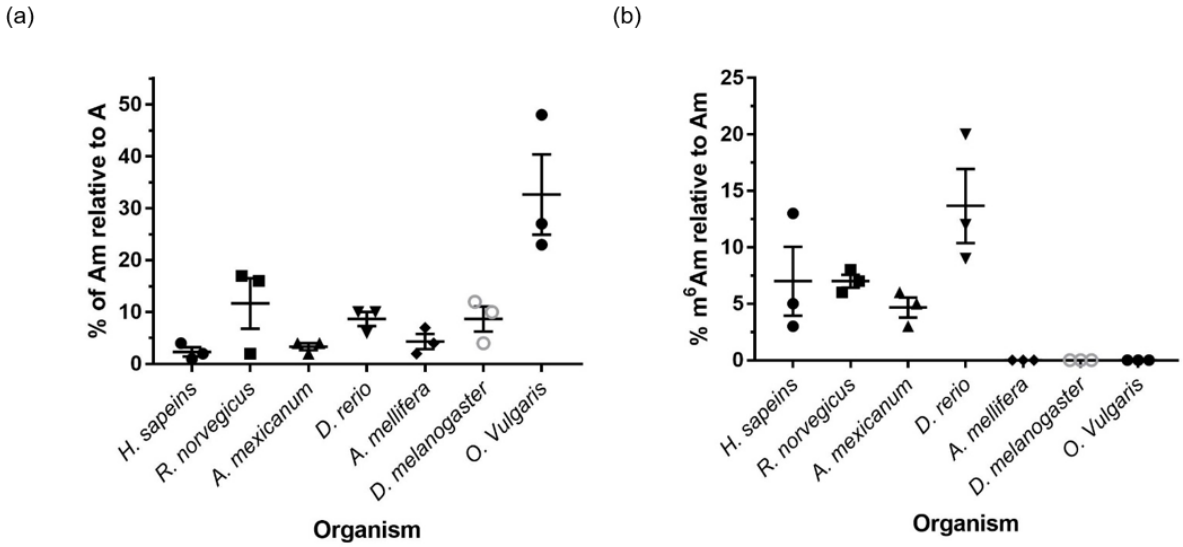


Figure 3.4. Biological variation in the different organisms. The mean and standard error of the mean is put onto the graphs along with the three data points. a) Graph shows the percentage of Am relative to A for the three biological replicates. b) Graph showing the percentage of m^6Am relative to Am for the three biological replicates.

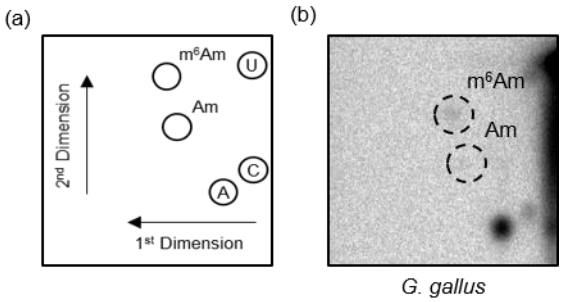


Figure 3.5. TLC showing the first nucleotide from poly(A) RNA from *G. gallus* hypothalamus. One replicate was performed. a) TLC with annotated with A, C and U nucleotides and Am and m^6Am modified nucleotides. b) Cap adjacent TLC on poly(A) RNA from *G. gallus* hypothalamus.

3.2.4 Do *D. melanogaster* snRNAs contain m^6Am?

From the work in this study the m^6Am modification was observed only in vertebrates. However the invertebrates used in this study all contain the CMTR1 cap1 methylase required for the observed Am modification and PCIF1, the methylase required for conversion of Am to the m^6Am modification (Table 3.1). The absence of detectable m^6Am in these species suggests that if the modification is present, it is on a very small population of transcripts that is beyond the TLC detection limit when all capped transcripts are assayed.

It has been reported that m^6Am is found on snRNAs (Mauer et al., 2019). snRNAs are found in the nucleus and have roles in splicing and nuclear processing of RNAs (Valadkhan and Gunawardane, 2013). The first nucleotide of U1 snRNA is adenosine in *D. melanogaster*. Furthermore it has been reported that when *Fto* is knocked out in HEK293 cells there is an increase in m^6Am on the U1 snRNA, from 0.1 m^6Am/Am to 0.8 m^6Am/Am in $Fto^{-/-}$ (Mauer et al., 2019). Thus it is a possibility that invertebrate species which contain PCIF1 but lack FTO might naturally possess m^6Am on these snRNAs. The first nucleotide of the U2 snRNA was investigated in wild type *D. melanogaster* and *cmtr1* and *cmtr2* double knockout *D. melanogaster*; to do this a modified version of site-specific cleavage and radioactive-labelling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET) was performed. The standard protocol of SCARLET analyses internal m^6A methylation. To detect the first nucleotide of the transcript a few changes were made which are detailed in chapter 2 (2.2.6) and a summary diagram is presented in Figure 3.7. To test these changes the modified SCARLET method was performed on RNA oligonucleotides, where the first nucleotides were known. The three oligonucleotides synthesised started with A, Am

and m^6Am and the oligonucleotides were heterogeneous as a result of the synthesis method used.

A labelling control was performed, where the oligonucleotides were labelled individually then mixed before P1 digestion which was compared to the oligonucleotides that were mixed before labelling and P1 digestion. In figure 3.6a the oligonucleotides labelled individually were then mixed together in ligation mix and the oligonucleotides mixed together before labelling are called A, Am, m^6Am. There was some labelling bias, A, Am, m^6Am had a ratio of 65.84:34.16 A:Am+m^6Am which was lower than the ligation mix which had a ratio of 56.88:43.12 A:Am+m^6Am. This suggests that A is preferentially labelled and that Am and m^6Am may be under represented in this work.

With the SCARLET method the transcript of interest is ligated on to a ssDNA, for the method to detect cap adjacent nucleotides a ssRNA was also tested. More RNA was ligated to the ssRNA though the efficiency of ligation for A, and the Am m^6Am mix were the same and was not different from the ligation mix that went into the experiment. The ligation mix had a ratio of 56.88:43.12 A:Am+m^6Am, the ratio using the ssRNA was 58.23:41.77 A:Am+m^6Am and the ssDNA ratio was 59.85:40.15 A:Am+m^6Am.

The modified SCARLET method on the *D. melanogaster* RNA was successful, however, there are some other nucleotides on the TLCs, and this is a result of the purification of the U1 snRNA ligated to the ssRNA. m^6Am was not found in the wildtype *D. melanogaster* snRNAs, the first nucleotide was found to be A and Am (Figure 3.6c). As expected A was the first nucleotide of the *cmtr1/cmtr2* double knockout RNA as the proteins are not present to add the modification (Figure 3.6d).

	Modification		Protein		
Organism	Am	m^6Am	CMTR1	PCIF1	FTO
H. sapiens	✓	✓	✓	✓	✓
R. norvegicus	✓	✓	✓	✓	✓
A. mexicanum	✓	✓	✓	✓	✓
D. rerio	✓	✓	✓	✓	✓
A. mellifera	✓	×	✓	✓	×
D. melanogaster	✓	×	✓	✓	×
O. vulgaris	✓	×	✓	✓	×
A. thaliana	×	×	×	×	×
S. cerevisiae	×	×	×	×	×

Table 3.1. Table detailing the presence of modified nucleotides and the associated methylases and demethylases. The table shows the presence of m^6Am and Am modifications and m^6Am associated proteins, CMTR1, PCIF1 and FTO, in the organisms looked at in the study as well as *Arabidopsis thaliana* and *Saccharomyces cerevisiae*.

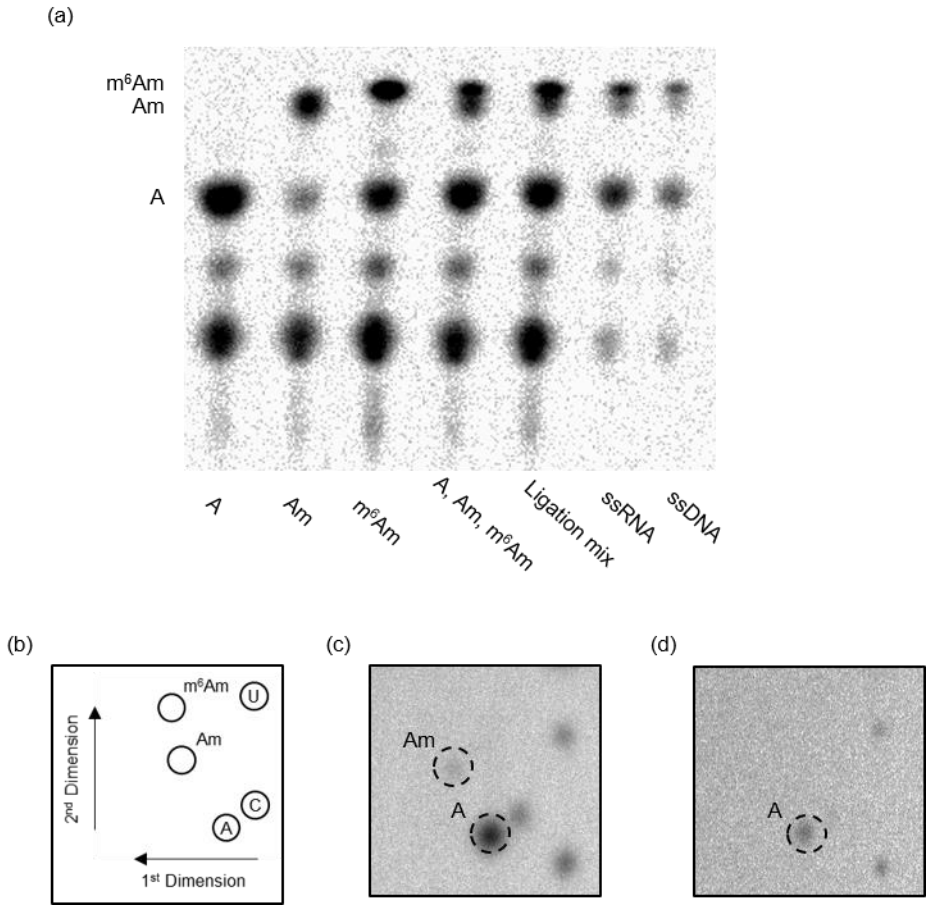
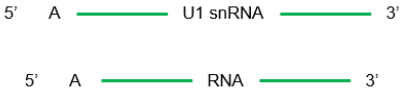
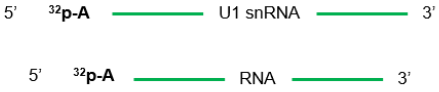


Figure 3.6. Optimisation of the SCARLET protocol and analysis of U1 snRNA. One replicate of each TLC was performed. a) Optimisation of the SCARLET protocol, oligonucleotides starting with A, Am and m^{6}Am were labelled individually to act as standards. A, Am, m^{6}Am, the oligonucleotides were mixed and labelled together to test labelling efficiency. Ligation mix, the oligonucleotides were labelled and then mixed together, these were the oligonucleotides which were put into the ligation and to act as a control for the SCARLET. The oligonucleotides were ligated to ssRNA using an RNA ligase and ssDNA using a DNA ligase. b) TLC with annotated A, C and U nucleotides and Am and m^{6}Am modified nucleotides. c) TLC showing the first nucleotide of the U1 snRNA from wildtype *D. melanogaster* RNA. d) TLC showing the first nucleotide of the U1 snRNA from *cmtr1/cmtr2* knockout *D. melanogaster* RNA.

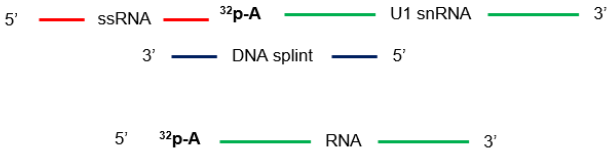
Decap



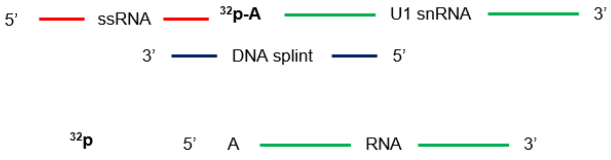
Radiolabel whole sample



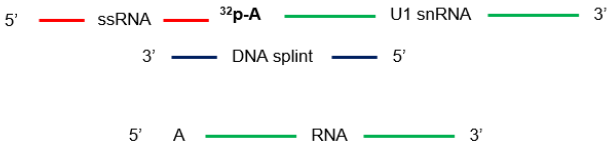
Splint ligation



Dephosphorylation



Clean up



P1 nuclease digestion

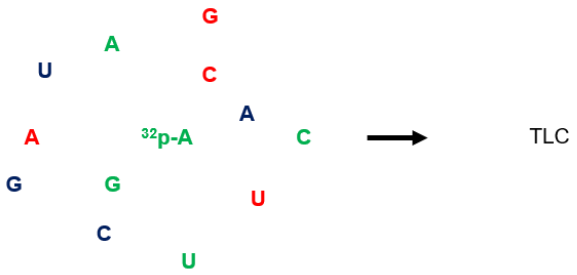


Figure 3.7. Diagram of the modified SCARLET method. First RNA enriched for the U1 snRNA was decapped and the 5′end radioactively labelled with 32p-ATP. The labelled U1 snRNA is ligated to a ssRNA oligonucleotide using a DNA splint with sequence homology to the U1 snRNA and the ssRNA. RNA that is not ligated is dephosphorylated removing the 32p from the 5′end. The RNA is then put through a spin column to remove the radiation before P1 nuclease digestion. The single nucleotides are run on a TLC, only the first nucleotide of the U1 snRNA should be labelled and will be detected in the final image.

3.2.5 Identifying orthologues of m^6Am associated proteins by evolutionary conservation of protein sequences

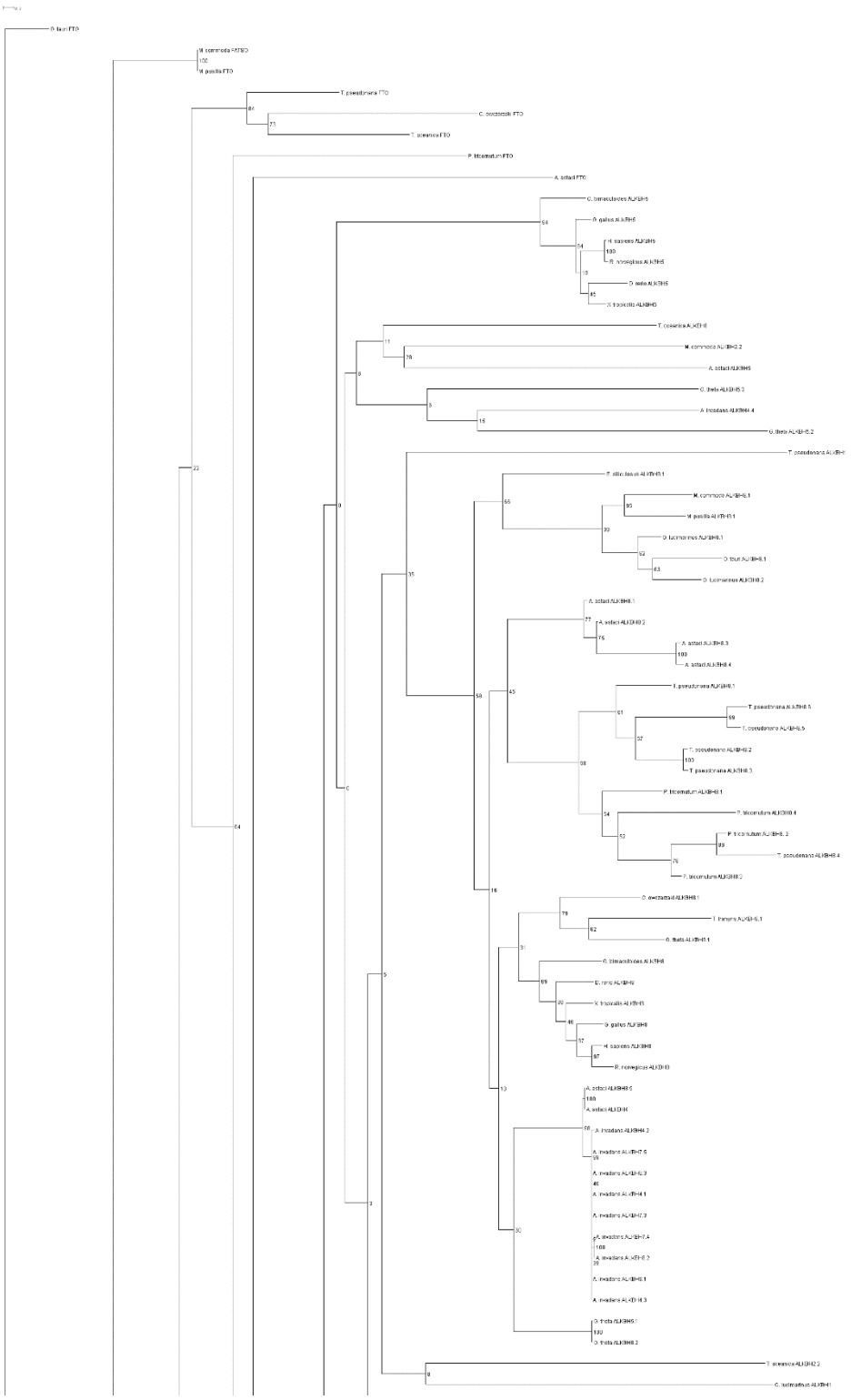
From the findings presented above and published data, m^6Am and FTO are only found in vertebrates (Table 3.1). A few papers have reported that FTO can be found in marine algae (Robbens et al., 2008, Fredriksson et al., 2008, Iyer et al., 2016). This raises many questions such as: do algae species that contain the FTO protein contain the m^6Am modification? How many species of algae have m^6Am associated proteins? How closely related are the algae proteins to human ones? Why do algae have m^6Am associated proteins along with vertebrates?

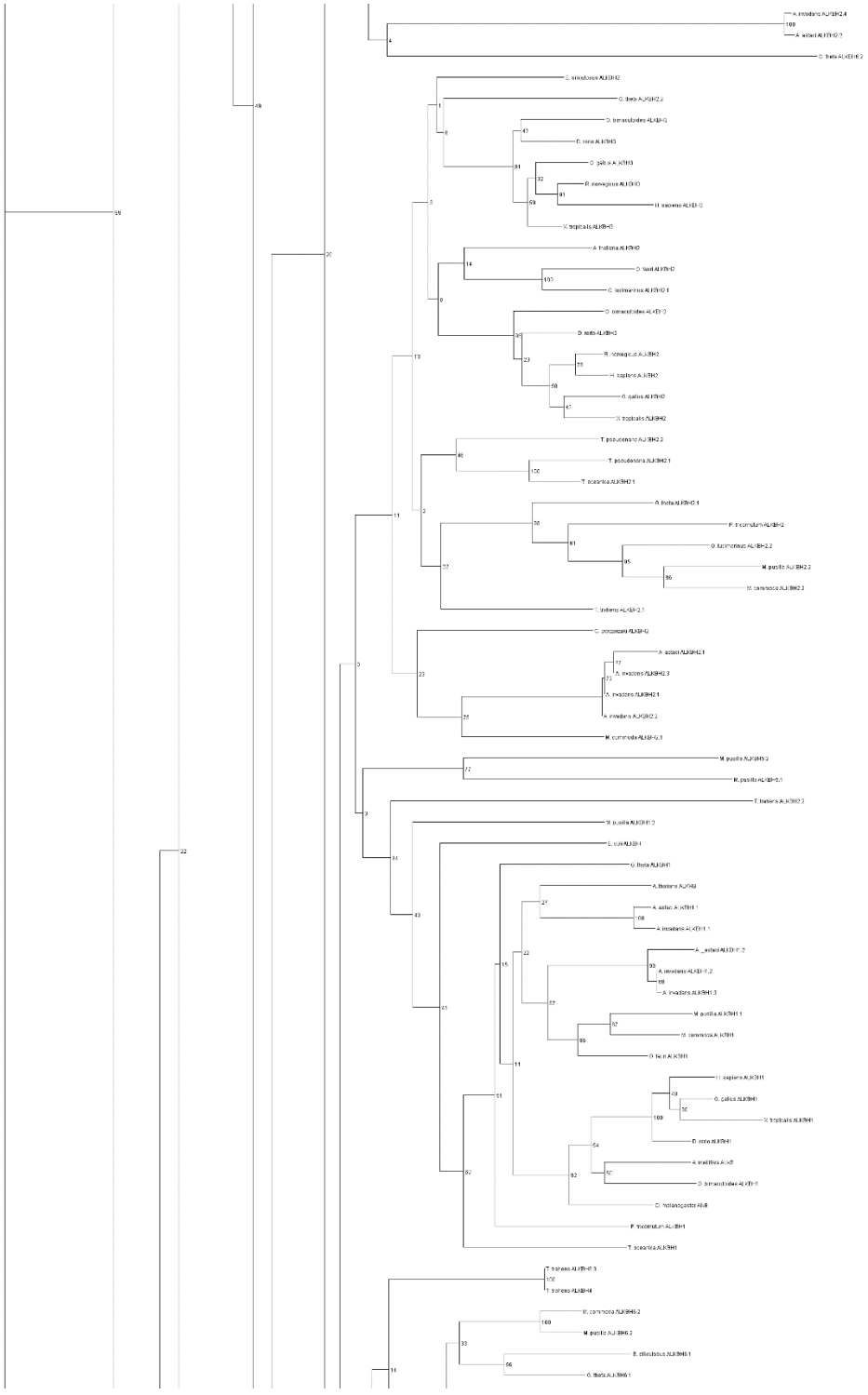
To understand how closely related the FTO protein in algae is to human FTO, all 9 human ALKB family members including FTO were searched for using protein blast to identify regions of similarity in sequenced algae genomes. The protein sequences of the organisms previously examined by TLC, were compared with the exception of *A. mexicanum* which was replaced by *X. tropicalis* as the genome is better annotated. In the FTO phylogenetic tree there are 14 clades (Figure 3.8). Many of the algae proteins are in clades with the animal proteins. The animal FTO proteins are in a clade on their own and appear not to be closely related to the algae proteins. The *E. siliculosus* FTO protein blast result appears to be more closely related to ALKBH4, 6 and 7 (Figure 3.8).

CMTR1 and CMTR2 were searched for using protein blast in the algae genomes, where FTO was found. In the majority of cases, both query protein sequences resulted in the same algae sequences being returned in the blast search, thus, in the phylogenetic tree the proteins are labelled CMTR without a suffix number. The phylogenetic tree has three clades and one protein on its own. The three clades are CMTR1, CMTR2 and algae proteins (Figure 3.9). There are a mix of algae proteins in

the clades with the animal CMTR proteins, however the animal and algae proteins are in different sub clades. This shows the cap methylase protein is less conserved between algae and animal sequences in the tree. There are a few exceptions *C. owczarzaki* in the same CMTR1 clade as the human proteins and *G. theta* in the same clade as the human CMTR2 protein.

There are seven clades in the PCIF1 phylogenetic tree (Figure 3.10). It has been reported that to methylate adenosine at the N^6 position a SAM binding amino acid motif is required [DNSH]PP[YFW] (Iyer et al., 2016). 32 out of the 50 proteins from the blast results for PCIF1 contained the SAM binding motif. On the phylogenetic tree the protein names in bold contain this SAM binding motif. The proteins with and without the motif are majoritively in separate clades. In the clade with the animal proteins there are no algae species. The tree of similarity shows that the algae proteins are more conserved with each other than the animal sequences. This may mean the function of the algae proteins are different from the animal ones, or may reflect divergence over time.


B. taurus FTO
M. commoda FTO
100
M. pasilla FTO
T. pseudonana FTO
84
C. owczarzaki FTO
77
T. oceanica FTO
P. tricornutum FTO
A. sativa FTO
C. bimaculoides ALKBH5
64
G. gallus ALKBH5
54
H. sapiens ALKBH5
100
R. norvegicus ALKBH5
18
D. rerio ALKBH5
45
X. tropicalis ALKBH5
T. oceanica ALKBH1
11
M. commoda ALKBH2
78
A. sativa ALKBH5
8
G. theta ALKBH5 3
5
A. thaliana ALKBH4 4
16
G. theta ALKBH5 2
T. pseudonana ALKBH1
22
P. tricornutum ALKBH1 1
M. commoda ALKBH5 1
95
M. pasilla ALKBH1 1
65
99
D. melanogaster ALKBH5 1
92
D. rerio ALKBH5 1
83
D. melanogaster ALKBH5 2
0
35
A. sativa ALKBH8 1
77
A. sativa ALKBH8 2
76
A. sativa ALKBH8 3
100
A. sativa ALKBH8 4
T. pseudonana ALKBH8 1
59
45
91
T. pseudonana ALKBH8 5
99
T. pseudonana ALKBH8 6
57
T. pseudonana ALKBH8 2
100
T. pseudonana ALKBH8 3
98
P. tricornutum ALKBH8 1
54
P. tricornutum ALKBH8 4
52
P. tricornutum ALKBH8 2
99
T. pseudonana ALKBH8 4
76
P. tricornutum ALKBH8 3
64
18
D. swezeyi ALKBH8 1
79
T. thermera ALKBH5 1
82
G. theta ALKBH5 1
31
C. bimaculoides ALKBH8
89
D. rerio ALKBH8
82
X. tropicalis ALKBH8
46
G. gallus ALKBH8
57
H. sapiens ALKBH8
67
R. norvegicus ALKBH8
5
10
A. sativa ALKBH8 5
100
A. sativa ALKBH8
66
A. thaliana ALKBH4 3
A. thaliana ALKBH7 5
58
A. thaliana ALKBH7 6
46
A. thaliana ALKBH4 1
A. thaliana ALKBH7 3
30
A. thaliana ALKBH7 4
100
A. thaliana ALKBH7 2
99
A. thaliana ALKBH4 2
A. thaliana ALKBH4 3
G. theta ALKBH5 1
100
G. theta ALKBH5 2
T. oceanica ALKBH2 2
9
C. bimaculoides ALKBH1


A. invadans ALDH2.4
A. salaxi ALDH2.1
G. theta ALDH2.2
D. tamaulipas ALDH2
D. rerio ALDH2
G. gallus ALDH2
R. norvegicus ALDH2
H. sapiens ALDH2
X. tropicalis ALDH2
T. pseudaethalis ALDH2.2
T. pseudaethalis ALDH2.1
T. bosavica ALDH2.1
G. theta ALDH2.1
P. tricornutum ALDH2
O. lucimarinus ALDH2.2
M. pusilla ALDH2.2
M. commoda ALDH2.2
T. bohlinii ALDH2.1
C. owczarzaki ALDH2
A. salaxi ALDH2.1
A. invadans ALDH2.3
A. invadans ALDH2.1
A. invadans ALDH2.2
M. commoda ALDH2.1
M. pusilla ALDH5.2
M. pusilla ALDH5.1
T. bohlinii ALDH2.3
M. pusilla ALDH1.2
L. con ALDH1
G. theta ALDH1
A. thaliana ALDH1
A. salaxi ALDH1.1
A. invadans ALDH1.1
A. salaxi ALDH1.2
A. invadans ALDH1.2
A. invadans ALDH1.3
M. pusilla ALDH1.1
M. commoda ALDH1
O. lucimarinus ALDH1
H. sapiens ALDH1
G. gallus ALDH1
X. tropicalis ALDH1
D. rerio ALDH1
A. thaliana ALDH
D. tamaulipas ALDH
O. volcani ALDH
P. tricornutum ALDH1
T. bosavica ALDH1
T. bohlinii ALDH4.3
T. bohlinii ALDH4
M. commoda ALDH6.2
M. pusilla ALDH6.2
E. siliculosus ALDH6.1
G. theta ALDH6.1

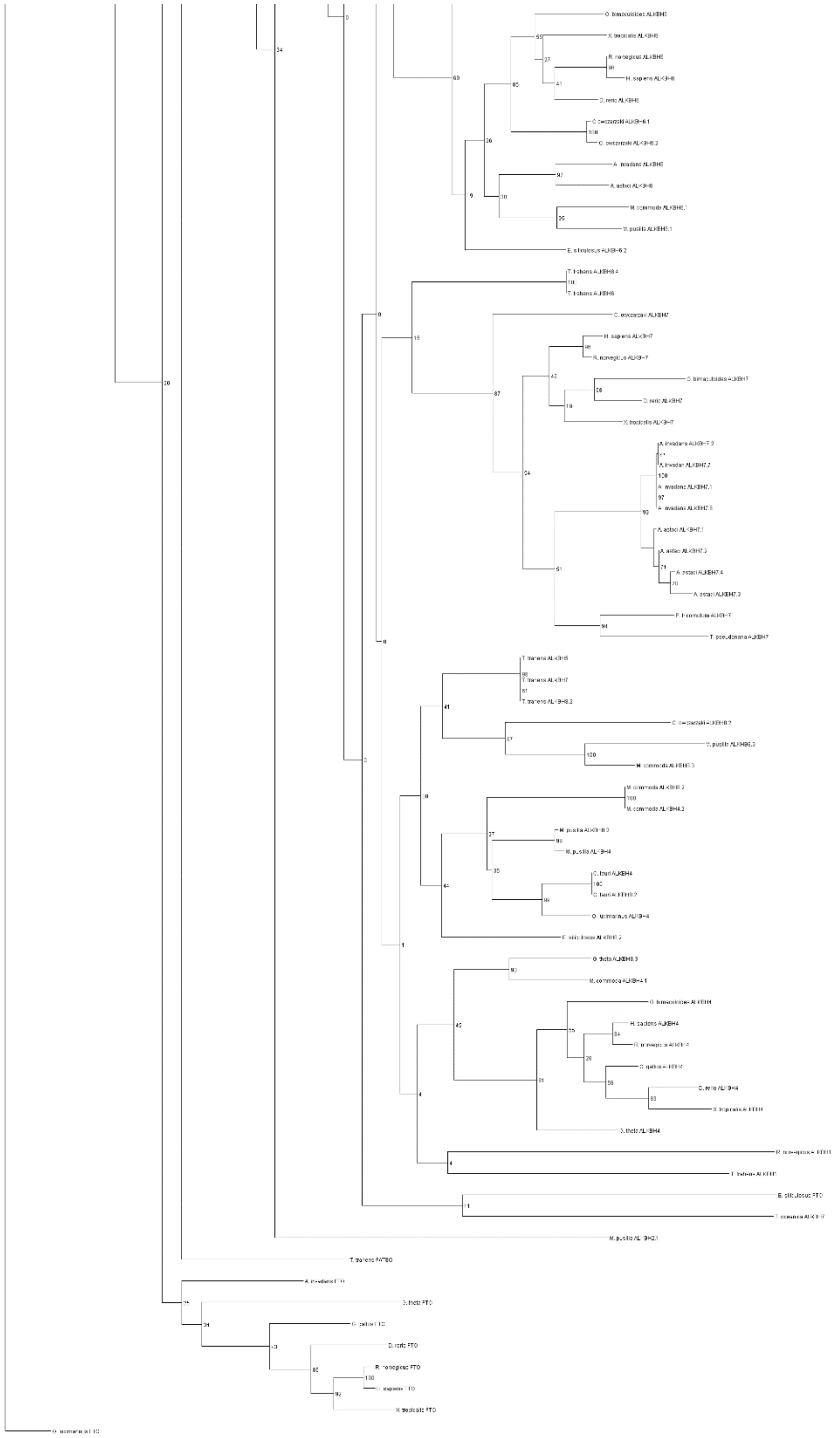


Figure 3.8. FTO tree of similarity for ALKB family members from animals and algae protein sequences. ALKBH1-8 and FTO protein sequences from the animals used in this study were collected and the human sequences were blasted against the algae proteomes. Accession numbers are presented in the appendix.

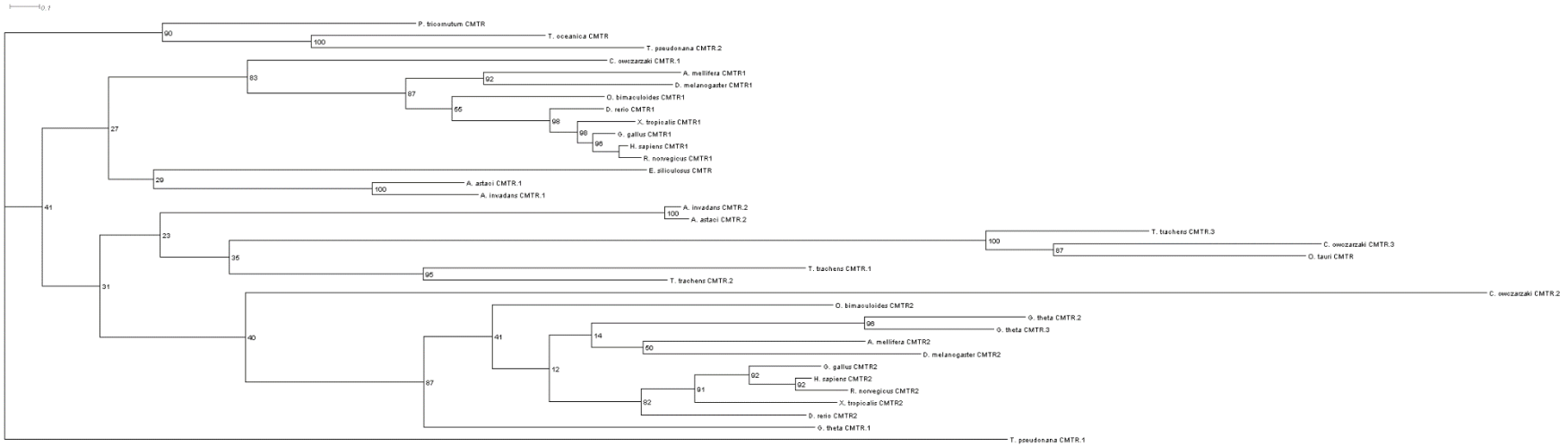


Figure 3.9. Tree of similarity for CMTR1 and CMRT2 orthologues. The tree contains the protein sequences from the animals used in this study, plus sequences from the 11 algae species that contained an FTO orthologue. Accession numbers are presented in the appendix.

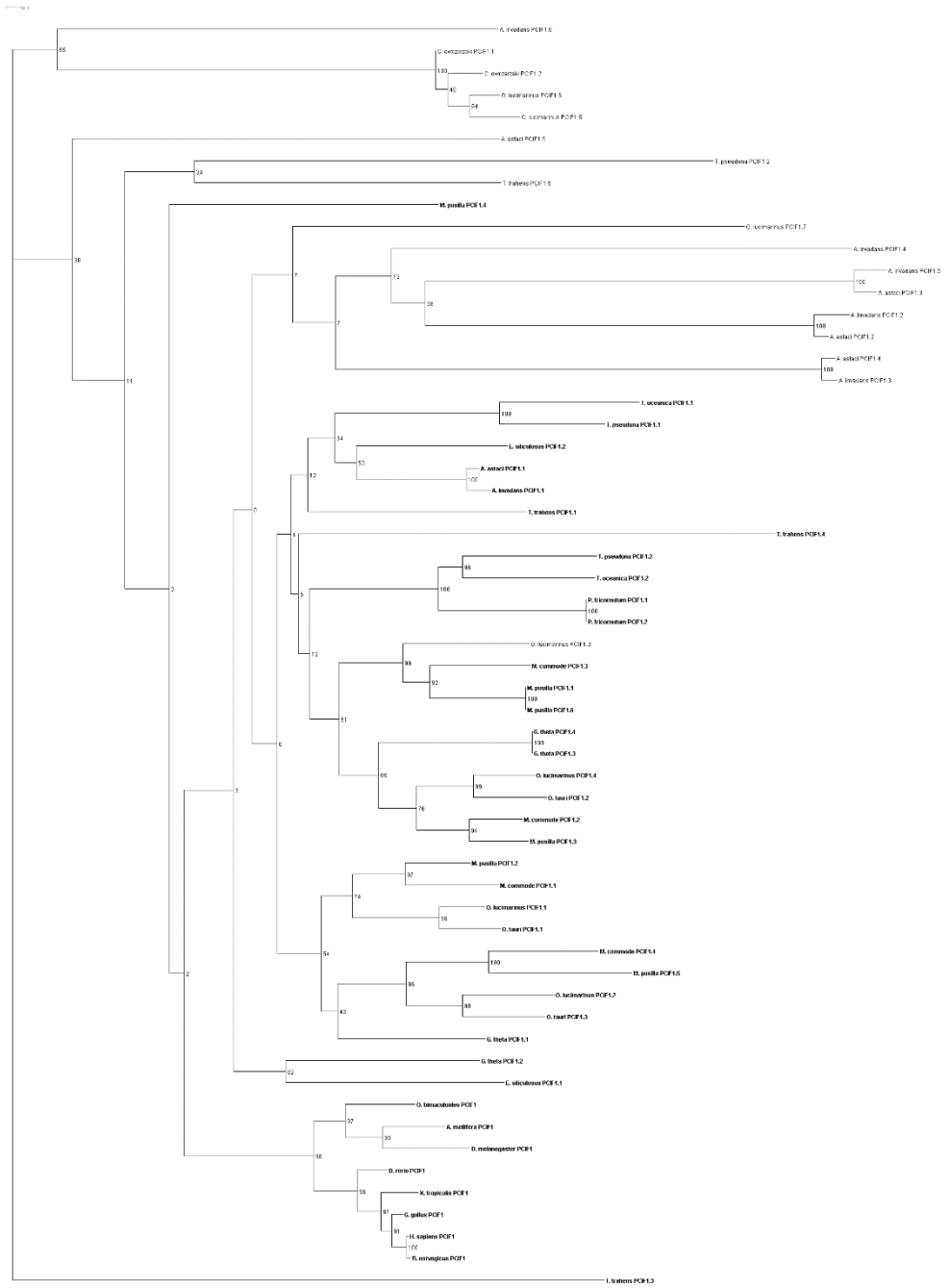


Figure 3.10. Tree of similarity for PCIF1 orthologues. The tree contains the protein sequences from the animal species used in this study and blast results of the human PCIF1 in the 11 marine algae species that contain a FTO orthologue. Names in bold denote proteins that contain the SAM binding motif. Accession numbers are presented in the appendix.

3.2.6 Searching for m^6Am in *E. siliculosus*

E. siliculosus is a multicellular brown marine algae. It has potential orthologues of FTO CMTR1.and PCIF1 all of which are required for m^6Am methylation. Thus this brown algae has orthologues of the three proteins known to be associated with m^6Am and the orthologues are closely related to species that contain m^6Am.

TLCs were performed on *E. siliculosus* poly(A) RNA to assess if m^6Am was present. Poly(A) RNA was picked because there is a higher proportion of m^6Am on poly(A) RNA compared to all capped RNAs. The poly(A) RNA was tested to see if there were any m^6Am on uncapped or fragmented RNA. No modified nucleotides could be detected in the poly(A) RNA without decapping the RNA first (Figure 3.10b). Next the first nucleotide of the poly(A) RNA was analysed, Am was found in all three samples tested, however it was at a low level with an average of 0.85% (Figure 3.11e). As with all of the organism tested in this chapter there is variation between the replicates. This may be partially attributed to technical and biological variation (Figure 3.11g). In one of the samples there is a dot which could potentially be m^6Am (Figure 3.11c). The dot is a more elongated than would be expected so more work is need to investigate this. The m^6Am was 29.44% relative to Am in the sample (Figure 3.11f).

m^6A modification was assayed for as it is known to be present in the majority of eukaryotes (Wang and Zhao, 2016). In the one repeat performed m^6A was not found in *E. siliculosus*. To further investigate this trees of similarity were produced to see if *E. siliculosus* had the proteins that are responsible for the m^6A modification. The protein sequences for METTL3, METTL14 and WTAP were retrieved for the organisms used in this study and the human and *Arabidopsis* proteins were blasted against the algae proteomes from the previous trees of similarity. The *E. siliculosus* genome

contains all three proteins, however these proteins have diverged from those in animal species.

The evolutionary tree with the METTL3 orthologues contains two clades and two proteins on their own. The clades can be divided into animal proteins and algae proteins. The *E. siliculosus* protein was found in a clade with other algae proteins (Figure 3.12). There are seven clades and two proteins on their own in the METTL14 evolutionary tree. 9 proteins from *E. siliculosus* shared some similarity to METTL14/ MTB when the human and *Arabidopsis* proteins were compared to the *E. siliculosus* proteome. None of the proteins are in the same clade as METTL14 or MTB. Potentially the proteins returned in the blast search have a m^6A methylase function. However there are 33 proteins in the METTL family so these proteins may methylate a different substrate (Figure 3.13). In the WTAP evolutionary tree there are two clades and three proteins on their own. The clade which contains the animal sequences also contains the *E. siliculosus* proteins (Figure 3.14). This suggests that the WTAP protein is not highly conserved in most marine algae.

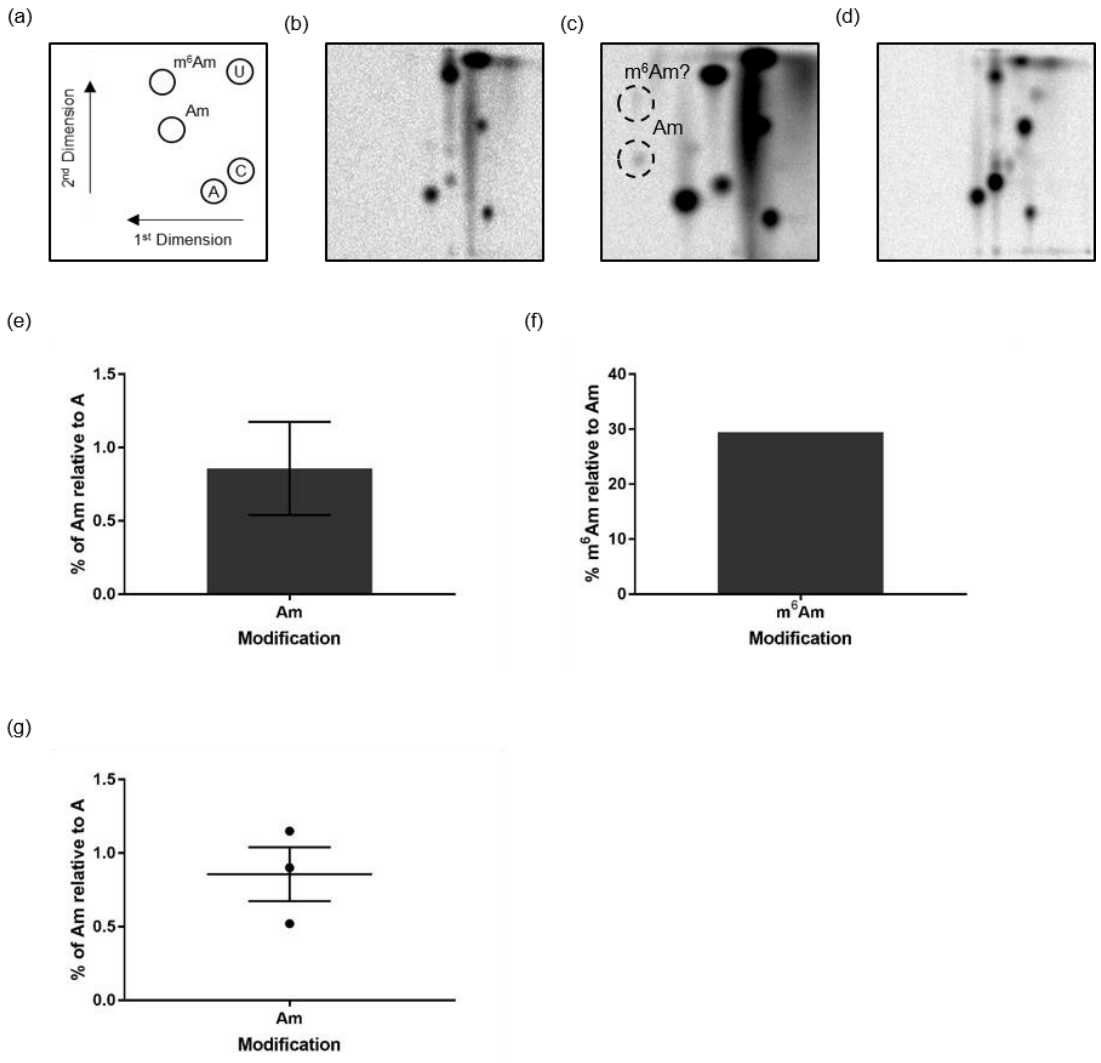


Figure 3.11. Representative TLCs detecting modified nucleotides in *E. siliculosus*. TLCs performed on three replicates. a) TLC with annotated with A, C and U nucleotides and Am and m^{6}Am modified nucleotides. b) TLC examining 5′ uncapped ends in poly(A) RNA. c) TLC detecting m^{6}Am in poly(A) RNA. d) TLC detecting m^{6}A in poly(A) RNA. e) Graph showing the average ratio of Am relative to A with mean and SEM. f) Graph showing the ratio of m^{6}Am relative to Am in the one sample that was found to contain m^{6}Am. g) Graph showing the percentage of Am relative to A for the three biological replicates, the mean and standard error of the mean are shown on the graph.

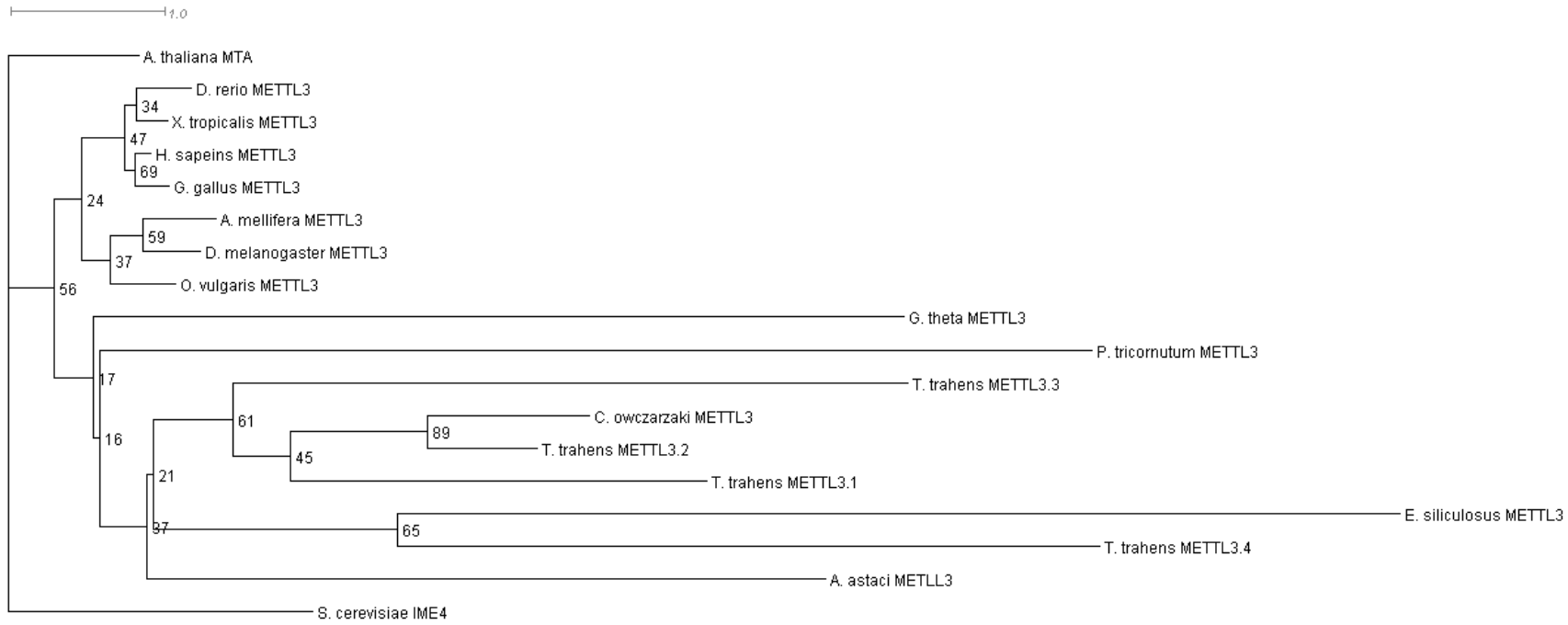


Figure 3.12. Tree of similarity for METTL3 protein orthologues. The tree contains the protein sequences from the animal used in this study, plus sequences from the 11 algae species that contained an FTO orthologue. Accession numbers are presented in the appendix.

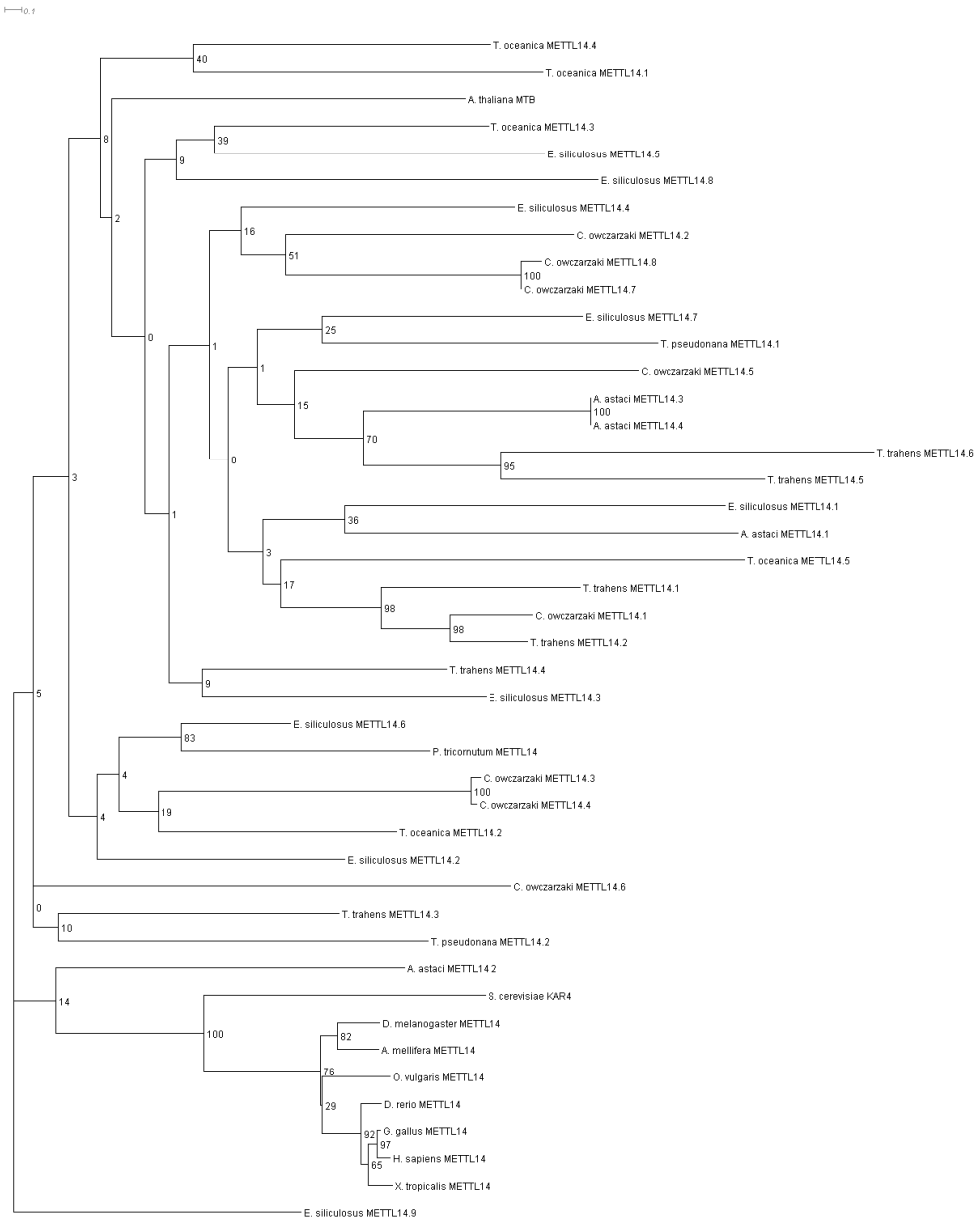


Figure 3.13. Tree of similarity for METTL14 protein orthologues. The tree contains the protein sequences from the animal used in this study, plus sequences from the 11 algae species that contained an FTO orthologue. Accession numbers are presented in the appendix.

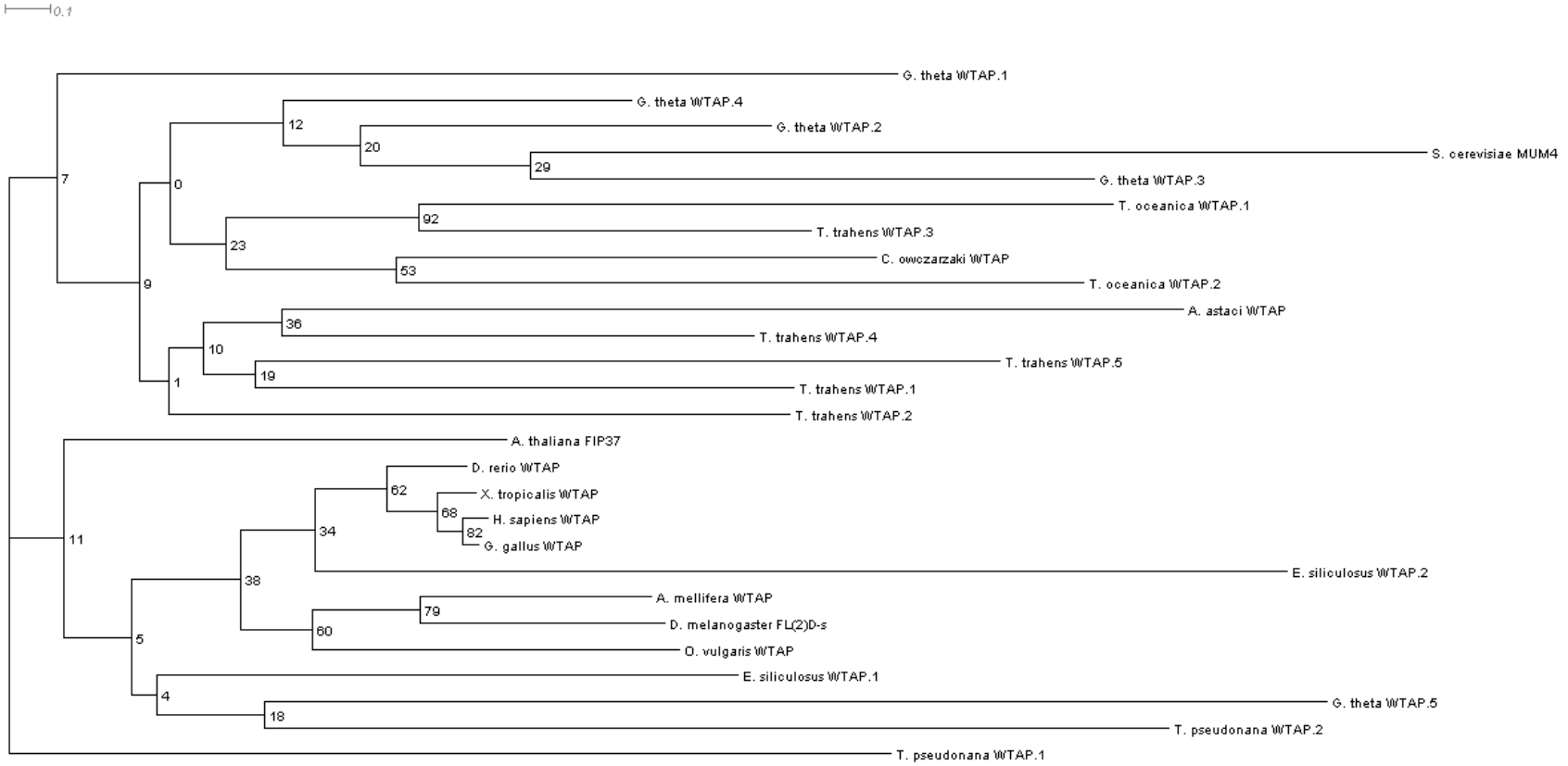

0.1
G. theta WTAP.1
G. theta WTAP.4
G. theta WTAP.2
S. cerevisiae MUM4
G. theta WTAP.3
T. oceanica WTAP.1
T. trahens WTAP.3
C. owczarzaki WTAP
T. oceanica WTAP.2
A. astaci WTAP
T. trahens WTAP.4
T. trahens WTAP.5
T. trahens WTAP.1
T. trahens WTAP.2
A. thaliana FIP37
D. rerio WTAP
X. tropicalis WTAP
H. sapiens WTAP
G. gallus WTAP
E. siliculosus WTAP.2
A. mellifera WTAP
D. melanogaster FL(2)D-s
O. vulgaris WTAP
E. siliculosus WTAP.1
G. theta WTAP.5
T. pseudonana WTAP.2
T. pseudonana WTAP.1
12
20
29
7
0
92
23
53
9
36
10
19
1
11
62
68
82
34
38
79
60
5
4
18

Figure 3.14. Tree of similarity for WTAP protein orthologues. The tree contains the protein sequences from the animal used in this study, plus the sequences from the 11 algae species that contained an FTO orthologue. Accession numbers are presented in the appendix

3.3 Discussion

In this study, rRNA removal approaches were tested and optimised, m^6Am was detected and quantified to determine evolutionary conservation using brain/head tissue from a diverse group of organisms. In addition, the evolutionary conservation of potential proteins involved in m^6Am methylation was elucidated and the modification was potentially detected in marine algae.

3.3.1 Analysis of rRNA removal

The poly(A) purification and rRNA depletion were the most efficient methods at removing rRNA from the samples. However, there are a number of disadvantages to these methods. For example, the m^6Am modification is at the 5′ end of the transcript, and poly(A) purification will pull out poly(A) tails so any degradation would result in loss of 5′ ends. Furthermore poly(A) purification does not work as well on *Drosophila* RNA which results in the need to use multiple methods to compare across species. The rRNA depletion kit is expensive and hybridising oligonucleotides have not been optimised for all species in this study, so this was not a feasible method for multiple samples. Terminator exonuclease was the method which was chosen. It removes uncapped RNA, like rRNAs and fragmented RNA and can be used on all of the samples. It also allowed the identification of m^6Am on a wider range of species of RNA increasing the chance of novel discovery.

It was apparent in the Agilent graphs that there is still some rRNA contaminating the samples after the terminator treatment. Legrand et al., showed that after two rounds of poly(A) selection, one snRNA depletion and two rRNA depletions, 7% of the deep sequencing RNA-seq reads were from rRNA (Legrand et al., 2017). Human rRNA

contains approximately 100 2′-O-methylation and pseudourodine sites, which could conceivably be detected. However, we see limited evidence of pseudourodine and only low levels of 2′-O-methylation from C and U (which can also be found at the cap1 position).

3.3.2 TLC as a method for modified nucleotide detection

The TLCs in figure 3.2 are not comparable to published studies. The TLCs in the paper by Kruse et al., 2011 are on poly(A) purified RNA from mouse tissue, the TLC on mouse brain is not similar to the poly(A) terminator treated RNA TLC in this study. For example Kruse et al,. found a ratio of 15:1 m^6Am:Am ratio (Kruse et al., 2011); in this study a ratio of 4.3:1 was found in rat brain treated in a similar manner. Figure 3.2 includes a TLC on the RNA left over from poly(A) purification (poly(A) minus) where 5' ends which had an OH were phosphorylated and the RNA terminator treated to removed uncapped transcripts. This TLC looks more similar to the TLCs using terminator exonuclease rat brain RNA found in this study than TLCs using mRNA..

It was shown that mass spectrometry can be used for RNA sequencing in 2006 (Thomas and Akoulitchev, 2006) and the method is increasingly being used to detect RNA modifications (Akichika et al., 2018, Sendinc et al., 2019). Sendinc et al., 2019 used mass spectrometry to examine different species for the m^6Am modification. They found it in *H. sapiens* MEL624 melanoma cell line*, M. musculus* C2C12 mice myoblast cell line, and *D. rerio* whole fish but not in *D. melanogaster* KC cell line, *C. elegans* whole worm and *S. pombe* ED666 (Sendinc et al., 2018). The work in this study supports this. In the tested organisms, m^6Am is only found in vertebrates. On MEL264 cells it has been quantified that the level of m^6Am compared to A and found 0.04% (Sendinc et al., 2018). When comparing this to the TLC on poly(A) RNA in figure 3.2,

the TLC in our study had 0.71% m^6Am when compared to A which is considerably higher than reported by the other authors. However the work here is on primary tissue which is considerably different from cell lines.

Akichika et al., 2018, who first identified the m^6Am methylase, investigated levels of modified nucleotides in RNA from HEK293T cells. They calculated that at the cap1 position there was no A, 8% Am and 92% m^6Am using mass spectrometry (Akichika et al., 2018). When analysing the TLC data from figure 3.2b to look at the percentages A, Am, and m^6Am, 99.1% was A, 0.18% Am and 0.71% m^6Am. It seems surprising that there would be no unmodified A in the HEK293T cells when every other published paper and the work in this study has always found adenosine as the first nucleotide in a high proportion of the transcripts in the samples tested.

In the three papers cited above and the research presented here, varying levels of cap-adjacent modified nucleotides have been reported. It is hard to compare the findings from these studies as they all use different methods to extract the RNA and purify the mRNA. Moreover, comparison between different organisms, different tissues and cell lines might not be biologically informative as, m^6Am might have a potentially different role in the regulation of mRNA in each of them and therefore might occur at different levels.

3.3.3 Evolution of m^6Am associated proteins

Through phylogenetic tree analysis proteins similar to FTO, PCIF1 and CMTR1 were found in certain marine algae. The number of algae where FTO has been found has been expanded from eight species (Robbens et al., 2008) to eleven. It was also discovered that the marine algae *E. siliculosus* contains the cap adjacent modification Am and potentially m^6Am. It is interesting why these organism have these proteins and

the modifications, as the next organisms on the evolutionary tree which contain them are vertebrates. It is not known whether they have gained the proteins through convergent evolution or horizontal transfer or if these proteins were present in a basal ancestor and have been lost in some early branching lineages. There is evidence that adenine methyl- transferases have been independently acquired by eukaryotes at least 13 times from prokaryotic restriction-modification and counter-restriction systems for DNA methylation of adenosine (Iyer et al., 2016). *E. siliculosus* has a very close relationship with bacteria, certain bacterial strains are required for its filamentous morphology and reproduction (Tapia et al., 2016). If all three genes were transferred as a block at the same time, then they must have subsequently relocated within the genome as the FTO, PCIF1, and CMTR1 genes are on different chromosomes.

In this work it appears that *E. siliculosus* does not have m^6A the internal adenosine methylation. However, the assay was only carried out on one biological repeat. *E. siliculosus* does not appear to have close orthologues to METTL3 or WTAP. It is unknown why, unlike in other eukaryotes m^6A is not present while Am and potentially m^6Am are present.

3.3.4 Proteins required for the m^6Am modification

CMTR1 is the cap1 methylase and PCIF1 is the N^6 methylase of Am. These proteins are required for the m^6Am modification. However m^6Am is not present in all organisms tested which have these two genes. The modification can be found in vertebrates that were tested but not invertebrates. In this study, FTO was only found in the vertebrates and marine brown algae, and these organisms are the ones with m^6Am. Is it possible that *in vivo* FTO as well as being the eraser may also have a role in writing the m^6Am modification, potentially interacting with PCIF1? FTO knock outs have been used in a

number of sequencing methods with varying results on the level of m^6Am (Mauer et al., 2016, Mauer and Jaffrey, 2018, Mauer et al., 2019).

In *in vitro* studies using PCIF1 and oligonucleotides with Am after the guanine cap, PCIF1 can methylate Am to m^6Am (Akichika et al., 2018). The motif responsible for SAM binding is [DNSH]PP[YFW] (Iyer et al., 2016), in vertebrates it is NPPF, which conforms to this motif. In the invertebrates tested, *D. melanogaster* has NPPH, *O. vulgaris* has NPPF and *A. mellifera* has NPPY, all of these motif should allow for SAM binding and thus potential N^6 methylation of adenosine. Possibly the structure of the protein is not allowing methylation to occur. To understand the function of PCIF1 and to understand why m^6Am is not present in invertebrates, a number of follow up experiments could be performed.

The first questions would address the ability of PCIF1 from *D. melanogaster* and *O. vulgaris* to bind to mRNA and to methylate Am *in vitro*. *D. melanogaster* would be picked as it has a different SAM binding motif to vertebrates and *O. vulgaris* as it has the same motif as vertebrates. Testing *O. vulgaris* and *D. melanogaster* PCIF1 in *in vitro* methylase assays would identify if a protein was blocking the methylation *in vivo* and if PCIF1 was able to bind and methylate the RNA. If methylation did not occur *in vitro* it would be interesting to see if the PCIF1 can bind to the guanine cap or cap1 modification using a gel shift assay. Another experiment that could be performed would be to introduce the *D. melanogaster* and *O. vulgaris Pcif1* genes into a *Pcif1* knockout human cell line. This would test if there were other proteins in vertebrates enabling methylation. It may be that PCIF1 can methylate *in vitro* but there are activator and scaffold proteins that act *in vivo* to increase the efficiency of the enzyme. To investigate this a protein pull down using PCIF1 would identify interacting proteins.

RIP-seq could be performed to understand what RNAs are being targeted and methylated by PCIF1. RIP-seq cross links proteins and RNA *in vivo* then RNA bound to PCIF1 would be immunoprecipitated and sequenced. It could be possible that if m^6Am methylation is taking place in invertebrates, its function is to target RNAs for rapid decay and so is not detectable on the TLCs.

3.3.5 The role of m^6Am

m^6Am is not specific to the brain. It can be found in many tissues. So its function is likely a more general post-transcriptional regulatory mechanism, although this regulation may play an important role in the function of each tissue. It has been shown to be found at its highest levels in the brain (15:1 m^6Am:Am), closely followed by testis (10:1) and at lower levels in the kidney (2.4:1) and liver (2:1) (Kruse et al., 2011).

RNA modifications including Am are involved in differentiating self RNA and non-self RNA. Animal RNA viruses use these modified nucleotides and cap structures to evade the immune response (Bokar, 2005, Ramanathan et al., 2016). RIG1 binds to non-self RNA and stimulates the interferon β response (Schuberth-Wagner et al., 2015), the presence of the Am cap1 prevents RIG1 binding, potentially m^6Am could also prevent the binding of RIG1.

The guanine cap structure protects RNA from degradation, recruits complexes involved in RNA processing, facilitates export of RNA from the nucleus and is involved in translation initiation (Galloway and Cowling, 2019). The proteins involved in these processes may be able to differentiate between GpppA, GpppAm and $Gppm^6Am$ adding another level of regulation to the system.

The half-life of certain RNAs in vertebrates are higher than in invertebrates. In *D. melanogaster* embryos, the mean half-life is 80 minutes and for neural specific

transcripts it is 76 minutes (Burow et al., 2015). In mammalian cells the mean half-life is 10 hours (Schwanhäusser et al., 2011). In this larger window more regulation may be needed to protect the RNA from degradation, and to allow the cell to respond to environmental changing the post transcriptional modifications found on RNA.

3.3.6 Potential role of modified nucleotides in the brain

Local translation is required for synaptic plasticity (Kang and Schuman, 1996), which strengthens synaptic connections and facilitates communication among neurones promoting memory formation (Richter, 2015). A single neuron can have more than 1000 synaptic connections and each synapse needs to be individually maintained and regulated in a specific manor in response to the physiological input each one receives (Iacoangeli and Tiedge, 2013). Potentially RNA modifications like Am may be involved in RNA stability at the synapse or directing RNA to the synapse ready for translation when plasticity occurs. Memory formation is of key importance in vertebrates and invertebrates alike. So m^6Am might provide an additional level of regulation to fine tune memory formation or it might be involved in different forms of synaptic plasticity which may contribute to different forms of memory.

Chapter 4- m^6Am in the Human Brain

4.1 Preface

Little research has examined levels of RNA methylation in human brain tissue and as yet there are no published studies using brain tissue from individuals with neurodegenerative diseases. Current studies using animal models or cell cultures have focused on internal m^6A methylation in processes such as neurogenesis and memory formation (Yoon et al., 2017, Widagdo et al., 2016). It is now recognised that m^6A is important for many biological processes including these. In contrast, little is known about RNA regulation by m^6Am and its potential to contribute to disease mechanisms. The work in this chapter focuses on the global levels of cap adjacent methylation, Am and m^6Am in human hippocampus. Samples from non-affected individuals and patients with Parkinson's disease and Lewy body dementia, two common neurodegenerative diseases, were analysed to investigate changes in total Am and m^6Am methylation.

There are several methods to detect and sequence transcripts which have m^6Am such as mi-CLIP and m^6Am-exo-seq. mi-CLIP uses antibody crosslinking at m^6A and m^6Am sites which causes AC to AT conversions and truncations. The conversions that happen near to the transcription start site (TSS) are likely to be m^6Am modifications (Linder et al., 2015). m^6Am-exo-seq fragments the RNA and the digests away uncapped fragments. The m^6Am fragments are pulled out using an m^6A antibody and sequenced (Sendinc et al., 2019). These two methods rely on the m^6A antibody binding to m^6Am. This also means that in m^6A sequencing data some of the peaks will be m^6Am. To investigate this, m^6A data from human grey matter, white matter and foetal brain was analysed to characterise potential m^6Am peaks.

4.2 Results

4.2.1 Detection of m^6Am in brain

As previously discussed in chapter 3, preforming TLC enables global changes in RNA methylation on all capped RNAs to be examined. TLC experiments were performed in biological triplicate and a ratio of A:Am and Am:m^6Am was calculated.

Figure 4.1 shows an annotated representation of the expected pattern of nucleotides after running the digested RNA in two dimensions. The TLC results showed that the Am nucleotide was present in all the hippocampal samples, i.e. the unaffected control, Parkinson's disease and Lewy body dementia (Figure 4.1b-d). The level of Am varied between the Lewy body dementia and Parkinson's samples (PD mean=3.8% and LBD mean=4.0%) compared to the unaffected hippocampal tissue (mean=2.2%) (Figure 4.1e). However, this variation in Am was found not to be significant (Kruskal Wallis, T= 1.77, N=9, p=0.446).

The m^6Am nucleotide was also present in all three hippocampal samples (Figure 4.1b-d). The levels of m^6Am methylation, was again found to vary between the samples. The control samples had a mean of 6.3%, the Parkinson's disease sample had 11.2% and the Lewy body dementia had an average of 8% (Figure 4.1f). This variation between the levels of m^6Am in the three samples was also not significantly different (Kruskal Wallis, T= 1.867, N=9, p=0.439).

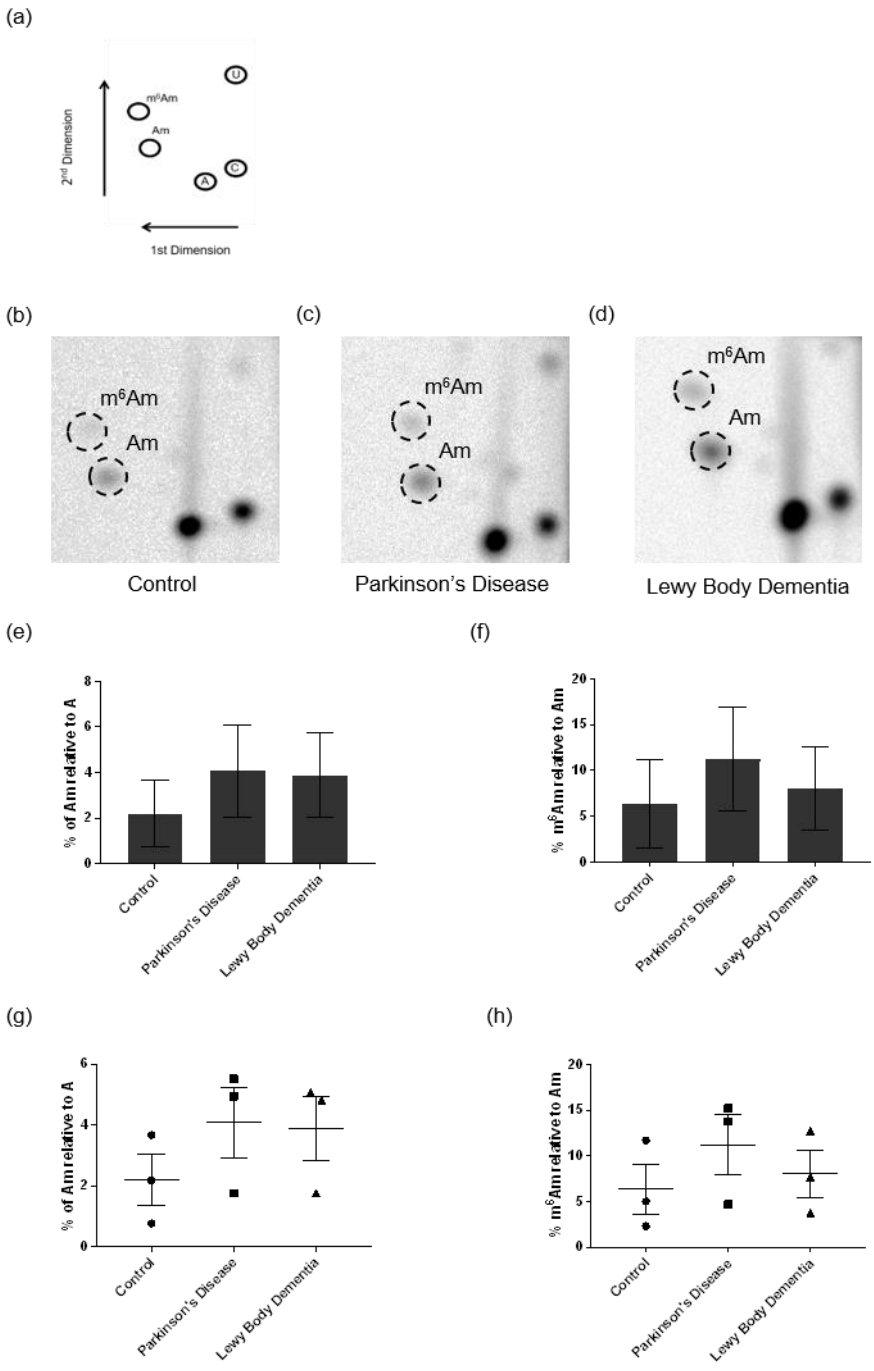


Figure 4.1. Representative TLCs examining cap methylation in RNA from non-disease control, Lewy body dementia and Parkinson's disease hippocampal brain samples. Graphs show the quantification and biological variation between samples. a) TLC annotated with A, C, U nucleotides and Am and m^{6}Am modified nucleotides. b) TLC on RNA from control *H. sapiens* hippocampus. c) TLC on RNA from Lewy body dementia *H. sapiens* hippocampus. d) TLC on RNA from Parkinson's disease *H. sapiens* hippocampus. e) Graph showing the average ratio

of Am relative to A showing mean with SEM. f) Graph showing the average ratio of m^6Am relative to Am showing mean with SEM. g) Biological variation in three individuals for each condition. The graph shows the percentage of Am relative to A for the three biological replicates with the mean and the standard error of the mean. h) Biological variation in three individuals for each condition. The graph shows the percentage of m^6Am relative to Am for the three biological replicates with the mean and the standard error of the mean.

For each category, i.e. unaffected, Parkinson's disease and Lewy body dementia, three replicates were performed to examine biological variation between individuals. The graphs in figure 4.1g-h presents the data points from each of the replicates and the standard error of the mean. On the graphs the variation between the samples is greater than the standard error of the mean. This is the case for both of the modified nucleotide ratios investigated. In the samples, the level of Am and m^6Am was proportionate, individual samples with higher levels of Am had higher levels of m^6Am. The control sample replicates are not clustered for both A:Am and Am:m^6Am ratios. Two of the Lewy body dementia replicates are similar for both ratios while the third one is lower. Two of the Parkinson's disease samples are the same for the A:Am ratio with a third lower replicate. For the Am:m^6Am ratio the samples of the Parkinson's disease replicates are more evenly spread out.

The variation between these samples may be a result of biological and technical variation. The sample number is low, being one 3 samples for each condition and in the disease and control groups the age, and gender of the people are not matched. There are also differences in the disease progression which is further discussed below. There will also be technical variation, with pipetting error and error in nanodrop readings of the RNA concentration.

When examining the pathology reports that are associated with the patient brains, it was found that the samples with higher levels of disease pathology had lower levels of methylation. Coincidentally the brains that had the worst pathology and the lowest levels of methylation in the disease states were the female brain samples. In the Parkinson's disease samples the lowest methylation was 98.24:1.76 A:Am and 92.27:4.73 Am:m^6Am, the sample came from a female brain and had tau fibres in the CA2 regions of the hippocampus and tau tangles and neuritic plaques in CA1 region.

In the other two hippocampal brain samples the pathology report stated that they were mildly affected, one had some cell loss in the CA1 and Lewy body staining in the CA3-4 regions. The other had minimal α-synuclein staining in the CA3-2 region of the hippocampus.

In the Lewy body with dementia samples the sample with the lowest methylation 98.24:1.76 A:Am and 96.25:3.75 Am:m^6Am was from a female brain, with the most pathology in the hippocampus but still mild. There was loss of CA1 neurons, with mild amyloid depositions and sparse Lewy bodies throughout the hippocampus. The other Lewy body dementia brain samples had sparse neruofibril tangles and Lewy bodies in the hippocampus. It may be that in this disease, pathology in a different brain region is then affecting the function and RNA methylation in the hippocampus.

In the healthy control brain samples, there is not a clear pattern. Although these samples had no disease diagnosis they were from individuals who were elderly, mean age 86.6 (range 79-94) years old. The brain sample with the highest level of methylation was the female brain 96.33:3.67 A:Am and 88.31:11.69 Am:m^6Am. At the time of death it was noted that the hippocampus of this brain had severe and frequent neurofibril tangles. The hippocampal sample with the lowest level of methylation 99.24:0.76 A:Am and 97.67:2.33 Am:m^6Am also had neurofibril tangles throughout the hippocampus. Such neurofibril tangles are associated with aging and are common in the elderly brain. As methylation levels did not show any correlation between the occurrences of neurofibril tangles, this would suggest there must be other factors which contribute to variation in modification levels in these brain samples than this natural ageing phenomena.

4.2.2 Detection of m^6Am from m^6A sequencing

m^6Am is found on the first nucleotide of the RNA transcript. To identify potential m^6Am peaks from whole genome m^6A sequencing data, modification sites at the 5′ end of the RNA where the starting nucleotide is adenosine, were selected. The sequencing reads were aligned to the HG38 using Tophat2 (Kim et al., 2013) and the m^6A peaks were identified using exomePeak calling algorithm (Zhang et al., 2008) and to identify putative m^6Am peaks, Homer (hypergeometric optimisation of motif enrichment) a tool for motif discovery and next generation sequencing analysis was used. Peaks that Homer annotated as TSS and that started with adenosine were annotated as putative m^6Am peaks. As a means to test for bias on the first nucleotide, the first nucleotide of the cDNA sequences for the transcripts that have an m^6A modification in the total data set were calculated. This was then compared to the first nucleotide of the peaks in the total data set and the peaks annotated with TSS. Figure 4.2 presents the workflow which was followed for this analysis. GO analysis was performed to identify if the transcripts which putatively have m^6Am were enriched for a specific processes or function. At each stage of the work flow there is a loss of peaks/ transcripts. Table 4.1 shows the number of transcripts and peaks at each stage for the three data sets foetal brain, grey matter and white matter.

From this analysis 233 peaks from the foetal brain data set were classed as m^6Am peaks. This is 2.83% of the total methylated transcripts. From the grey matter data set, 138 peaks were classified as m^6Am this is 2.31% of the unique transcripts. From the white matter data set, 110 peaks were classified as m^6Am which is 1.32% of the unique transcripts. Tables of the gene names of the putatively m^6Am methylated transcripts can be found in the appendix. These values are lower than what the published papers have reported for human malignant melanoma cell lines. For example, Sendic et al.,

2019 found there to be 1500 m^6Am peaks in their exo-m^6Am-seq on RNA extracted from MEL624 cells.

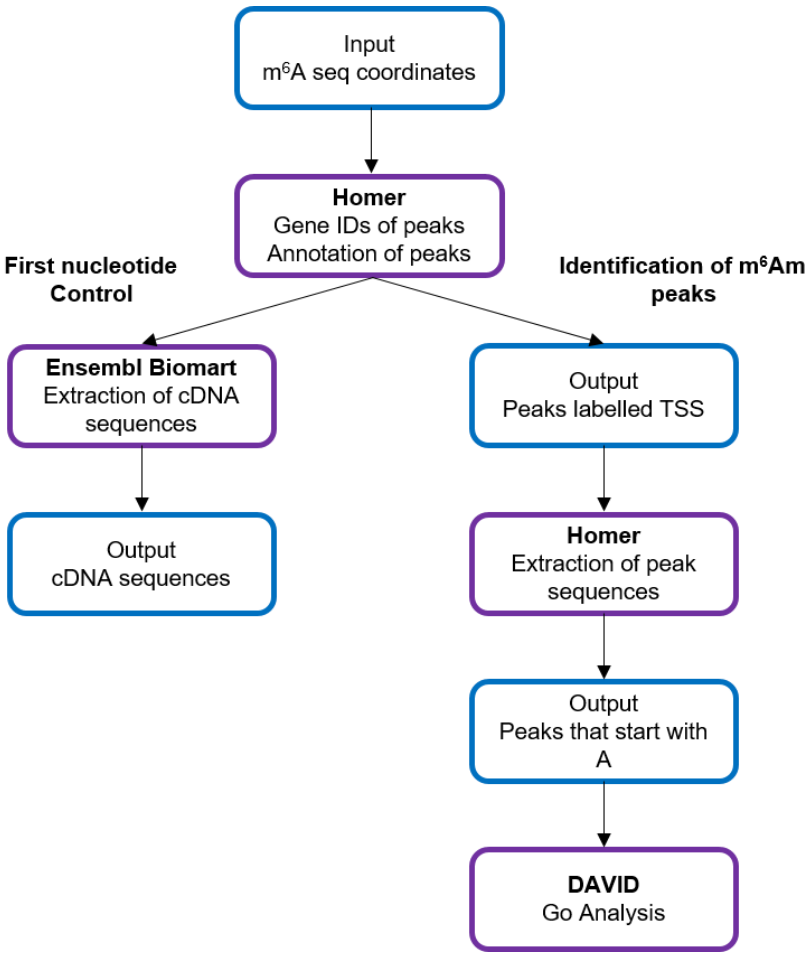


Figure 4.2. Work flow of bioinformatic analysis to detect m^{6}Am sites. The location of the m^{6}A peak sites were annotated and peaks annotated with at the TSS were identified. The TSS peaks that started with an adenosine were labelled as putative m^{6}Am peaks and were analysed for enrichment of GO terms compared to the total data set. As a control to assess sequencing bias on the first nucleotide, the first nucleotide of cDNA sequences for the genes in the total data were compared to the first nucleotide of the TSS peaks.

		Foetal brain	Grey matter	White matter
Number of peaks from m^6A/m^6Am sequencing data				
	Total peaks	17215	10726	17329
	Unique gene IDs	8234	5966	8310
	TSS peaks	1261	776	517
	Unique gene IDs	1138	705	463
	TSS A peaks (m^6Am)	237	140	112
	Unique gene IDs (m^6Am)	233	138	110
Go Analysis	TSS A (m^6Am) unique gene IDs	207	119	90
Number of cDNA transcripts				
	Total cDNA transcripts	72050	51148	73972
	Unique gene IDs	8217	5954	8924

Table 4.1. Number of peaks and unique gene IDs at the different stages of the m^6Am sequencing analysis work flow. Listed are the number of peaks, cDNA transcripts and unique gene IDs identified at each stage of the work flow for the three data sets foetal brain, grey matter and white matter.

Figure 4.3a,c,e show the distribution of m^6A peaks in the categories given by Homer. In all three samples the largest number of peaks were found in the exons, and less than 10% of the peaks were found at the TSS. The same distribution of annotations were found in all three samples. m^6Am was found at a much lower frequency than m^6A as would be expected because m^6Am modification occurs only once per transcript while internal m^6A could be present multiple times per transcript. Early work showed that m^6A could be found at an average of 1.8 nucleotides per 1000 bases or 4-5 methyl groups per molecule in poly-A containing mRNA of BHK-21 hamster cells (Dubin and Taylor, 1975). From the sequencing data here, it is not possible to tell whether the m^6A peaks are on the same or different RNA molecules. The average number of m^6A peaks per gene ID is 2.09 in the human foetal sequencing, 1.80 in the grey matter sequencing and 2.09 in the white matter sequencing similar to the early values for the hamster mRNA.

The analysis preformed called TSS peaks that begin with an adenosine putatively m^6Am peaks. This requires the first nucleotide of the peaks to be the same as the first nucleotide of the fragmented RNA and the first nucleotide of the transcript. During the sequencing process there are many steps where bias can occur and the first nucleotide can change. To understand this bias, the first nucleotide of the TSS peaks were compared to the first nucleotide of the cDNA transcripts of the respective genes found in the total data set. If there was no bias from sequencing it would be expected that the starting nucleotides of these two groups would be similar. If there was bias as a result of the sequencing the first nucleotides would be more similar to the total peaks first nucleotide distribution.

The cDNA starting nucleotide is similar across the 3 data sets ~ 30% adenosine, ~20% cytosine, ~35% guanine and ~10% thymine (Figure 4.3b,d,f). The starting nucleotides

for the total peaks varies between the three data sets, with the white matter having the biggest difference in starting nucleotide compared to the other two. The starting nucleotide for the total peaks for foetal brain is 24% adenosine, 27% cytosine, 29% guanine and 20% thymine. For the grey matter 22% adenosine, 29% cytosine, 29% guanine and 20% thymine. And for the white matter 28% adenosine, 24% cytosine, 24% guanine and 24% thymine (Figure 4.3b,d,f).

The percentage starting nucleotide distribution for the TSS peaks are a value between the first nucleotides of the total peaks and the first nucleotides cDNA. Foetal brain and grey matter are again more similar to each other than to the white matter. The TSS peaks starting nucleotides for foetal brain is 19% adenosine, 30% cytosine, 35% guanine and 16% thymine. For the grey matter 18% adenosine, 33% cytosine, 36% guanine and 14% thymine. And for the white matter 23% adenosine, 28% cytosine, 26% guanine and 23% thymine (Figure 4.3b,d,f). This suggests that the first nucleotide of the TSS peaks may have been partially biased by the sequencing method.

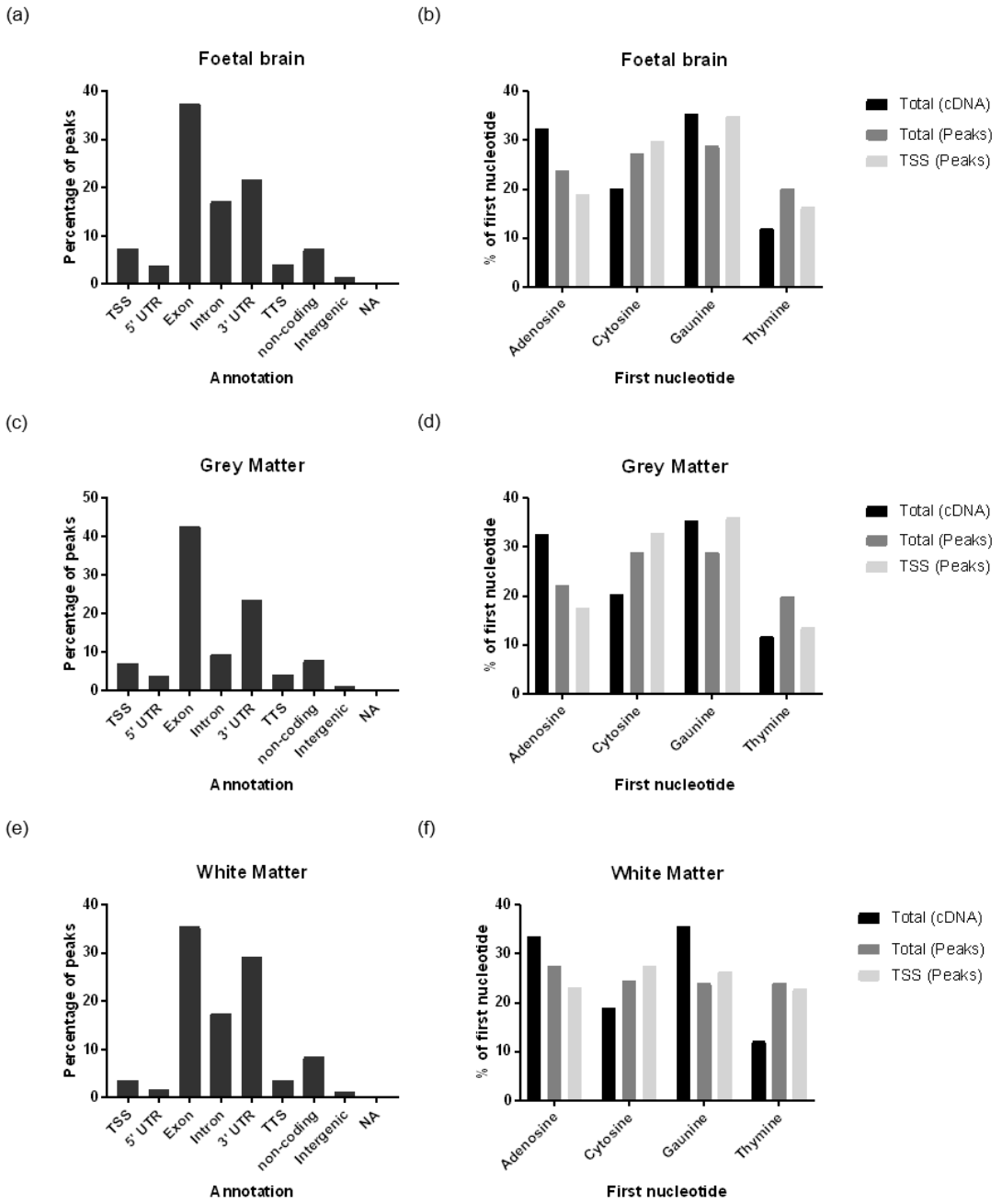


Figure 4.3. Peak annotations and starting nucleotides of cDNA transcripts of the genes in the total dataset, the total peaks and TSS peaks in the foetal brain, white matter and grey matter m^{6}A sequencing. a) Graph showing the percentage of peaks for each of the transcript location annotations for human-foetal m^{6}A-seq data set. b) Graph showing the starting nucleotide for the cDNA for genes in the total data set, the total peaks and the TSS peaks from the foetal brain sequencing. c) Graph showing the percentage of peaks for each of the transcript location annotations for grey matter m^{6}A-seq data set. d) Graph showing the starting nucleotide for the

cDNA for genes in the total data set, the total peaks and the TSS peaks from the grey matter sequencing. e) Graph showing the percentage of peaks for each of the transcript location annotations for white matter m^6A-seq data set. f) Graph showing the starting nucleotide for the cDNA for genes in the total data set, the total peaks and the TSS peaks from the white matter sequencing.

4.2.2.1 *De novo* motif analysis

De novo motif analysis was performed to examine if there was an enrichment for the known m^6A motifs GGACU and or m^6Am motif BCA (where B=C,G or T) or another motif where the methyltransferase, PCIF1, might bind. The BCA motif is very similar to the first part of the transcription initiation motif YYANWY (Y=C or T, N= any base, W= A or T). The BCA motif will not be found in the analysis, because the genomic sequence before the TSS peak needs to be analysed. Potentially a logo similar to the second half of the transcription initiation motif, ANWY, may be found and this would start with the modified m^6Am. *De novo* motif analysis was performed on the total dataset of peaks, TSS peaks, and putative m^6Am peaks compared to randomly selected GC matched sequences. In addition motif analysis was performed on the putative m^6Am peaks compared to the sequences of the total peaks.

The *de novo* motif enrichment for the total peaks in the three data sets were all similar to the m^6A motif with a conserved GGAC (Figure 4.4). The first nucleotide shows some variation with, the motif in the foetal brain and white matter data sets starts with a U and the grey matter starts with a C. Motif analysis of the TSS peaks is similar to the motif for all the modification sites with the conserved GGAC. Foetal brain and grey matter were found to be GGACU while white matter was UGGAC.

The putative m^6Am peaks are a very small population of peaks (Table 4.1). The motif for the grey matter data is the only statistically significant motif and is UGGAC similar to the m^6A motif. The motif for white matter and foetal brain do not resemble the m^6A or m^6Am motif or each other. Comparing the putative m^6Am peaks to the total peaks were also not statistically significant. However the foetal brain motif is similar to the white matter TSS A motif and the grey matter motif is similar to the foetal brain TSS A

motif (Figure 4.4). As the motifs do not start with an adenosine, they are unlikely to be an m^6Am motif. However, they could be a binding site motif for PCIF1.

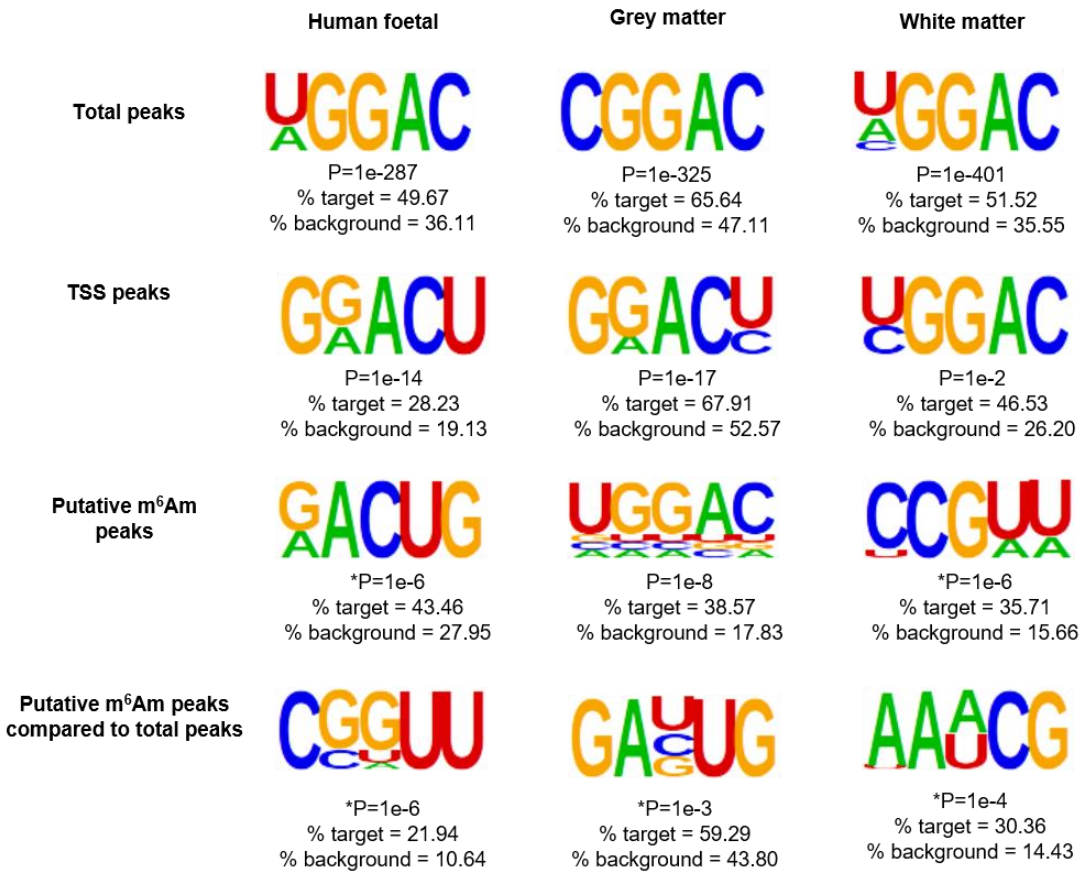


Figure 4.4. *De novo* motif analysis of potential m^{6}Am peaks. The figure shows the motifs with associated p value, percentage of target sequences with the motif and percentage of background sequences with the motif. The size of the letters is proportional to the number of that base found in the position in the motif.

4.2.2.2 GO analysis of potential m^{6}Am transcripts

Functional Gene Ontology (GO) analysis and reactome analysis was performed to compare the enrichment of GO terms between putative m^{6}Am methylated transcripts to all methylated transcripts in the total data set. Tables 4.2-4.7 presents the top 20 enriched terms with the lowest p values, (<0.05) for the GO and reactome analysis.

The GO term and reactome results for the foetal brain is presented in Table 4.2 and Table 4.3 respectively. The foetal brain data set is enriched for cell components and binding of RNA and protein. The GO terms with the lowest p-values are cellular competent organisation or biogenesis (p=2.68e-8) and cellular component organisation (9.39e-08). Four of the top twenty terms are nuclear, nuclear part (p=9.24e-07), nucleoplasm (p=1.75e-06), nuclear lumen (p=7.92e-06), and nucleus (p=8.35e-05). Other terms which are enriched relate RNA binding (p=1.42e-06), protein binding (p=1.22e-06) and cellular protein localisation (p=9.92e-05). Only one brain specific term was enriched, neuron projection morphogenesis (p=6.34e-05), which is the term with the highest fold enrichment of 3.265. However, many of the other terms indicated are processes required for normal function.

The pathways enriched in the reactome analysis relate more to neural development, as would be expected in the developing foetal brain. Axon guidance (p=1.40e-03) and MTOR pathway (p=2.32e-04) are needed for cell proliferation in the brain (Ryskalin et al., 2017), and EPHA mediated growth cone collapse (p=0.041), is involved in cell retraction during the process of axon guidance (Suetterlin and Drescher, 2014, Shigeoka et al., 2013).

The GO term for the putative m^{6}Am peaks in the grey matter mostly relate to metabolic functions. These include processes associated with regulation of macromolecule metabolic processes (p=0.002), negative regulation of phosphorous metabolic process

(p=0.003) and regulation of protein metabolic process (p=0.003). The top GO term is molecular function regulator (p=1.94e-04) and there are other associated GO terms such as enzyme regulator activity (p=3.53e-04) and regulation of protein stability (p=0.001). The GO term with the highest fold enrichment is lipid homeostasis with an 18.7 fold increase. Lipid homeostasis is very important in brain tissue, as it influences both synaptic function and action potential propagation and misregulation can lead to neurodegeneration (Wang et al., 2002a).

In both the reactome and GO analysis tyrosine kinase receptor protein is highlighted (p=0.023 and p=0.003 respectively). Another membrane process enriched for in the reactome analysis is basigin interactions with a fold enrichment of 20 (Table 4.5). Basigin, also called CD147 or EMMPRIN is a transmembrane glycoprotein. Basigin is involved in intracellular interactions and is involved in the synthesis of several matrix metalloproteinases which degrade the components of endothelial basal lamina (Grass and Toole, 2015). It is also essential for transport of nutrients and migration of inflammatory leukocytes (Muramatsu, 2016).

In contrast, the GO terms for white matter putatively m^6Am methylated transcripts are strikingly different. There are three GO terms which are highly enriched that come up multiple times. These are endocytosis, interleukin-6 response and chromosome (Table 4.6). Along with regulation of endocytosis (p=0.012), the GO terms viral processes p=0.001) and interspecies interaction between organisms (p=0.009) are also enriched. Interleukin 6 is a major cytokine in the brain and is known to participate in neurogenesis (Erta et al., 2012). The GO term negative regulation of interleukin-6 secretion has a fold enrichment of 96%. In addition, the processes regulation of mRNA and RNA stability (p=0.033 and p=0.036 respectively) were highlighted. In the reactome analysis

the only pathway which was enriched for is AUF1 (hnRNP D0) (p=0.014) which has a role binding and destabilizing mRNA.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of genes
GO:0071840	Cellular component organization or biogenesis	2.68e-8	1.582	5.90E-07	95
GO:0016043	Cellular component organization	9.39e-08	1.571	2.96E-04	92
GO:0044428	Nuclear part	9.24e-07	1.771	4.36E-04	65
GO:0005515	Protein binding	1.22e-06	1.354	3.79E-04	117
GO:0003723	RNA binding	1.42e-06	2.337	7.47E-04	37
GO:0005654	Nucleoplasm	1.75e-06	1.919	8.28E-04	53
GO:0051128	Regulation of cellular component organization	3.92e-06	2.020	0.012	45
GO:0031981	Nuclear lumen	7.92e-06	1.752	0.004	58
GO:0019899	Enzyme binding	2.14e-05	2.104	0.011	36
GO:0006996	Organelle organization	4.20e-05	1.640	0.124	59
GO:0044446	Intracellular organelle part	4.69e-05	1.344	0.022	102
GO:0048812	Neuron projection morphogenesis	6.34e-05	3.265	0.181	17
GO:0010467	Gene expression	6.96e-05	1.474	0.197	75
GO:0070013	Intracellular organelle lumen	6.99e-05	1.579	0.032	63
GO:0044260	Cellular macromolecule metabolic process	8.08e-05	1.306	0.225	105
GO:0005634	Nucleus	8.35e-05	1.385	0.039	90
GO:0005737	Cytoplasm	8.40e-05	1.496	0.021	72
GO:0034613	Cellular protein localization	9.92e-05	2.086	0.269	32
GO:0070727	Cellular macromolecule localization	1.15e-04	2.069	0.304	32
GO:0043233	Organelle lumen	1.16e-04	1.553	0.053	63

Table 4.2. GO Terms associated with transcripts that putatively have the m^6Am modification in foetal brain. The table contains the top 20 GO terms with p-value, fold enrichment, Bonferroni adjusted p-value and number of genes with that GO term.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of genes
hsa04150	mTOR signalling pathway	2.32e-04	10.465	0.032	6
hsa04360	Axon guidance	1.40e-03	5.576	0.178	7
R-HSA-3371571	HSF1-dependent transactivation	0.013	16.858	0.964	3
hsa04141	Protein processing in endoplasmic reticulum	0.024	3.592	0.966	6
hsa04071	Sphingolipid signalling pathway	0.029	4.215	0.984	5
R-HSA-3928663	EPHA-mediated growth cone collapse	0.041	9.151	1.000	3

Table 4.3. Reactome analysis for the transcripts that putatively have the m^6Am modification in foetal brain. The table contains the top 20 reactome terms with p-value, fold enrichment, Bonferroni adjusted p-value and number of genes with that term.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of genes
GO:0098772	Molecular function regulator	1.94e-04	2.547	0.085	20
GO:0019899	Enzyme binding	2.34e-04	2.287	0.102	23
GO:0030234	Enzyme regulator activity	3.53e-04	2.858	0.150	16
GO:0050790	Regulation of catalytic activity	5.72e-04	2.013	0.792	26
GO:0005515	Protein binding	6.33e-04	1.327	0.142	67
GO:0055088	Lipid homeostasis	1.22e-03	18.720	0.555	4
GO:0030054	Cell junction	1.44e-03	2.332	0.026	18
GO:0031647	Regulation of protein stability	1.50e-03	5.604	0.983	7
GO:1902580	Single-organism cellular localization	2.35e-03	2.484	0.998	15
GO:0060255	Regulation of macromolecule metabolic process	2.47e-03	1.444	0.999	46
GO:0019222	Regulation of metabolic process	2.56e-03	1.420	0.999	48
GO:0010605	Negative regulation of macromolecule metabolic process	2.60e-03	1.883	0.999	24
GO:0007169	Transmembrane receptor protein tyrosine kinase signalling pathway	2.64e-03	3.098	0.999	11
GO:0010563	Negative regulation of phosphorus metabolic process	2.73e-03	3.355	0.999	10
GO:0045936	Negative regulation of phosphate metabolic process	2.73e-03	3.355	0.999	10
GO:0031324	Negative regulation of cellular metabolic process	2.88e-03	1.868	1.000	24
GO:0051049	Regulation of transport	3.15e-03	2.033	1.000	20
GO:0051246	Regulation of protein metabolic process	3.24e-03	1.816	1.000	25
GO:0009892	Negative regulation of metabolic process	3.41e-03	1.809	1.000	25
GO:0043549	Regulation of kinase activity	3.50e-03	2.780	1.000	12

Table 4.4. GO Terms associated with the transcripts that putatively have the m^6Am modification in grey matter. The table contains the top 20 GO terms with p-value, fold enrichment, Bonferroni adjusted p-value and number of genes with that GO term.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of genes
hsa04360	Axon guidance	7.21e-03	6.298	0.584	5
R-HSA-210991	Basigin interactions	8.88e-03	20.547	0.788	3
2.7.10.1	receptor protein-tyrosine kinase	0.023	11.905	0.433	3

Table 4.5. Reactome analysis for the transcripts that putatively have the m^6Am modification in grey matter. The table contains the top 20 reactome terms with p-value, fold enrichment, Bonferroni adjusted p-value and number of genes with that term.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of Genes
GO:0016032	Viral process	6.25e-04	5.343	0.223	8
GO:0044764	Multi-organism cellular process	8.04e-03	2.594	1.000	11
GO:0044403	Symbiosis, encompassing mutualism through parasitism	9.03e-03	2.550	1.000	11
GO:0044419	Interspecies interaction between organisms	9.03e-03	2.550	1.000	11
GO:0044454	Nuclear chromosome part	0.010	3.748	0.890	7
GO:0030100	Regulation of endocytosis	0.012	5.489	1.000	5
GO:0048259	Regulation of receptor-mediated endocytosis	0.013	7.960	1.000	4
GO:0000228	Nuclear chromosome	0.014	3.437	0.963	7
GO:0045806	Negative regulation of endocytosis	0.019	13.645	1.000	3
GO:1900165	Negative regulation of interleukin-6 secretion	0.021	95.515	1.000	2
GO:0005694	Chromosome	0.031	2.378	0.999	9
GO:0043488	Regulation of mRNA stability	0.033	5.619	1.000	4
GO:0043487	Regulation of RNA stability	0.035	5.458	1.000	4
GO:0043232	Intracellular non-membrane-bounded organelle	0.036	1.441	1.000	25
GO:0044427	Chromosomal part	0.039	2.462	1.000	8
GO:0098687	Chromosomal region	0.040	3.813	1.000	5
GO:0032635	Interleukin-6 production	0.040	9.243	1.000	3
GO:0032675	Regulation of interleukin-6 production	0.040	9.243	1.000	3
GO:0005938	Cell cortex	0.041	3.756	1.000	5
GO:0000785	Chromatin	0.044	3.035	1.000	6

Table 4.6. GO Terms associated for the transcripts that putatively have the m^6Am modification in white matter. The table contains the top 20 GO terms with p-value, fold enrichment, Bonferroni adjusted p-value and number of genes with that GO term.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of genes
R-HSA-450408	AUF1 (hnRNP D0) binds and destabilizes mRNA	0.014	15.652	0.812	3

Table 4.7. Reactome analysis for transcripts that putatively have the m^6Am modification in white matter. The table contains the top 20 reactome terms with p-value, fold enrichment, Bonferroni adjusted p-value and number of genes with that term.

4.3 Discussion

4.3.1 Human brain function and disease

Disease models are challenged to replicate the complexity of human neurodegenerative disease and the effect such diseases have on human experience and cognition. Disease models are artificially created and often only replicate one aspect of the disease (Ransohoff, 2018). In many diseases the immune system plays an important role. In these rodent studies the immune system may show different responses, in part because of the sterile conditions the mice are raised in (Tao and Reese, 2017). The neuroanatomy in rodents also differs from that of people so is not an accurate reflection of processing in CNS circuits. Using human brain tissue including disease samples is informative for the understanding cognitive function and dysfunction.

Working with human tissue has its limitations. Many factors can influence neurodegenerative diseases and natural brain decline that comes with age (Peters, 2006, Boot et al., 2013, Hubble et al., 1993). It is hard to control-match diseased tissue samples because of the natural aging process which happens differently to each individual. Working with RNA, there is a high risk of RNA degrading. It may be hours before the tissue is removed and frozen, this time is recorded. However, by using the Terminator Exonuclease enzyme, fragmented RNA will be digested. Furthermore RNA with short half-lives may have been degraded between the time of death and when the brain was frozen. It is not known to what extent this degradation will affect the level of methylation on the TLCs versus what is present in living tissue.

4.3.2 m^6Am in neurodegenerative diseases

The findings of the TLCs demonstrated that capped RNA species are methylated in non-diseased hippocampal brain tissue as well as tissue from individuals with Parkinson's disease and Lewy body dementia. There was variation between the cap adjacent RNA methylation in the two disease states and the healthy control samples; this variation was not statistically significant. However when looking at the samples and the pathology, the samples with the most severe pathology have the least RNA methylation. Potentially levels of methylation change during the course of the diseases.

The TLCs experiments using the human brain will have detected all capped RNA species. This included lncRNAs which have recently been suggested to contribute to Parkinson's disease pathology (Elkouris et al., 2019). 40% of lncRNAs are distributed in the CNS (Mercer et al., 2010). They are important for normal brain development and cell proliferation as well as synaptic function (Klattenhoff et al., 2013, Mercer et al., 2010, Kondrashov et al., 2005). In Parkinson's disease, lncRNAs regulate key disease genes such as *PINK1*, *UCHL1*, *LRKK2*, and *SNCA* (Elkouris et al., 2019, Carrieri et al., 2015, Coupland et al., 2016). The conserved neuronal lncRNA *MALTA1* has also been shown to upregulate α-synuclein, the key protein found in the pathological cellular aggregates (Zhang et al., 2016).

It is known that lncRNAs are methylated with m^6A and as a RNA species with a 7' methyl guanine cap they could potentially be m^6Am methylated. The effect of RNA methylation on lncRNAs will be different depending on the individual lncRNA and potentially where along the transcript it is methylated and what protein interacts with the methylation. Two examples that have internal m^6A methylation, are the lncRNAs *RPII* and *MATAL1*. m^6A methylation of *RPII* results in the lncRNA accumulating in the nucleus where it can up regulate ZEB1 through post transcriptional regulation. In

colorectal cancer cells this causes metastasis (Wu et al., 2019). m^6A methylation of *MATLA1* changes the secondary structure of the RNA allowing for HNRNPC to bind. Binding triggers HNRNPC to translocate out of the nucleus affecting the cell cycle (Zhou et al., 2016, Yang et al., 2013).

In Parkinson's disease and Lewy body dementia there is some evidence to support that as well as the Lewy body pathology there is also an increase in neuroinflamation (Tansey and Romero-Ramos, 2019, Surendranathan et al., 2015). For example, α-synuclein, which is abnormally aggregated in both Parkinson's disease and Lewy body Dementia is reported to trigger the innate immune response (Allen Reish and Standaert, 2015). Interferon-β signalling activates pathways which form the innate and adaptive immune responses (Ivashkiv and Donlin, 2014). The increase in Am methylation, may be as a result of the diseases increasing Interferon-β signalling and the innate immune response causing the methylation of "self" RNA (Freund et al., 2019). CMTR1 increases 3 fold after interferon induction resulting in an increase in cap1 methylation of RNA. (Su et al., 2002, Kato et al., 2003, Guerra et al., 2003, Geiss et al., 2003).

m^6Am levels were also found to be increased in the disease states. In sporadic Parkinson's disease there is some evidence suggesting that α-synuclein mRNA levels do not increase though there is an increase in the protein (Quinn et al., 2012, Su et al., 2017). m^6Am may increase translation of methylated transcripts, potentially transcripts like α-synuclein might be methylated in the disease leading to the increase in protein. There is conflicting evidence whether m^6Am increases translation, however it may be tissue and cell context dependant.

4.3.3 Detection of m^{6}Am from sequencing data

In the m^{6}Am from m^{6}A sequencing analysis, assumptions were made that the TSS peaks that start with an adenosine are m^{6}Am peaks. The first assumption was that the peaks were at the start of the transcript. Homer peak annotation TSS, is defined as peaks -1kb to +100bp from the genes defined TSS. The broad range of peaks that were annotated as TSS means that potentially these peaks will include alternative start sites from both focused and dispersed TSS. Initiation from focused transcription starts sites is from a single nucleotide or within a cluster of several nucleotides, whereas initiation from dispersed transcription start site, starts from several weak transcription start sites over a broad region of about 50 to 100 nucleotides (Juven-Gershon and Kadonaga, 2010, Kruse et al., 2011). Conversely this broad definition of TSS may increase the chance that 5′ m^{6}A sites are being detected instead of m^{6}Am.

A second assumption was that the first nucleotide of the peaks was the first nucleotide of the transcript. There is positional bias through the fragmentation process which is biased against the ends of the sequence and potentially the nucleotides the break happens between. There is also potential sequence specific bias of the ligation of the adapters and PCR steps. The trimming process to remove the adapter sequence after sequencing is not completely accurate and will introduce error to the first nucleotide of the transcript (Finotello et al., 2014). There will be false positives and false negatives along with the m^{6}Am peaks.

A third assumption is the antibody binds to m^{6}Am with a high efficiency. In the m^{6}A sequencing the Abcam m^{6}A antibody (abs151230) was used. Work completed by Dr. Bodi showed that this antibody had the best binding efficiency to m^{6}A out of the three tested. Its ability to bind m^{6}Am, with the guanine cap structure so close, is not known.

Only four papers currently report m^6Am sequencing (Linder et al., 2015, Sendinc et al., 2019, Engel et al., 2018, Schwartz et al., 2014). These studies used different methods and examined different cell types. Exo-m^6Am-seq uses an exonuclease to remove all fragments that are not capped and the m^6A antibody to isolate the m^6Am methylated peaks. In the sequencing data of MEL624 cells 1521 peaks were found (Sendinc et al., 2019). Linder et al,. 2015 were the first to detect the BCA motif. Using mi-CLIP which causes an AC-AT conversion where the antibody binds. In HEK293 cells Linder et al., found 531 BCA motifs. Shwartz et al., 2015 used bioinformatics analysis on m^6A to find peaks that were 50 bp up and down from the TSS or 50bp up 25bp down from TATA box. 7255 peaks were found from their data sets. Engel et al., 2018 found peaks that did not change in WTAP KO cells and report 1801 putative m^6Am peaks.

The number of putative peaks from this analysis were 237 (1.38% of the total peaks), 140 (1.30% of the total peaks), 112 (0.64% of the total peaks) for the foetal brain, grey matter and white matter data sets respectively. The number of peaks that were found in this analysis are much fewer than in previously published papers. The numbers of total TSS peaks is more similar to the published numbers. The lower number of peaks may be as a reflection of what is occurring *in vivo*. The work in the previous studies was performed on cell lines while the current study was performed on primary human brain tissue.

The putative m^6Am sites detected in this study would need to be validated to confirm the analysis was accurate. A whole transcriptome m^6Am sequencing approach such as Exo-m^6Am-seq would be a potential approach, the peaks from this sequencing could then be compared to the putative peaks. However, there is the possibility of false positives from the Exo-m^6Am-seq, as it can detect m^6A modifications near to the cap.

Peaks could also be validated through a specific transcript approach such as SCARLET.

4.3.4 The function of m^6Am in the human brain

The function of m^6Am in the brain is likely dependent on the cell type. Grey matter is made from the cell bodies and dendrites. The cell body is where much of the cellular metabolism occurs and as, neurons have huge energy requirements phosphate metabolism is very important (Barness, 1961). This is reflected in the pathways that are enriched for in the GO terms which implicate metabolic processes. White matter is made from the axons of neurones and glial cells. Glial cell functions appear enriched for in the m^6Am putative peaks from the white matter. For instance, terms related to the immune system and interleukin-6 were enriched and these processes happen in microglia (Yan et al., 2019, Liu and Quan, 2018). Furthermore, endocytosis was also enriched and, astrocytes perform endocytosis removing neurotransmitters from the synaptic cleft (Henn and Hamberger, 1971).

Cap adjacent RNA methylation may play a role in disease (Dimitrova et al., 2019). There is an overall average increase in RNA methylation in the disease samples that were analysed. This suggests that misregulation of m^6Am could be involved in disease progression. However since the highest levels of RNA methylation were found in samples with the least severe pathology, the increase in m^6Am may be a response to disease processes that change throughout disease progression.

4.3.5 Future work

There are still questions that could be asked to better understand the role of m^6Am in brain function and disease. The work here has shown there is an increase in

methylation in diseased brain tissues. It would be informative to carry out Exo-m^{6}Am-seq to see if there was a change in the transcripts being methylated in healthy and disease tissues. Potentially RNAs for disease associated proteins and lncRNAs could be being methylated. To understand disease progression, levels of methylation and methylated transcripts detected at different stages of the disease.

This work provides evidence supporting the presence of m^{6}Am in the brain and different transcripts in different cell populations being methylated. To understand further the function of m^{6}Am in brain, the methylation response to synaptic activation could be investigated and transcripts at the synapse could be sequenced to see if after plasticity if global cap adjacent methylation levels, and methylation of individual transcripts change.

Chapter 5- Identifying m^6Am Modifications by DCP2 Enzymatic Treatment

5.1. Preface

The enzyme DCP2 plays a key role in the turnover of mRNA by catalysing the cleavage of the cap structure to allow 5′-3′ RNA degradation (Wang et al., 2002b, Parker and Song, 2004). It has been reported that m^6Am prevents the DCP2 enzyme from decapping the RNA transcript (Mauer et al., 2016). If this finding can be confirmed, and m^6Am does provide resistance to DCP2, it would provide a method for enriching m^6Am capped mRNA which could be combined with sequencing to identify these protected modified transcripts. This would allow for the identification of sequences with m^6Am, so providing an alternative approach to antibody dependant methodologies.

5.2. Results

5.2.1. Expressing soluble DCP2 protein

The production of soluble DCP2 protein, that is active and can decap RNA, had to be optimised. A plasmid was constructed, based upon the NEB pMAL protein expression vector, in which the human *DCP2* coding sequence was fused in frame downstream of the 42 kDa maltose binding protein. The map of this plasmid can be seen in figure 5.1. This plasmid pMAL-c5x-hDCP2 was transformed into *E. coli* strain Dh5α. A standard protocol for expressing protein was then tested (Chapter 2, 2.2.12). The protein size expected was 90kDa, the 48kDa DCP2 fused to the 42kDa maltose binding protein. The expression was induced for 1, 2 and 3 hours and expression assessed by polyacrylamide gel electrophoresis (PAGE). Using quantity one software (Bio-Rad) the band of induced protein was measured and the percentage of induced protein was found to be 27.2%, 43.3%, 29.4% for 1, 2, and 3 hours of induction, respectively. The highest of recombinant protein expression was acheived at 2 hours. Very little of this protein from the 2 hour time point was soluble (Figure 5.2a).

In an attempt to improve expression levels and solubility, the plasmid was transformed into Rosetta-gami *E. coli*. This strain enhances disulfide bond formation and enhances expression of eukaryotic proteins that contain codons rarely used in *E. coli*. To promote protein solubility, two concentrations of IPTG was tested, the concentration recommended when using the plasmid (0.3mM IPTG) and half that concentration. Less IPTG was tested as the slower the protein is made the more likely it is to be folded properly and to be soluble and active(Vera et al., 2007). The Rosetta-gami cells produced higher levels of protein than the Dh5α cells. Using 0.3mM IPTG only 5% of

the DCP2 protein produced was found in the crude extract. Using the 0.15mM IPTG 40% of the DCP2 protein produced was soluble (Figure 5.2b).

40% soluble protein is still low so it was decided to test expression in SoluBL21 cells and to reduce the IPTG and temperature and leave the cells for a longer induction period (Picard-Jean et al., 2018, Liu et al., 2008). With this method three different medias were used to test different growth rates: M9 minimal media with glucose; M9 minimal media with glycerol; and, lysogeny broth (LB). M9 minimal media with glycerol produced the slowest cell growth rate and the cells grown in this media did not produce much protein. However, 60% of what was produced was soluble. The cells grown in M9 minimal media with glucose produced more protein and more of the protein was soluble, 65%. The best media was LB. This produced the most soluble protein, with 80% of the DCP2 produced being found in the crude extract (Figure 5.2c).

It was found that growing the cells in LB media at 18°C for 20hrs, with 0.4mM IPTG and 2% ethanol produced the highest proportion of soluble protein. To purify the protein the His tag that was added to the coding sequence before cloning was used. A large amount of protein was produced. There were some contaminating proteins even after purification but these were assumed to be at a lower proportion than DCP2 (Figure 5.3).

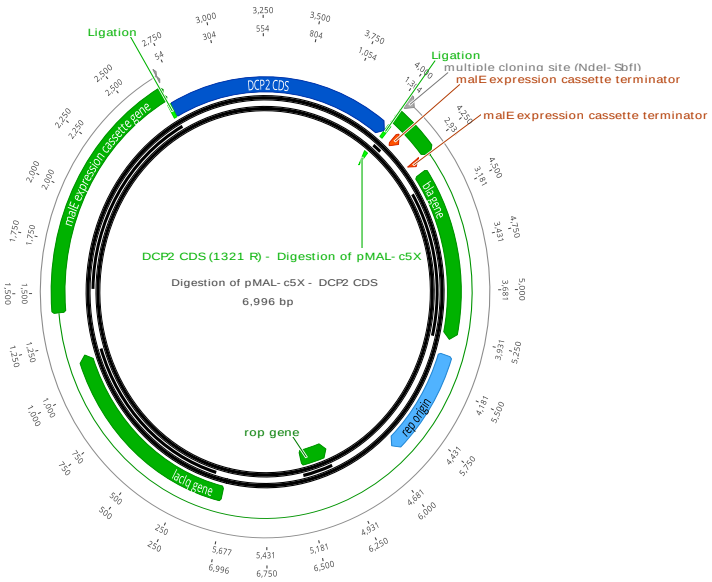


Figure 5.1. Map of the pMAL-c5x-hDCP2 construct. The human DCP2 coding sequence was cloned into the pMAL-c5x plasmid using Gibson assembly. The restriction enzyme Xmn1 was used to cut the pMAL-c5x plasmid in the multiple cloning site.

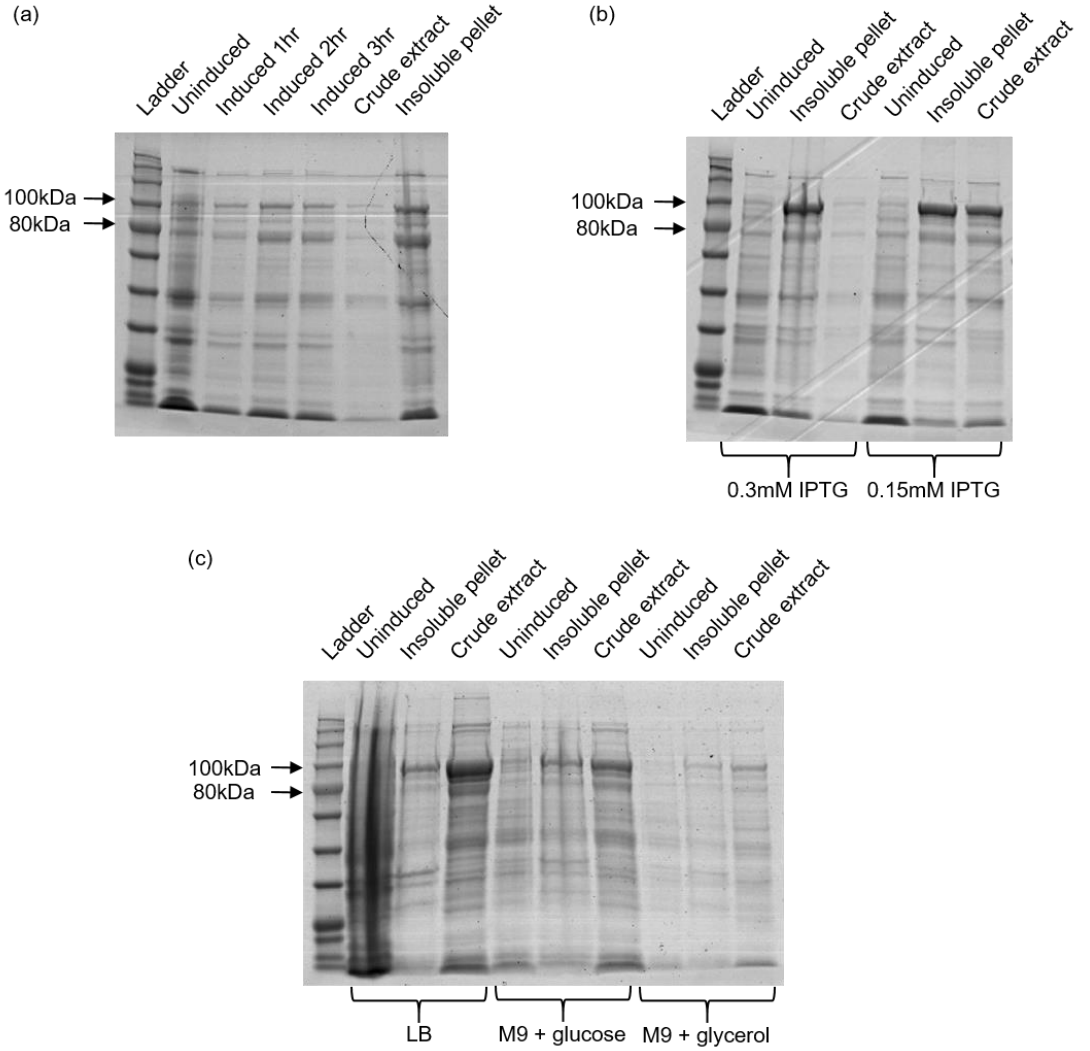


Figure 5.2. Protein gels from the pilot experiments optimising soluble DCP2 protein production. A band at 90kDa is expected to be seen after induction. The DCP2 protein is 48kDa and this is fused to amylose binding protein which is 42kDa. a) Pilot experiment in DH5α cells, cells were induced with 0.3mM IPTG. The protein gel shows uninduced cell lysate, cell lysate from cells induced for 1, 2, and 3 hours, the crude extract and insoluble pellet from the 2 hour cell lysate. b) Pilot experiment in Rosetta-gami cells, cells were induced with two concentrations of IPTG 0.3mM and 0.15mM. The protein gel shows uninduced, insoluble pellet and crude extract. c) Pilot experiment in SoluBL21 cells. The protein gel shows uninduced, insoluble pellet and crude extract for the cells using three different medias, LB with 0.4mM IPTG and 2% ethanol, M9 minimal media with glucose and 0.4mM IPTG and M9 minimal media with glycerol and 1mM IPTG.

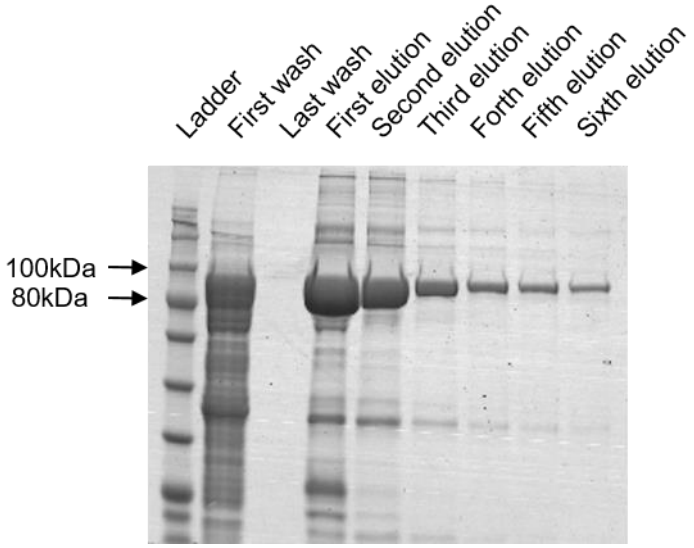


Figure 5.3. Protein gel showing DCP2 protein purification elutions. The DCP2 protein was purified on a gravity flow nickel resin column. The protein gel shows the first and last washes, and the 6 elutions.

5.2.2. Using DCP2 to decap *R. norvegicus* brain RNA

TLCs were used to assess the first nucleotide of capped and uncapped transcripts from rat brain before and after DCP2 treatment. Capped RNA that had been poly-A and terminator treated was used in this experiment. *R. norvegicus* brain RNA contains Am and m^6Am (A:Am 99.8:0.2, Am:m^6Am 18.7:81.3) (Figure 5.4b). The buffer that the RNA was treated in did not appear to decap or degrade the RNA although there are some 5′ P and 5′ OH ends. None of these were Am or m^6Am (Figure 5.4c).

In the capped sequences after DCP2 treatment there was not any Am or m^6Am (Figure 5.3d). This suggests that DCP2 was preferentially decapping transcripts with cap1 and m^6Am which is contrary to the published literature. DCP2 is reported to leave a 5′ P, although when tested, the enzyme also left 5′ OH ends. The 5′ OH ends have a ratio of 99:1 A:Am and 13:87 Am:m^6Am (Figure 5.4e). The decapped transcripts with 5′OH have a slightly higher proportion of Am and m^6Am than the rat brain RNA before treatment. This again suggests that DCP2 is preferentially decapping transcripts that start with cap1 and m^6Am.

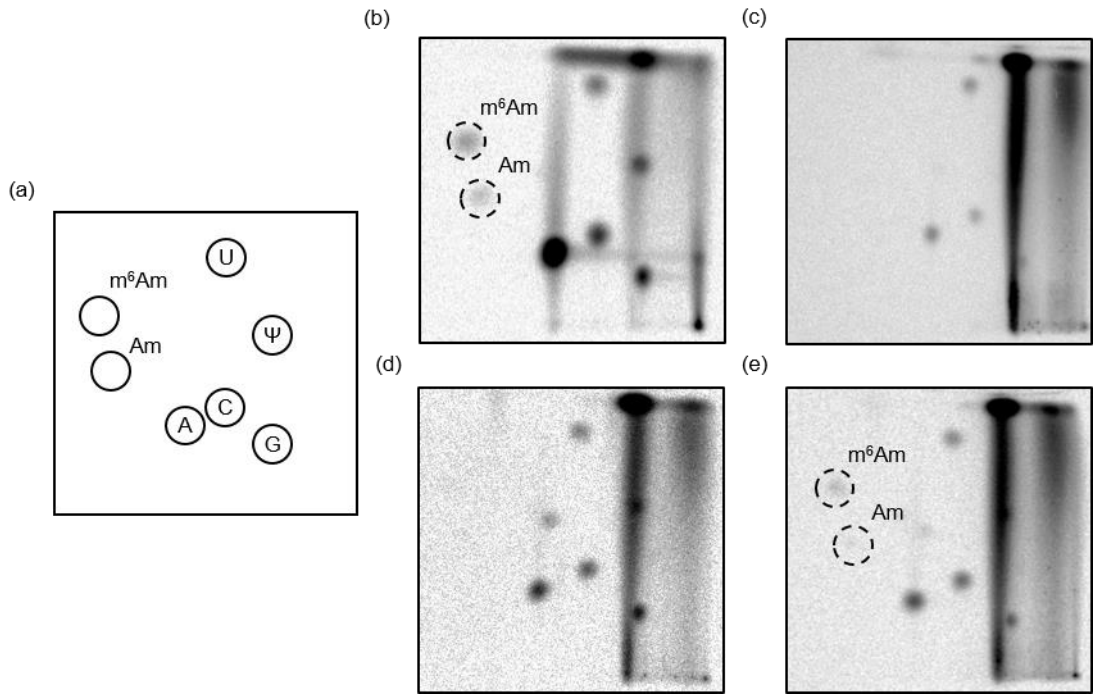


Figure 5.4. TLCs showing the first nucleotide of control and DCP2 treated RNA**.** These TLCs were completed in one replicate. a) Annotated nucleotides A, C, G, and U and modified nucleotides Am, m^6Am and pseudouridine that can be seen on the TLCs. b) Control TLC of rat brain RNA before DCP2 treatment. c) Control TLC showing transcripts that start with a 5′ phosphate or 5′OH after buffer treatment. d) TLC showing the first nucleotides of transcripts which were not decapped by DCP2. e) TLC showing the first nucleotide of transcripts that have been decapped by DCP2 starting with a 5′OH.

5.3. Discussion

5.3.1. Sequencing transcripts with m^{6}Am

When this work was started there were not any published methods for sequencing transcripts that started with m^{6}Am. In the last four years, three methods have been developed by other labs. The first method uses bioinformatics to identify m^{6}Am peaks from m^{6}A MeRIP-seq data (Schwartz et al., 2014). m^{6}A peaks are aligned to the genome and the 5′ peaks are annotated as m^{6}Am peaks. The second method is to remove an m^{6}A writer or eraser or m^{6}Am writer or eraser and compare the peaks before and after (Mauer et al., 2016, Engel et al., 2018). The third method is called Exo-m^{6}Am-seq (Sendinc et al., 2019). This method fragments the RNA, uses an exonuclease to digest uncapped fragments and then a m^{6}A antibody to pull out capped fragments with an N^6-methyladenosine. The intrinsic issue with all three methods is that they select for N^6-methyladenosine and not m^{6}Am. There is a chance that m^{6}A close to the 5′ cap is being detected in these methods. Had the approach to use DCP2 been successful, this would be the only method to select for the whole m^{6}Am nucleotide.

The final TLC of this piece of work is currently missing. This is the TLC assessing the first nucleotide of the transcripts that were decapped by DCP2 and starting with a 5′ OH and a 5′ P. Even without the complete picture, we are confident that the paper previously published saying DCP2 does not decap transcripts that start with m^{6}Am is incorrect. In the TLC that assesses the transcripts that were not decapped there is no m^{6}Am and in the decapped transcript starting with a 5′ OH there is m^{6}Am (Figure 5.4). This result means that DCP2 cannot be used as the first step of m^{6}Am enriched sequencing.

5.3.2. Technical differences with published findings

DCP2 decapping is a well-studied process. It is known that DCP2 removes the guanine cap leaving a 5′ phosphate so that Xrn1 an exoribonuclease can digest the RNA (Wang et al., 2002b, Jones et al., 2012). In this work a 5′ OH at the start of the RNA after decappng was detected which was not expected. One explanation may be that the DCP2 protein was not completely pure and, there are other proteins present in elutions of DCP2 protein used in the experiments (Figure 5.3). Potentially there are exonucleases and other enzymes that could affect the results and aid the DCP2 enzyme in decapping or digesting the RNA immediately after decapping. To test for this potential exonuclease activity a negative control would have been to treat RNA with a his tagged maltose binding protein. This would have contained the other proteins that came through in the purification unless these proteins are interacting with DCP2.

The buffer the experiment was performed in may also potentially affect the result. In the work previously performed by Mauer et al., 2016 they did not test specifically where the DCP2 enzyme was cutting. RNA 5′ Pyrophosphohydrolase (RPPH) can have two functions depending on the ions present in the buffers. For example, if Cl^{2+} ions are present the enzyme will remove triphosphates from the 5′ of RNA and if SO_4^{2-} ions are present the enzyme removes the 7′ methyl guanine cap. The catalytic domain of RppH and DCP2 are structurally homologous (Wurm et al., 2017). The buffer used in my work has mostly Cl^{2+} ions from NaCl, $MgCl_2$ and $MnCl_2$. Previous work looking at DCP2 decapping used OAc^- ions through MgOAc and KOAc supplemented with $MnCl_2$ (Wang et al., 2002b, Stevens and Maupin, 1987, Liu et al., 2008). The buffer used in this work was the same as in Mauer et al., 2016. The difference in buffers between the previous work and this study may play an important role in the activity of the enzyme.

5.3.3. DCP2 *in vivo*

The work completed here investigated the activity of DCP2 *in vitro* decapping RNA on its own. *In vivo* DCP2 works as part of a large complex of proteins. DCP2 interacts with DCP1, EDC4, EDC1 and EDC3. EDC1, EDC3, and DCP1 are activators of DCP2. They change the protein from an open conformational shape to a closed and catalytically active one (Valkov et al., 2016, Mugridge et al., 2018, Chang et al., 2014). DCP1 binds to the N-terminal domain stabilising the closed conformation. It has been reported that DCP2 is only active when bound to DCP1 and EDC1 (Wurm et al., 2017). EDC4 acts as a scaffold for the DCP2/DCP1 protein complex as the interaction between these two proteins is weak (Chang et al., 2014).

Some work has been performed in yeast showing that DCP1 is required for *in vitro* and *in vivo* decapping (Sakuno et al., 2004, She et al., 2008). However, there is much work including the work presented here that shows that DCP2 can decap RNA on its own (Wang et al., 2002b, Stevens and Maupin, 1987). EDC1 and DCP1 increase the efficiency but DCP2 may decap RNA at a reduced rate in its open conformation.

5.3.4. Regulation of translation by DCP2

If DCP2 is preferentially decapping m^6Am and cap1, it may be regulating translation. There is evidence that in rat brain, all RNA that is bound to polysomes has a cap1 modification including m^6Am (Ven Murthy, 1982). Cap1 may denote transcripts which are, or will shortly be, translated. DCP2 may be regulating translation through decapping and starting the degradation process for transcripts that are no longer needed.

In the cytoplasm along with the decapping enzymes there is cytoplasmic recapping. In the cytoplasm there are RNA guanylyltransferases and 5′-phosphatases which will add a guanine cap onto 5′ phosphate RNA (Otsuka et al., 2009). CMTR2 can also be found in the cytoplasm along with these enzymes that recap RNA transcripts, and it has been hypothesised to add the cap1 modification as well as the cap2 modification. Potentially if cap1 is required for polysome binding in rat brain, then the cap1 modification will need to be dynamic and added to transcripts that are required to be translated. This may happen in the cytoplasm in some cases. The whole of the cap may be removed and put back on or CMTR2 may add the 2′-O-methylation to the first nucleotide.

There are many decapping enzymes found in the cell such as DXO and Nudix proteins Nudt3 and Nudt16 (Jiao et al., 2013, Song et al., 2010, Grudzien-Nogalska et al., 2016). DCP2 has been shown to modestly contribute to bulk mRNA decapping (Song et al., 2010) and is only expressed in some tissues, including human and mouse brain (Song et al., 2010) where m^6Am is known to be enriched (Kruse et al., 2011, Dominissini et al., 2012). It may be that different populations of RNA denoted by different cap structures are degraded by different pathways with different decapping enzymes.

Chapter 6: Identification of m^6Am Binding Proteins

6.1 Preface

Modification by m^6A is known to regulate RNA processes by causing alterations to RNA secondary structure (Spitale et al., 2015, Liu et al., 2015) and through the interaction with m^6A specific binding proteins (Yang et al., 2018). The effect of m^6Am on mRNA itself is debated. It has been claimed and counter claimed to increase RNA stability and translatability (Mauer et al., 2016, Akichika et al., 2018, Sendinc et al., 2019, Boulias et al., 2019). However, it is not known whether any effect on stability or translatability (if indeed true) is through RNA conformational changes or through interaction with RNA binding proteins. Currently there are only two proteins reported to be associated with m^6Am. These proteins are FTO a demethylase or "eraser" (Mauer et al., 2016) and PCIF1 a m^6Am methylase or "writer" (Akichika et al., 2018). There are no known "reader" proteins that interact with m^6Am. To understand the function of this modification, it is necessary to know which proteins bind (or are blocked by) the m^6Am modification. To uncover the identity of any such proteins a protein pull down was undertaken using RNA oligonucleotides with different cap structures, cap0 (m^7GpppA), cap1 ($m^7GpppAm$), and m^6Am (m^7Gpppm^6Am).

6.2 Results

6.2.1 Optimisation of substrate synthesis

Three different oligonucleotides were synthesised, which started with the different cap structures cap0 (m^7GpppA), cap1 (m^7GpppAm), and m^6Am (m^7Gpppm^{6}Am). The m^6Am oligonucleotide should detect m^6Am reader and eraser proteins; the cap1 oligonucleotide should detect the m^6Am writer protein; and the cap0 oligonucleotide will detect the Am writer.

There are two methods that could be utilised to make such oligonucleotides, chemical synthesis or *in vitro* transcription. Chemical synthesis results in an OH at the 5′ end. The *Vaccinia* capping enzyme requires a di or tri phosphate at the 5′ end which can be added with two enzymes, PNK and nucleoside 5′ diphosphate kinase. T7 transcription leaves a nucleotide triphosphate at the starting position, so we sought to develop an *in vitro* transcription and enzymatic modification protocol to make these substrates.

T7 polymerase was used to transcribe the oligonucleotides, the *Vaccinia* virus capping system was used to add the 7′ methyl guanine cap, and the RNA Cap 2′-O-Methyltransferase was used to add the 2′-O-methylation to the first nucleotide. To achieve adenosine as the first nucleotide a modified T7 promoter was used, this was the only adenosine in the sequence. The second nucleotide of the sequence was cytosine which can be seen on the TLCs in figure 6.1. Biotinylated uridine was added to the transcription mixture, and was transcribed into the biotinylated polyuridine tail which bound to the streptavidin beads used in the protein pulldowns.

To transcribe the cap0 oligonucleotide adenosine was used in the T7 transcription, and the *Vaccinia* virus capping system added the methyl 7 guanine cap. To transcribe the cap1 oligonucleotide 2′-O-Me ATP was initially used, however it was found that the T7

polymerase could not incorporate the 2′-O-Me ATP nucleotide into the oligonucleotide (Figure 6.1b). As a result the cap0 oligonucleotide was methylated using RNA Cap 2′-O-Methyltransferase. To transcribe the m^6Am oligonucleotide N^6-methyl-ATP was used in the transcription mixture and the 2′-O-methylation to make the m^6Am was added by RNA Cap 2′-O-Methyltransferase. The *Vaccinia* virus capping system added the methyl 7 guanine cap.

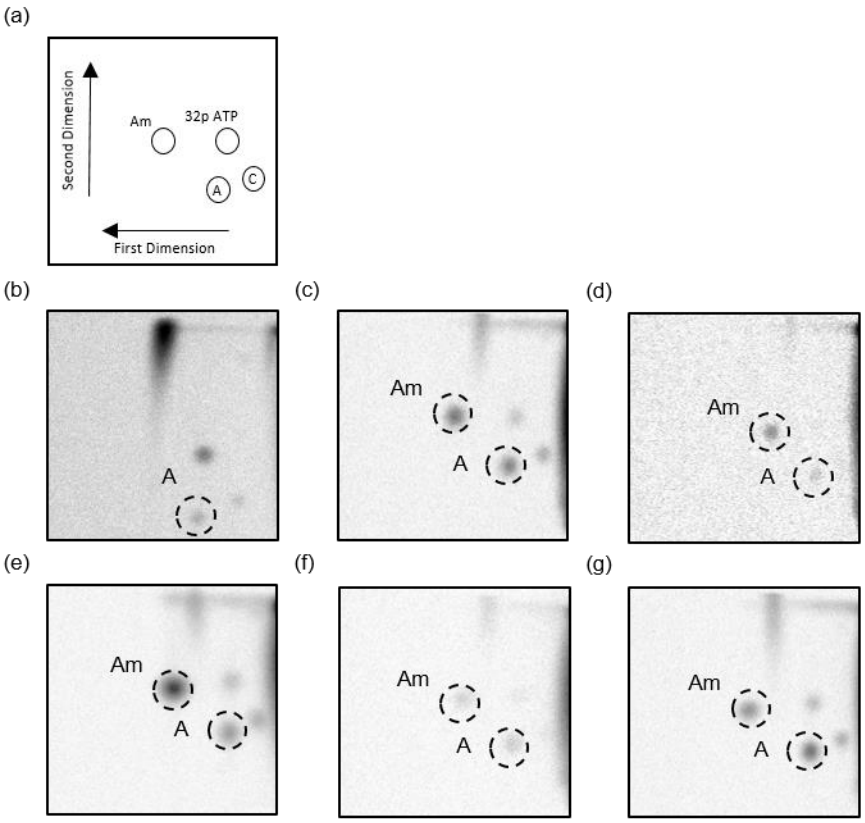
The *Vaccinia* capping system and RNA Cap 2´-O-Methyltransferase are not 100% efficient. The efficiency is further decreased because the oligonucleotides that are being capped and methylated are 28 nucleotides long. The capping and methylation efficiency was optimised by the addition of 2mM SAM, and 2mM $MgCl_2$ was tested. The amount of the two enzymes used to add the 7′ methyl guanine cap and the 2′-O-methylation, *Vaccinia* capping system and RNA Cap 2´-O-Methyltransferase, was increased from 1ng of each enzyme in the reaction with 37.5ng of RNA was increased to 1ng of each enzyme with 18.75ng and 12.5ng of RNA (Fuchs et al., 2016). In the text, this ratio is referred to as 1:37 ng enzymes:RNA.

Figure 6.1a indicates the position of modified nucleotide Am, the nucleotides A and C as well as 32p ATP derivative on a TLC. A spot at the 32p ATP position indicates 32pATP is left over from labelling and shows the clean-up was not efficient. Figure 6.1 presents the first nucleotide of the transcribed oligonucleotides to assess the capping and 2′-O-methylation efficiency.

Using the standard reaction volumes of *Vaccinia* virus capping system and RNA cap 2′-O-methyltransferase gave a capping efficiency of 25% and methylation efficiency of 59% (Figure 6.1c). The capping efficiency was calculated by working out the input RNA and the RNA after everything without a cap was removed. This value will be skewed

because of the RNA lost when the reactions were cleaned up using phenol chloroform and the RNA precipitated. The methyltransferase efficiency was calculated by doing cap adjacent TLCs then using the Quantity One software (Bio-Rad) to calculate the ratio of A:Am.

The reaction with the highest capping efficiency was when a final concentration of 2mM $MgCl_2$ was added to a ratio of 1:37 ng enzymes:RNA, which resulted in 34% of the RNA being capped. This reaction though had the lowest methylation efficiency at 41% of the adenosines being methylated (Figure 6.1g). The reactions with an increased enzyme 1:18.75 and 1:12.5 ng enzymes:RNA, resulted in 83% of transcripts being methylated. The capping efficiency for these reactions was 28% and 19% respectively (Figure 6.1c,d). Adding additional SAM to the reaction lowered both capping efficiency and methylation compared to the standard reaction (Figure 6.1f). The transcription reaction which was subsequently used in this study was a compromise between capping and methylation efficiency and was 1ng of each enzyme to 18.75 ng RNA with no other additives.



	Enzyme:RNA (ng)	Additions	Capping efficiency	Methylation efficiency (A:Am)
c)	1:37.5	-	25%	41:59
d)	1:18.75	-	28%	17:83
e)	1:12.5	-	19%	17:83
f)	1:37.5	0.4mM SAM	10%	56:44
g)	1:37.5	2mM $MgCl_2$	34%	59:41

Figure 6.1. Optimisation of capping and 2′-O-methylation of the transcribed cap1 oligonucleotide*.* TLCs of the first nucleotide of the cap1 oligonucleotide generated through methylation of the cap0 oligonucleotide. Below a table showing the capping efficiency and methylation efficiency of the conditions tested. The same volume of both the *Vaccinia* capping system enzyme and the RNA 2′-O-methyltransferase was used, the ratio of both these enzymes to the RNA is referred to as enzymes:RNA. These TLCs were carried out once. a) Annotation

of the position of dots indicating the different modification states. b) TLC showing that T7 polymerase cannot incorporate 2′-O-Me ATP nucleotide into RNA. c) TLC showing the first nucleotide of the oligonucleotides methylated and capped with the enzymes:RNA ratio of 1:37.5. d) TLC showing the first nucleotide of the oligonucleotides methylated and capped with the enzymes:RNA ratio of 1:18.75. e) TLC showing the first nucleotide of the oligonucleotides methylated and capped with the enzymes:RNA ratio of 1:12.5. f) TLC showing the first nucleotide of the oligonucleotides methylated and capped with the enzymes:RNA ratio of 1:37.5 with the addition of 0.4mM SAM. g) TLC showing the first nucleotide of the oligonucleotides methylated and capped with the enzymes:RNA ratio of 1:37.5 with the addition of 2mM $MgCl_2$ (final concentration).

6.2.2 Proteomic analysis of protein pull downs

Cap0, cap1 and m^6Am RNA oligonucleotides (Figure 6.2) were incubated in rat brain protein lysate with protease inhibitors, pulled down using streptavidin coated magnetic beads and prepared for mass spectrometry analysis. Initially pull downs were performed on three replicates but the proteomics gave poor data due to the carryover of detergent. The procedure was repeated using the same method but with different buffers to minimise detergent carry over (chapter 2, 2.2.14) and this time four replicates were performed. This was more successful than the first attempt but still had some detergent contamination and did not contain enough input material for high quality mass spectrometry analysis. The pull downs of interacting proteins was carried out at University of Nottingham and the enriched proteins were sent to the Dominissini group in Tel Aviv for mass spectrometry analysis.

560 proteins were detected from the four replicates of the second pulldowns prior to the data processing and quality control. The data was analysed using Label Free Quantification (LFQ) intensities which normalises the mass spectrometry intensity value across samples. The replicate values were then grouped and the proteins filtered. 236 proteins were detected in 3 out of the 4 replicates in all the conditions cap1, cap2, m^6Am and bead only. Missing values were imputed using a normal distribution algorithm and an ANOVA was performed to assess which proteins were significantly enriched compared to the bead only control condition. The details of the proteins and p-values for these proteins are presented in table 6.1.

Only eight proteins were found to be significantly bound to the RNA oligonucleotides (ANOVA, S0= 0.1, FDR=0.05) which is an indication of proteins binding to cap0, cap1 and m^6Am oligonucleotides rather than one modification state only (Table 6.1). Six of the eight proteins, SRSF3, HNRNPR, U2AF2, SNRNP70, HNRNPU, U2AF1, are

involved in mRNA splicing, which is consistent with the function previously suggested for the cap structure (Ajiro et al., 2016, Sutandy et al., 2018, Braunschweig et al., 2013, Geuens et al., 2016). The function of the other two proteins EWSR1 and LENG8 is not yet well understood. EWSR1 contains a RNA binding domain and has been linked to various cellular processes including gene expression, cell signalling, and RNA processing and transport (Yoon et al., 2018, Ribadeneyra and Amin, 2016). LENG8 is thought to be involved in receptor signalling of leukocyte receptors (Nagase et al., 2001).

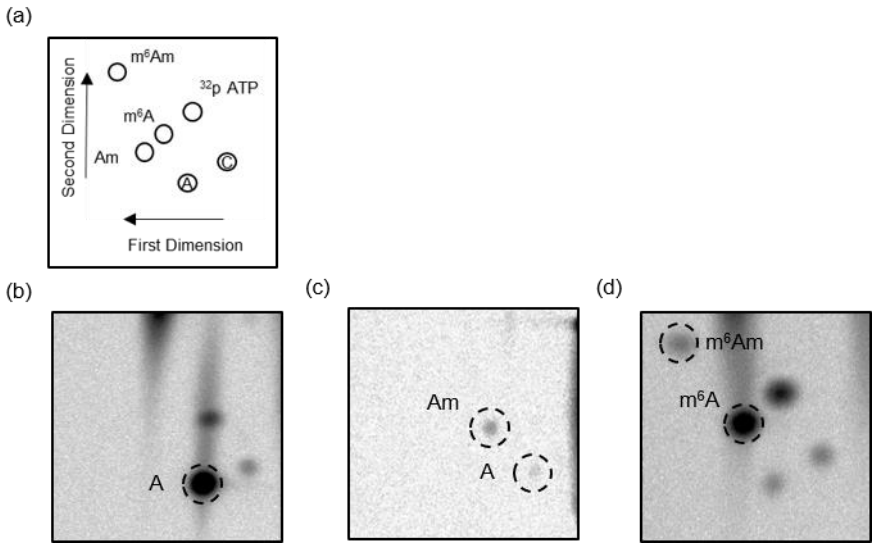


Figure 6.2. TLC showing the first nucleotide of the three transcribed oligonucleotides. a) TLC annotated with the A and C nucleotides, Am, m^6A and m^6Am modified nucleotides and ^{32}p ATP. b) TLC showing the first nucleotide of the cap0 oligonucleotide. c) TLC showing the first nucleotide of the cap1 oligonucleotide. d) TLC showing the first nucleotide of the m^6Am oligonucleotide.

Uniprot ID	Protein Name	Peptides (N)	Coverage (%)	p-Value	Description
Spliceosome proteins					
B2RZ74	SNRNP70	21	39.9	0.0072	Small nuclear ribonucleoprotein 70 (U1)
A0A0U1RRV7	SRSF3	16	67.7	0.029	Serine/arginine-rich splicing factor 3-like
Q7TP17	U2AF1	9	33.9	0.01	U2 small nuclear RNA auxiliary factor 1
F2Z3T9	U2AF2	14	40	0.000	U2 small nuclear RNA auxiliary factor 2
HNRNP components					
Q566E4	HNRNPR	32	48.9	0.011	Heterogeneous nuclear ribonucleoprotein R
Q6IMY8	HNRNPU	46	49.6	0.01	Heterogeneous nuclear ribonucleoprotein U
Unclassified					
A0A0G2K850	EWSR1	7	16.1	0.008	RNA-binding protein EWS-like
D3ZX52	LENG8	25	32.3	0.006	Leukocyte receptor cluster (LRC) member 8

Table 6.1. Table of proteins which were found to be bound to the three RNA oligonucleotides. The table contains the Uniprot ID, protein name and description with the number of peptides detected by mass spectrometry and the percentage of the protein the peptides. The significance (p-value) was determined using ANOVA (S0= 0.1, FDR=0.05).

To explore the mass spectrometry data further, the stringency of the cut off was reduced. For this analysis both protein pull downs were combined. Mass spectrometry intensity values for the proteins bound to the bead only control were subtracted from the mass spectrometry intensity values of the proteins bound to the RNA oligonucleotides. The data was graded and if there was a positive value in 75% of the replicated conditions, the proteins were annotated as bound to that cap structure. If there was a positive value in 50% or less of the replicates, the proteins were annotated as possibly binding. If there was not a positive number in any of the replicates the protein was annotated as not binding to that RNA cap structure. To determine which proteins bound to the different cap structures, proteins which were categorised as bound to the cap0, cap1 or m^6Am individually.

37 proteins potentially bind cap0 (Table 6.2). These proteins can be generally classified under the heading of RNA processing, e.g. associated with translation, ribosomes, or splicing. Mitochondrial; microtubules and brain related proteins were also found to bind the cap0 structure. First, CMTR1 the cap 1 methylase, which binds to nascent RNA transcripts and adds the 2′-O-methylation was detected. This shows that the method works and can detect cap binding proteins. Pre-mRNA splicing proteins LSM2, LSM4 and LSM8 were identified as well as SERBP1 and RPL23 which interact with one another; SERBP1 is proposed to be involved in mRNA stability and RPL23 is a ribosomal protein. Furthermore EIF2S2 was identified which plays a role in translation forming a ternary complex with GTP and initiator tRNA part of the 43S preinitiation complex. KIF2A, which is associated with mitotic spindles, and SEC13, is involved with the nuclear pore were also pulled down. It may be that these proteins bind to RNA transcripts to transport them out of the nucleus.

17 proteins were found to potentially bind cap1 (Table 6.3). The proteins are mainly HNRNP components, spliceosomal proteins and helicase proteins. HNRNPH3, HNRNPA2B1, CHERP, U2AF1I4, PUF60 and DDX46 are involved in splicing and interact together. EIF5B and RPL37A were also identified. EIF5B is an initiation factor and RPL37A is a ribosomal protein. It is known that EIF5B stimulates 48S initiation complex formation during ribosomal scanning for the start codon (Pisareva and Pisarev, 2014)

25 proteins potentially were found to bind the m^6Am oligonucleotide (Table 6.4), many of which are microtubules associated proteins. This is of particular interest as the m^6Am methylase, PCIF1, is associated with microtubules (Human Protein Atlas). This also highlights a potential function in the cytoplasm as microtubules are important for cytoskeleton intracellular transport and change in the structure of synapses (Dent, 2017, Dent and Baas, 2014). Like with the other cap structures, other potential binding proteins include spliceosome proteins and spliceosome associated proteins such as SRSF4, SRSF6 and SRSF10. In addition, BCLAF1 which acts as a splicing regulator of SRSF10 (Zhou et al., 2014) were also found to potentially bind m^6Am.

In all of the protein pull downs, brain related proteins were identified. For example Myelin Basic Protein (MBP) had the highest percentage coverage of peptides in the cap0 pull down, 94.1%. In the cap1 pull down a different variant of MBP had a coverage of 93.3%. In the m^6Am pull down two subunits of the ATP1 sodium/potassium transport ATPase were identified.

Uniprot ID	Protein Name	Peptides (N)	Coverage (%)	Description
Spliceosome associated proteins				
Q9Y333	LSM2	6	82.1	U6 snRNA-associated Sm-like protein LSm2
V9GZ56	LSM4	5	23.1	LSM4 homolog, U6 small nuclear RNA associated
O95777	LSM8	6	74	LSM8 homolog, U6 small nuclear RNA associated
mRNA Processing				
Q8N1G2	CMTR1	4	5.3	Cap-specific mRNA (nucleoside-2′-O-)-methyltransferase 1
Mitochondrial Proteins				
G3V6D3	ATP5B	14	43.7	ATP synthase subunit beta, mitochondrial
F1LP05	ATP5A1	22	45.9	ATP synthase subunit alpha, mitochondrial
D3ZAF6	ATP5J2	3	25	ATP synthase subunit f, mitochondrial
D3ZD09	COX6B1	4	54.7	Cytochrome C Oxidase Subunit 6B1
Q68FX0	IDH3B	6	17.1	Isocitrate dehydrogenase [NAD] subunit beta, mitochondrial
Q09073	SLC25A5	20	54.4	ADP/ATP translocase 2
Ribosomal				
P62832	RPL23	2	20	Ribosomal protein L23
Microtubules				
O00139-2	KIF2A	15	22.6	Kinesin-like protein
E9PC69	MARK2	8	15	Serine/threonine-protein kinase
Brain related proteins				
V9GYM8	ARHGEF2	8	9.6	Rho guanine nucleotide exchange factor 2
P07340	ATP1B1	7	23.4	Sodium/potassium-transporting ATPase subunit beta-1
Q5XIL1	ATP6V1H	6	19.5	ATPase, H+ transporting, lysosomal V1 subunit H
P13233	CNP	25	51.9	2′,3′-cyclic-nucleotide 3′-phosphodiesterase
A0A0C4DH22	EPB41L1	9	15.9	Erythrocyte membrane protein band 4.1-like
P02688	MBP	28	94.1	Myelin basic protein
A0A0H2UHP9	RAB6A	2	11.1	Ras-related protein Rab-6A

B1WBX6	SMAP2	10	28.5	Small ArfGAP2
G3V733	SYN2	8	22.2	Synapsin-2
Translation				
P20042	EIF2S2	12	29.1	Eukaryotic translation initiation factor 2, subunit 2 beta
Unclassified				
G3V6U3	ALG2	7	18.3	Asparagine-linked glycosylation 2 homolog
P54311	GNB1	3	13.5	Guanine nucleotide-binding protein beta 1
P02091	HBB	7	50.3	Haemoglobin subunit beta
Q96QR8	PURB	14	53.8	7.7Transcriptional activator protein Pur-beta
Q52KS1	PFKM	8	10.6	ATP-dependent 6-phosphofructokinase
D3ZRG7	RBM26	25	24.1	RNA binding motif protein 26
D3ZKC9	VIRMA	12	9.3	Similar to mKIAA1429 (virma) protein
E2RUH2	RNH1	3	7.7	Ribonuclease inhibitor
Q5XFW8	SEC13	7	36	Protein SEC13 homolog
Q8NC51	SERBP1	10	34.3	Plasminogen activator inhibitor 1 RNA-binding protein
A0A087WX29	TARDBP	7	31.7	TAR DNA binding protein

Table 6.2. Proteins which were found to be bound to the cap0 RNA oligonucleotide (m^7GpppA). The table contains the Uniprot ID, protein name and description with the number of peptides detected by mass spectrometry and the percentage coverage of that protein.

Uniprot ID	Protein Name	Peptides (N)	Coverage (%)	Description
Spliceosome proteins				
Q7TP17	U2AF1L4	7	44.5	Splicing factor U2AF 26 kDa subunit
Spliceosome associated proteins				
D3ZAX5	CHERP	8	12.3	Calcium homeostasis endoplasmic reticulum protein
A0A0H2UHZ6	PUF60	21	45.1	Poly(U)-binding-splicing factor PUF60
A0A0G2K8K0	SFPQ	5	9.6	Splicing factor proline/glutamine-rich
HNRNP components				
F1LNF1	HNRNPA2B1	20	58.8	Heterogeneous nuclear ribonucleoproteins A2/B1
P31942	HNRNPH3	7	15.5	Heterogeneous nuclear ribonucleoproteins H3
Brain related proteins				
Q8VIN2	ANXA7	11	26.6	Annexin A7
I7FKL4	MBP	32	93.3	Myelin basic protein
Translation				
A0A0H2UHV4	EIF5B	7	7.1	Eukaryotic translation initiation factor 5B
Ribosomal				
C9J4Z3	RPL37A	2	39.7	Ribosomal protein L37a
Helicase				
Q62780	DDX46	56	48.4	Probable ATP-dependent RNA helicase
D4ADZ6	HELZ	8	4.5	Helicase with zinc finger
Nucleoprotein				
A0A0G2KAT3	RPA1	11	27	Replication protein A subunit

Unclassified

Q02413	DSG1	13	18.8	Desmoglein 1
A0A0G2K3B0	GRB14	1	2	Growth factor receptor-bound protein 14
D3ZKQ4	RABL6	13	15.7	Putative uncharacterized protein
A0A0H2UHA6	TMEM33	3	13	Transmembrane protein 33

Table 6.3. Proteins that were found to be bound to the cap1 RNA oligonucleotide (m^7GpppAm). The table contains the Uniprot ID, protein name and description with the number of peptides detected by mass spectrometry and the percentage coverage of that protein.

Uniprot ID	Protein Name	Peptides (N)	Coverage (%)	Description
Spliceosome proteins				
D3Z8R4	RBM25	19	23.6	RNA binding motif protein 25-like 1
Q5U214	SNRPA	3	18.5	Small nuclear ribonucleoprotein polypeptide A
G3V6S8	SRSF6	11	27.4	Serine/arginine-rich splicing factor 6
Spliceosome associated proteins				
B1WC16	BCLAF1	14	19.4	BCL2 associated transcription factor 1
G3V798	SCAF4	3	3	Splicing factor, arginine/serine-rich 15
Q4KM38	SRSF10	10	39	Serine/arginine-rich splicing factor 10
Microtubules				
Q66HA6	ARL8B	3	15.1	ADP-ribosylation factor-like protein 8B
D3ZZM9	CCSAP	5	25	Centriole, cilia and spindle-associated protein
D3ZT16	MAP6D1	3	27.2	Putative uncharacterized protein RGD1560102_predicted
D3ZNX9	MAPK1IP1L	1	7.1	Mitogen-activated protein kinase 1 interacting protein 1-like
F1LVV4	RCC2	21	48.7	Regulator of chromosome condensation 2
P85108	TUBB2A	31	80.9	Tubulin beta-2A chain
B4F7C2	TUBB4B	33	65.7	Tubulin beta-4B chain
P69897	TUBB5	30	70	Tubulin beta-5 chain
Mitochondria				
P31399	ATP5H	4	31.7	ATP synthase subunit d, mitochondrial
Q64428	HADHA	5	9.4	Trifunctional enzyme subunit alpha, mitochondrial
Brain related proteins				
P06686	ATP1A2	36	31.8	Sodium/potassium-transporting ATPase subunit alpha-2
Q5M9H4	ATP1B2	4	12.4	Sodium/potassium-transporting ATPase subunit beta-2
O35095	NCDN	4	7.1	Neurochondrin
P63012	RAB3A	7	40	Ras-related protein Rab-3A
B0BNJ1	SRI	5	21.2	OC683667 protein; Sorcin

D4A6N8	TBR1	11	21.9	T-box, brain, 1
Unclassified				
Q5BJT0	ARGLU1	4	12.8	Arginine and glutamate-rich protein 1
P19139	CSNK2A1	4	14.1	Casein kinase II subunit alpha
P67874	CSNK2B	2	14.2	Casein kinase II subunit beta
A0A0G2K4R7	RBMS1	6	22.3	RNA-binding motif, single-stranded-interacting protein 1
D3Z8R4	RBMS25L1	19	23.6	RNA-binding motif protein 25-like 1
D4A0U3	ZFP638	17	11.4	Zinc finger, matrin-like (Predicted), isoform CRA_a

Table 6.4. Proteins that bound to the m^6Am RNA oligonucleotide (m^7Gpppm^6Am). The table contains the Uniprot ID, protein name and description with the number of peptides detected by mass spectrometry and the percentage coverage of that protein.

6.2.3 Comparison of proteomic analysis

A database search was performed to investigate if any of the proteins that had been identified in this study were known to bind mRNA and the cap structure. Data sets showing three methods for looking at cell wide protein RNA interactions were chosen. One, Orthogonal Organic Phase Separation (OOPS) (Queiroz et al., 2019); two, Phenol Toluol extraction (PTex) (Urdaneta et al., 2019); three, enhanced RNA interactome capture (eRIC) (Perez-Perri et al., 2018).The eRIC method used carboxylated magnetic beads that have a short oligonucleotide attached that will bind to the poly-A tail of mRNA that has been cross-linked to proteins in the cell. In contrast, the OOPS and PTex methods use the physicochemical properties of the RNA protein complexes to isolate them, followed by RNase treatment to isolate the RNA binding proteins. All three methods used mass spec to identify the proteins.

The OOPs method identified 1784 proteins, the PTex method 3034 proteins and the eRIC method 708 proteins. Although there are important differences in this study compared to the published studies, for example this work is on the RNA cap not the whole transcript, of the 87 proteins identified, 25 (34%) of these were also detected in all three of the published studies. Only the eRIC, identified the same mitochondrial proteins as our findings and this may be a result of the magnetic beads that were used in both studies. Table 6.5 details the proteins from the statistically significant, cap0, cap1 and m^6Am oligonucleotide protein pull downs that were found in all three of the published studies.

Uniprot ID	Protein Name	Peptides (N)	Coverage (%)	Description
Statistically Significant				
A0A0G2K850	EWSR1	7	16.1	RNA-binding protein EWS-like
Q566E4	HNRNPR	32	48.9	Heterogeneous nuclear ribonucleoprotein R
Q6IMY8	HNRNPU	46	49.6	Heterogeneous nuclear ribonucleoprotein U
B2RZ74	SNRNP70	21	39.9	Small nuclear ribonucleoprotein 70 (U1)
A0A0U1RRV7	SRSF3	16	67.7	Serine/arginine-rich splicing factor 3-like
F2Z3T9	U2AF2	14	40	U2 small nuclear RNA auxiliary factor 2
Cap0				
O00139-2	KIF2A	15	22.6	Kinesin-like protein
Q9Y333	LSM2	6	82.1	U6 snRNA-associated Sm-like protein LSm2
P62832	RPL23	2	20	Ribosomal protein L23
Q8NC51	SERBP1	10	34.3	Plasminogen activator inhibitor 1 RNA-binding protein
A0A087WX29	TARDBP	7	31.7	TAR DNA binding protein
Cap1				
D3ZAX5	CHERP	8	12.3	Calcium homeostasis endoplasmic reticulum protein
Q62780	DDX46	56	48.4	Probable ATP-dependent RNA helicase
A0A0H2UHV4	EIF5B	7	7.1	Eukaryotic translation initiation factor 5B
D4ADZ6	HELZ	8	4.5	Helicase with zinc finger
F1LNF1	HNRNPA2B1	20	58.8	Heterogeneous nuclear ribonucleoproteins A2/B1
P31942	HNRNPH3	7	15.5	Heterogeneous nuclear ribonucleoproteins H3
A0A0H2UHZ6	PUF60	21	45.1	Poly(U)-binding-splicing factor PUF60
A0A0G2KAT3	RPA1	11	27	Replication protein A subunit
A0A0G2K8K0	SFPQ	5	9.6	Splicing factor proline/glutamine-rich

m^6Am				
B1WC16	BCLAF1	14	19.4	BCL2 associated transcription factor 1
A0A0G2K4R7	RBMS1	6	22.3	RNA-binding motif, single-stranded-interacting protein 1
F1LVV4	RCC2	21	48.7	Regulator of chromosome condensation 2
G3V6S8	SRSF6	11	27.4	Serine/arginine-rich splicing factor 6
Q4KM38	SRSF10	10	39	Serine/arginine-rich splicing factor 10

Table 6.5. Proteins that were identified to bind to the RNA cap modifications were also found to bind to RNA in database searches. The proteins were present in the OOPS, eRIC and pTEX methods. The table contains the Uniprot ID, protein name and description with the number of peptides detected by mass spectrometry and the percentage coverage of that protein.

6.2.4 Gene ontology of modification binding proteins

Functional Gene Ontology (GO) analysis was performed on the proteins indicated to be potentially bound to the RNA oligonucleotides. It was found that each of the different cap structures were indicated to be enriched for specific GO terms. The proteins that bound to cap0 were found to be enriched for terms relating to organelles and vesicles and the transport of RNA (Table 6.6). A number of GO terms were associated with ATP metabolic processes which relate to mitochondrial proteins identified. These ATP associated GO terms have the highest fold enrichment, the proportion of proteins with these GO terms is much higher in the cap0 proteins. Myelin sheath (p=9.40e-10) was found to be the most significantly enriched term. This is interesting as myelin sheath is not usually associated with mRNA processing per se. Instead myelin sheaths are composed of glial Schwann and oligodendrocytes cells which wrap around the processes of neurones and the facilitate action potential conduction along axons.

Analysis of the cap0 binding proteins using reactome database mostly reflects the terms that came up in the GO analysis (Table 6.7). The most significantly enriched pathway is RNA degradation (p=6.98e-04), which is not evident in GO analysis. There is a second term which is mRNA decay by 5′ to 3′ exoribonuclease (p=0.36). The second most significantly enriched term is oxidative phosphorylation (p=0.004), there is also formation of ATP by chemiosmotic coupling (p=0.043). This is reflected strongly in the GO analysis, with terms relating to membrane and organelle were highly enriched where ATP processes occur. The term splicing (p=0.0335) was indicated in the reactome analysis, but not in the top twenty enriched GO terms.

The GO terms associated with the proteins that were found to bind to cap1 oligonucleotides relate to RNA, mRNA binding, and the nucleus (Table 6.8). Proteins were associated with the terms for RNA processing, in particular splicing (p=1.04e-4),

with splicing via transesterification reactions with bulged adenosine as nucleophile, via spliceosome and via transesterification reactions were evident (p=8.25e-04, p=8.35e-04). These showed the highest fold enrichment values at ~19. Two terms were identified in the cap1 protein reactome analysis which are mRNA Splicing - Major Pathway (p=3.06e-15) and Spliceosome (p=1.76e-04) (Table 6.9).

The GO terms associated with the proteins that were found to bind to m^6Am modification are enriched for nucleotide and compound binding (Table 6.10). Vesicle and organelle were also highly enriched for, extracellular vesicle which had the highest fold enrichment at ~60. Seven of the top twenty terms related to organelles, the most enriched term was intracellular membrane bound organelle (p=3.74e-05). Microtubule and microtubule based processes were the 26^{th} and 29^{th} most significant enriched GO terms with p-values of 0.00095 and 0.0015 respectively. The most enriched reactome term for the m^6Am binding proteins was spliceosome (p=0.001). The next most enriched terms were ion transport by P-type ATPase (p=0.002) and ion homeostasis (p=0.003). These terms also have some of the highest fold enrichment 38.88 and 35.21 respectively (Table 6.11).

GO and reactome analysis for all the capping structures/modifications did not indicate an enrichment for any brain related term. This might be expected as the proteins that bind m^6Am and the other cap structures would not necessarily be brain specific proteins. However, brain related proteins were pulled down and the role of the Am and m^6Am modification might change because of their cellular context.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of proteins
GO:0043209	Myelin sheath	9.40E-10	26.44	1.11E-07	9
GO:0044822	Poly(A) RNA binding	5.92E-07	5.78	6.10E-05	13
GO:0032991	Macromolecular complex	1.32E-06	2.54	2.11E-05	23
GO:0003723	RNA binding	5.64E-06	4.22	1.32E-03	14
GO:0044422	Organelle part	6.72E-06	2.03	1.08E-04	26
GO:0043227	Membrane-bounded organelle	1.66E-05	1.57	4.83E-03	31
GO:0044446	Intracellular organelle part	2.10E-05	2.00	6.11E-03	25
GO:0043226	Organelle	2.21E-05	1.48	3.54E-04	32
GO:0043234	Protein complex	3.03E-05	2.59	8.81E-03	19
GO:0005737	Cytoplasm	3.53E-05	1.66	1.02E-02	29
GO:0031988	Membrane-bounded vesicle	3.57E-05	2.87	1.04E-02	17
GO:0070062	Extracellular exosome	5.29E-05	3.18	1.53E-02	15
GO:0090662	ATP hydrolysis coupled transmembrane transport	5.61E-05	51.84	5.60E-02	4
GO:1903561	Extracellular vesicle	5.64E-05	3.16	1.63E-02	15
GO:0043230	Extracellular organelle	5.76E-05	3.16	1.67E-02	15
GO:1901363	Heterocyclic compound binding	5.96E-05	2.06	1.38E-02	23
GO:0031982	Vesicle	6.67E-05	2.74	1.93E-02	17
GO:0097159	Organic cyclic compound binding	7.62E-05	2.03	1.77E-02	23
GO:0046034	ATP metabolic process	7.95E-05	46.21	1.53E-02	4
GO:0016469	Proton-transporting two-sector ATPase complex	1.22E-04	40.09	3.50E-02	4

Table 6.6. GO Terms associated with the proteins that bound to the cap0 oligonucleotide**.** The table contains the top 20 GO terms with p-value, fold enrichment, bonferroni corrected p-value and number of proteins with that GO term.

Term	Description	P-value	Fold enrichment	Bonferroni	No. of proteins
rno03018	RNA degradation	6.98e-04	21.00	0.043	4
Rno00190	Oxidative phosphorylation	0.004	11.96	0.200	4
Rno03040	Spliceosome	0.035	9.50	0.894	3
R-RNO-430039	mRNA decay by 5′ to 3′exoribonuclease	0.036	52.39	0.902	2
R-RNO-163210	Formation of ATP by chemiosmotic coupling	0.043	43.65	0.938	2
Rno05012	Parkinson's disease	0.043	8.50	0.937	3
Rno01100	Metabolic pathways	0.049	2.35	0.958	7

Table 6.7. Reactome analysis for the proteins that bound to the cap0 oligonucleotide. The table contains the top 20 reactome terms with p-value, fold enrichment, bonferroni corrected p-value and number of proteins with that term.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of Proteins
GO:0044822	Poly(A) RNA binding	2.26E-07	8.64	9.71E-06	10
GO:0003723	RNA binding	4.33E-07	6.28	8.23E-06	11
GO:0003676	Nucleic acid binding	6.51E-05	3.18	1.37E-03	12
GO:0008380	RNA splicing	1.04E-04	18.03	5.75E-02	5
GO:0006397	mRNA processing	1.93E-04	15.35	1.04E-01	5
GO:0000166	Nucleotide binding	2.08E-04	15.21	8.90E-03	5
GO:0006396	RNA processing	2.44E-04	9.15	1.30E-01	6
GO:0016071	mRNA metabolic process	5.41E-04	11.71	2.65E-01	5
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	8.25E-04	19.66	3.75E-01	4
GO:0000398	mRNA splicing, via spliceosome	8.25E-04	19.66	3.75E-01	4
GO:0000375	RNA splicing, via transesterification reactions	8.35E-04	19.57	3.79E-01	4
GO:1901363	Heterocyclic compound binding	9.38E-04	2.26	9.38E-02	13
GO:0097159	Organic cyclic compound binding	1.08E-03	2.22	1.08E-01	13
GO:0016604	Nuclear body	1.96E-03	14.67	2.79E-01	4
GO:1901265	Nucleoside phosphate binding	2.91E-03	3.45	5.94E-02	8
GO:0044428	Nuclear part	3.11E-03	2.95	2.13E-01	9
GO:0005634	Nucleus	3.20E-03	2.11	2.72E-01	12
GO:0030529	Intracellular ribonucleoprotein complex	4.71E-03	27.24	2.54E-01	3
GO:0036094	Small molecule binding	5.67E-03	3.13	4.50E-01	8
GO:0043227	Membrane-bounded organelle	5.97E-03	1.54	2.27E-01	15

Table 6.8. GO Terms associated with the proteins that bound to the cap1 oligonucleotide. The table contains the top 20 GO terms with p-value, fold enrichment, bonferroni corrected p-value and number of proteins with that GO term.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of Proteins
R-RNO-72163	mRNA Splicing - Major Pathway	3.06E-05	21.83	1.13E-03	5
rno03040	Spliceosome	1.76E-04	28.49	1.58E-03	4

Table 6.9. Reactome analysis for the proteins that bound to the cap1 oligonucleotide. The table contains the top 20 reactome terms with p-value, fold enrichment, bonferroni corrected p-value and number of proteins with that term.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of Proteins
GO:0000166	Nucleotide binding	5.42E-07	4.28	9.32E-05	15
GO:1901265	Nucleoside phosphate binding	5.42E-07	4.28	9.32E-05	15
GO:0036094	Small molecule binding	1.52E-06	3.94	2.62E-04	15
GO:0043231	Intracellular membrane-bounded organelle	3.74E-05	1.73	8.45E-03	24
GO:1903561	Extracellular vesicle	4.08E-05	56.98	3.14E-03	4
GO:0043229	Intracellular organelle	4.29E-05	1.61	9.69E-03	25
GO:0044424	Intracellular part	7.10E-05	1.46	1.60E-02	26
GO:0005634	Nucleus	9.24E-05	2.18	2.08E-02	19
GO:0044446	Intracellular organelle part	1.17E-04	2.03	2.62E-02	20
GO:0044422	Organelle part	1.74E-04	1.98	3.13E-03	20
GO:0005622	Intracellular	1.84E-04	1.41	4.09E-02	26
GO:0044822	poly(A) RNA binding	2.24E-04	4.85	3.79E-02	9
GO:1901363	Heterocyclic compound binding	2.88E-04	2.10	4.83E-02	18
GO:0043226	Organelle	3.45E-04	1.47	6.19E-03	25
GO:0097159	Organic cyclic compound binding	3.50E-04	2.07	5.84E-02	18
GO:0003723	RNA binding	3.62E-04	3.93	6.04E-02	10
GO:0050684	Regulation of mRNA processing	3.74E-04	26.80	1.84E-01	4
GO:0043227	Membrane-bounded organelle	3.84E-04	1.55	8.35E-02	24
GO:0017111	Nucleoside-triphosphatase activity	4.16E-04	6.55	6.91E-02	7
GO:0016462	Pyrophosphatase activity	5.83E-04	6.15	9.54E-02	7

Table 6.10. GO Terms associated with the proteins that bound to the m^6Am oligonucleotide. The table contains the top 20 GO terms with p-value, fold enrichment, bonferroni corrected p-value and number of proteins with that GO term.

GO Term	Description	P-value	Fold enrichment	Bonferroni	No. of proteins
rno03040	Spliceosome	0.001	16.28	0.057	4
R-RNO-936837	Ion transport by P-type ATPases	0.002	38.88	0.068	3
R-RNO-5578775	Ion homeostasis	0.003	35.21	0.083	3
rno04911	Insulin secretion	0.009	19.31	0.321	3
rno04540	Gap junction	0.009	18.87	0.333	3
rno04964	Proximal tubule bicarbonate reclamation	0.036	50.32	0.804	2
rno04145	Phagosome	0.040	8.60	0.836	3
rno05168	Herpes simplex infection	0.049	7.69	0.892	3

Table 6.11. Reactome analysis for the proteins that bound to the m^6Am oligonucleotide. The table contains the top 20 reactome terms with p-value, fold enrichment, bonferroni corrected p-value and number of proteins with that term.

6.3 Discussion

6.3.1 Transcription of oligonucleotides

Synthesising oligonucleotides is not an efficient process. Only a third of the substrates were found to be capped. A further problem with this method was that there was a mix of substrates. In the cap1 oligonucleotides there is a mix of A and Am and in the m^6Am oligonucleotide there is a mix of m^6A and m^6Am. More controls sould have been completed. The A pull down can act as a control for the Am pull down. Potentially proteins that bind in both pull downs could bind both cap structure or could be binding the A oligonucleotide.

There is an alternative method which may be superior for future validation studies. This method uses synthesised oligonucleotides which start with A, Am or m^6Am. The 5′ OH of the oligonucleotide is changed into a 5′ diphosphate using nucleoside 5′ diphosphate kinase. The guanine cap is added to this using the *Vaccinia* capping enzyme. The m^6Am nucleotide cannot be purchased. However, colleagues in the School of Chemistry have synthesised the m^6Am phosophoramidite and used this to make RNA oligonucleotides that start with this modification. The 5′ diphosphate kinase is no longer commercially available and so the enzyme would have to be expressed. Alternatively, as part of a separate collaboration our colleagues in the School of Chemistry have recently adapted a method for converting a nucleoside to a nucleotide triphosphate, and it should be possible to extend this methodology to add a triphosphate to an RNA oligonucleotide. It may also be possible to add the m^7G cap to the oligonucleotides when synthesising them (Mauer et al., 2016).

The purity of the oligonucleotides needs to be improved. The enzyme DXO could be used to select for nucleotides that start with Am (Picard-Jean et al., 2018). There is

currently no way to select for m^6Am using an enzymatic approach, although it was hoped that DCP2 would facilitate this (Chapter 5).

6.3.2 Protein pull downs

The protein pull downs only identified 8 significantly enriched proteins. This was likely in part due to insufficient input material being used, and experiments were further compromised by detergent contamination. The synthesis of the oligonucleotides is difficult to scale up as such a large amount of input material and enzymes are required. Potentially if the enzymes were expressed and produced in the lab more could be used with a lower cost. To improve the protein pull downs the use of Ni Sepharose excel histidine-tagged protein purification resin may potentially reduce detergent contamination, and binding of proteins to the beads themselves.

Another consideration is that the proteins and RNA were not cross linked. This means potentially that transitory interactions would not be detected. However, because RNA and proteins were not cross linked proteins and RNA in close proximity that were not interacting, should not have been pulled down and this should have helped reduce the number of false positives.

6.3.3 Writers and erasers

CMTR1, the cap1 methylase, was identified binding to the cap1 oligonucleotides. PCIF1 and FTO were not found in the pull down experiments. It could be that the interaction between Am and PCIF1 and m^6Am and FTO are transitory so would not be detected in this experimental setup.

If the cell lysate buffer did not rupture the nuclei of the cells, the cell lysate may be predominantly cytoplasmic, FTO and PCIF1 might not have been detected as they are

predominantly nuclear proteins. There is conflicting evidence about the location of FTO in cell, it has been reported that in the hypothalamus FTO can be found in the cytoplasm of neurons under starvation conditions (Vujovic et al., 2013). However two papers in mouse embryonic fibroblast (MEF) cells show that FTO can consistently be found in the cytoplasm (Aas et al., 2017, Gulati et al., 2014). *In vivo* hypothalamus tissue will be different to MEFs in culture. If the first paper is correct, then the cap binding complex may protect m^6Am from demethylation in the nucleus and only under stress conditions when FTO translocates to the cytoplasm is m^6Am demethylated.

6.3.4 m^6Am reader proteins

The protein pull downs completed in this study show that different proteins are binding to different cap structures cap0, cap1, and m^6Am. Potentially some of these proteins are binding just to the m^6Am modification and hence regulating the metabolism of the mRNA transcript in a modification dependant way. Most of the proteins that were pulled down are cytoplasmic proteins, but there were also some nuclear proteins. The m^6Am modified oligonucleotides have identified brain associated proteins like Sodium/potassium-transporting ATPase subunits which are responsible for maintenance of the electrochemical gradient in neurons. It is not clear why these proteins are potentially binding m^6Am methylated RNA oligonucleotides. However it may be that highly abundant proteins are being carried through the protein pull downs.

Among the proteins that bound to the m^6Am oligonucleotide, four proteins are associated for the GO term extracellular vesicle. There is some evidence that mRNA can travel between neurones and between the hematopoietic system and neurones (Pastuzyn et al., 2018, Ridder et al., 2014). This area has not been intensively studied

but potentially m^6Am could be involved in regulating the RNA that travels between the cells in extracellular vesicles.

m^6Am may also have a role in more general cell processes. When looking at the proteins that bind to m^6Am there are a number of spliceosome and slicing associated proteins. There are also three proteins with known RNA binding domains. These proteins may be m^6Am reader proteins and with further validation, new functions of these proteins may be found.

6.3.5 Future studies

Epitranscriptomics is a relatively new field and work has focused on the internal methylation m^6A. There are many reader proteins associated with this modification. Currently there are no recorded reader proteins associated with Am or 5Cm. This is either the result of lack of research in this area or that m^6A is unique in having reader proteins. The modifications themselves may be what affects the RNA such as Am blocking binding of IFIT and RIG1.

The proteins that were found to be bound to the m^6Am oligonucleotide are all potential m^6Am reader type proteins. The interaction of these proteins should be validated through gel electrophoresis mobility shift assay (EMSA) in which protein nucleic acid complexes are detected because they move at a slower rate than just nucleic acid alone. Further validation of these protein would mean that the m^6A modification is not unique in interacting with a large number of proteins affecting RNA metabolism and cell processes and that other modified nucleotides do the same.

Chapter 7- General Discussion

7.1 Global changes in cap adjacent methylation

In this study, the global level of Am and m^6Am methylation was found to both differ between species and between individuals of the species. Moreover, the levels of Am and m^6Am did not show patterns correlating to where in the evolutionary tree each organism can be found. This may result from the variety of regulatory mechanisms found between species. For example, dosage compensation equalises expression of genes between different sexes. In *D. melanogaster* m^6A methylation translationally represses male specific lethal 2 (*msl-2*) to prevent dosage compensation in females through m^6A guided splicing of sex lethal (*sxl)* (Haussmann et al., 2016). Potentially cap adjacent methylation may regulate expression in similar processes.

Currently there are a number of theories regarding the function of m^6Am and the effect this modification has on mRNA. This modification has been debated to increase RNA stability and translatability (Mauer et al., 2016, Wei et al., 2018, Sendinc et al., 2019, Akichika et al., 2018). There is limited evidence that the marine algae *E. siliculosus* contains the m^6Am modification. As this is a relatively simple organism compared to higher species such as *Homo sapiens*, it may be that the m^6Am modification has the same effects on mRNA at the molecular level, but in each cell and in each organism different transcripts may be methylated with different proteins binding the modification resulting in distinct cellular and physiological effects.

The variation observed between individuals of the same species may be a result of genetic and environmental differences. Each individual will react at the molecular, cellular and physiological levels to internal and external stimuli with variation in the mechanisms used. The innate immune system is found in all multicellular organisms. Its origin can be traced back to the first eukaryotic cells (Buchmann, 2014). It is highly conserved between vertebrates, with elements found in plants and invertebrates

(Boller and He, 2009, Jones et al., 2016). For example, one current known consequence of cap1 methylation which includes the Am modification, is in preventing innate immune response RNA degradation. Cap1 methylation inhibits the binding of RIG1 and IFIT1 proteins so that the innate immune system is not triggered (Abbas et al., 2017, Schuberth-Wagner et al., 2015). If the immune response had been activated or indeed any other system which may influence cap adjacent modification in the brain this may have led to differences between individuals.

7.2 Molecular function of cap adjacent methylation

Proteins that bind the Am and m^6Am modifications will help to define the function of the modification. Some processes will be ubiquitous and others molecular consequences might be cell type specific. Work associated with m^6A calls these proteins that bind to the modification “readers”. Currently m^6A is the only known RNA modification to have “reader” or binding proteins, this may be because biologically it is the only modification with proteins which act as reader effector proteins, or there is a lack of research into other RNA modifications such as cap adjacent Am and m^6Am.

In the present study, observations have been made which have led to hypothesises about the molecular function of the cap adjacent modifications. The data presented here suggests that Am and m^6Am modifications may distinguish sub populations of mRNA and influence their fates. In this study it has been shown that DCP2 potentially preferentially decaps RNAs with Am or m^6Am. This is in contrast with Jaffrey et al., 2016 but is consistent with the work of Sendinc et al., 2019 and Akichika et al., 2018 which found m^6Am does not increase mRNA stability. The decapping enzyme DXO has been shown to exclusively decap cap0 transcripts lacking any modification to the first nucleotide (Picard-Jean et al., 2018). This suggests that RNA sub populations

which have different cap adjacent modifications may have different regulation and degradation pathways.

The guanine cap is known to be needed for efficient splicing and the assembly of pre splicing complexes (Patzelt et al., 1987). The work in this study highlights splicing proteins bound to both Am and m^6Am. It may be that the different proteins bind the cap modifications so that splicing occurs at the correct time in transcription. The cap binding complex is reported to be bound to RNA throughout processing events in the nucleus. This complex is made of two core proteins CDC20 and CDC80, other proteins that make up this complex change throughout RNA processing (Giacometti et al., 2017, Gonatopoulos-Pournatzis and Cowling, 2014). It could be that some of these splicing associated proteins from the protein pulldowns in this study may join this complex resulting in the occurrence of specific splicing events.

7.2.1 Am binding proteins indicate the molecular effect of Am methylation

My findings highlight that putative Am binding proteins were associated with translation and the removal of RNA secondary structure. The helicases DDC46 and HELZ were found to bind to the modification in pull down experiments performed in this study. It is known that DHX15 binds CMTR1 to remove secondary structure of 5′ RNA so that RNA processing and translation can occur (Toczydlowska-Socha et al., 2018). These helicases binding Am may have a similar function.

Am is only one of the four possible cap1 modifications. It would be interesting to see if different proteins bind to Am, Cm, Gm and Um through a protein pull down approach. Transcripts starting with Cm could be particularly interesting, as only 17% of transcripts are reported to start with a C (Schibler et al., 1977). Transcripts that start with C and

are followed by 4–14 uninterrupted pyrimidine residues are called terminal oligopyrimidine tract (TOP) messages. These transcripts encode proteins required for ribosome biogenesis, and the synthesis of these proteins is tightly controlled and sensitive to growth rate (Hamilton et al., 2006). Potentially because of this they may be regulated through different pathways and so have different RNA binding proteins.

7.2.2 m^6Am binding proteins indicate the molecular effect of m^6Am methylation

The findings from this study indicate that proteins associated with molecular processes such as splicing and nucleotide binding were bound to the m^6Am modification. Other proteins found to bind to m^6Am were associated with microtubules and extracellular vesicles (EVs). This suggests that Am and m^6Am provide different functions at the molecular level.

Microtubules are how proteins, metabolites and RNA are transported throughout the cell (Franker and Hoogenraad, 2013). In neurons, specific RNAs are transported and held at the synapse so that local translation can occur during synaptic plasticity. To test whether m^6Am is specifically involved in synaptic transport, RNA from the synapse could be isolated and exo-m^6Am-seq and limited nucleotide sequencing could be performed to identify the transcripts with m^6Am.

There is increasing evidence for intercellular communication between neurons, astrocytes and microglia through transport of exRNA cargo types in extracellular vesicles in both healthy and diseased brain (Paolicelli et al., 2019, Wei et al., 2017, Murillo et al., 2019). Modifications on the RNAs found in the EVs has not been studied. Potentially the m^6Am modification could provide a mechanism for tagging RNAs to be

incorporated in to these vesicles. To investigate this hypothesis further, vesicles could be isolated using ultracentrifugation and the RNA and specific modification identified.

As both microtubules and EVs have a function in protein and RNA transport, it is therefore possible that in the brain, that in addition to the regulation of molecular processes, m^6Am is a protein binding tag for transcripts to be moved throughout and between cells.

7.2.3 Transcripts modified by m^6Am suggest cellular functions

The work in this study found transcripts putatively methylated with m^6Am have cell specific functions. In grey matter the methylated transcripts had functions associated with the negative regulation of phosphate metabolism, the transcripts encoded protein kinases, the phosphorylation of proteins is important for cell function in general as well as specifically for neuronal function and synaptic plasticity (Walaas and Greengard, 1991). In the white matter sequencing, the putatively m^6Am methylated transcripts were found to be associated with processes such as endocytosis and interleukin signalling, which is an important process in microglia and the immune system (Gleeson, 2014, Erta et al., 2012).

It may be that in addition to differences across tissue and cell types, different mRNAs are methylated in different subcellular locations of the cell. It is well known that subcellular localisation of mRNA is important for local translation at the synapse (Zappulo et al., 2017, Middleton et al., 2019). However in most cells mRNA shows patterns of specific localisation in the cell, they are not uniformly abundant everywhere which is known as an “asymmetric distribution”. Of the 25% of the Drosophila transcriptome, 3370 genes, that were analysed in Drosophila embryos at different

stages, 70% of these mRNA species were asymmetrically distributed (Lécuyer et al., 2007). The largest percentage of the transcripts (55%) were located to the poles of the cells. m^6Am may be found on different mRNAs stabilising or increasing translation of key transcripts for the process that are occurring in that particular subcellular compartment.

New technologies such as High-Density Spatial Transcriptomics (HDST) have been developed which could assist in identifying methylated mRNAs found in subcellular locations. This method involves tissue sections mounted to glass slides coated with individually barcoded oligoDT. Light microscopy can be used to visualise the tissue as normal before the cells are permeabilised so that the RNA in the cell can bind to the tagged oligoDT which contains unique barcodes for each region of the slide. RNA sequencing methods can subsequently be performed and the location tags in the sequencing will identify the location of the transcript in the tissue (Vickovic et al., 2019).

7.3 Cap adjacent methylation and neurodegenerative disease

In the Parkinson's disease and Lewy body with dementia brain samples studied, an increase in both Am and m^6Am methylation was found. In the disease samples, it was evident that there was a correlation between the severity of the pathology and the level of methylation. However the small number of individuals investigated limits interpretation. This correlation could be further examined by investigating the role of methylation in disease progression and which transcripts are being methylated.

In the early stages of Parkinson's disease synaptic function becomes impaired as a result of α-synuclein oligomers and protofibrils at the synapse (Ingelsson, 2016, Li et al., 2001, Garcia-Reitböck et al., 2010, Prots et al., 2018). If RNA methylation is part of a mechanism regulating local translation and RNA availability at synapses, the

increase in Am and m^6Am in the disease samples could relate to aberrant α-synuclein protein production and hence pathological processes. Alternatively, increased methylation could relate to a need to upregulate the production of plasticity proteins involved in synaptic function which have been displaced from the synapses owing to the presence of α-synuclein aggregates.

In Parkinson's disease there has also been reported aberrant lncRNA expression. It have been shown that both *JHDMID-AS1* and *TAL1* lncRNA are down regulated in Parkinson's disease (Zhou et al., 2018). It is not fully understood the effect on disease of this misregulation. However, *JHDMID-AS1* is upregulated in cancer cells and, promotes tumorigenesis through upregulating angiogenesis and triggering inflammation (Kondo et al., 2017). TAL1 is a transcription factor involved with microglial aging (Wehrspaun et al., 2015). The down regulation *TAL1* lncRNA may mean an increase in the TAL1 protein which is known to play crucial roles in neurodegeneration and brain ageing through neuroimmune functions (Galatro et al., 2017). It is possible that the aberrant expression of lncRNAs at later stages of disease may be influenced by RNA methylation such as m^6Am changing the fate of the RNA.

Although the primary pathology of Parkinson's disease and Lewy body dementia are Lewy bodies composed of α-synuclein which spread throughout the brain, chronic inflammation is associated with the later stages of Parkinson's disease (Tansey and Goldberg, 2010, Tansey and Romero-Ramos, 2019). The increase in cap adjacent methylation in the Parkinson's disease and Lewy body dementia samples might potentially be as a result of the cells methylating "self" RNA so the innate immune response does not degrade these RNAs. It would be insightful to understand which transcripts are methylated with Am and m^6Am during disease progression to understand the changes that occur in specific cell types and at the synapse.

7.4 Cap adjacent methylation and memory formation

For cap adjacent methylation to be involved in memory formation, it is likely to have a role in translation and in local translation at activated synapses. Eukaryotic initiation factors bind to the 5′ of the RNA scanning for the translation initiation site and recruiting the ribosomal proteins (Pestova and Kolupaeva, 2002). In this work, EIF5B was found to bind to Am which links the modification to translation. Evidence from studies using X-ray crystallography indicate that the first two nucleotides of the transcript affect EIF4E binding. A Niedzwiecka et al., reported that when the first nucleotide after the guanine cap is a guanine there is a disordered arrangement of the EIF4E protein (Niedzwiecka et al., 2002). However when the first nucleotide is an adenosine the C terminal loop interacts with the first nucleotide (Tomoo et al., 2002). Therefore, potentially Am and m^6Am modified nucleotides might affect the binding of EIF4E and hence have a role in the regulation of translation.

7.5 Future directions of research into cap adjacent methylation research

This thesis presents observations that support m^6Am and Am having roles in mRNA processing which may impact on cell specific and subcellular specific function. Misregulation of cap adjacent methylation may have a role in many brain diseases such as Parkinson's disease and Lewy body dementia. To enable future research which would confirm and extend these preliminary findings, the detection of the m^6Am modification on individual transcripts needs to be improved.

Currently, the only sequencing method used to detect m^6Am that is not bioinformatics based is exo-m^6Am-seq (Sendinc et al., 2019). However this method still requires the use of antibodies, the m^6A antibody used binds to N^6-methyladenosine so both m^6Am and m^6A at the 5′ end of the mRNA will be immunoprecipitated. To detect the N^6-

methyladenosine and the 2′-O-methyl ribose of the m^6Am modification exo-m^6Am-seq could be paired with a new method for limited nucleotide sequencing which is currently in the process of being optimised which detects cap1 methylation (N Archer unpublished). In this strategy, cap1 modification are detected by a modified RTL-P approach (Dong et al., 2012), to enable a transcriptome wide mapping of the cap adjacent modifications by a simple “before and after” differential gene expression comparison of RNA reverse transcribed under normal compared to limited dNTP concentrations. If transcripts have a peak in the exo-m^6Am-seq and are lost under limited nucleotides reverse transcription, the transcript would most likely have m^6Am as the starting nucleotide.

Nanopore sequencing is another method which could potentially be used to detect m^6Am. As the RNA goes through the channel, the current across the channel changes for each nucleotide. As a result the m^6Am modification would be detected as a whole not as the two separate methylations that are present on the adenosine. This method should also be able to quantify how many of the transcripts of a specific gene have the m^6Am modification. This would lead into a number of new questions such as: are all transcripts methylated in the same way or is there variation? Nanopore is a new technology and as a result it currently has a high error rate and low throughput for RNA sequencing. The technology is improving and it is being used for sequencing m^6A in published papers with error correction (Zhao et al., 2019).

7.5.1 Function of m^6Am in healthy and diseased cells

Understanding the molecular function of cap adjacent methylation will inform the understanding of the function at the cellular and physiological levels. Cap adjacent methylation will most likely be context dependant. The function will depend on the cell

type and what physiological processes are occurring to, and in, the cell. This needs to be understood before explorations of its role in disease. Many processes that are biologically misregulated have a negative effect on cell function. In terms of cap adjacent modifications, this may be through the methylation of inappropriate transcripts, insufficient methylation or misregulation of the proteins that bind the modification.

m^6A methylation has been linked to neurological diseases. Aberrant translation at synapses has been linked to autism, fragile X and other intellectual disorders (Bassell and Warren, 2008, Kelleher and Bear, 2008, Santini et al., 2012). m^6A has a role in local translation at the synapses and so misregulation of such processes potentially has a role in these diseases (Merkurjev et al., 2018). A down regulation of m^6A methylation has been shown to lead to dopaminergic cell death in the 6-OHDA injected Parkinson's disease rat model. The cell death was a result of increased oxidative stress, increased NMDAR1 a subunit of N-methyl-D-aspartate (NMDA) receptor, and increased calcium influx (Chen et al., 2019).

7.6 Concluding remarks

To conclude, this thesis set out to investigate the role of m^6Am in brain. The four aims of this thesis were: 1) to investigate where in evolution m^6Am could be found to aid in the identification of candidate methylase proteins; 2) to assess if there was a perturbation in global m^6Am levels in human neurodegenerative disease; 3) to characterise which transcripts are m^6Am methylated in human tissue; 4) to identify proteins which bind to the m^6Am modification.

I have shown that the m^6Am modification is found in vertebrates and potentially marine brown algae. The m^6Am methylase, PCIF1, was discovered by Akichika et al., 2018.

With this discovery the m^6Am methylase and demethylase were correlated with the m^6Am modification and it was found that invertebrates contained PCIF1 but not the modification. I have observed that in the neurodegenerative diseases, Parkinson's disease and Lewy body dementia, there is variation in m^6Am and Am methylation in hippocampal tissue compared to control samples.

I have putatively shown which transcripts have the m^6Am modification using bioinformatics approaches, and I have shown that they are associated with cell specific functions. The novel sequencing method using DCP2 was not effective, as DCP2 appears to decap transcripts with Am and m^6Am preferentially. This did however highlight that there may be different degradation pathways for sub populations of mRNAs with different cap structures. Finally I was able to discover m^6Am and Am binding proteins which potentially will change the fate of the mRNA that they bind. I have shown that the proteins that bind the two modifications have different functions, Am binding proteins are associated with mRNA processing events while m^6Am binding proteins tend to be associated with potential intra and inter cellular mRNA trafficking. The findings of this thesis advances our understanding of cap adjacent methylation in brain and provides a firm foundation for future studies unravelling disease associated modification mechanisms.

References

2017. Method of the Year 2016: Epitranscriptome analysis. *Nat Meth,* 14**,** 1-1.

AAS, A., ISAKSON, P., BINDESBØLL, C., ALEMU, E. A., KLUNGLAND, A. & SIMONSEN, A. 2017. Nucleocytoplasmic Shuttling of FTO Does Not Affect Starvation-Induced Autophagy. *PLOS ONE,* 12**,** e0168182.

ABBAS, Y. M., LAUDENBACH, B. T., MARTÍNEZ-MONTERO, S., CENCIC, R., HABJAN, M., PICHLMAIR, A., DAMHA, M. J., PELLETIER, J. & NAGAR, B. 2017. Structure of Human IFIT1 with Capped RNA Reveals Adaptable mRNA Binding and Mechanisms for Sensing N1 and N2 Ribose 2′-O Methylations. *Proceedings of the National Academy of Sciences,* 114**,** E2106-E2115.

ABDELHAMID, R. F., PLESSY, C., YAMAUCHI, Y., TAOKA, M., DE HOON, M., GINGERAS, T. R., ISOBE, T. & CARNINCI, P. 2014. Multiplicity of 5′ Cap Structures Present on Short RNAs. *PLOS ONE,* 9**,** e102895.

ADAMOWICZ, D. H., ROY, S., SALMON, D. P., GALASKO, D. R., HANSEN, L. A., MASLIAH, E. & GAGE, F. H. 2017. Hippocampal α-Synuclein in Dementia with Lewy Bodies Contributes to Memory Impairment and Is Consistent with Spread of Pathology. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 37**,** 1675-1684.

AJIRO, M., JIA, R., YANG, Y., ZHU, J. & ZHENG, Z.-M. 2016. A Genome Landscape of SRSF3-regulated Splicing Events and Gene Expression in Human Osteosarcoma U2OS Cells. *Nucleic acids research,* 44**,** 1854-1870.

AKICHIKA, S., HIRANO, S., SHICHINO, Y., SUZUKI, T., NISHIMASU, H., ISHITANI, R., SUGITA, A., HIROSE, Y., IWASAKI, S., NUREKI, O. & SUZUKI, T. 2018. Cap-specific Terminal N 6 -methylation of RNA by an RNA Polymerase II–associated Methyltransferase. *Science***,** eaav0080.

ALBERTIN, C. B., SIMAKOV, O., MITROS, T., WANG, Z. Y., PUNGOR, J. R., EDSINGER-GONZALES, E., BRENNER, S., RAGSDALE, C. W. & ROKHSAR, D. S. 2015. The Octopus Genome and the Evolution of Cephalopod Neural and Morphological Novelties. *Nature,* 524**,** 220.

ALLEN REISH, H. E. & STANDAERT, D. G. 2015. Role of α-synuclein in Inducing Innate and Adaptive Immunity in Parkinson Disease. *Journal of Parkinson's disease,* 5**,** 1-19.

AMINOFF, E. M., KVERAGA, K. & BAR, M. 2013. The role of the Parahippocampal Cortex in Cognition. *Trends in cognitive sciences,* 17**,** 379-390.

AZEVEDO, F., CARVALHO, L., GRINBERG, L., FARFEL, J. M., FERRETTI, R., LEITE, R., FILHO WILSON, J., LENT, R. & HERCULANO-HOUZEL, S. 2009. Equal Numbers of Neuronal and Nonneuronal Cells make the Human Brain an Isometrically Scaled-up Primate Brain. *Journal of Comparative Neurology,* 513**,** 532-541.

BARNESS, L. A. 1961. Phosphorus Metabolism of the brain. *The Journal of Pediatrics,* 58**,** 752.

BASSELL, G. J. & WARREN, S. T. 2008. Fragile X Syndrome: Loss of Local mRNA Regulation Alters Synaptic Development and Function. *Neuron,* 60**,** 201-214.

BASU, J. & SIEGELBAUM, S. A. 2015. The Corticohippocampal Circuit, Synaptic Plasticity, and Memory. *Cold Spring Harbor Perspectives in Biology,* 7.

BELANGER, F., STEPINSKI, J., DARZYNKIEWICZ, E. & PELLETIER, J. 2010. Characterization of hMTr1, a Human Cap1 2 '-O-Ribose Methyltransferase. *Journal of Biological Chemistry,* 285**,** 33037-33044.

BOCCALETTO, P., MACHNICKA, M. A., PURTA, E., PIĄTKOWSKI, P., BAGIŃSKI, B., WIRECKI, T. K., DE CRÉCY-LAGARD, V., ROSS, R., LIMBACH, P. A., KOTTER, A., HELM, M. & BUJNICKI, J. M. 2017. MODOMICS: a Database of RNA Modification Pathways. 2017 Update. *Nucleic Acids Research,* 46**,** D303-D307.

BODI, Z., ZHONG, S. L., MEHRA, S., SONG, J., GRAHAM, N., LI, H. Y., MAY, S. & FRAY, R. G. 2012. Adenosine Methylation in Arabidopsis mRNA is Associated with the 3 ' end and Reduced Levels Cause Developmental Defects. *Frontiers in Plant Science,* 3.

BOKAR, J. A. 2005. The Biosynthesis and Functional Roles of Methylated Nucleosides in Eukaryotic mRNA. *In:* GROSJEAN, H. (ed.) *Fine-Tuning of RNA Functions by Modification and Editing.* Berlin, Heidelberg: Springer Berlin Heidelberg.

BOKAR, J. A., SHAMBAUGH, M. E., POLAYES, D., MATERA, A. G. & ROTTMAN, F. M. 1997. Purification and cDNA Cloning of the AdoMet-binding Subunit of the Human mRNA (N6-adenosine)-methyltransferase. *RNA (New York, N.Y.),* 3**,** 1233-1247.

BOLLER, T. & HE, S. Y. 2009. Innate Immunity in Plants: an Arms Race between Pattern Recognition Receptors in Plants and Effectors in Microbial Pathogens. *Science (New York, N.Y.),* 324**,** 742-744.

BOOT, B. P., ORR, C. F., AHLSKOG, J. E., FERMAN, T. J., ROBERTS, R., PANKRATZ, V. S., DICKSON, D. W., PARISI, J., AAKRE, J. A., GEDA, Y. E., KNOPMAN, D. S., PETERSEN, R. C. & BOEVE, B. F. 2013. Risk Factors for Dementia with Lewy Bodies: a Case-Control Study. *Neurology,* 81**,** 833-840.

BOULIAS, K., TOCZYDŁOWSKA-SOCHA, D., HAWLEY, B. R., LIBERMAN, N., TAKASHIMA, K., ZACCARA, S., GUEZ, T., VASSEUR, J.-J., DEBART, F., ARAVIND, L., JAFFREY, S. R. & GREER, E. L. 2019. Identification of the m6Am Methyltransferase PCIF1 Reveals the Location and Functions of m6Am in the Transcriptome. *Molecular Cell,* 75**,** 631-643.e8.

BRAAK, H., TREDICI, K. D., RÜB, U., DE VOS, R. A. I., JANSEN STEUR, E. N. H. & BRAAK, E. 2003. Staging of Brain Pathology Related to Sporadic Parkinson's Disease. *Neurobiology of Aging,* 24**,** 197-211.

BRAUNSCHWEIG, U., GUEROUSSOV, S., PLOCIK, A. M., GRAVELEY, BRENTON R. & BLENCOWE, BENJAMIN J. 2013. Dynamic Integration of Splicing within Gene Regulatory Pathways. *Cell,* 152**,** 1252-1269.

BRIDI, J. C. & HIRTH, F. 2018. Mechanisms of α-Synuclein Induced Synaptopathy in Parkinson's Disease. *Frontiers in Neuroscience,* 12.

BUCHMANN, K. 2014. Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals. *Frontiers in immunology,* 5**,** 459-459.

BUFFINGTON, S. A., HUANG, W. & COSTA-MATTIOLI, M. 2014. Translational Control in Synaptic Plasticity and Cognitive Dysfunction. *Annual Review of Neuroscience,* 37**,** 17-38.

BUROW, D. A., UMEH-GARCIA, M. C., TRUE, M. B., BAKHAJ, C. D., ARDELL, D. H. & CLEARY, M. D. 2015. Dynamic Regulation of mRNA Decay During Neural Development. *Neural development,* 10**,** 11-11.

CAJIGAS, I. J., TUSHEV, G., WILL, T. J., DIECK, S. T., FUERST, N. & SCHUMAN, E. M. 2012. The Local Transcriptome in the Synaptic Neuropil Revealed by Deep Sequencing and High-Resolution Imaging. *Neuron,* 74**,** 453-466.

CAMICIOLI, R., MOORE, M. M., KINNEY, A., CORBRIDGE, E., GLASSBERG, K. & KAYE, J. A. 2003. Parkinson's Disease is Associated with Hippocampal Atrophy. *Movement Disorders,* 18**,** 784-790.

CARRIERI, C., FORREST, A. R. R., SANTORO, C., PERSICHETTI, F., CARNINCI, P., ZUCCHELLI, S. & GUSTINCICH, S. 2015. Expression Analysis of the long non-coding RNA antisense to Uchl1 (AS Uchl1) during Dopaminergic Cells' Differentiation in vitro and in Neurochemical Models of Parkinson's Disease. *Frontiers in Cellular Neuroscience,* 9.

CHANG, C.-T., BERCOVICH, N., LOH, B., JONAS, S. & IZAURRALDE, E. 2014. The Activation of the Decapping Enzyme DCP2 by DCP1 occurs on the EDC4 Scaffold and Involves a Conserved loop in DCP1. *Nucleic Acids Research,* 42**,** 5217-5233.

CHEN, X., YU, C., GUO, M., ZHENG, X., ALI, S., HUANG, H., ZHANG, L., WANG, S., HUANG, Y., QIE, S. & WANG, J. 2019. Down-Regulation of m6A mRNA Methylation Is Involved in Dopaminergic Neuronal Death. *ACS Chemical Neuroscience,* 10**,** 2355-2363.

CHESNOKOVA, E., BAL, N. & KOLOSOV, P. 2017. Kinases of eIF2a Switch Translation of mRNA Subset during Neuronal Plasticity. *International journal of molecular sciences,* 18**,** 2213.

CHOI, B.-K., CHOI, M.-G., KIM, J.-Y., YANG, Y., LAI, Y., KWEON, D.-H., LEE, N. K. & SHIN, Y.-K. 2013. Large α-synuclein Oligomers Inhibit Neuronal SNARE-mediated Vesicle Docking. *Proceedings of the National Academy of Sciences,* 110**,** 4087-4092.

CHONG, S. Y. C., ROSENBERG, S. S., FANCY, S. P. J., ZHAO, C., SHEN, Y.-A. A., HAHN, A. T., MCGEE, A. W., XU, X., ZHENG, B., ZHANG, L. I., ROWITCH, D. H., FRANKLIN, R. J. M., LU, Q. R. & CHAN, J. R. 2012. Neurite Outgrowth Inhibitor Nogo-A Establishes Spatial Segregation and Extent of Oligodendrocyte Myelination. *Proceedings of the National Academy of Sciences,* 109**,** 1299.

CHUNG, W.-S., ALLEN, N. J. & EROGLU, C. 2015. Astrocytes Control Synapse Formation, Function, and Elimination. *Cold Spring Harbor Perspectives in Biology,* 7.

CHUNG, W.-S., VERGHESE, P. B., CHAKRABORTY, C., JOUNG, J., HYMAN, B. T., ULRICH, J. D., HOLTZMAN, D. M. & BARRES, B. A. 2016. Novel Allele-dependent Role for APOE in Controlling the Rate of Synapse Pruning by Astrocytes. *Proceedings of the National Academy of Sciences,* 113**,** 10186.

CLANCY, M. J., SHAMBAUGH, M. E., TIMPTE, C. S. & BOKAR, J. A. 2002. Induction of Sporulation in Saccharomyces cerevisiae Leads to the Formation of N6-methyladenosine in mRNA: a Potential Mechanism for the Activity of the IME4 Gene. *Nucleic Acids Research,* 30**,** 4509-4518.

COLGAN, D. F. & MANLEY, J. L. 1997. Mechanism and Regulation of mRNA Polyadenylation. *Genes & Development,* 11**,** 2755-2766.

CORDEN, J. L. 1990. Tails of RNA Polymerase II. *Trends in Biochemical Sciences,* 15**,** 383-387.

COSTA-MATTIOLI, M., SOSSIN, W. S., KLANN, E. & SONENBERG, N. 2009. Translational Control of Long-Lasting Synaptic Plasticity and Memory. *Neuron,* 61**,** 10-26.

COUPLAND, K. G., KIM, W. S., HALLIDAY, G. M., HALLUPP, M., DOBSON-STONE, C. & KWOK, J. B. J. 2016. Role of the Long Non-Coding RNA MAPT-AS1 in Regulation of Microtubule Associated Protein Tau (MAPT) Expression in Parkinson's Disease. *PLOS ONE,* 11**,** e0157924.

CUEVAS, J. 2007. The Somatic Nervous System. *In:* ENNA, S. J. & BYLUND, D. B. (eds.) *xPharm: The Comprehensive Pharmacology Reference.* New York: Elsevier.

DANG, W., XIE, Y., CAO, P., XIN, S., WANG, J., LI, S., LI, Y. & LU, J. 2019. N6-Methyladenosine and Viral Infection. *Frontiers in Microbiology,* 10.

DENT, E. W. 2017. Of Microtubules and Memory: Implications for Microtubule Dynamics in Dendrites and Spines. *Molecular biology of the cell,* 28**,** 1-8.

DENT, E. W. & BAAS, P. W. 2014. Microtubules in Neurons as Information Carriers. *Journal of Neurochemistry,* 129**,** 235-239.

DESROSIE, R., FRIDERIC, K. & ROTTMAN, F. 1974. Identification of Methylated Nucleosides in Messenger-RNA from Novikoff Hepatoma-Cells. *Proceedings of the National Academy of Sciences of the United States of America,* 71**,** 3971-3975.

DIJKSTRA, A. A., INGRASSIA, A., DE MENEZES, R. X., VAN KESTEREN, R. E., ROZEMULLER, A. J. M., HEUTINK, P. & VAN DE BERG, W. D. J. 2015. Evidence for Immune Response, Axonal Dysfunction and Reduced Endocytosis in the Substantia Nigra in Early Stage Parkinson's Disease. *PLOS ONE,* 10**,** e0128651.

DIMITROVA, D. G., TEYSSET, L. & CARRÉ, C. 2019. RNA 2'-O-Methylation (Nm) Modification in Human Diseases. *Genes,* 10**,** 117.

DINIZ, L. P., MATIAS, I. C. P., GARCIA, M. N. & GOMES, F. C. A. 2014. Astrocytic Control of Neural Circuit Formation: Highlights on TGF-beta Signaling. *Neurochemistry International,* 78**,** 18-27.

DOMINISSINI, D., MOSHITCH-MOSHKOVITZ, S., SCHWARTZ, S., SALMON-DIVON, M., UNGAR, L., OSENBERG, S., CESARKAS, K., JACOB-HIRSCH, J., AMARIGLIO, N., KUPIEC, M., SOREK, R. & RECHAVI, G. 2012. Topology of the Human and Mouse m6A RNA Methylomes Revealed by m6A-seq. *Nature,* 485**,** 201-206.

DONG, Z.-W., SHAO, P., DIAO, L.-T., ZHOU, H., YU, C.-H. & QU, L.-H. 2012. RTL-P: a Sensitive Approach for Detecting Sites of 2′-O-methylation in RNA Molecules. *Nucleic Acids Research,* 40**,** e157-e157.

DU, H., ZHAO, Y., HE, J., ZHANG, Y., XI, H., LIU, M., MA, J. & WU, L. 2016. YTHDF2 Destabilizes m6A-containing RNA through Direct Recruitment of the CCR4–NOT Deadenylase Complex. *Nature Communications,* 7**,** 12626.

DUBIN, D. T. & TAYLOR, R. H. 1975. The Methylation State of Poly A-Containing Messenger RNA From Cultured Hamster Cells. *Nucleic Acids Research,* 2**,** 1653-1668.

EDENS, B. M., VISSERS, C., SU, J., ARUMUGAM, S., XU, Z., SHI, H., MILLER, N., ROJAS RINGELING, F., MING, G.-L., HE, C., SONG, H. & MA, Y. C. 2019. FMRP Modulates Neural Differentiation through m6A-Dependent mRNA Nuclear Export. *Cell Reports,* 28**,** 845-854.e5.

EDGAR, R. C. 2004a. MUSCLE: a Multiple Sequence Alignment Method with Reduced Time and Space Complexity. *BMC Bioinformatics,* 5**,** 113.

EDGAR, R. C. 2004b. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Research,* 32**,** 1792-1797.

ELKOURIS, M., KOUROUPI, G., VOURVOUKELIS, A., PAPAGIANNAKIS, N., KALTEZIOTI, V., MATSAS, R., STEFANIS, L., XILOURI, M. & POLITIS, P. K. 2019. Long Non-coding RNAs Associated With Neurodegeneration-Linked Genes Are Reduced in Parkinson's Disease Patients. *Frontiers in Cellular Neuroscience,* 13.

ENGEL, M., EGGERT, C., KAPLICK, P. M., EDER, M., RÖH, S., TIETZE, L., NAMENDORF, C., ARLOTH, J., WEBER, P., REX-HAFFNER, M., GEULA, S., JAKOVCEVSKI, M., HANNA, J. H., LESHKOWITZ, D., UHR, M., WOTJAK, C. T., SCHMIDT, M. V., DEUSSING, J. M., BINDER, E. B. & CHEN, A. 2018. The Role of m6A/m-RNA Methylation in Stress Response Regulation. *Neuron,* 99**,** 389-403.e9.

ENGLUND, C., UV, A. E., CANTERA, R., MATHIES, L. D., KRASNOW, M. A. & SAMAKOVLIS, C. 1999. adrift, a Novel bnl-induced Drosophila Gene, Required for Tracheal Pathfinding into the CNS. *Development,* 126**,** 1505.

ERTA, M., QUINTANA, A. & HIDALGO, J. 2012. Interleukin-6, a Major Cytokine in the Central Nervous System. *International journal of biological sciences,* 8**,** 1254-1266.

FAN, H., SAKURABA, K., KOMURO, A., KATO, S., HARADA, F. & HIROSE, Y. 2003. PCIF1, a Novel Human WW Domain-containing Protein, Interacts with the Phosphorylated RNA Polymerase II. *Biochemical and Biophysical Research Communications,* 301**,** 378-385.

FINOTELLO, F., LAVEZZO, E., BIANCO, L., BARZON, L., MAZZON, P., FONTANA, P., TOPPO, S. & DI CAMILLO, B. 2014. Reducing Bias in RNA Sequencing Data: a Novel Approach to Compute Counts. *BMC bioinformatics,* 15 Suppl 1**,** S7-S7.

FLEXNER, L. B., FLEXNER, J. B. & STELLAR, E. 1965. Memory and Cerebral Protein Synthesis in Mice as Affected by Graded Amounts of Puromycin. *Experimental Neurology,* 13**,** 264-&.

FRANKER, M. A. M. & HOOGENRAAD, C. C. 2013. Microtubule-based Transport – Basic Mechanisms, Traffic Rules and Role in Neurological Pathogenesis. *Journal of Cell Science,* 126**,** 2319-2329.

FREDRIKSSON, R., HÄGGLUND, M., OLSZEWSKI, P. K., STEPHANSSON, O., JACOBSSON, J. A., OLSZEWSKA, A. M., LEVINE, A. S., LINDBLOM, J. & SCHIÖTH, H. B. 2008. The Obesity Gene, FTO, Is of Ancient Origin, Up-Regulated during Food Deprivation and Expressed in Neurons of Feeding-Related Nuclei of the Brain. *Endocrinology,* 149**,** 2062-2071.

FREUND, I., EIGENBROD, T., HELM, M. & DALPKE, H. A. 2019. RNA Modifications Modulate Activation of Innate Toll-Like Receptors. *Genes,* 10.

FUCHS, A.-L., NEU, A. & SPRANGERS, R. 2016. A General Method for Rapid and Cost-Efficient Large-Scale Production of 5' Capped RNA. *RNA (New York, N.Y.),* 22**,** 1454-1466.

FURUICHI, Y. 2015. Discovery of m(7)G-cap in Eukaryotic mRNAs. *Proceedings of the Japan Academy. Series B, Physical and biological sciences,* 91**,** 394-409.

GALATRO, T. F., HOLTMAN, I. R., LERARIO, A. M., VAINCHTEIN, I. D., BROUWER, N., SOLA, P. R., VERAS, M. M., PEREIRA, T. F., LEITE, R. E. P., MÖLLER, T., WES, P. D., SOGAYAR, M. C., LAMAN, J. D., DEN DUNNEN, W., PASQUALUCCI, C. A., OBA-

SHINJO, S. M., BODDEKE, E. W. G. M., MARIE, S. K. N. & EGGEN, B. J. L. 2017. Transcriptomic Analysis of Purified Human Cortical Microglia Reveals Age-associated Changes. *Nature Neuroscience,* 20**,** 1162.

GALLOWAY, A. & COWLING, V. H. 2019. mRNA Cap Regulation in Mammalian Cell Function and Fate. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1862**,** 270-279.

GARCIA-REITBÖCK, P., ANICHTCHIK, O., BELLUCCI, A., IOVINO, M., BALLINI, C., FINEBERG, E., GHETTI, B., DELLA CORTE, L., SPANO, P., TOFARIS, G. K., GOEDERT, M. & SPILLANTINI, M. G. 2010. SNARE Protein Redistribution and Synaptic Failure in a Transgenic Mouse Model of Parkinson's Disease. *Brain,* 133**,** 2032-2044.

GEISS, G. K., CARTER, V. S., HE, Y., KWIECISZEWSKI, B. K., HOLZMAN, T., KORTH, M. J., LAZARO, C. A., FAUSTO, N., BUMGARNER, R. E. & KATZE, M. G. 2003. Gene Expression Profiling of the Cellular Transcriptional Network Regulated by Alpha/Beta Interferon and Its Partial Attenuation by the Hepatitis C Virus Nonstructural 5A Protein. *Journal of Virology,* 77**,** 6367.

GEUENS, T., BOUHY, D. & TIMMERMAN, V. 2016. The hnRNP family: Insights into their Role in Health and Disease. *Human genetics,* 135**,** 851-867.

GIACOMETTI, S., BENBAHOUCHE, N. E. H., DOMANSKI, M., ROBERT, M.-C., MEOLA, N., LUBAS, M., BUKENBORG, J., ANDERSEN, J. S., SCHULZE, W. M., VERHEGGEN, C., KUDLA, G., JENSEN, T. H. & BERTRAND, E. 2017. Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate. *Cell Reports,* 18**,** 2635-2650.

GLEESON, P. A. 2014. The Role of Endosomes in Innate and Adaptive Immunity. *Seminars in Cell & Developmental Biology,* 31**,** 64-72.

GOEDERT, M. 2015. Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled Aβ, tau, and α-synuclein. *Science*.

GOLOMB, J., DE LEON, M. J., KLUGER, A., GEORGE, A. E., TARSHISH, C. & FERRIS, S. H. 1993. Hippocampal Atrophy in Normal Aging: An Association With Recent Memory Impairment. *JAMA Neurology,* 50**,** 967-973.

GONATOPOULOS-POURNATZIS, T. & COWLING, V. H. 2014. Cap-Binding Complex (CBC). *Biochemical Journal,* 458**,** 185.

GRASS, G. D. & TOOLE, B. P. 2015. How, with Whom and When: an Overview of CD147-mediated Regulatory Networks Influencing Matrix Metalloproteinase Activity. *Bioscience reports,* 36**,** e00283-e00283.

GRECO, M., SÁEZ, C. A., BROWN, M. T. & BITONTI, M. B. 2014. A Simple and Effective Method for High Quality Co-Extraction of Genomic DNA and Total RNA from Low Biomass Ectocarpus siliculosus, the Model Brown Alga. *PLOS ONE,* 9**,** e96470.

GRUDZIEN-NOGALSKA, E., JIAO, X., SONG, M.-G., HART, R. P. & KILEDJIAN, M. 2016. Nudt3 is an mRNA Decapping Enzyme that Modulates Cell Migration. *RNA,* 22**,** 773-781.

GUERRA, S., LÓPEZ-FERNÁNDEZ, L. A., PASCUAL-MONTANO, A., MUÑOZ, M., HARSHMAN, K. & ESTEBAN, M. 2003. Cellular Gene Expression Survey of Vaccinia Virus Infection of Human HeLa Cells. *Journal of Virology,* 77**,** 6493.

GULATI, P., AVEZOV, E., MA, M., ANTROBUS, R., LEHNER, P., O'RAHILLY, S. & YEO, GILES S. H. 2014. Fat Mass and Obesity-related (FTO) Shuttles between the Nucleus and Cytoplasm. *Bioscience Reports,* 34**,** e00144.

HALINE-VAZ, T., LIMA SILVA, T. C. & ZANCHIN, N. I. T. 2008. The Human Interferon-regulated ISG95 Protein Interacts with RNA Polymerase II and Shows Methyltransferase Activity. *Biochemical and Biophysical Research Communications,* 372**,** 719-724.

HALLIDAY, G. M. & MCCANN, H. 2010. The Progression of Pathology in Parkinson's Disease. *Annals of the New York Academy of Sciences,* 1184**,** 188-195.

HAMILTON, T. L., STONELEY, M., SPRIGGS, K. A. & BUSHELL, M. 2006. TOPs and their Regulation. *Biochemical Society Transactions,* 34**,** 12.

HAUSSMANN, I. U., BODI, Z., SANCHEZ-MORAN, E., MONGAN, N. P., ARCHER, N., FRAY, R. G. & SOLLER, M. 2016. m6A Potentiates Sxl Alternative pre-mRNA Splicing for Robust Drosophila Sex Determination. *Nature,* 540**,** 301.

HECK, A. M. & WILUSZ, C. J. 2019. Small Changes, Big Implications: The Impact of m6A RNA Methylation on Gene Expression in Pluripotency and Development. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1862**,** 194402.

HECKLE, W. L., FENTON, R. G., WOOD, T. G., MERKEL, C. G. & LINGREL, J. B. 1977. Methylated Nucleosides in Globin mRNA from Mouse Nucleated Erythroid Cells. *Journal of Biological Chemistry,* 252**,** 1764-1770.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell,* 38**,** 576-589.

HENN, F. A. & HAMBERGER, A. 1971. Glial Cell Function: Uptake of Transmitter Substances. *Proceedings of the National Academy of Sciences of the United States of America,* 68**,** 2686-2690.

HESS, M. E., HESS, S., MEYER, K. D., VERHAGEN, L. A. W., KOCH, L., BRONNEKE, H. S., DIETRICH, M. O., JORDAN, S. D., SALETORE, Y., ELEMENTO, O., BELGARDT, B. F., FRANZ, T., HORVATH, T. L., RUTHER, U., JAFFREY, S. R., KLOPPENBURG, P. & BRUNING, J. C. 2013. The Fat Mass and Obesity Associated Gene (FTO) Regulates Activity of the Dopaminergic Midbrain Circuitry. *Nature Neuroscience,* 16**,** 1042-U96.

HIROSE, Y., IWAMOTO, Y., SAKURABA, K., YUNOKUCHI, I., HARADA, F. & OHKUMA, Y. 2008. Human Phosphorylated CTD-interacting Protein, PCIF1, Negatively Modulates Gene Expression by RNA polymerase II. *Biochemical and Biophysical Research Communications,* 369**,** 449-455.

HOLZ RW, F. S. 1999. Synaptic Transmission. *In:* SIEGEL GJ, A. B., ALBERS RW (ed.) *Basic Neurochemistry: Molecular, Cellular and Medical Aspects. 6th edition.* Philadelphia: Lippincott-Raven.

HONGAY, C. F. & ORR-WEAVER, T. L. 2011. Drosophila Inducer of MEiosis 4 (IME4) is Required for Notch Dignaling During Oogenesis. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 14855-14860.

HOOZEMANS, J. J. M., VAN HAASTERT, E. S., EIKELENBOOM, P., DE VOS, R. A. I., ROZEMULLER, J. M. & SCHEPER, W. 2007. Activation of the Unfolded Protein Response in Parkinson's Disease. *Biochemical and Biophysical Research Communications,* 354**,** 707-711.

HOOZEMANS, J. J. M., VAN HAASTERT, E. S., NIJHOLT, D. A. T., ROZEMULLER, A. J. M., EIKELENBOOM, P. & SCHEPER, W. 2009. The Unfolded Protein Response Is Activated in Pretangle Neurons in Alzheimer's Disease Hippocampus. *The American Journal of Pathology,* 174**,** 1241-1251.

HSIN, J.-P. & MANLEY, J. L. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & Development,* 26**,** 2119-2137.

HSU, P. J., ZHU, Y., MA, H., GUO, Y., SHI, X., LIU, Y., QI, M., LU, Z., SHI, H., WANG, J., CHENG, Y., LUO, G., DAI, Q., LIU, M., GUO, X., SHA, J., SHEN, B. & HE, C. 2017. Ythdc2 is an N6-methyladenosine Binding Protein that Regulates Mammalian Spermatogenesis. *Cell Research,* 27**,** 1115.

HUBBLE, J. P., CAO, T., HASSANEIN, R. E. S., NEUBERGER, J. S. & ROLLER, W. C. 1993. Risk Factors for Parkinson's Disease. *Neurology,* 43**,** 1693.

HUSON, D. H., RICHTER, D. C., RAUSCH, C., DEZULIAN, T., FRANZ, M. & RUPP, R. 2007. Dendroscope: an Interactive Viewer for Large Phylogenetic Trees. *BMC Bioinformatics,* 8**,** 460.

HUSON, D. H. & SCORNAVACCA, C. 2012. Dendroscope 3: an Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology,* 61**,** 1061-1067.

IACOANGELI, A. & TIEDGE, H. 2013. Translational Control at the Synapse: Role of RNA Regulators. *Trends in Biochemical Sciences,* 38**,** 47-55.

INESTA-VAQUERA, F. & COWLING, V. H. 2017. Regulation and function of CMTR1-dependent mRNA cap methylation. *Wiley Interdisciplinary Reviews: RNA,* 8**,** e1450.
INGELSSON, M. 2016. Alpha-Synuclein Oligomers—Neurotoxic Molecules in Parkinson's Disease and Other Lewy Body Disorders. *Frontiers in Neuroscience,* 10.
IRVINE, G. B., EL-AGNAF, O. M., SHANKAR, G. M. & WALSH, D. M. 2008. Protein Aggregation in the Brain: the Molecular Basis for Alzheimer's and Parkinson's Diseases. *Molecular medicine (Cambridge, Mass.),* 14**,** 451-464.
IVASHKIV, L. B. & DONLIN, L. T. 2014. Regulation of Type I Interferon Responses. *Nature reviews. Immunology,* 14**,** 36-49.
IYER, L. M., ZHANG, D. & ARAVIND, L. 2016. Adenine Methylation in Eukaryotes: Apprehending the Complex Evolutionary History and Functional Potential of an Epigenetic Modification. *BioEssays : news and reviews in molecular, cellular and developmental biology,* 38**,** 27-40.
JÄKEL, S. & DIMOU, L. 2017. Glial Cells and Their Function in the Adult Brain: A Journey through the History of Their Ablation. *Frontiers in cellular neuroscience,* 11**,** 24-24.
JIA, G., FU, Y., ZHAO, X., DAI, Q., ZHENG, G., YANG, Y., YI, C., LINDAHL, T., PAN, T., YANG, Y.-G. & HE, C. 2011. N6-Methyladenosine in Nuclear RNA is a Major Substrate of the Obesity-Associated FTO. *Nat Chem Biol,* 7**,** 885-887.
JIA, G. F., YANG, C. G., YANG, S. D., JIAN, X., YI, C. Q., ZHOU, Z. Q. & HE, C. 2008. Oxidative Demethylation of 3-Methylthymine and 3-Methyluracil in Single-Stranded DNA and RNA by Mouse and Human FTO. *Febs Letters,* 582**,** 3313-3319.
JIAO, X., CHANG, JEONG H., KILIC, T., TONG, L. & KILEDJIAN, M. 2013. A Mammalian Pre-mRNA 5′ End Capping Quality Control Mechanism and an Unexpected Link of Capping to Pre-mRNA Processing. *Molecular Cell,* 50**,** 104-115.
JONES, C. I., ZABOLOTSKAYA, M. V. & NEWBURY, S. F. 2012. The 5′ → 3′ exoribonuclease XRN1/Pacman and its Functions in Cellular Processes and Development. *Wiley Interdisciplinary Reviews: RNA,* 3**,** 455-468.
JONES, J. D. G., VANCE, R. E. & DANGL, J. L. 2016. Intracellular Innate Immune Surveillance Devices in Plants and Animals. *Science,* 354**,** aaf6395.
JOVE, R. & MANLEY, J. L. 1984. In vitro Transcription from the Adenovirus 2 Major Late Promoter Utilizing Templates Truncated at Promoter-Proximal Sites. *Journal of Biological Chemistry,* 259**,** 8513-8521.
JUVEN-GERSHON, T. & KADONAGA, J. T. 2010. Regulation of Gene Expression via the Core Promoter and the Basal Transcriptional Machinery. *Developmental Biology,* 339**,** 225-229.
KANG, H. J. & SCHUMAN, E. M. 1996. A Requirement for Local Protein Synthesis in Neurotrophin-Induced Hippocampal Synaptic Plasticity. *Science,* 273**,** 1402-1406.
KAPUR, M., MONAGHAN, C. E. & ACKERMAN, S. L. 2017. Regulation of mRNA Translation in Neurons—A Matter of Life and Death. *Neuron,* 96**,** 616-637.
KATO, A., HOMMA, T., BATCHELOR, J., HASHIMOTO, N., IMAI, S., WAKIGUCHI, H., SAITO, H. & MATSUMOTO, K. 2003. Interferon-alpha/beta Receptor-mediated Selective Induction of a Gene Cluster by CpG Oligodeoxynucleotide 2006. *BMC immunology,* 4**,** 8-8.
KELLEHER, R. J., III & BEAR, M. F. 2008. The Autistic Neuron: Troubled Translation? *Cell,* 135**,** 401-406.
KIM, D., PERTEA, G., TRAPNELL, C., PIMENTEL, H., KELLEY, R. & SALZBERG, S. L. 2013. TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions. *Genome Biology,* 14**,** R36.
KLATTENHOFF, CARLA A., SCHEUERMANN, JOHANNA C., SURFACE, LAUREN E., BRADLEY, ROBERT K., FIELDS, PAUL A., STEINHAUSER, MATTHEW L., DING, H., BUTTY, VINCENT L., TORREY, L., HAAS, S., ABO, R., TABEBORDBAR, M., LEE, RICHARD T., BURGE, CHRISTOPHER B. & BOYER, LAURIE A. 2013. Braveheart, a Long Noncoding RNA Required for Cardiovascular Lineage Commitment. *Cell,* 152**,** 570-583.

KOBAYASHI, I. 2001. Behavior of Restriction–modification Systems as Selfish Mobile Elements and their Impact on Genome Evolution. *Nucleic Acids Research,* 29**,** 3742-3756.

KONDO, A., NONAKA, A., SHIMAMURA, T., YAMAMOTO, S., YOSHIDA, T., KODAMA, T., ABURATANI, H. & OSAWA, T. 2017. Long Noncoding RNA JHDM1D-AS1 Promotes Tumor Growth by Regulating Angiogenesis in Response to Nutrient Starvation. *Molecular and Cellular Biology,* 37**,** e00125-17.

KONDRASHOV, A. V., KIEFMANN, M., EBNET, K., KHANAM, T., MUDDASHETTY, R. S. & BROSIUS, J. 2005. Inhibitory Effect of Naked Neural BC1 RNA or BC200 RNA on Eukaryotic in vitro Translation Systems is Reversed by Poly(A)-binding Protein (PABP). *Journal of Molecular Biology,* 353**,** 88-103.

KRUSE, S., ZHONG, S., BODI, Z., BUTTON, J., ALCOCER, M. J. C., HAYES, C. J. & FRAY, R. 2011. A Novel Synthesis and Detection Method for Cap-Associated Adenosine Modifications in Mouse mRNA. *Scientific Reports,* 1**,** 126.

LANGSTROM, N. S., ANDERSON, J. P., LINDROOS, H. G., WINBLAND, B. & WALLACE, W. C. 1989. Alzheimer's Disease-associated Reduction of Polysomal mRNA Translation. *Molecular Brain Research,* 5**,** 259-269.

LARSEN, K. E., SCHMITZ, Y., TROYER, M. D., MOSHAROV, E., DIETRICH, P., QUAZI, A. Z., SAVALLE, M., NEMANI, V., CHAUDHRY, F. A., EDWARDS, R. H., STEFANIS, L. & SULZER, D. 2006. α-Synuclein Overexpression in PC12 and Chromaffin Cells Impairs Catecholamine Release by Interfering with a Late Step in Exocytosis. *The Journal of Neuroscience,* 26**,** 11915-11922.

LÉCUYER, E., YOSHIDA, H., PARTHASARATHY, N., ALM, C., BABAK, T., CEROVINA, T., HUGHES, T. R., TOMANCAK, P. & KRAUSE, H. M. 2007. Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function. *Cell,* 131**,** 174-187.

LEGRAND, C., TUORTO, F., HARTMANN, M., LIEBERS, R., JACOB, D., HELM, M. & LYKO, F. 2017. Statistically Robust Methylation Calling for Whole-Transcriptome Bisulfite Sequencing Reveals Distinct Methylation Patterns for Mouse RNAs. *Genome Research*.

LI, J., UVERSKY, V. N. & FINK, A. L. 2001. Effect of Familial Parkinson's Disease Point Mutations A30P and A53T on the Structural Properties, Aggregation, and Fibrillation of Human α-Synuclein. *Biochemistry,* 40**,** 11604-11613.

LI, M., ZHAO, X., WANG, W., SHI, H., PAN, Q., LU, Z., PEREZ, S. P., SUGANTHAN, R., HE, C., BJØRÅS, M. & KLUNGLAND, A. 2018. Ythdf2-mediated m6A mRNA Clearance Modulates Neural Development in Mice. *Genome Biology,* 19**,** 69.

LINDER, B., GROZHIK, A. V., OLARERIN-GEORGE, A. O., MEYDAN, C., MASON, C. E. & JAFFREY, S. R. 2015. Single-Nucleotide-Resolution Mapping of m6A and m6Am throughout the Transcriptome. *Nat Meth,* 12**,** 767-772.

LIU, J., YUE, Y., HAN, D., WANG, X., FU, Y., ZHANG, L., JIA, G., YU, M., LU, Z., DENG, X., DAI, Q., CHEN, W. & HE, C. 2013a. A METTL3–METTL14 Complex Mediates Mammalian Nuclear RNA N6-adenosine Methylation. *Nature Chemical Biology,* 10**,** 93.

LIU, N., DAI, Q., ZHENG, G. Q., HE, C., PARISIEN, M. & PAN, T. 2015. N-6-methyladenosine-Dependent RNA Structural Switches Regulate RNA-Protein Interactions. *Nature,* 518**,** 560-564.

LIU, N., PARISIEN, M., DAI, Q., ZHENG, G., HE, C. & PAN, T. 2013b. Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA (New York, N.Y.),* 19**,** 1848-1856.

LIU, S. W., JIAO, X., WELCH, S. & KILEDJIAN, M. 2008. Chapter 1 Analysis of mRNA Decapping. *Methods in Enzymology.* Academic Press.

LIU, X. & QUAN, N. 2018. Microglia and CNS Interleukin-1: Beyond Immunological Concepts. *Frontiers in neurology,* 9**,** 8-8.

LODISH, H., BERK, A. & ZIPURSKY, L. 2000. Molecular Cell Biology. *In:* FREEMAN, W. H. (ed.) 4th ed. New York.

LOTHARIUS, J. & BRUNDIN, P. 2002. Pathogenesis of Parkinson's Disease: Dopamine, Vesicles and α-synuclein. *Nature Reviews Neuroscience,* 3**,** 932-942.
LV, J., ZHANG, Y., GAO, S., ZHANG, C., CHEN, Y., LI, W., YANG, Y.-G., ZHOU, Q. & LIU, F. 2018. Endothelial-Specific m6A Modulates Mouse Hematopoietic Stem and Progenitor Cell Development via Notch Signaling. *Cell Research,* 28**,** 249-252.
LYONS, S. M., FAY, M. M. & IVANOV, P. 2018. The Role of RNA Modifications in the Regulation of tRNA Cleavage. *FEBS Letters,* 592**,** 2828-2844.
MACK, S., KANDEL, E. R., JESSELL, T. M., SCHWARTZ, J. H., SIEGELBAUM, S. A. & HUDSPETH, A. J. 2013. *Principles of Neural Science, Fifth Edition*, McGraw-Hill Education.
MAGUIRE, E. A., INTRAUB, H. & MULLALLY, S. L. 2016. Scenes, Spaces, and Memory Traces: What Does the Hippocampus Do? *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry,* 22**,** 432-439.
MANNERS, O., BAQUERO-PEREZ, B. & WHITEHOUSE, A. 2019. m6A: Widespread Regulatory Control in Virus Replication. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1862**,** 370-381.
MARTIN, L. J., PAN, Y., PRICE, A. C., STERLING, W., COPELAND, N. G., JENKINS, N. A., PRICE, D. L. & LEE, M. K. 2006. Parkinson's Disease α-Synuclein Transgenic Mice Develop Neuronal Mitochondrial Degeneration and Cell Death. *The Journal of Neuroscience,* 26**,** 41-50.
MATTHEWS, M. A. & DUNCAN, D. 1971. A Quantitative Study of Morphological changes Accompanying the Initiation and Progress of Myelin Production in the Dorsal Funiculus of the Rat Spinal Cord. *Journal of Comparative Neurology,* 142**,** 1-22.
MAUER, J. & JAFFREY, S. R. 2018. FTO, m6Am, and the hypothesis of reversible epitranscriptomic mRNA modifications. *FEBS Letters,* 592**,** 2012-2022.
MAUER, J., LUO, X., BLANJOIE, A., JIAO, X., GROZHIK, A. V., PATIL, D. P., LINDER, B., PICKERING, B. F., VASSEUR, J.-J., CHEN, Q., GROSS, S. S., ELEMENTO, O., DEBART, F., KILEDJIAN, M. & JAFFREY, S. R. 2016. Reversible methylation of m6Am in the 5′ cap controls mRNA stability. *Nature,* advance online publication.
MAUER, J., SINDELAR, M., DESPIC, V., GUEZ, T., HAWLEY, B. R., VASSEUR, J.-J., RENTMEISTER, A., GROSS, S. S., PELLIZZONI, L., DEBART, F., GOODARZI, H. & JAFFREY, S. R. 2019. FTO Controls Reversible m6Am RNA Methylation during snRNA Biogenesis. *Nature Chemical Biology,* 15**,** 340-347.
MERCER, T. R., QURESHI, I. A., GOKHAN, S., DINGER, M. E., LI, G., MATTICK, J. S. & MEHLER, M. F. 2010. Long Noncoding RNAs in Neuronal-glial Fate Specification and Oligodendrocyte Lineage Maturation. *BMC Neuroscience,* 11**,** 14.
MERKURJEV, D., HONG, W.-T., IIDA, K., OOMOTO, I., GOLDIE, B. J., YAMAGUTI, H., OHARA, T., KAWAGUCHI, S.-Y., HIRANO, T., MARTIN, K. C., PELLEGRINI, M. & WANG, D. O. 2018. Synaptic N6-methyladenosine (m6A) Epitranscriptome Reveals functional Partitioning of Localized Transcripts. *Nature Neuroscience,* 21**,** 1004-1014.
MEYER, K. D., SALETORE, Y., ZUMBO, P., ELEMENTO, O., MASON, C. E. & JAFFREY, S. R. 2012. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3 ' UTRs and near Stop Codons. *Cell,* 149**,** 1635-1646.
MIDDLETON, S. A., EBERWINE, J. & KIM, J. 2019. Comprehensive Catalog of Dendritically Localized mRNA Isoforms from Sub-cellular Sequencing of Single Mouse Neurons. *BMC Biology,* 17**,** 5.
MONNE, L. 1948. Functioning of the Cytoplasm. *Advances in Enzymology and Related Subjects of Biochemistry,* 8**,** 1-69.
MUGRIDGE, J. S., TIBBLE, R. W., ZIEMNIAK, M., JEMIELITY, J. & GROSS, J. D. 2018. Structure of the Activated Edc1-Dcp1-Dcp2-Edc3 mRNA Decapping Complex with Substrate Analog Poised for Catalysis. *Nature Communications,* 9**,** 1152.
MURAMATSU, T. 2016. Basigin (CD147), a Multifunctional Transmembrane Glycoprotein with Various Binding Partners. *Journal of biochemistry,* 159**,** 481-490.
MURILLO, O. D., THISTLETHWAITE, W., ROZOWSKY, J., SUBRAMANIAN, S. L., LUCERO, R., SHAH, N., JACKSON, A. R., SRINIVASAN, S., CHUNG, A., LAURENT, C. D.,

KITCHEN, R. R., GALEEV, T., WARRELL, J., DIAO, J. A., WELSH, J. A., HANSPERS, K., RIUTTA, A., BURGSTALLER-MUEHLBACHER, S., SHAH, R. V., YERI, A., JENKINS, L. M., AHSEN, M. E., CORDON-CARDO, C., DOGRA, N., GIFFORD, S. M., SMITH, J. T., STOLOVITZKY, G., TEWARI, A. K., WUNSCH, B. H., YADAV, K. K., DANIELSON, K. M., FILANT, J., MOELLER, C., NEJAD, P., PAUL, A., SIMONSON, B., WONG, D. K., ZHANG, X., BALAJ, L., GANDHI, R., SOOD, A. K., ALEXANDER, R. P., WANG, L., WU, C., WONG, D. T. W., GALAS, D. J., VAN KEUREN-JENSEN, K., PATEL, T., JONES, J. C., DAS, S., CHEUNG, K.-H., PICO, A. R., SU, A. I., RAFFAI, R. L., LAURENT, L. C., ROTH, M. E., GERSTEIN, M. B. & MILOSAVLJEVIC, A. 2019. exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. *Cell,* 177**,** 463-477.e15.

MURPHY, D. D., RUETER, S. M., TROJANOWSKI, J. Q. & LEE, V. M.-Y. 2000. Synucleins Are Developmentally Expressed, and α-Synuclein Regulates the Size of the Presynaptic Vesicular Pool in Primary Hippocampal Neurons. *The Journal of Neuroscience,* 20**,** 3214-3220.

MURTHY, V. N. 1998. Synaptic Plasticity: Step-wise Strengthening. *Current Biology,* 8**,** R650-R653.

NAGARAJAN, V. K., JONES, C. I., NEWBURY, S. F. & GREEN, P. J. 2013. XRN 5'→3' Exoribonucleases: Structure, Mechanisms and Functions. *Biochimica et biophysica acta,* 1829**,** 590-603.

NAGASE, T., KIKUNO, R. & OHARA, O. 2001. Prediction of the Coding Sequences of Unidentified Human Genes. XXI. The Complete Sequences of 60 New cDNA Clones from Brain Which Code for Large Proteins. *DNA Research,* 8**,** 179-187.

NAOI, M. & MARUYAMA, W. 1999. Cell Death of Dopamine Neurons in Aging and Parkinson's Disease. *Mechanisms of Ageing and Development,* 111**,** 175-188.

NEVE, J., PATEL, R., WANG, Z., LOUEY, A. & FURGER, A. M. 2017. Cleavage and Polyadenylation: Ending the Message Expands Gene Regulation. *RNA biology,* 14**,** 865-890.

NICHOLAS, K. B., NICHOLAS, H. & DEERFIELD, D. 1997. GeneDoc: analysis and visualization of genetic variation. *EMBnet News*.

NIEDZWIECKA, A., MARCOTRIGIANO, J., STEPINSKI, J., JANKOWSKA-ANYSZKA, M., WYSLOUCH-CIESZYNSKA, A., DADLEZ, M., GINGRAS, A.-C., MAK, P., DARZYNKIEWICZ, E., SONENBERG, N., BURLEY, S. K. & STOLARSKI, R. 2002. Biophysical Studies of eIF4E Cap-binding Protein: Recognition of mRNA 5′ Cap Structure and Synthetic Fragments of eIF4G and 4E-BP1 Proteins. *Journal of Molecular Biology,* 319**,** 615-635.

OHIRA, T. & SUZUKI, T. 2016. Precursors of tRNAs are Stabilized by Methylguanosine Cap Structures. *Nature Chemical Biology,* 12**,** 648.

OSTROFF, L. E., FIALA, J. C., ALLWARDT, B. & HARRIS, K. M. 2002. Polyribosomes Redistribute from Dendritic Shafts into Spines with Enlarged Synapses During LTP in Developing Rat Hippocampal Slices. *Neuron,* 35**,** 535-545.

OTSUKA, Y., KEDERSHA, N. L. & SCHOENBERG, D. R. 2009. Identification of a Cytoplasmic Complex That Adds a Cap onto 5′-Monophosphate RNA. *Molecular and Cellular Biology,* 29**,** 2155-2167.

PAOLICELLI, R. C., BERGAMINI, G. & RAJENDRAN, L. 2019. Cell-to-cell Communication by Extracellular Vesicles: Focus on Microglia. *Neuroscience,* 405**,** 148-157.

PARKER, R. & SONG, H. 2004. The Enzymes and Control of Eukaryotic mRNA Turnover. *Nature Structural & Molecular Biology,* 11**,** 121-127.

PASTUZYN, E. D., DAY, C. E., KEARNS, R. B., KYRKE-SMITH, M., TAIBI, A. V., MCCORMICK, J., YODER, N., BELNAP, D. M., ERLENDSSON, S., MORADO, D. R., BRIGGS, J. A. G., FESCHOTTE, C. & SHEPHERD, J. D. 2018. The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell,* 172**,** 275-288.e18.

PATIL, D. P., CHEN, C.-K., PICKERING, B. F., CHOW, A., JACKSON, C., GUTTMAN, M. & JAFFREY, S. R. 2016. m6A RNA Methylation Promotes XIST-mediated Transcriptional Repression. *Nature,* 537**,** 369.

PATZELT, E., THALMANN, E., HARTMUTH, K., BLAAS, D. & KUECHLER, E. 1987. Assembly of pre-mRNA Splicing Complex is Cap Dependent. *Nucleic acids research,* 15**,** 1387-1399.

PEREANU, W., SPINDLER, S., CRUZ, L. & HARTENSTEIN, V. 2007. Tracheal Development in the Drosophila Brain is Constrained by Glial Cells. *Developmental Biology,* 302**,** 169-180.

PEREZ-PERRI, J. I., ROGELL, B., SCHWARZL, T., STEIN, F., ZHOU, Y., RETTEL, M., BROSIG, A. & HENTZE, M. W. 2018. Discovery of RNA-binding Proteins and Characterization of their Dynamic Responses by Enhanced RNA Interactome Capture. *Nature Communications,* 9**,** 4408.

PERRY, R. P. & KELLEY, D. E. 1974. Existence of Methylated Messenger-RNA in Mouse L Cells. *Cell,* 1**,** 37-42.

PESTOVA, T. V. & KOLUPAEVA, V. G. 2002. The Roles of Individual Eukaryotic Translation Initiation Factors in Ribosomal Scanning and Initiation Codon Selection. *Genes & Development,* 16**,** 2906-2922.

PETERS, R. 2006. Ageing and the brain. *Postgraduate medical journal,* 82**,** 84-88.

PFEIFFER, B. E. & HUBER, K. M. 2006. Current Advances in Local Protein Synthesis and Synaptic Plasticity. *The Journal of Neuroscience,* 26**,** 7147-7150.

PICARD-JEAN, F., BRAND, C., TREMBLAY-LÉTOURNEAU, M., ALLAIRE, A., BEAUDOIN, M. C., BOUDREAULT, S., DUVAL, C., RAINVILLE-SIROIS, J., ROBERT, F., PELLETIER, J., GEISS, B. J. & BISAILLON, M. 2018. 2'-O-methylation of the mRNA Cap Protects RNAs from Decapping and Degradation by DXO. *PLOS ONE,* 13**,** e0193804.

PING, X. L., SUN, B. F., WANG, L., XIAO, W., YANG, X., WANG, W. J., ADHIKARI, S., SHI, Y., LV, Y., CHEN, Y. S., ZHAO, X., LI, A., YANG, Y., DAHAL, U., LOU, X. M., LIU, X., HUANG, J., YUAN, W. P., ZHU, X. F., CHENG, T., ZHAO, Y. L., WANG, X. Q., DANIELSEN, J. M. R., LIU, F. & YANG, Y. G. 2014. Mammalian WTAP is a Regulatory Subunit of the RNA N6-methyladenosine Methyltransferase. *Cell Research,* 24**,** 177-189.

PISAREVA, V. P. & PISAREV, A. V. 2014. eIF5 and eIF5B Together Stimulate 48S Initiation Complex Formation during Ribosomal Scanning. *Nucleic Acids Research,* 42**,** 12052-12069.

POWER, J. H. T., BARNES, O. L. & CHEGINI, F. 2017. Lewy Bodies and the Mechanisms of Neuronal Cell Death in Parkinson's Disease and Dementia with Lewy Bodies. *Brain Pathology,* 27**,** 3-12.

PROTS, I., GROSCH, J., BRAZDIS, R.-M., SIMMNACHER, K., VEBER, V., HAVLICEK, S., HANNAPPEL, C., KRACH, F., KRUMBIEGEL, M., SCHÜTZ, O., REIS, A., WRASIDLO, W., GALASKO, D. R., GROEMER, T. W., MASLIAH, E., SCHLÖTZER-SCHREHARDT, U., XIANG, W., WINKLER, J. & WINNER, B. 2018. α-Synuclein Oligomers Induce Early Axonal Dysfunction in Human iPSC-based Models of Synucleinopathies. *Proceedings of the National Academy of Sciences,* 115**,** 7813-7818.

QUEIROZ, R. M. L., SMITH, T., VILLANUEVA, E., MARTI-SOLANO, M., MONTI, M., PIZZINGA, M., MIREA, D.-M., RAMAKRISHNA, M., HARVEY, R. F., DEZI, V., THOMAS, G. H., WILLIS, A. E. & LILLEY, K. S. 2019. Comprehensive Identification of RNA–protein Interactions in any Organism using Orthogonal Organic Phase Separation (OOPS). *Nature Biotechnology,* 37**,** 169-178.

QUINN, J. G., COULSON, D. T. R., BROCKBANK, S., BEYER, N., RAVID, R., HELLEMANS, J., IRVINE, G. B. & JOHNSTON, J. A. 2012. α-Synuclein mRNA and Soluble α-synuclein Protein Levels in Post-mortem Brain from Patients with Parkinson's Disease, Dementia with Lewy Bodies, and Alzheimer's Disease. *Brain Research,* 1459**,** 71-80.

RAMANATHAN, A., ROBB, G. B. & CHAN, S.-H. 2016. mRNA Capping: Biological Functions and Applications. *Nucleic Acids Research,* 44**,** 7511-7526.
RANSOHOFF, R. M. 2018. All (Animal) Models (of Neurodegeneration) are Wrong. Are they also Useful? *The Journal of Experimental Medicine,* 215**,** 2955.
RAO, D. N., DRYDEN, D. T. F. & BHEEMANAIK, S. 2013. Type III Restriction-modification Enzymes: a Historical Perspective. *Nucleic Acids Research,* 42**,** 45-55.
RASMUSSEN, E. B. & LIS, J. T. 1993. In vivo Transcriptional Pausing and Cap Formation on Three Drosophila Heat Shock Genes. *Proceedings of the National Academy of Sciences,* 90**,** 7923.
RIBADENEYRA, C. & AMIN, S. 2016. EWSR1 Potentially Functions as an RNA Chaperone in the Stress Response. *The FASEB Journal,* 30**,** 594.3-594.3.
RICHTER, J. D. 2015. RNA and the Synapse. *RNA (New York, N.Y.),* 21**,** 716-717.
RICHTER, JOEL D. & COLLER, J. 2015. Pausing on Polyribosomes: Make Way for Elongation in Translational Control. *Cell,* 163**,** 292-300.
RIDDER, K., KELLER, S., DAMS, M., RUPP, A.-K., SCHLAUDRAFF, J., DEL TURCO, D., STARMANN, J., MACAS, J., KARPOVA, D., DEVRAJ, K., DEPBOYLU, C., LANDFRIED, B., ARNOLD, B., PLATE, K. H., HÖGLINGER, G., SÜLTMANN, H., ALTEVOGT, P. & MOMMA, S. 2014. Extracellular Vesicle-Mediated Transfer of Genetic Information between the Hematopoietic System and the Brain in Response to Inflammation. *PLOS Biology,* 12**,** e1001874.
RITCHIE JOSEPH, M. 1982. On the Relation between Ribre Diameter and Conduction Velocity in Myelinated Nerve Fibres. *Proceedings of the Royal Society of London. Series B. Biological Sciences,* 217**,** 29-35.
ROBBENS, S., ROUZÉ, P., COCK, J. M., SPRING, J., WORDEN, A. Z. & VAN DE PEER, Y. 2008. The FTO Gene, Implicated in Human Obesity, Is Found Only in Vertebrates and Marine Algae. *Journal of Molecular Evolution,* 66**,** 80-84.
ROBERTUS, J. D., LADNER, J. E., FINCH, J. T., RHODES, D., BROWN, R. S., CLARK, B. F. C. & KLUG, A. 1974. Structure of Yeast Phenylalanine tRNA at 3 Å Resolution. *Nature,* 250**,** 546-551.
RYSKALIN, L., LAZZERI, G., FLAIBANI, M., BIAGIONI, F., GAMBARDELLA, S., FRATI, A. & FORNAI, F. 2017. mTOR-Dependent Cell Proliferation in the Brain. *BioMed research international,* 2017**,** 7082696-7082696.
SAKUNO, T., ARAKI, Y., OHYA, Y., KOFUJI, S., TAKAHASHI, S., HOSHINO, S.-I. & KATADA, T. 2004. Decapping Reaction of mRNA Requires Dcp1 in Fission Yeast: Its Characterization in Different Species from Yeast to Human. *The Journal of Biochemistry,* 136**,** 805-812.
SANTINI, E., HUYNH, T. N., MACASKILL, A. F., CARTER, A. G., PIERRE, P., RUGGERO, D., KAPHZAN, H. & KLANN, E. 2012. Exaggerated Translation causes Synaptic and Behavioural Aberrations Sssociated with Autism. *Nature,* 493**,** 411.
SCHIBLER, U., KELLEY, D. E. & PERRY, R. P. 1977. Comparison of Methylated Sequences in Messenger RNA and Heterogeneous Nuclear RNA from Mouse L Cells. *Journal of Molecular Biology,* 115**,** 695-714.
SCHMITT, M. E., BROWN, T. A. & TRUMPOWER, B. L. 1990. A Rapid and Simple Method for Preparation of RNA from Saccharomyces cerevisiae. *Nucleic Acids Research,* 18**,** 3091-3092.
SCHUBERTH-WAGNER, C., LUDWIG, J., BRUDER, ANN K., HERZNER, A.-M., ZILLINGER, T., GOLDECK, M., SCHMIDT, T., SCHMID-BURGK, JONATHAN L., KERBER, R., WOLTER, S., STÜMPEL, J.-P., ROTH, A., BARTOK, E., DROSTEN, C., COCH, C., HORNUNG, V., BARCHET, W., KÜMMERER, BEATE M., HARTMANN, G. & SCHLEE, M. 2015. A Conserved Histidine in the RNA Sensor RIG-I Controls Immune Tolerance to N1-2′O-Methylated Self RNA. *Immunity,* 43**,** 41-51.
SCHUMAN, E. M., DYNES, J. L. & STEWARD, O. 2006. Synaptic Regulation of Translation of Dendritic mRNAs. *Journal of Neuroscience,* 26**,** 7143-7146.

SCHWANHÄUSSER, B., BUSSE, D., LI, N., DITTMAR, G., SCHUCHHARDT, J., WOLF, J., CHEN, W. & SELBACH, M. 2011. Global quantification of mammalian gene expression control. *Nature,* 473**,** 337.

SCHWARTZ, S., AGARWALA, S. D., MUMBACH, M. R., JOVANOVIC, M., MERTINS, P., SHISHKIN, A., TABACH, Y., MIKKELSEN, T. S., SATIJA, R., RUVKUN, G., CARR, S. A., LANDER, E. S., FINK, G. R. & REGEV, A. 2013. High-Resolution Mapping Reveals a Conserved, Widespread, Dynamic mRNA Methylation Program in Yeast Meiosis. *Cell,* 155**,** 1409-1421.

SCHWARTZ, S., MUMBACH, MAXWELL R., JOVANOVIC, M., WANG, T., MACIAG, K., BUSHKIN, G. G., MERTINS, P., TER-OVANESYAN, D., HABIB, N., CACCHIARELLI, D., SANJANA, NEVILLE E., FREINKMAN, E., PACOLD, MICHAEL E., SATIJA, R., MIKKELSEN, TARJEI S., HACOHEN, N., ZHANG, F., CARR, STEVEN A., LANDER, ERIC S. & REGEV, A. 2014. Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5′ Sites. *Cell Reports,* 8**,** 284-296.

SENDINC, E., VALLE-GARCIA, D., DHALL, A., CHEN, H., HENRIQUES, T., NAVARRETE-PEREA, J., SHENG, W., GYGI, S. P., ADELMAN, K. & SHI, Y. 2019. PCIF1 Catalyzes m6Am mRNA Methylation to Regulate Gene Expression. *Molecular Cell*.

SENDINC, E., VALLE-GARCIA, D., DHALL, A., CHEN, H., HENRIQUES, T., SHENG, W., ADELMAN, K. & SHI, Y. 2018. PCIF1 Catalyzes m6Am mRNA Methylation to Regulate Gene Expression. *bioRxiv***,** 484931.

SHE, M., DECKER, C. J., SVERGUN, D. I., ROUND, A., CHEN, N., MUHLRAD, D., PARKER, R. & SONG, H. 2008. Structural Basis of Dcp2 Recognition and Activation by Dcp1. *Molecular Cell,* 29**,** 337-349.

SHEPPARD, P., CHOLERIS, E. & GALEA, L. 2019. Structural Plasticity of the Hippocampus in Response to Estrogens in Female Rodents. *Molecular Brain,* 12.

SHI, H., WANG, X., LU, Z., ZHAO, B. S., MA, H., HSU, P. J., LIU, C. & HE, C. 2017. YTHDF3 Facilitates Translation and Decay of N6-methyladenosine-modified RNA. *Cell Research,* 27**,** 315.

SHI, H., WEI, J. & HE, C. 2019. Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Molecular Cell,* 74**,** 640-650.

SHI, H., ZHANG, X., WENG, Y.-L., LU, Z., LIU, Y., LU, Z., LI, J., HAO, P., ZHANG, Y., ZHANG, F., WU, Y., DELGADO, J. Y., SU, Y., PATEL, M. J., CAO, X., SHEN, B., HUANG, X., MING, G.-L., ZHUANG, X., SONG, H., HE, C. & ZHOU, T. 2018. m6A Facilitates Hippocampus-dependent Learning and Memory through YTHDF1. *Nature,* 563**,** 249-253.

SHIGEOKA, T., LU, B. & HOLT, C. E. 2013. RNA-based mechanisms underlying axon guidance. *The Journal of Cell Biology,* 202**,** 991.

SHUMAN, S. 2000. Structure, Mechanism, and Evolution of the mRNA Capping Apparatus. *Progress in Nucleic Acid Research and Molecular Biology.* Academic Press.

SIMABUCO, F. M., PAVAN, I. C. B., PESTANA, N. F., CARVALHO, P. C., BASEI, F. L., CAMPOS GRANATO, D., PAES LEME, A. F. & ZANCHIN, N. I. T. 2019. Interactome Analysis of the Human Cap-specific mRNA (nucleoside-2′-O-)-methyltransferase 1 (hMTr1) Protein. *Journal of Cellular Biochemistry,* 120**,** 5597-5611.

SLOAN, K. E., WARDA, A. S., SHARMA, S., ENTIAN, K.-D., LAFONTAINE, D. L. J. & BOHNSACK, M. T. 2017. Tuning the Ribosome: The Influence of rRNA Modification on Eukaryotic Ribosome Biogenesis and Function. *RNA biology,* 14**,** 1138-1152.

SONG, M.-G., LI, Y. & KILEDJIAN, M. 2010. Multiple mRNA Decapping Enzymes in Mammalian Cells. *Molecular Cell,* 40**,** 423-432.

SPITALE, R. C., FLYNN, R. A., ZHANG, Q. C., CRISALLI, P., LEE, B., JUNG, J. W., KUCHELMEISTER, H. Y., BATISTA, P. J., TORRE, E. A., KOOL, E. T. & CHANG, H. Y. 2015. Structural Imprints in vivo Decode RNA Regulatory Mechanisms. *Nature,* 519**,** 486-+.

STAMATAKIS, A. 2014. RAxML Version 8: a Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics,* 30**,** 1312-1313.

STEVENS, A. & MAUPIN, M. K. 1987. A 5′ → 3′ Exoribonuclease of Saccharomyces cerevisiae: Size and Novel Substrate Specificity. *Archives of Biochemistry and Biophysics,* 252**,** 339-347.

STEWARD, O. & REEVES, T. M. 1988. Protein-Synthetic Machinery Beneath Postsynaptic Sites on CNS Neurons - Association between Polyribosomes and other Organelles at the Synaptic Site. *Journal of Neuroscience,* 8**,** 176-184.

SU, A. I., PEZACKI, J. P., WODICKA, L., BRIDEAU, A. D., SUPEKOVA, L., THIMME, R., WIELAND, S., BUKH, J., PURCELL, R. H., SCHULTZ, P. G. & CHISARI, F. V. 2002. Genomic Analysis of the Host Response to Hepatitis C Virus Infection. *Proceedings of the National Academy of Sciences,* 99**,** 15669.

SU, X., FISCHER, D. L., LI, X., BANKIEWICZ, K., SORTWELL, C. E. & FEDEROFF, H. J. 2017. Alpha-Synuclein mRNA Is Not Increased in Sporadic PD and Alpha-Synuclein Accumulation Does Not Block GDNF Signaling in Parkinson's Disease and Disease Models. *Molecular Therapy,* 25**,** 2231-2235.

SUETTERLIN, P. & DRESCHER, U. 2014. Target-independent ephrina/EphA-mediated Axon-axon Repulsion as a Novel Element in Retinocollicular Mapping. *Neuron,* 84**,** 740-752.

SUN, H., ZHANG, M., LI, K., BAI, D. & YI, C. 2018. Cap-specific, Terminal N6-methylation by a Mammalian m6Am Methyltransferase. *Cell Research*.

SURENDRANATHAN, A., ROWE, J. B. & O'BRIEN, J. T. 2015. Neuroinflammation in Lewy body dementia. *Parkinsonism & Related Disorders,* 21**,** 1398-1406.

SUTANDY, F. X. R., EBERSBERGER, S., HUANG, L., BUSCH, A., BACH, M., KANG, H.-S., FALLMANN, J., MATICZKA, D., BACKOFEN, R., STADLER, P. F., ZARNACK, K., SATTLER, M., LEGEWIE, S. & KÖNIG, J. 2018. In vitro iCLIP-based Modeling Uncovers how the Splicing Factor U2AF2 Relies on Regulation by Cofactors. *Genome Research,* 28**,** 699-713.

SVORC, P. 2018. Introductory Chapter: Autonomic Nervous System - What We Know About It. *In:* SVORC, P. (ed.) *Autonomic Nervous System.* intechopen.

TAKEUCHI, T., DUSZKIEWICZ, A. J. & MORRIS, R. G. M. 2014. The Synaptic Plasticity and Memory Hypothesis: Encoding, Storage and Persistence. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 369.

TANSEY, M. G. & GOLDBERG, M. S. 2010. Neuroinflammation in Parkinson's Disease: Its Role in Neuronal Death and Implications for Therapeutic Intervention. *Neurobiology of Disease,* 37**,** 510-518.

TANSEY, M. G. & ROMERO-RAMOS, M. 2019. Immune System Responses in Parkinson's disease: Early and Dynamic. *European Journal of Neuroscience,* 49**,** 364-383.

TAO, L. & REESE, T. A. 2017. Making Mouse Models That Reflect Human Immune Responses. *Trends in Immunology,* 38**,** 181-193.

TAOKA, M., NOBE, Y., YAMAKI, Y., SATO, K., ISHIKAWA, H., IZUMIKAWA, K., YAMAUCHI, Y., HIROTA, K., NAKAYAMA, H., TAKAHASHI, N. & ISOBE, T. 2018. Landscape of the Complete RNA Chemical Modifications in the Human 80S Ribosome. *Nucleic Acids Research,* 46**,** 9289-9298.

TAPIA, J. E., GONZÁLEZ, B., GOULITQUER, S., POTIN, P. & CORREA, J. A. 2016. Microbiota Influences Morphology and Reproduction of the Brown Alga Ectocarpus sp. *Frontiers in Microbiology,* 7.

THOMAS, B. & AKOULITCHEV, A. V. 2006. Mass Spectrometry of RNA. *Trends in Biochemical Sciences,* 31**,** 173-181.

TIRUMURU, N. & WU, L. 2019. HIV-1 Envelope Proteins Up-regulate N6-methyladenosine Levels of Cellular RNA Independently of Viral replication. *Journal of Biological Chemistry,* 294**,** 3249-3260.

TOCZYDLOWSKA-SOCHA, D., ZIELINSKA, M. M., KURKOWSKA, M., ASTHA, ALMEIDA, C. F., STEFANIAK, F., PURTA, E. & BUJNICKI, J. M. 2018. Human RNA cap1 Methyltransferase CMTr1 Cooperates with RNA Helicase DHX15 to Modify RNAs with Highly Structured 5' termini. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 373**,** 20180161.

TOMOO, K., SHEN, X., OKABE, K., NOZOE, Y., FUKUHARA, S., MORINO, S., ISHIDA, T., TANIGUCHI, T., HASEGAWA, H., TERASHIMA, A., SASAKI, M., KATSUYA, Y., KITAMURA, K., MIYOSHI, H., ISHIKAWA, M. & MIURA, K.-I. 2002. Crystal Structures of 7-methylguanosine 5′-triphosphate (m7GTP)- and P1-7-methylguanosine-P3-adenosine-5′,5′-triphosphate (m7GpppA)-bound Human Full-length Eukaryotic Initiation Factor 4E: Biological Importance of the C-terminal Flexible Region. *Biochemical Journal,* 362**,** 539-544.

URDANETA, E. C., VIEIRA-VIEIRA, C. H., HICK, T., WESSELS, H.-H., FIGINI, D., MOSCHALL, R., MEDENBACH, J., OHLER, U., GRANNEMAN, S., SELBACH, M. & BECKMANN, B. M. 2019. Purification of Cross-linked RNA-protein Complexes by Phenol-toluol Extraction. *Nature Communications,* 10**,** 990.

VALADKHAN, S. & GUNAWARDANE, L. S. 2013. Role of Small Nuclear RNAs in Eukaryotic Gene Expression. *Essays In Biochemistry,* 54**,** 79.

VALKOV, E., MUTHUKUMAR, S., CHANG, C.-T., JONAS, S., WEICHENRIEDER, O. & IZAURRALDE, E. 2016. Structure of the Dcp2–Dcp1 mRNA-decapping Complex in the Activated Conformation. *Nature Structural &Amp; Molecular Biology,* 23**,** 574.

VAN DIJK, R. M., HUANG, S.-H., SLOMIANKA, L. & AMREIN, I. 2016. Taxonomic Separation of Hippocampal Networks: Principal Cell Populations and Adult Neurogenesis. *Frontiers in neuroanatomy,* 10**,** 22-22.

VAUGHAN, R. A. & FOSTER, J. D. 2013. Mechanisms of Dopamine Transporter Regulation in Normal and Disease States. *Trends in Pharmacological Sciences,* 34**,** 489-496.

VEN MURTHY, M. R. 1982. Blocked and Methylated 5′-Terminal Cap Structures of Rat Brain Messenger Ribonucleic Acids. *Journal of Neurochemistry,* 38**,** 28-40.

VERA, A., GONZÁLEZ-MONTALBÁN, N., ARÍS, A. & VILLAVERDE, A. 2007. The Conformational Quality of Insoluble Recombinant Rroteins is Enhanced at Low Growth Temperatures. *Biotechnology and Bioengineering,* 96**,** 1101-1106.

VICKOVIC, S., ERASLAN, G., SALMÉN, F., KLUGHAMMER, J., STENBECK, L., ÄIJÖ, T., BONNEAU, R., BERGENSTRÅHLE, L., FERNANDÉZ NAVARRO, J., GOULD, J., RONAGHI, M., FRISÉN, J., LUNDEBERG, J., REGEV, A. & STÅHL, P. L. 2019. High-density Spatial Transcriptomics Arrays for in situ Tissue Profiling. *bioRxiv***,** 563338.

VUJOVIC, P., STAMENKOVIC, S., JASNIC, N., LAKIC, I., DJURASEVIC, S. F., CVIJIC, G. & DJORDJEVIC, J. 2013. Fasting Induced Cytoplasmic Fto Expression in Some Neurons of Rat Hypothalamus. *PLOS ONE,* 8**,** e63694.

WALAAS, S. I. & GREENGARD, P. 1991. Protein phosphorylation and neuronal function. *Pharmacological Reviews,* 43**,** 299.

WALTERS, B. J., MERCALDO, V., GILLON, C. J., YIP, M., NEVE, R. L., BOYCE, F. M., FRANKLAND, P. W. & JOSSELYN, S. A. 2017. The Role of The RNA Demethylase FTO (Fat Mass and Obesity-Associated) and mRNA Methylation in Hippocampal Memory Formation. *Neuropsychopharmacology*.

WANG, E. T., SANDBERG, R., LUO, S., KHREBTUKOVA, I., ZHANG, L., MAYR, C., KINGSMORE, S. F., SCHROTH, G. P. & BURGE, C. B. 2008. Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature,* 456**,** 470.

WANG, L., SCHUSTER, G. U., HULTENBY, K., ZHANG, Q., ANDERSSON, S. & GUSTAFSSON, J.-A. 2002a. Liver X Receptors in the Central Nervous System: from Lipid Homeostasis to Neuronal Degeneration. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 13878-13883.

WANG, X., LU, Z. K., GOMEZ, A., HON, G. C., YUE, Y. N., HAN, D. L., FU, Y., PARISIEN, M., DAI, Q., JIA, G. F., REN, B., PAN, T. & HE, C. 2014. N-6-Methyladenosine-Dependent Regulation of messenger RNA Stability. *Nature,* 505**,** 117-+.

WANG, X., ZHAO, B. S., ROUNDTREE, I. A., LU, Z. K., HAN, D. L., MA, H. H., WENG, X. C., CHEN, K., SHI, H. L. & HE, C. 2015. N-6-Methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell,* 161**,** 1388-1399.

WANG, Y. & ZHAO, J. C. 2016. Update: Mechanisms Underlying N6-Methyladenosine Modification of Eukaryotic mRNA. *Trends in Genetics,* 32**,** 763-773.

WANG, Z., JIAO, X., CARR-SCHMID, A. & KILEDJIAN, M. 2002b. The hDcp2 Protein is a Mammalian mRNA Decapping Enzyme. *Proceedings of the National Academy of Sciences,* 99**,** 12663.

WEHRSPAUN, C. C., HAERTY, W. & PONTING, C. P. 2015. Microglia Recapitulate a Hematopoietic Master Regulator Network in the Aging Human Frontal Cortex. *Neurobiology of aging,* 36**,** 2443.e9-2443.e20.

WEI, C.-M., GERSHOWITZ, A. & MOSS, B. 1975. N6, O2'-dimethyladenosine a Novel Methylated Ribonucleoside next to the 5' terminal of Animal Cell and Virus mRNAs. *Nature,* 257**,** 251-253.

WEI, C. M., GERSHOWITZ, A. & MOSS, B. 1976. 5'-Terminal and Internal Methylated Nucleotide-Sequences in HELA-Cell messenger-RNA. *Biochemistry,* 15**,** 397-401.

WEI, J., LIU, F., LU, Z., FEI, Q., AI, Y., HE, P. C., SHI, H., CUI, X., SU, R., KLUNGLAND, A., JIA, G., CHEN, J. & HE, C. 2018. Differential m6A, m6Am, and m1A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Molecular Cell,* 71**,** 973-985.e5.

WEI, Z., BATAGOV, A. O., SCHINELLI, S., WANG, J., WANG, Y., EL FATIMY, R., RABINOVSKY, R., BALAJ, L., CHEN, C. C., HOCHBERG, F., CARTER, B., BREAKEFIELD, X. O. & KRICHEVSKY, A. M. 2017. Coding and Noncoding Landscape of Extracellular RNA Released by Human Glioma Stem Cells. *Nature Communications,* 8**,** 1145.

WEN, J., LV, R., MA, H., SHEN, H., HE, C., WANG, J., JIAO, F., LIU, H., YANG, P., TAN, L., LAN, F., SHI, Y. G., HE, C., SHI, Y. & DIAO, J. 2018. Zc3h13 Regulates Nuclear RNA m6A Methylation and Mouse Embryonic Stem Cell Self-Renewal. *Molecular Cell,* 69**,** 1028-1038.e6.

WERNER, M., PURTA, E., KAMINSKA, K. H., CYMERMAN, I. A., CAMPBELL, D. A., MITTRA, B., ZAMUDIO, J. R., STURM, N. R., JAWORSKI, J. & BUJNICKI, J. M. 2011. 2 '-O-ribose Methylation of cap2 in Human: Function and Evolution in a Horizontally Mobile Family. *Nucleic Acids Research,* 39**,** 4756-4768.

WIDAGDO, J., ZHAO, Q.-Y., KEMPEN, M.-J., TAN, M. C., RATNU, V. S., WEI, W., LEIGHTON, L., SPADARO, P. A., EDSON, J., ANGGONO, V. & BREDY, T. W. 2016. Experience-Dependent Accumulation of N(6)-Methyladenosine in the Prefrontal Cortex Is Associated with Memory Processes in Mice. *The Journal of Neuroscience,* 36**,** 6771-6777.

WU, B., SU, S., PATIL, D. P., LIU, H., GAN, J., JAFFREY, S. R. & MA, J. 2018. Molecular Basis for the Specific and Multivariant Recognitions of RNA Substrates by Human hnRNP A2/B1. *Nature Communications,* 9**,** 420.

WU, Y., YANG, X., CHEN, Z., TIAN, L., JIANG, G., CHEN, F., LI, J., AN, P., LU, L., LUO, N., DU, J., SHAN, H., LIU, H. & WANG, H. 2019. m6A-induced lncRNA RP11 Triggers the Dissemination of Colorectal cancer Cells via Upregulation of Zeb1. *Molecular Cancer,* 18**,** 87.

WURM, J. P., HOLDERMANN, I., OVERBECK, J. H., MAYER, P. H. O. & SPRANGERS, R. 2017. Changes in Conformational Equilibria Regulate the Activity of the Dcp2 Decapping Enzyme. *Proceedings of the National Academy of Sciences,* 114**,** 6034.

XIAO, W., ADHIKARI, S., DAHAL, U., CHEN, Y. S., HAO, Y. J., SUN, B. F., SUN, H. Y., LI, A., PING, X. L., LAI, W. Y., WANG, X., MA, H. L., HUANG, C. M., YANG, Y., HUANG, N., JIANG, G. B., WANG, H. L., ZHOU, Q., WANG, X. J., ZHAO, Y. L. & YANG, Y. G. 2016. Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell,* 61**,** 507-519.

XU, C., WANG, X., LIU, K., ROUNDTREE, I. A., TEMPEL, W., LI, Y. J., LU, Z. K., HE, C. & MIN, J. R. 2014. Structural Basis for Selective Binding of m(6)A RNA by the YTHDC1 YTH Domain. *Nature Chemical Biology,* 10**,** 927-929.

YABLONOVITCH, A. L., DENG, P., JACOBSON, D. & LI, J. B. 2017. The Evolution and Adaptation of A-to-I RNA Editing. *PLoS genetics,* 13**,** e1007064-e1007064.

YAN, X., LI, F., MAIXNER, D. W., YADAV, R., GAO, M., ALI, M. W., HOOKS, S. B. & WENG, H.-R. 2019. Interleukin-1beta Released by Microglia Initiates the Enhanced

Glutamatergic Activity in the Spinal Dorsal Horn during Paclitaxel-associated Acute Pain Syndrome. *Glia,* 67**,** 482-497.
YANG, F., YI, F., HAN, X., DU, Q. & LIANG, Z. 2013. MALAT-1 interacts with hnRNP C in Cell Cycle Regulation. *FEBS Letters,* 587**,** 3175-3181.
YANG, Y., HSU, P. J., CHEN, Y.-S. & YANG, Y.-G. 2018. Dynamic transcriptomic m6A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Research,* 28**,** 616-624.
YEO, M., LIN, P. S., DAHMUS, M. E. & GILL, G. N. 2003. A Novel RNA Polymerase II C-terminal Domain Phosphatase That Preferentially Dephosphorylates Serine 5. *Journal of Biological Chemistry,* 278**,** 26078-26085.
YOON, K.-J., RINGELING, F. R., VISSERS, C., JACOB, F., POKRASS, M., JIMENEZ-CYRUS, D., SU, Y., KIM, N.-S., ZHU, Y., ZHENG, L., KIM, S., WANG, X., DORÉ, L. C., JIN, P., REGOT, S., ZHUANG, X., CANZAR, S., HE, C., MING, G.-L. & SONG, H. 2017. Temporal Control of Mammalian Cortical Neurogenesis by m6A Methylation. *Cell,* 171**,** 877-889.e17.
YOON, Y., PARK, H., KIM, S., NGUYEN, P. T., HYEON, S. J., CHUNG, S., IM, H., LEE, J., LEE, S. B. & RYU, H. 2018. Genetic Ablation of EWS RNA Binding Protein 1 (EWSR1) Leads to Neuroanatomical Changes and Motor Dysfunction in Mice. *Experimental neurobiology,* 27**,** 103-111.
ZAPPULO, A., VAN DEN BRUCK, D., CIOLLI MATTIOLI, C., FRANKE, V., IMAMI, K., MCSHANE, E., MORENO-ESTELLES, M., CALVIELLO, L., FILIPCHYK, A., PEGUERO-SANCHEZ, E., MÜLLER, T., WOEHLER, A., BIRCHMEIER, C., MERINO, E., RAJEWSKY, N., OHLER, U., MAZZONI, E. O., SELBACH, M., AKALIN, A. & CHEKULAEVA, M. 2017. RNA Localization is a Key Determinant of Neurite-Enriched Proteome. *Nature Communications,* 8**,** 583.
ZHANG, C., CHEN, Y., SUN, B., WANG, L., YANG, Y., MA, D., LV, J., HENG, J., DING, Y., XUE, Y., LU, X., XIAO, W., YANG, Y.-G. & LIU, F. 2017. m6A Modulates Haematopoietic Stem and Progenitor Cell Specification. *Nature,* 549**,** 273-276.
ZHANG, Q.-S., WANG, Z.-H., ZHANG, J.-L., DUAN, Y.-L., LI, G.-F. & ZHENG, D.-L. 2016. Beta-asarone Protects against MPTP-induced Parkinson's Disease via Regulating Long Non-coding RNA MALAT1 and Inhibiting α-synuclein Protein Expression. *Biomedicine & Pharmacotherapy,* 83**,** 153-159.
ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology,* 9**,** R137.
ZHAO, L., ZHANG, H., KOHNEN, M. V., PRASAD, K. V. S. K., GU, L. & REDDY, A. S. N. 2019. Analysis of Transcriptome and Epitranscriptome in Plants Using PacBio Iso-Seq and Nanopore-Based Direct RNA Sequencing. *Frontiers in Genetics,* 10.
ZHAO, X., YANG, Y., SUN, B.-F., SHI, Y., YANG, X., XIAO, W., HAO, Y.-J., PING, X.-L., CHEN, Y.-S., WANG, W.-J., JIN, K.-X., WANG, X., HUANG, C.-M., FU, Y., GE, X.-M., SONG, S.-H., JEONG, H. S., YANAGISAWA, H., NIU, Y., JIA, G.-F., WU, W., TONG, W.-M., OKAMOTO, A., HE, C., DANIELSEN, J. M. R., WANG, X.-J. & YANG, Y.-G. 2014. FTO-dependent Demethylation of N6-methyladenosine Regulates mRNA Splicing and is Required for Adipogenesis. *Cell Research,* 24**,** 1403-1419.
ZHENG, G. Q., DAHL, J. A., NIU, Y. M., FEDORCSAK, P., HUANG, C. M., LI, C. J., VAGBO, C. B., SHI, Y., WANG, W. L., SONG, S. H., LU, Z. K., BOSMANS, R. P. G., DAI, Q., HAO, Y. J., YANG, X., ZHAO, W. M., TONG, W. M., WANG, X. J., BOGDAN, F., FURU, K., FU, Y., JIA, G. F., ZHAO, X., LIU, J., KROKAN, H. E., KLUNGLAND, A., YANG, Y. G. & HE, C. 2013. ALKBH5 is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Molecular Cell,* 49**,** 18-29.
ZHONG, S. L., LI, H. Y., BODI, Z., BUTTON, J., VESPA, L., HERZOG, M. & FRAY, R. G. 2008. MTA is an Arabidopsis Messenger RNA Adenosine Methylase and Interacts with a Homolog of a Sex-specific Splicing Factor. *Plant Cell,* 20**,** 1278-1288.
ZHOU, K. I., PARISIEN, M., DAI, Q., LIU, N., DIATCHENKO, L., SACHLEBEN, J. R. & PAN, T. 2016. N(6)-Methyladenosine Modification in a Long Noncoding RNA Hairpin

Predisposes Its Conformation to Protein Binding. *Journal of molecular biology,* 428**,** 822-833.

ZHOU, X., LI, X., CHENG, Y., WU, W., XIE, Z., XI, Q., HAN, J., WU, G., FANG, J. & FENG, Y. 2014. BCLAF1 and its Splicing Regulator SRSF10 Regulate the Tumorigenic Potential of Colon Cancer Cells. *Nature Communications,* 5**,** 4581.

ZHOU, Y., GU, C., LI, J., ZHU, L., HUANG, G., DAI, J. & HUANG, H. 2018. Aberrantly Expressed Long Noncoding RNAs and Genes in Parkinson's Disease. *Neuropsychiatric disease and treatment,* 14**,** 3219-3229.

ZÜST, R., CERVANTES-BARRAGAN, L., HABJAN, M., MAIER, R., NEUMAN, B. W., ZIEBUHR, J., SZRETTER, K. J., BAKER, S. C., BARCHET, W., DIAMOND, M. S., SIDDELL, S. G., LUDEWIG, B. & THIEL, V. 2011. Ribose 2′-O-methylation Provides a Molecular Signature for the Distinction of Self and Non-self mRNA Dependent on the RNA Sensor Mda5. *Nature Immunology,* 12**,** 137.

Appendix

Accession numbers for proteins in evolutionary trees

Table of accession numbers for the protein sequences found in the FTO tree of similarity.

Species name	Protein ID	Name code
Homo Sapiens	AAI26228.1	DNMT1
	NP_006011.2	ALKBH1
	NP_001001655.1	ALKBH2
	AAH67257.1	ALKBH3
	NP_060091.1	ALKBH4
	NP_060228.3	ALKBH5
	NP_942567.1	ALKBH6
	NP_115682.1	ALKBH7
	NP_001287939.1	ALKBH8
	NP_001073901.1	FTO
Rattus norvegicus	NP_001102188.1	ALKBH1
	NP_001119745.1	ALKBH2
	NP_001014202.1	ALKBH3
	NP_001099390.1	ALKBH4
	NP_001178572.1	ALKBH5
	NP_001120922.1	ALKBH6
	NP_001102854.1	ALKBH7
	NP_001178838.1	ALKBH8
	NP_001034802.1	FTO
Gallus gallus	NP_001026723.1	ALKBH1
	NP_001264426.1	ALKBH2
	NP_001269306.1	ALKBH3
	XP_415758.3	ALKBH4
	NP_001244130.1	ALKBH5
	XP_004938883.1	ALKBH8
	NP_001172076.1	FTO
Xenopus tropicalis	NP_001072394.1	ALKBH1
	XP_002932107.2	ALKBH2
	XP_017948748.1	ALKBH3
	XP_002942310.1	ALKBH4
	NP_001008038.1	ALKBH5
	NP_001072221.1	ALKBH6
	XP_002940667.2	ALKBH7
	XP_012812512.1	ALKBH8

	NP_001017273.1	FTO
Danio rerio	NP_001018527.1	ALKBH1
	NP_001243613.1	ALKBH2
	XP_017209591.1	ALKBH3
	NP_001313299.1	ALKBH4
	NP_001070855.1	ALKBH5
	NP_001005390.1	ALKBH6
	NP_001017609.1	ALKBH7
	XP_009290865.1	ALKBH8
	XP_021327673.1	FTO
Drosophila melanogaster	NP_996458.1	ALKB
Arabidopsis thaliana	Q9SA98.2	ALKBH
	Q9SIE0.2	ALKBH2
Escherichia coli	AAA16409.1	ALKBH
Micromonas commodas	5812	ALKBH1_1
	7616	ALKBH1_2
	3751	ALKBH1_3
	6795	ALKBH1_4
	7785	ALKBH4_1
	8368	ALKBH4_2
	7634	ALKBH6_1
	9587	ALKBH6_2
	5984	ALKBH8_1
	8368	ALKBH8_2
	9146	ALKBH8_3
	XP_002502764.1	FATSO
Micromonas_pusilla	1792	ALKBH1_1
	8013	ALKBH1_2
	5817	ALKBH1_3
	4240	ALKBH1_4
	1267	ALKBH4
	4406	ALKBH5_1
	4978	ALKBH5_2
	9141	ALKBH6_1
	3030	ALKBH6_2
	7306	ALKBH8_1
	1267	ALKBH8_2
	2325	ALKBH8_3
	ABY26524.1	FTO
Ostreococcus_tauri	69332	ALKBH1
	71234	ALKBH2
	70119	ALKBH4

	70940	ALKBH8_1
	70119	ALKBH8_2
	OUS43030.1	FTO
Thecamonas_trahens	KNC51118	ALKBH1
	KNC47619	ALKBH2_1
	KNC49768	ALKBH2_2
	KNC52673	ALKBH4
	KNC47622	ALKBH5
	KNC54709	ALKBH6
	KNC47622	ALKBH7
	KNC51592	ALKBH8_1
	KNC47622	ALKBH8_2
	KNC52673	ALKBH8_3
	KNC54709	ALKBH8_4
	XP_013753876.1	FATSO
Aphanomyces_invadans	ETV95596	ALKBH8_1
	ETV95594	ALKBH8_2
	ETV95595	ALKBH8_3
	ETV94055	ALKBH7_1
	ETV94056	ALKBH7_2
	ETV95596	ALKBH7_3
	ETV95594	ALKBH7_4
	ETV94055	ALKBH7_5
	ETV94056	ALKBH7_6
	ETW00557	ALKBH6
	ETV95596	ALKBH4_1
	ETV95594	ALKBH4_2
	ETV95595	ALKBH4_3
	ETV91041	ALKBH4_4
	ETW00546	ALKBH2_1
	ETW00547	ALKBH2_2
	ETW00548	ALKBH2_3
	ETW03157	ALKBH2_4
	ETW03537	ALKBH1_1
	ETW08398	ALKBH1_2
	ETW08399	ALKBH1_3
	ETV97029.1	FTO
Ectocarpus siliculosus	30278	ALKBH2
	26840	ALKBH6_1
	16605	ALKBH6_2
	25164	ALKBH8_1
	22752	ALKBH8_2

	CBJ31658.1	FTO
Guillardia theta	EKX47012	ALKBH1
	EKX50518	ALKBH2_1
	EKX54295	ALKBH2_2
	EKX39153	ALKBH4
	EKX38899	ALKBH5_1
	EKX41932	ALKBH5_2
	EKX53162	ALKBH5_3
	EKX35943	ALKBH6_1
	EKX41469	ALKBH6_2
	EKX50807	ALKBH8_1
	EKX38899	ALKBH8_2
	EKX51400	ALKBH8_3
	EKX51248.1	FTO
Capsaspora owczarzaki	KJE88374	ALKBH2
	KJE93517	ALKBH6_1
	KJE93518	ALKBH6_2
	KJE89644	ALKBH7
	KJE88559	ALKBH8_1
	KJE97982	ALKBH8_2
	KJE93176.1	FTO
Aphanomyces astaci	ETV72525	ALKBH1_1
	ETV89649	ALKBH1_2
	ETV84764	ALKBH2_1
	ETV74663	ALKBH2_2
	ETV81296	ALKBH4
	ETV82698	ALKBH5
	ETV84751	ALKBH6
	ETV66352	ALKBH7_1
	ETV66350	ALKBH7_2
	ETV66351	ALKBH7_3
	ETV66349	ALKBH7_4
	ETV73688	ALKBH8_1
	ETV73687	ALKBH8_2
	ETV73689	ALKBH8_3
	ETV73690	ALKBH8_4
	ETV81296	ALKBH8_5
	XP_009841387.1	FTO
Thalassiosira oceanica	73010	ALKBH1
	103875	ALKBH2_1
	79524	ALKBH2_2
	76266	ALKBH7

	101442	ALKBH8
	EJK46618.1	FTO
Thalassiosira pseudonana	8160	ALKBH1
	6486	ALKBH2_1
	30163	ALKBH2_2
	31380	ALKBH7
	25611	ALKBH8_1
	38448	ALKBH8_2
	17983	ALKBH8_3
	16715	ALKBH8_4
	30569	ALKBH8_5
	30568	ALKBH8_6
	EED95400.1	FTO
Ostreococcus lucimarinus	28948	ALKBH1
	43704	ALKBH2_1
	26042	ALKBH2_2
	88555	ALKBH4
	27202	ALKBH8_1
	36778	ALKBH8_2
	XP_001420808.1	FTO
Phaeodactylum tricornutum	46782	ALKBH1
	44026	ALKBH2
	41221	ALKBH7
	42894	ALKBH8_1
	9446	ALKBH8_2
	4653	ALKBH8_3
	24759	ALKBH8_4
	41429	FTO

Table of Accession numbers of the protein sequences in the PCIF1 tree of similarity.

Species name	Protein ID	Name code
Homo sapiens	BAC45238.1	PCIF1
	AAI26228.1	DNMT1
Rattus norvegicus	NP_001102075.1	PCIF1
Gallus gallus	XP_015152050	PCIF1
Danio rerio	XP_688220	PCIF1
Xenopus tropicalis	NP_001017341.1	PCIF1
Apis mellifera	XP_624144.2	PCIF1
Drosophila melanogaster	AAK93556	PCIF1
Octopus bimaculoides	KOF73944.1	PCIF1
Micromonas commode	6112	PCIF1_1
	1250	PCIF1_2
	10103	PCIF1_3
	653	PCIF1_4
Micromonas pusilla	4019	PCIF1_1
	7427	PCIF1_2
	6679	PCIF1_3
	374	PCIF1_4
	4493	PCIF1_5
Ostreococcus tauri	76150	PCIF1_1
	75522	PCIF1_2
	74487	PCIF1_3
Thecamonas trahens	KNC52772	PCIF1_1
	KNC47011	PCIF1_2
	KNC55176	PCIF1_3
	KNC51585	PCIF1_4
	KNC56123	PCIF1_5
Aphanomyces invadans	ETV98440	PCIF1_1
	ETW01278	PCIF1_2
	ETW01934	PCIF1_3
	ETW08457	PCIF1_4
	ETV95686	PCIF1_5
	ETW05309	PCIF1_6
Ectocarpus siliculosus	18400	PCIF1_1
	19650	PCIF1_2
Guillardia theta	102834	PCIF1_1
	135823	PCIF1_2
	158575	PCIF1_3
	144863	PCIF1_4
Capsaspora owczarzaki	KJE89362	PCIF1_1
	KJE93372	PCIF1_2

Aphanomyces astaci	ETV78266	PCIF1_1
	ETV88207	PCIF1_2
	ETV81173	PCIF1_3
	ETV73782	PCIF1_4
	ETV65050	PCIF1_5
Thalassiosira oceanica	95295	PCIF1_1
	72686	PCIF1_2
Thalassiosira pseudona	20873	PCIF1_1
	24679	PCIF1_2
Osterococcus lucimarinus	ABO97717	PCIF1_1
	ABO96253	PCIF1_2
	ABP00094	PCIF1_3
	ABO97203	PCIF1_4
	ABO94474	PCIF1_5
	ABP00214	PCIF1_6
	ABO98018	PCIF1_7
Phaeodactylum tricornutum	43163	PCIF1_1
	32323	PCIF1_2

Table of accession numbers for the protein sequences in the CMTR1 and CMTR2 tree of similarity.

Species name	Protein ID	Name code
Homo sapeins	AAI26228.1	DNMT1
	NP_055865.1	CMTR1
	NP_001311306.1	CMTR2
Gallus gallus	XP_015139279.1	CMTR1
	XP_025010064.1	CMTR2
Xenopus tropicalis	NP_001123814.1	CMTR1
	XP_004914385.1	CMTR2
Danio rerio	NP_956427.1	CMTR1
	NP_998598.1	CMTR2
Rattus norvegicus	NP_001014053.1	CMTR1
	NP_001099656.1	CMTR2
Drosphila melanogaster	Q9W4N2.1	CMTR1
	Q9UAS6.2	CMTR2
Apis mellifera	GB54512-RA	CMTR1
	GB41204-RA	CMTR2
Octopus bimaculoides	Ocbimv22034299m	CMTR1
	Ocbimv22001585m	CMTR2
Arabidopsis thaliana	Q9LHQ7.1	CMTR1
	Q5HZ60.1	CMTR2
Ostreococcus tauri	71744	CMTR
Thecamonas trachens	KNC52404	CMTR_1
	KNC46990	CMTR_2
	KNC49934	CMTR_3
Aphanomyces invadans	ETV98300	CMTR_1
	ETV98181	CMTR_2
Ectocarpus Siliculosus	18033	CMTR
Guillardia theta	EKX39952	CMTR_1
	EKX43947	CMTR_2
	EKX40861	CMTR_3
Aphanomyces astaci	ETV72857	CMTR_1
	ETV68077	CMTR_2
Thalassiosira oceanica	97085	CMTR
Thalassiosira pseudonana	23316	CMTR_1
	35647	CMTR_2
Phaeodactylum tricornutum	49487	CMTR_3
capsaspora owczarzaki	KJE95291	CMTR_1
	KJE88375	CMTR_2
	KJE94795	CMTR_3

Table of accession numbers for the protein sequences in the METTL3 tree of similarity.

Species name	Protein ID	Name code
Homo Sapiens	AAI26228.1	DNMT1
	NP_062826.2	METTL3
Gallus gallus	NP_001026319.1	METTL3
Xenopus tropicalis	AAI35882.1	METTL3
Danio rerio	NP_997945.1	METTL3
Drosophila melanogaster	NP_651204.1	METTL3
Apis mellifera	GB42237-RA	METTL3
Octopus vulgaris	XP_029648418.1	METTL3
Arabidopsis thaliana	OAP00159.1	MTA
***Arabidopsis thaliana* MTA blast results**		
Thecamonas trahens	KNC50938	METTL3_1
	KNC49976	METTL3_2
	KNC54976	METTL3_3
	KNC55102	METTL3_4
Guillardia theta	KJE94137	METTL3_1
Capsaspora owczarzaki	KJE94137	METTL3
Aphanomyces astaci	RHY55423	METTL3
***Homo sapiens* METTL3 blast results**		
Thecamonas trahens	KNC49976	METTL3_5
	KNC50938	METTL3_6
Ectocarpus siliculosus	CBJ28593	METTL3
Guillardia theta	KJE94137	Mettl3_2
	KJE92722	Mettl3_3
Phaeodactylum tricornutum	Phatr3_J48623.p1	METTL3

Table of accession numbers for the protein sequences in the METTL14 tree of similarity.

Species name	Protein ID	Name code
Homo Sapiens	AAI26228.1	DNMT1
	NP_066012.1	METTL14
Gallus gallus	NP_001026319.1	METTL14
Xenopus tropicalis	NP_001007920.1	METTL14
Danio rerio	NP_996954.1	METTL14
Drosophila melanogaster	NP_609205.1	METTL14
Apis Mellifera	GB56003-RA	METTL14
Octopus Vulgaris	XP_029651206.1	METTL14
Arabidopsis thaliana	Q94AI4.1	METTL14
***Arabidopsis thaliana* MTB blast results**		
Thecamonas trahens	KNC56370	METTL14_4
	KNC49829	METTL14_5
Ectocarpus siliculosus	CBN74272	METTL14_1
	CBJ27181	METTL14_2
	CBN75084	METTL14_3
	CBJ49082	METTL14_4
	CBN78712	METTL14_5
	CBJ30775	METTL14_6
Capsaspora owczarzaki	KJE94509	METTL14_6
	KJE91141	METTL14_7
	KJE91142	METTL14_8
Aphanomyces astaci	ETV81548	METTL14_3
	ETV81549	METTL14_4
Thalassiosira oceanica	EJK48470	METTL14_2
	EJK72642	METTL14_3
	EJK75108	METTL14_4
	EJK71166	METTL14_5
Thalassiosira pseudonana	EED96492	METTL14
***Homo sapiens* METTL4 blast results**		
Thecamonas trahens	KNC50938	METTL14_1
	KNC49976	METTL14_2
	KNC48060	METTL14_3
Ectocarpus Siliculosus	CBN77290	METTL14_7
	CBJ26655	METTL14_8
	CBN78849	METTL14_9
Capsaspora owczarzaki	KJE94137	METTL14_1
	KJE97149	METTL14_2
	KJE90461	METTL14_3
	KJE90462	METTL14_4

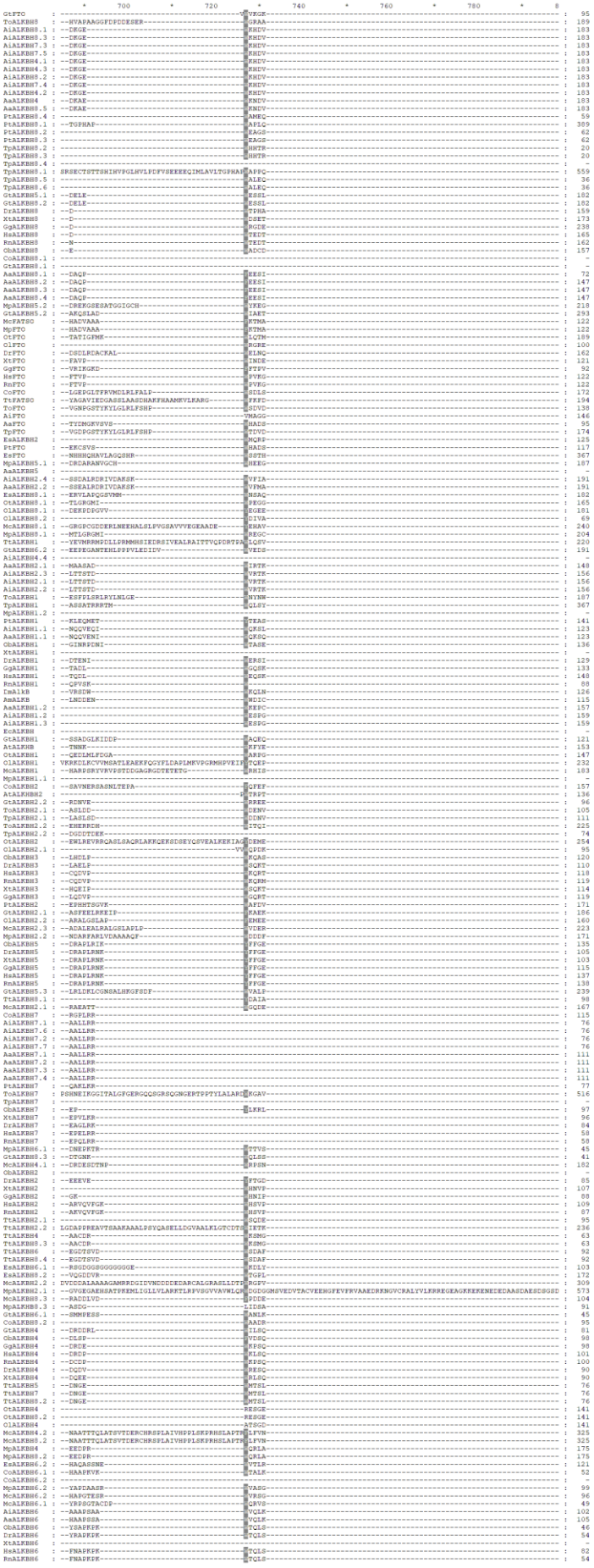
	KJE94640	METTL14_5
Aphanomyces astaci	RHY84196	METTL14_1
	RHY93064	METTL14_2
Thalassiosira oceanica	EJK74817	METTL14_1
Phaeodactylum tricornutum	Phatr3_EG01492.p1	METTL14_1
	Phatr3_J47183.p1	METTL14_2

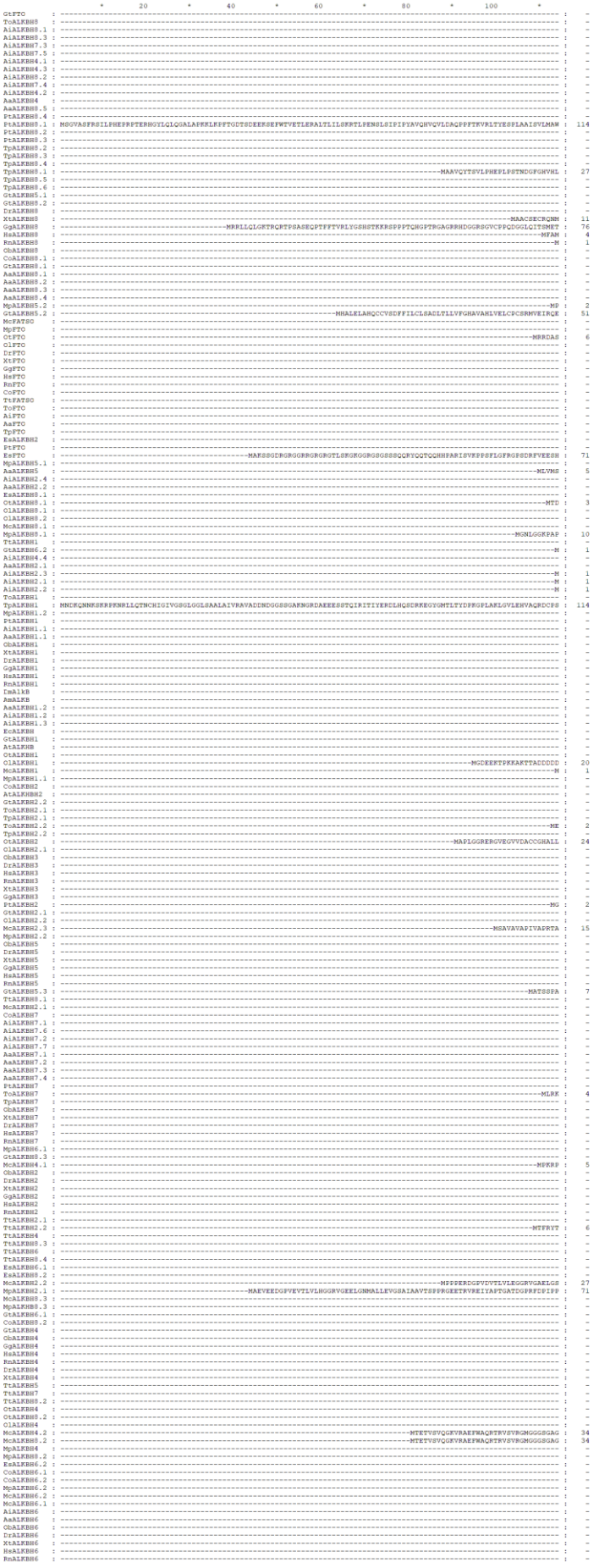
Table of accession numbers for the protein sequences in the WTAP tree of similarity.

Species name	Protein ID	Name code
Homo Sapiens	AAI26228.1	DNMT1
	NP_004897.2	WTAP
Gallus gallus	XP_025004601.1	WTAP
Xenopus tropicalis	NP_989199.1	WTAP
Danio rerio	NP_956147.1	WTAP
Drosophila melanogaster	CAB46726.1	FL(2)D-s
Apis Mellifera	GB45098-RA	WTAP
Octopus Vulgaris	XP_029648421.1	WTAP
Arabidopsis thaliana	OAP02666.1	FIP37
***Arabidopsis thaliana* FIP37 blast results**		
Thecamonas trahens	KNC48060	WTAP_3
	KNC55224	WTAP_4
	KNC54469	WTAP_5
Ectocarpus siliculosus	CBJ32722	WTAP_1
Guillardia theta	EKX43582	WTAP_1
	EKX37553	WTAP_2
	EKX52680	WTAP_3
	EKX40631	WTAP_4
	EKX48572	WTAP_5
Capsaspora owczarzaki	KJE92368	WTAP
Aphanomyces astaci	RHY90130	WTAP
Thalassiosira pseudonana	EED88922	WTAP_3
	EED91866	WTAP_4
***Homo sapiens* WTAP blast results**		
Thecamonas trahens	KNC47376	WTAP_1
	KNC47847	WTAP_2
Ectocarpus siliculosus	CBJ32722	WTAP_2
	CBN75865	WTAP_3
Thalassiosira oceanica	EJK73639	WTAP_1
	EJK63037	WTAP_2

Protein Alignments

Protein Alignment used in the FTO tree of similarity. The alignments were made using muscle and displayed using Genedoc. There are four levels of shading, primary which is 100% match between the amino acids; secondary which is 80% match between the amino acids; tertiary which is 60% match between the amino acids; and quaternary which is less than 60%.

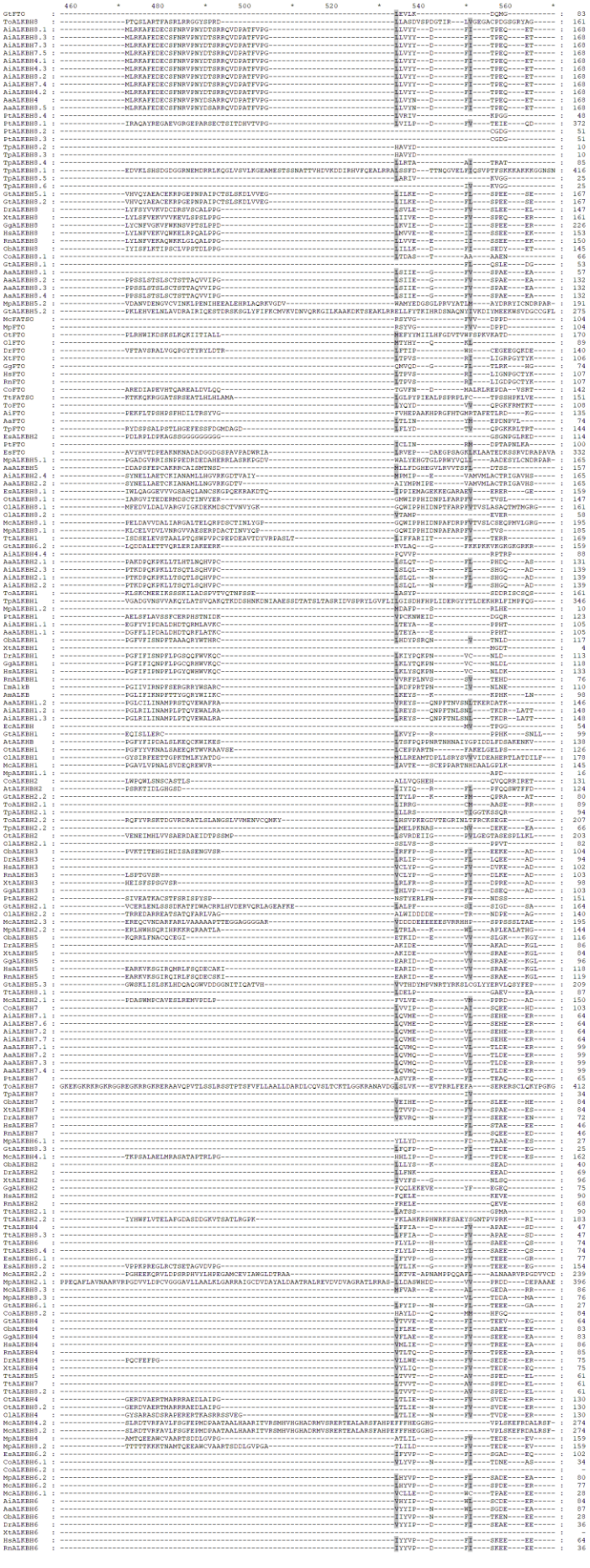


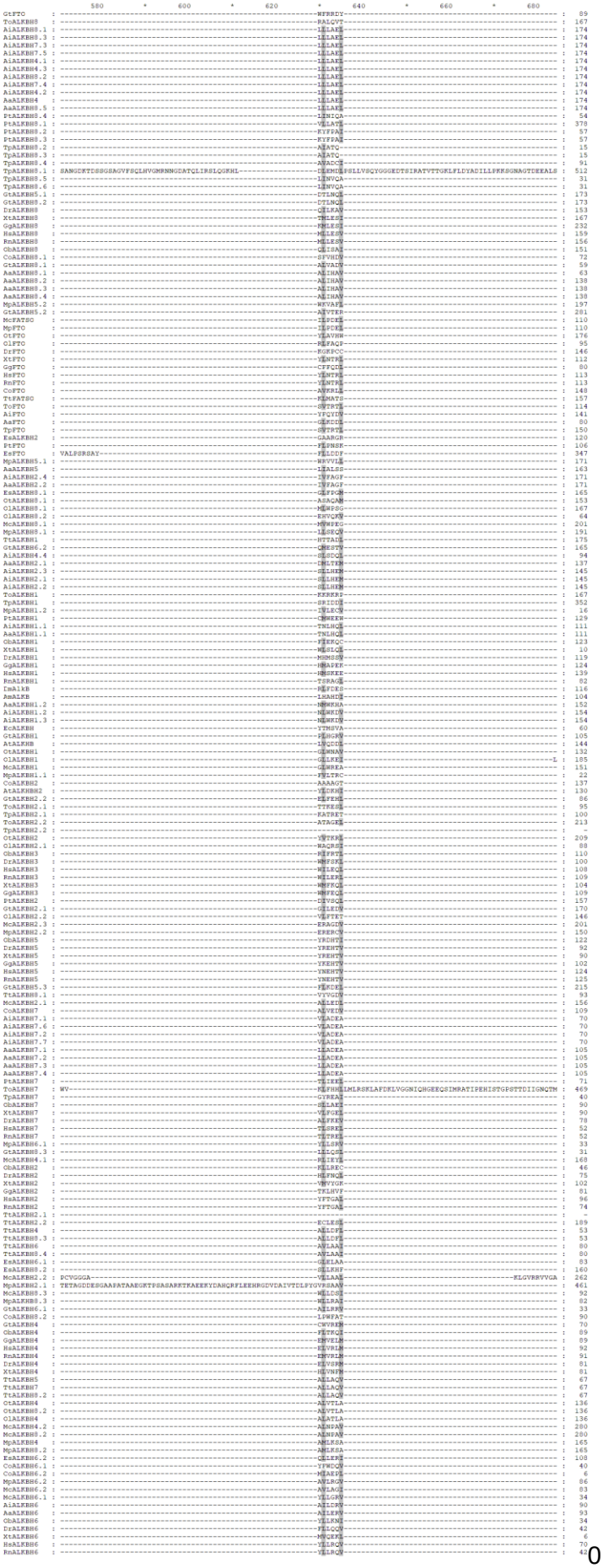


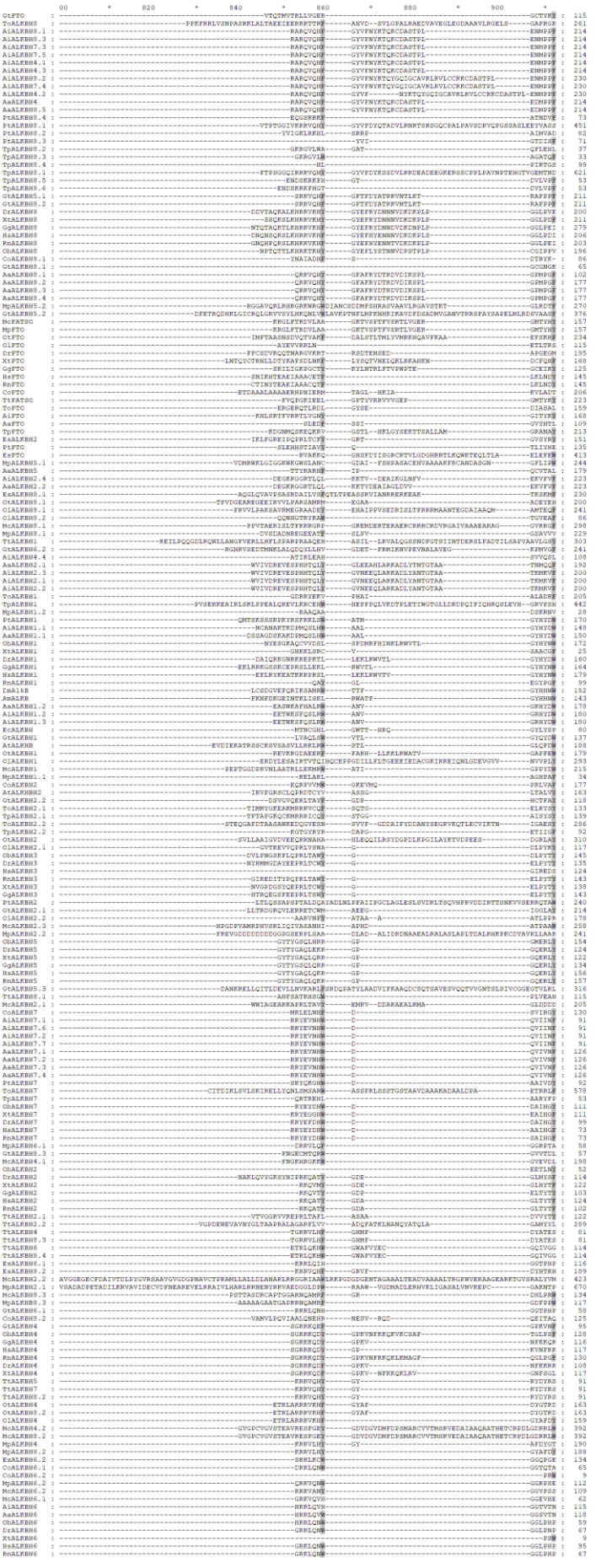


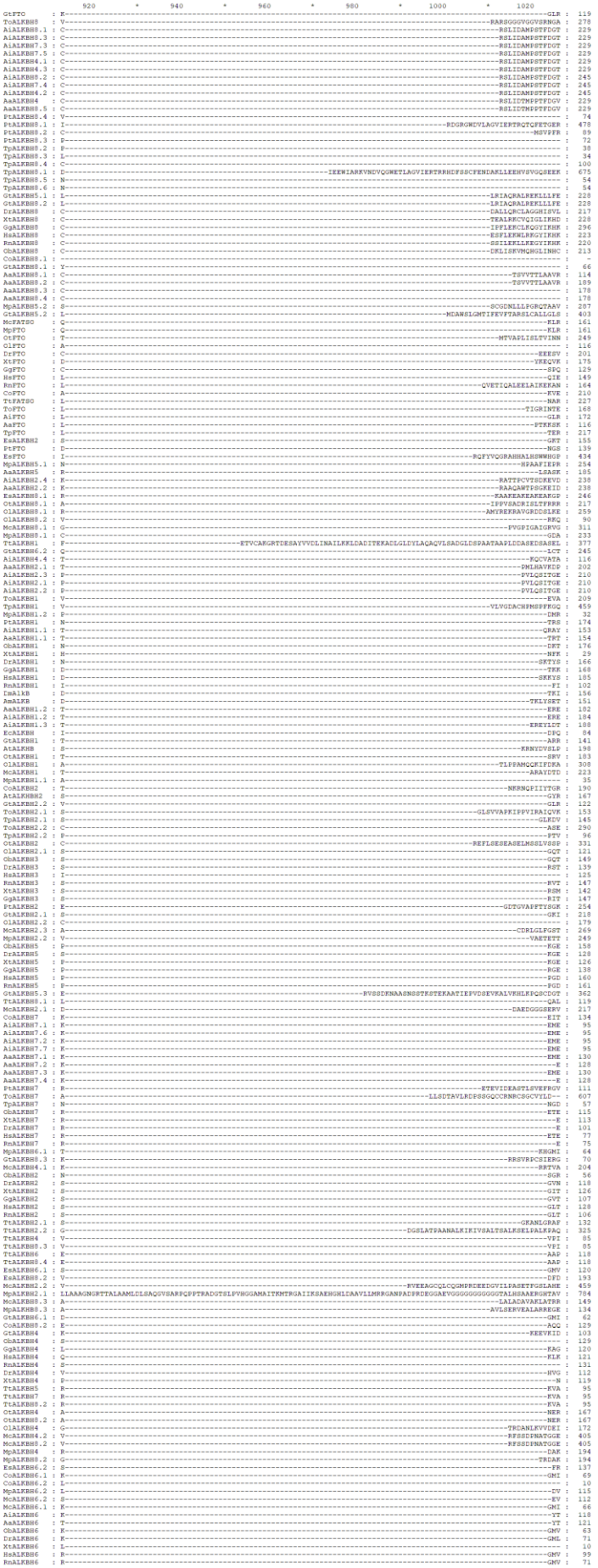


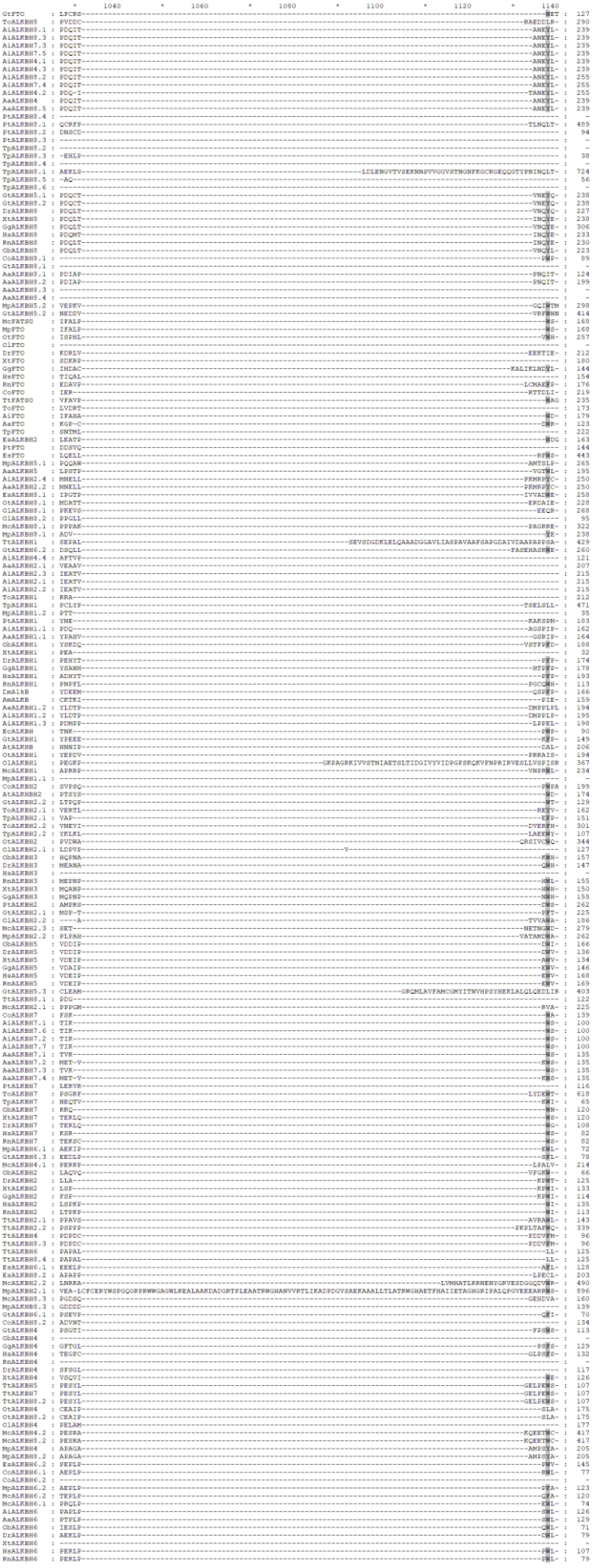


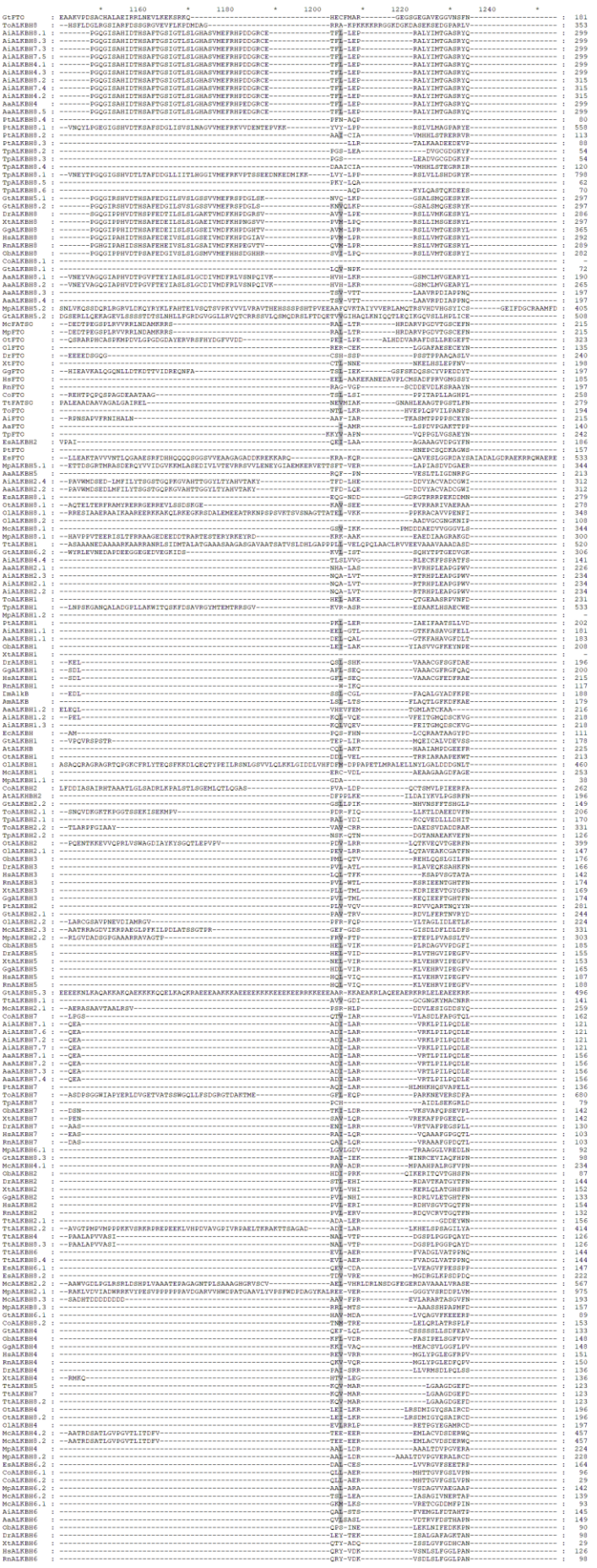




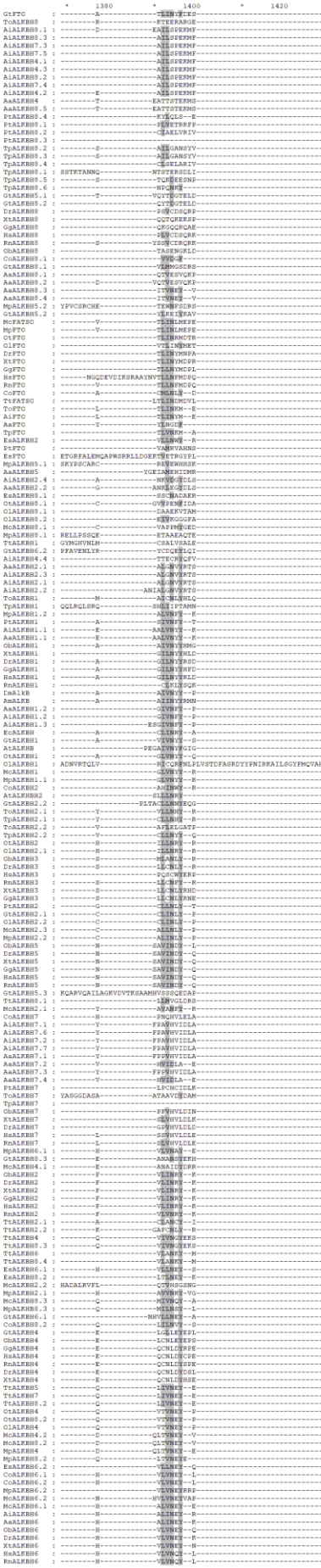


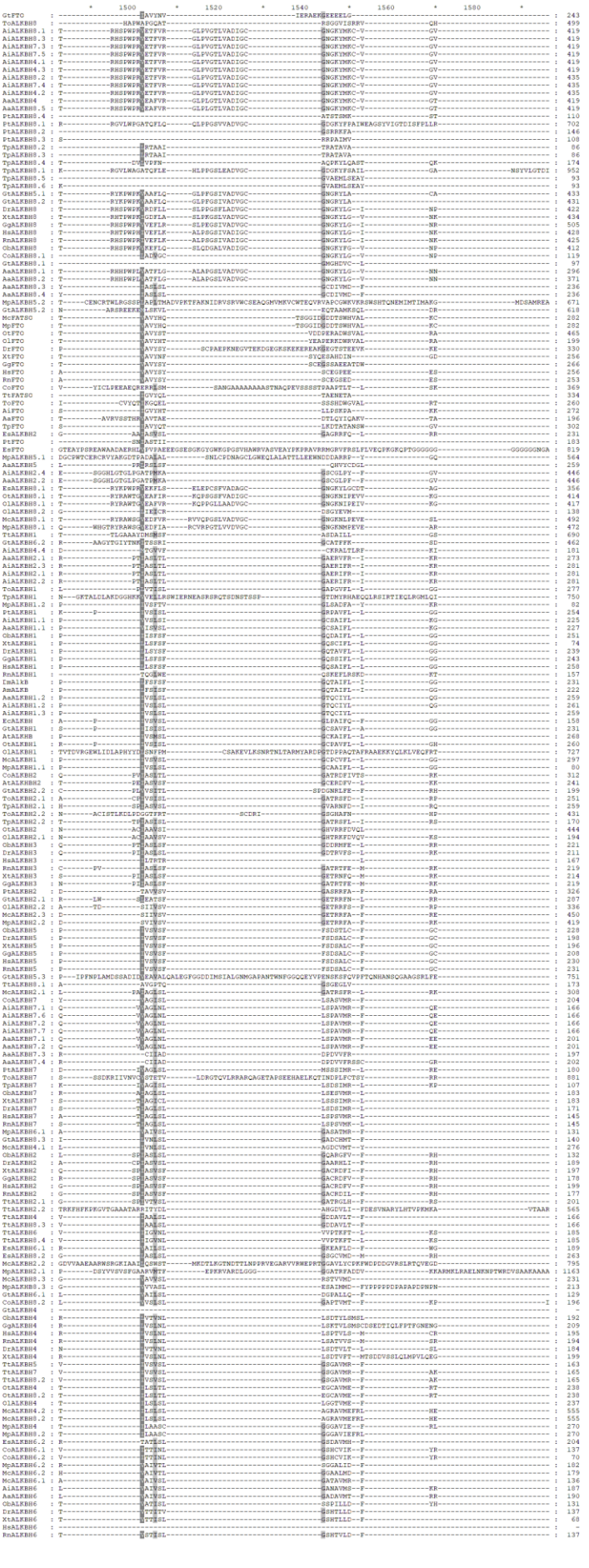


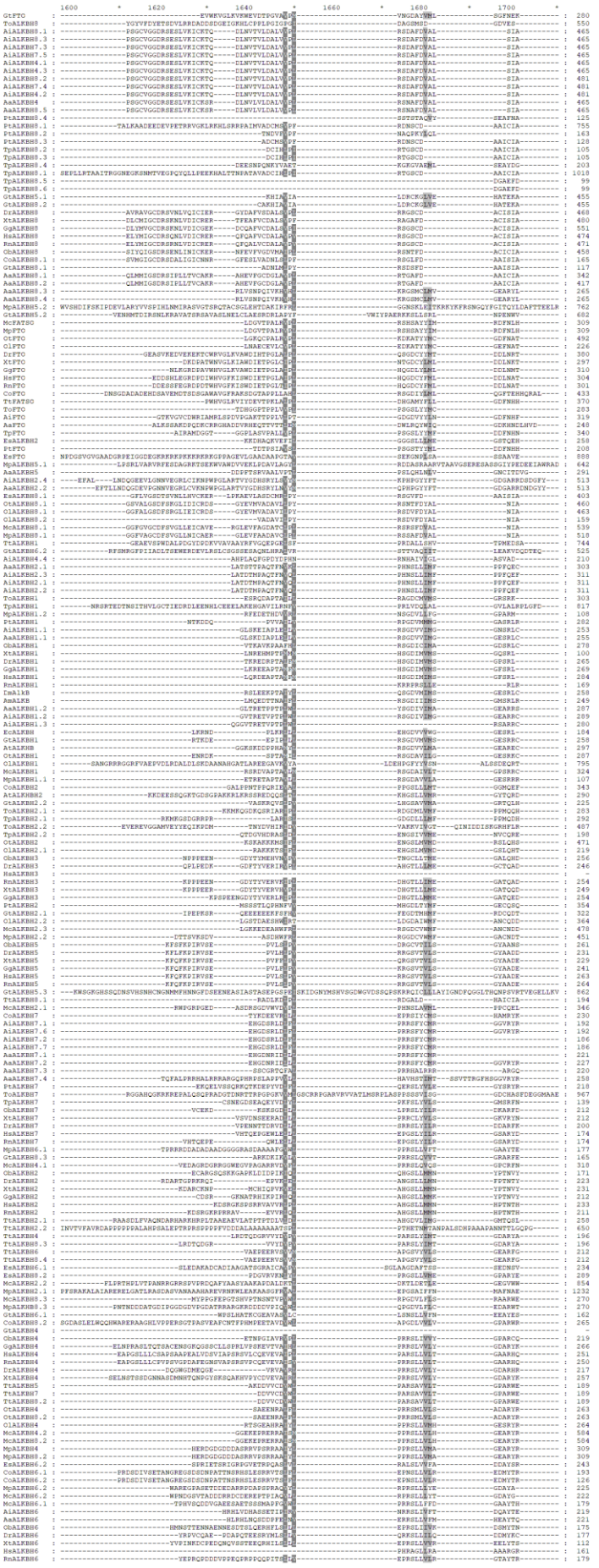


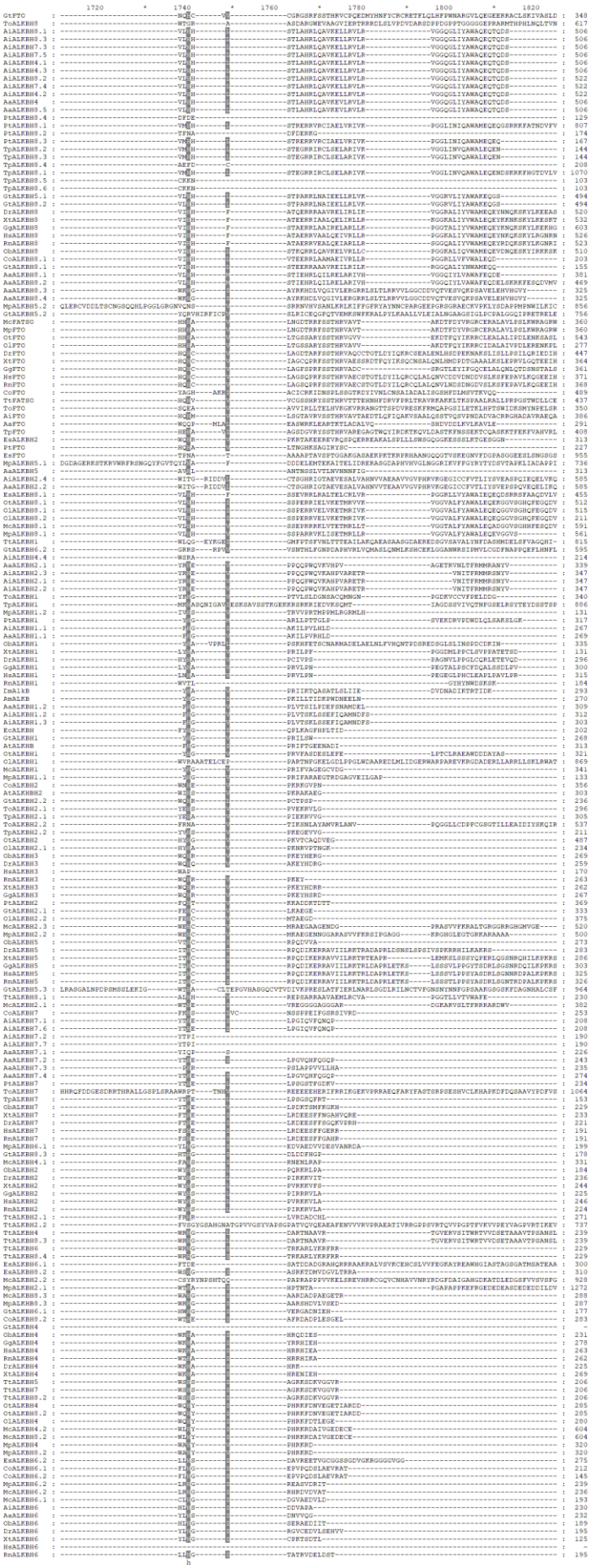


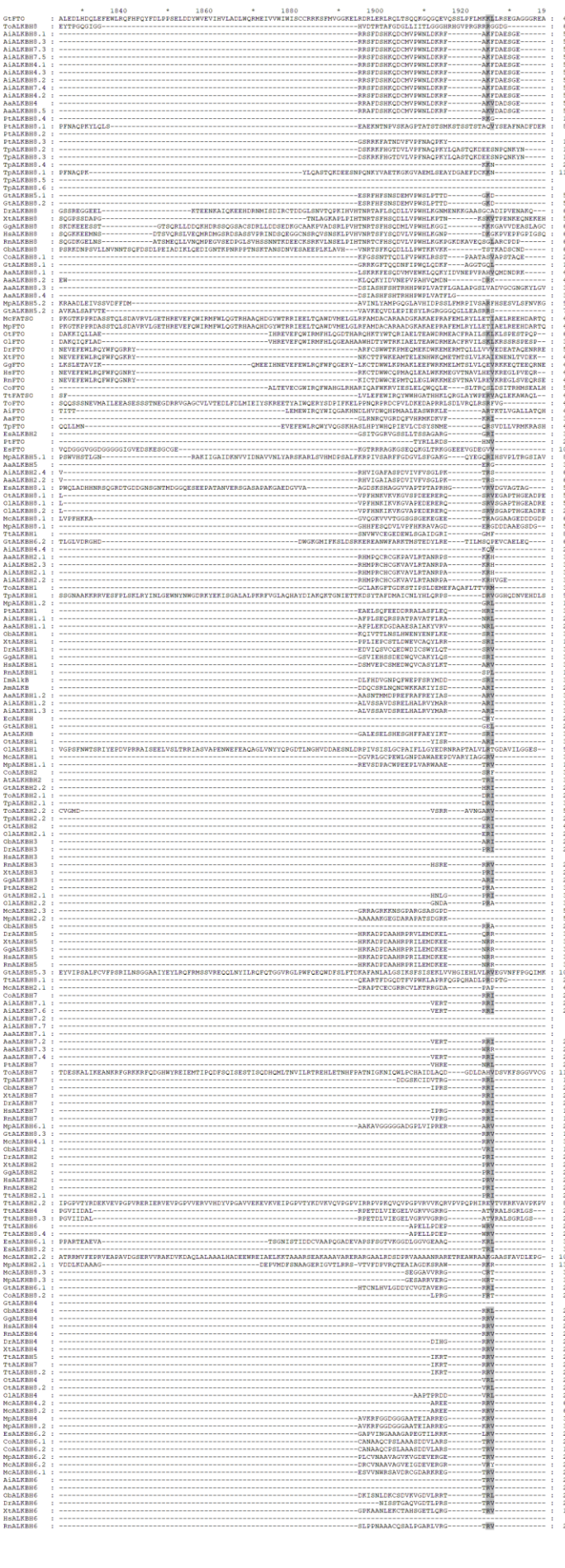


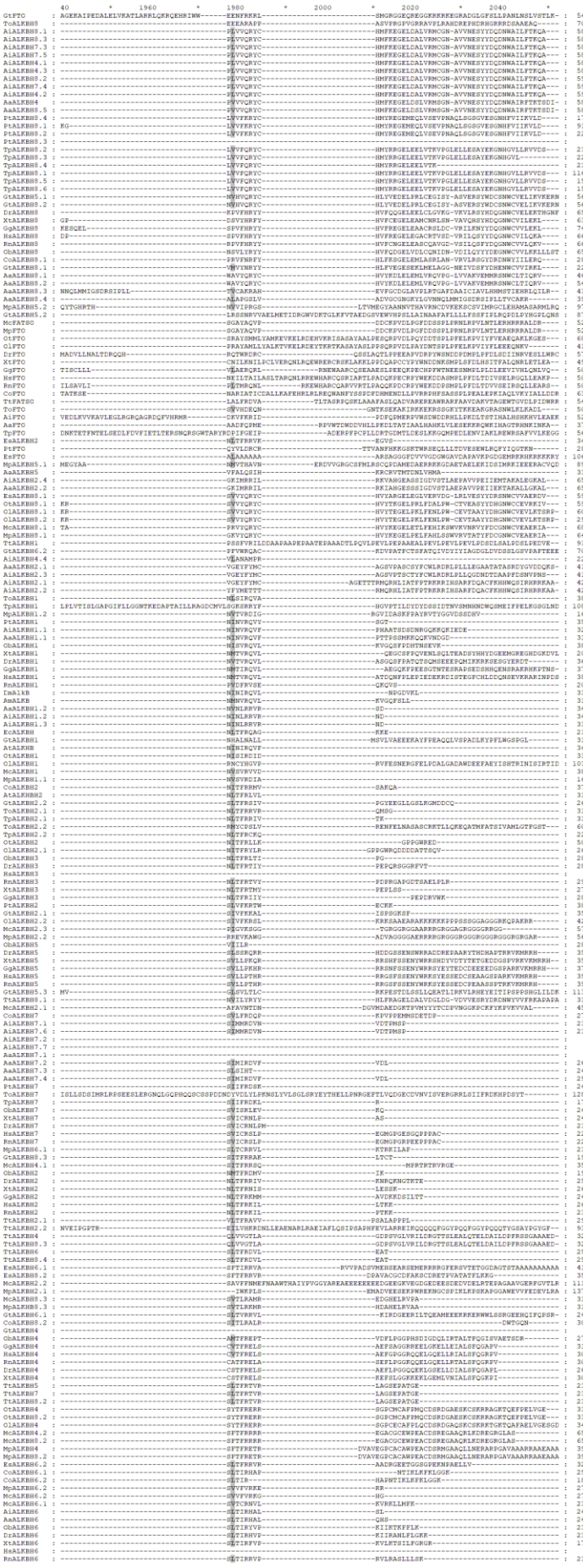


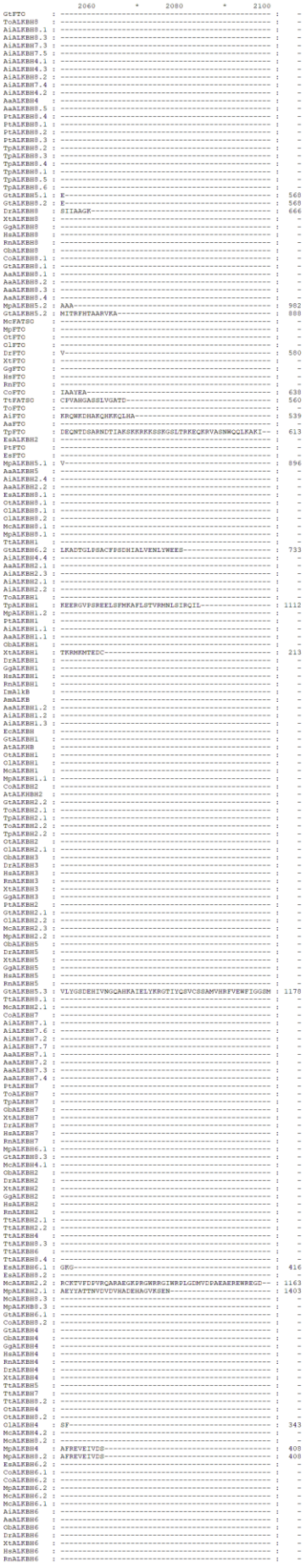




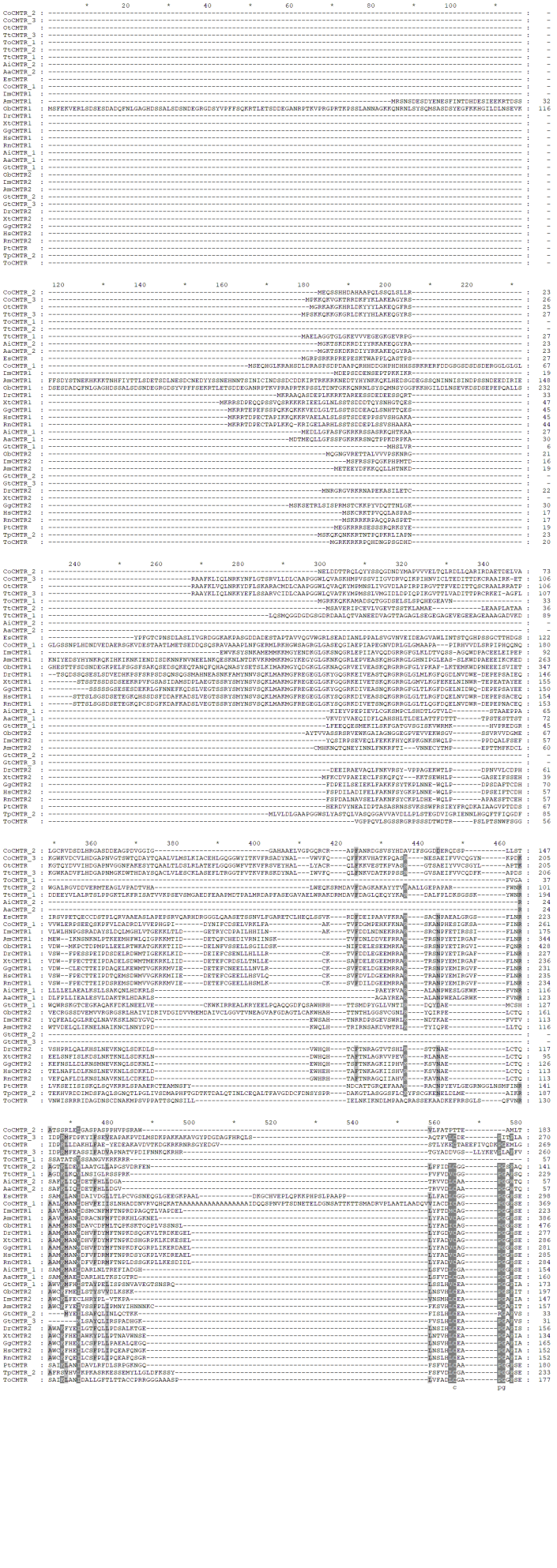




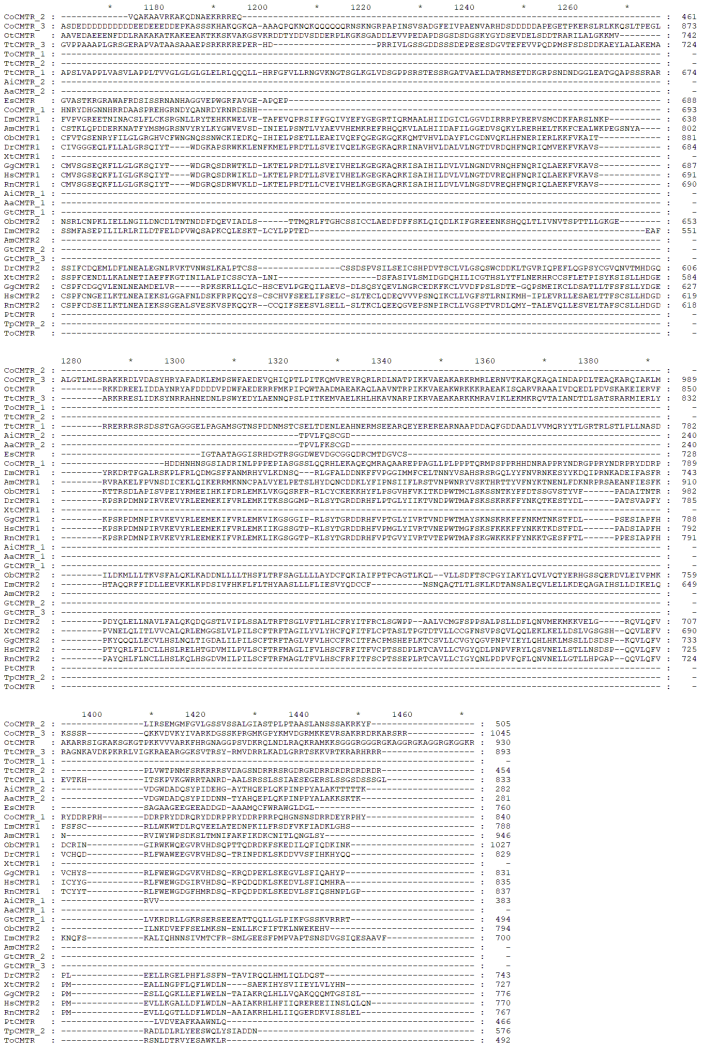



2060
2080
2100
GtFTO
TcALKBH8
AlALKBH8.1
AlALKBH8.3
AlALKBH7.3
AlALKBH7.5
AlALKBH4.1
AlALKBH4.3
AlALKBH9.2
AlALKBH7.4
AlALKBH4.2
AaALKBH4
AaALKBH8.5
PtALKBH8.4
PtALKBH9.1
PtALKBH8.2
PtALKBH8.3
TpALKBH9.2
TpALKBH8.3
TpALKBH8.4
TpALKBH9.1
TpALKBH8.5
TpALKBH8.6
GtALKBH5.1
GtALKBH8.2
DrALKBH8
XtALKBH8
GgALKBH8
HsALKBH8
RnALKBH8
ObALKBH8
CoALKBH8.1
GtALKBH8.1
AaALKBH8.1
AaALKBH8.2
AaALKBH8.3
AaALKBH8.4
MpALKBH5.2
GtALKBH5.2
McFATSO
MpFTO
OtFTO
OlFTO
DrFTO
XtFTO
GgFTO
HsFTO
RnFTO
CoFTO
TtFATSO
TcFTO
AiFTO
AaFTO
TpFTO
EsALKBH2
PtFTO
EsFTO
MpALKBH5.1
AaALKBH5
AlALKBH2.4
AaALKBH2.2
EsALKBH9.1
OtALKBH8.1
OlALKBH8.1
OlALKBH8.2
McALKBH8.1
MpALKBH8.1
TtALKBH1
GtALKBH6.2
AlALKBH4.4
AaALKBH2.1
AlALKBH2.3
AlALKBH2.1
AlALKBH2.2
TcALKBH1
TpALKBH1
MpALKBH1.2
PtALKBH1
AlALKBH1.1
AaALKBH1.1
ObALKBH1
XtALKBH1
DrALKBH1
GgALKBH1
HsALKBH1
RnALKBH1
ImALKB
AmALKB
AaALKBH1.2
AlALKBH1.2
AlALKBH1.3
EcALKBH
GtALKBH1
AtALKHB
OtALKBH1
OlALKBH1
McALKBH1
MpALKBH1.1
CoALKBH2
AtALKHBH2
GtALKBH2.2
TcALKBH2.1
TpALKBH2.1
TcALKBH2.2
TpALKBH2.2
OtALKBH2
OlALKBH2.1
ObALKBH3
DrALKBH3
HsALKBH3
RnALKBH3
XtALKBH3
GgALKBH3
PtALKBH2
GtALKBH2.1
OlALKBH2.2
McALKBH2.3
MpALKBH2.2
ObALKBH5
DrALKBH5
XtALKBH5
GgALKBH5
HsALKBH5
RnALKBH5
GtALKBH5.3
TtALKBH8.1
McALKBH2.1
CoALKBH7
AlALKBH7.1
AlALKBH7.6
AlALKBH7.2
AlALKBH7.7
AaALKBH7.1
AaALKBH7.2
AaALKBH7.3
AaALKBH7.4
PtALKBH7
TcALKBH7
TpALKBH7
ObALKBH7
XtALKBH7
DrALKBH7
HsALKBH7
RnALKBH7
MpALKBH6.1
GtALKBH8.3
McALKBH4.1
ObALKBH2
DrALKBH2
XtALKBH2
GgALKBH2
HsALKBH2
RnALKBH2
TtALKBH2.1
TtALKBH2.2
TtALKBH4
TtALKBH8.3
TtALKBH6
TtALKBH8.4
EsALKBH6.1
EsALKBH9.2
McALKBH2.2
MpALKBH2.1
McALKBH8.3
MpALKBH8.3
GtALKBH6.1
CoALKBH8.2
GtALKBH4
ObALKBH4
GgALKBH4
HsALKBH4
RnALKBH4
DrALKBH4
XtALKBH4
TtALKBH5
TtALKBH7
TtALKBH8.2
OtALKBH4
OtALKBH8.2
OlALKBH4
McALKBH4.2
McALKBH8.2
MpALKBH4
MpALKBH8.2
EsALKBH6.2
CoALKBH6.1
CoALKBH6.2
MpALKBH6.2
McALKBH6.2
McALKBH6.1
AlALKBH6
AaALKBH6
ObALKBH6
DrALKBH6
XtALKBH6
HsALKBH6
RnALKBH6

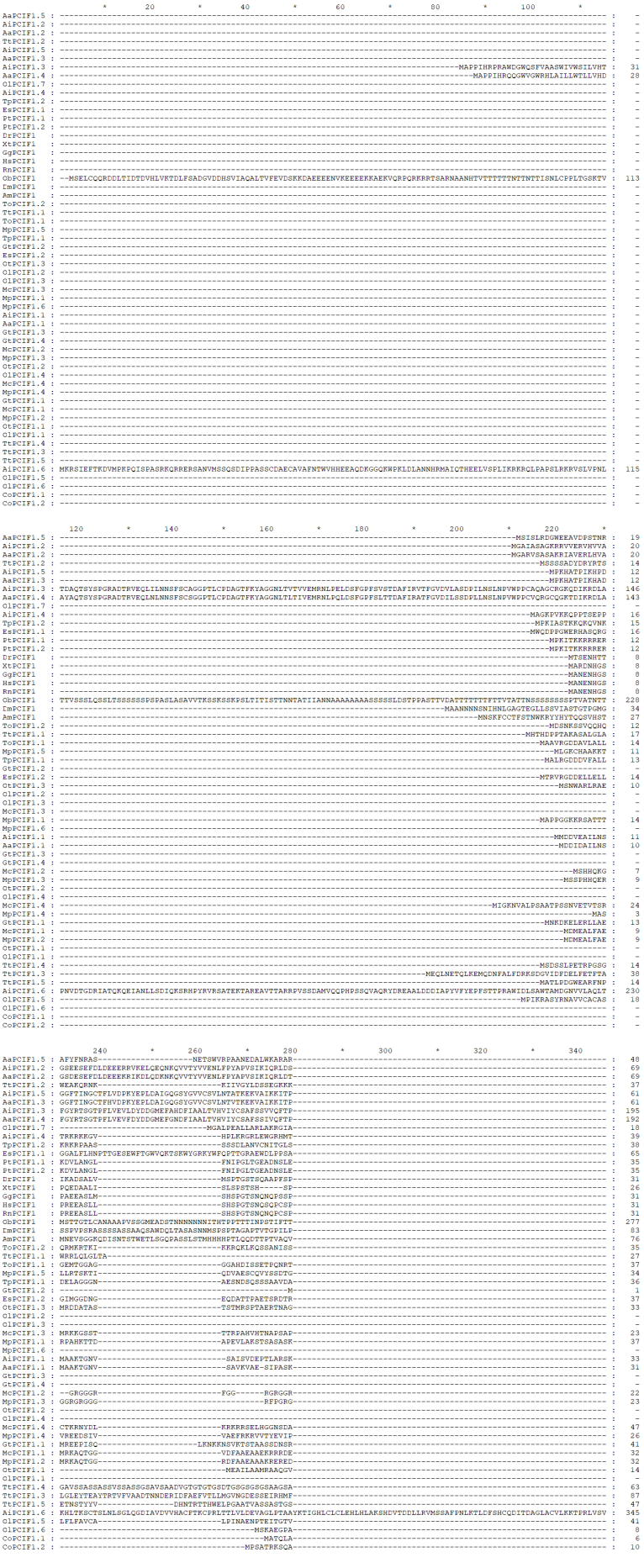
Protein Alignment used in the CMTR1 and CMTR2 tree of similarity. The alignments were made using muscle and displayed using Genedoc. There are four levels of shading, primary which is 100% match between the amino acids; secondary which is 80% match between the amino acids; tertiary which is 60% match between the amino acids; and quaternary which is less than 60%.

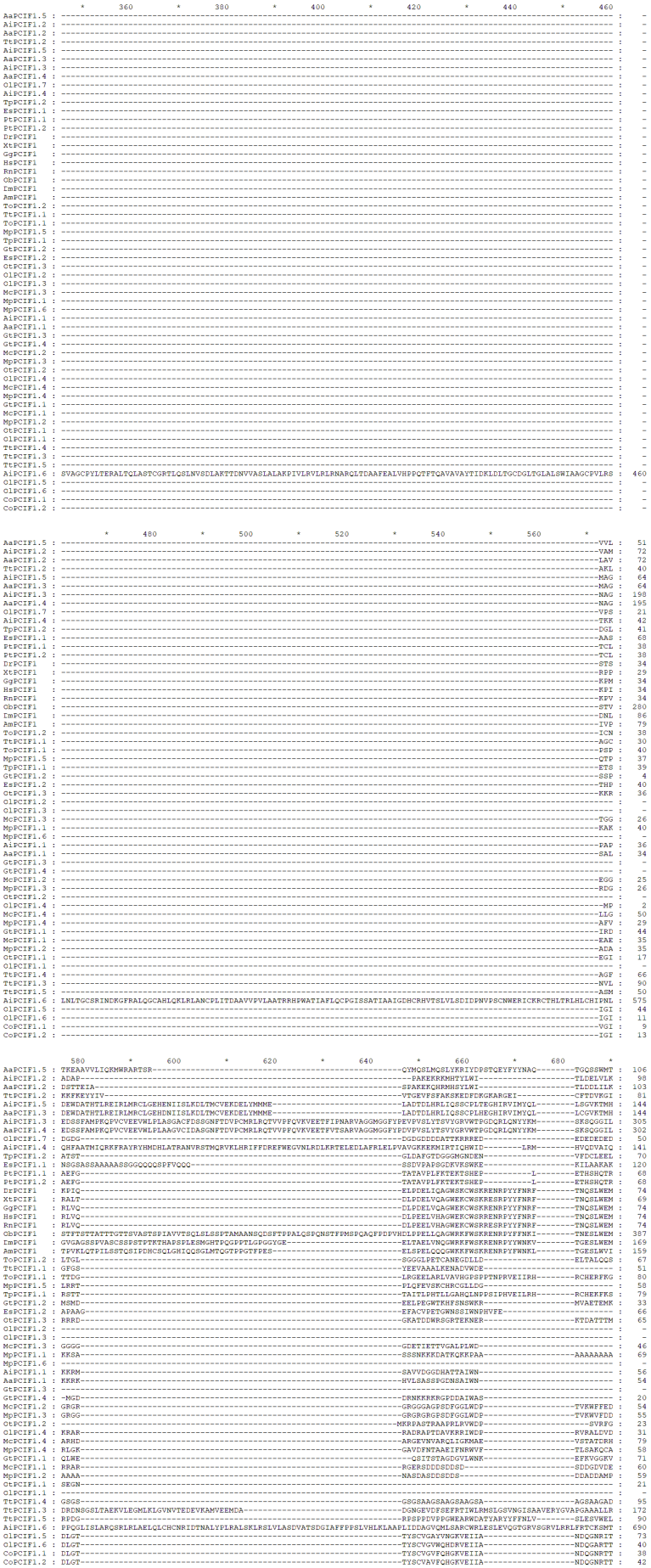


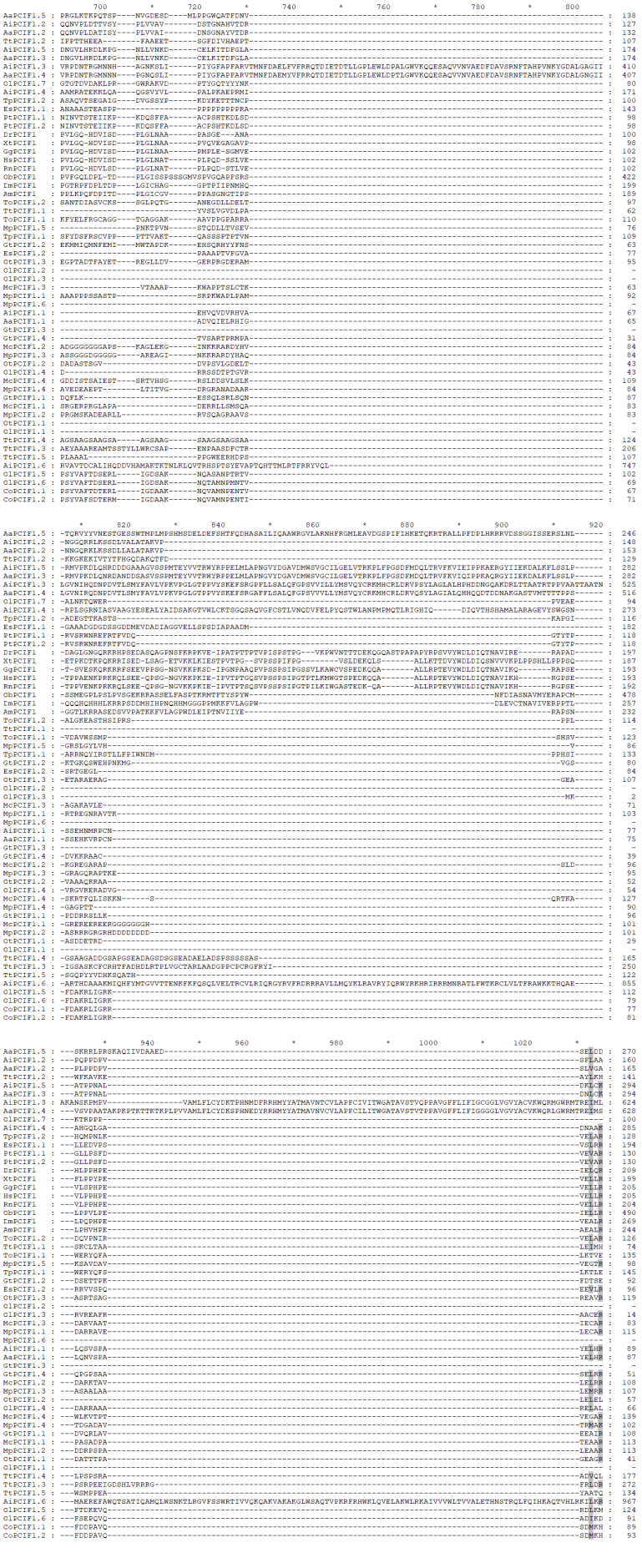


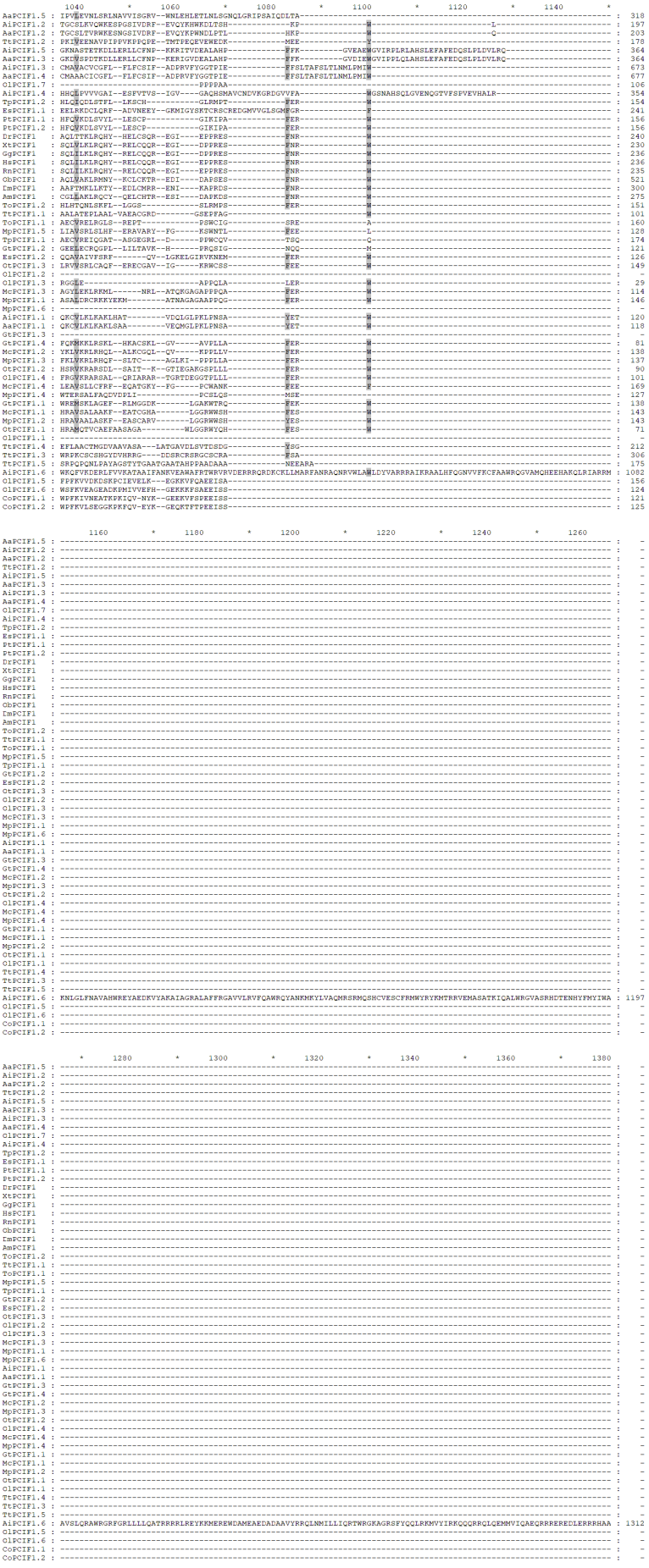


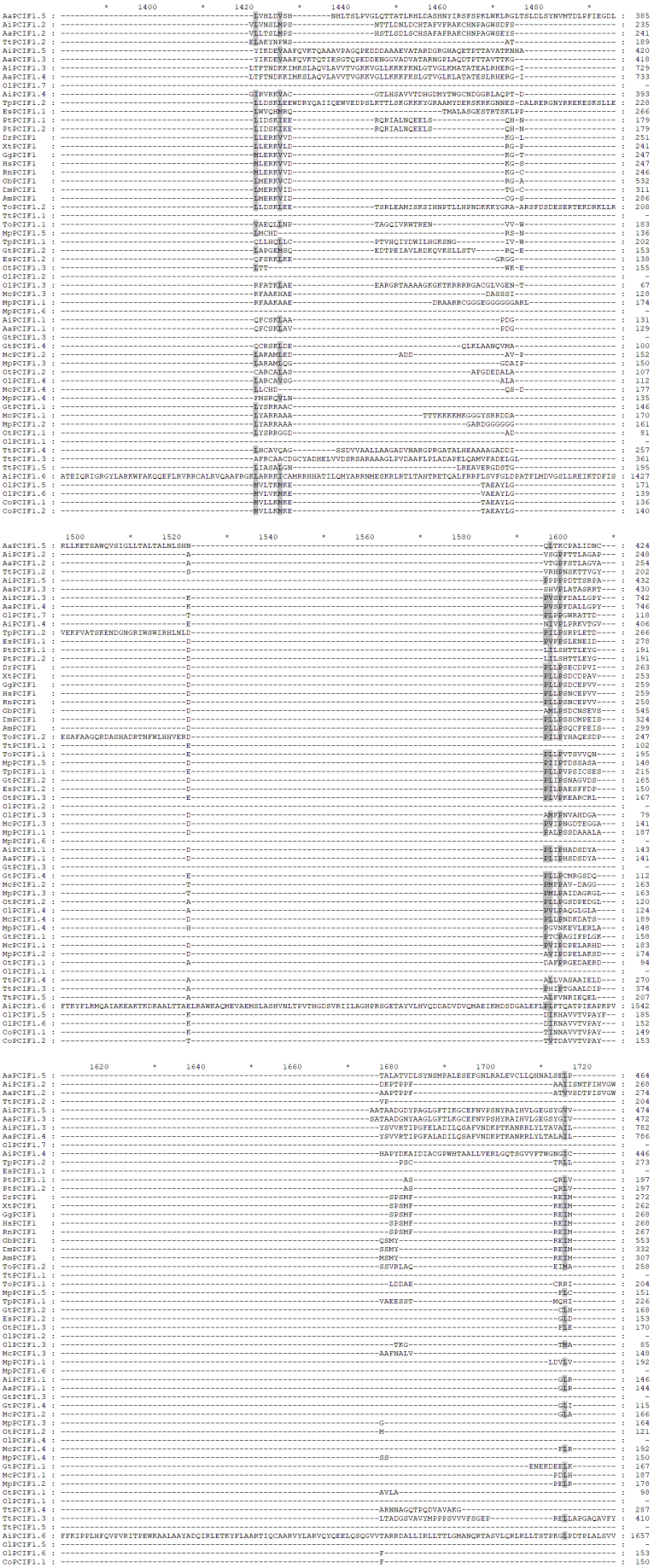
Protein Alignment used in the PCIF1 tree of similarity. The alignments were made using muscle and displayed using Genedoc. There are four levels of shading, primary which is 100% match between the amino acids; secondary which is 80% match between the amino acids; tertiary which is 60% match between the amino acids; and quaternary which is less than 60%.

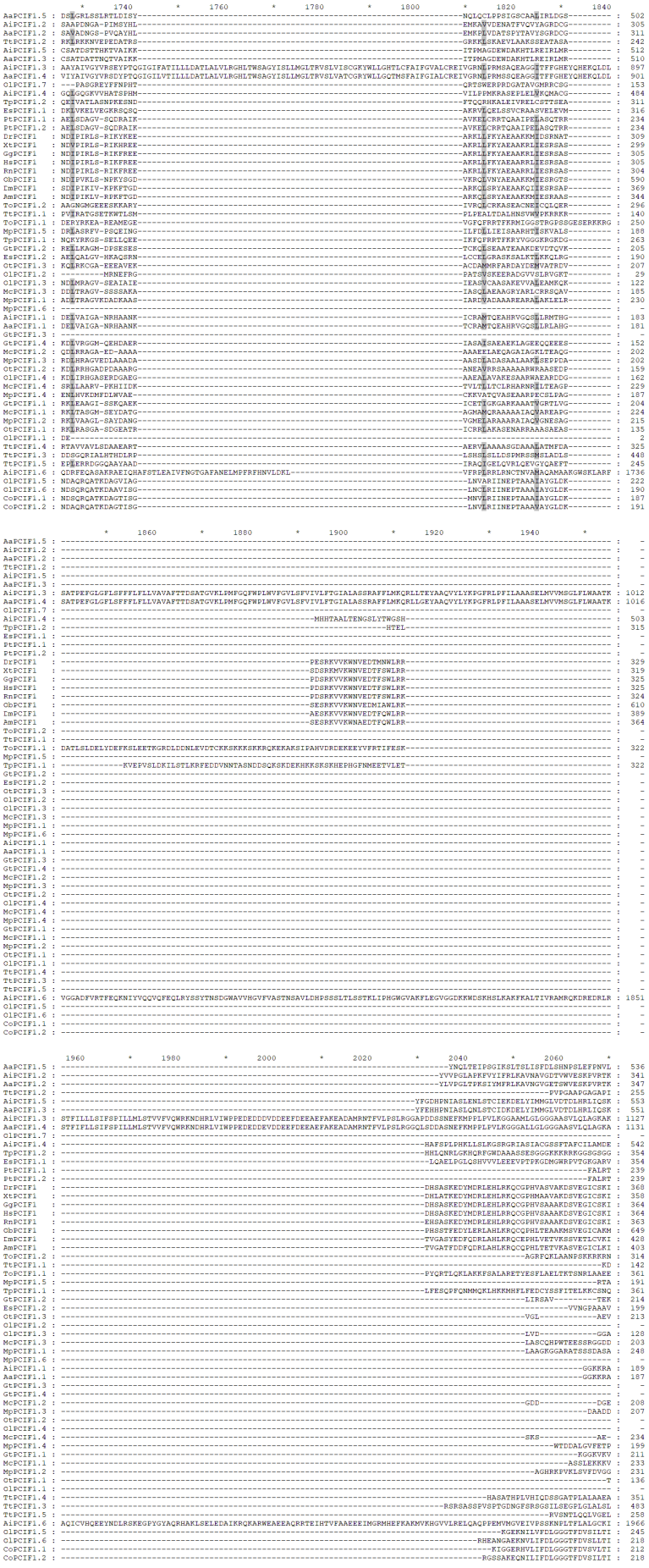


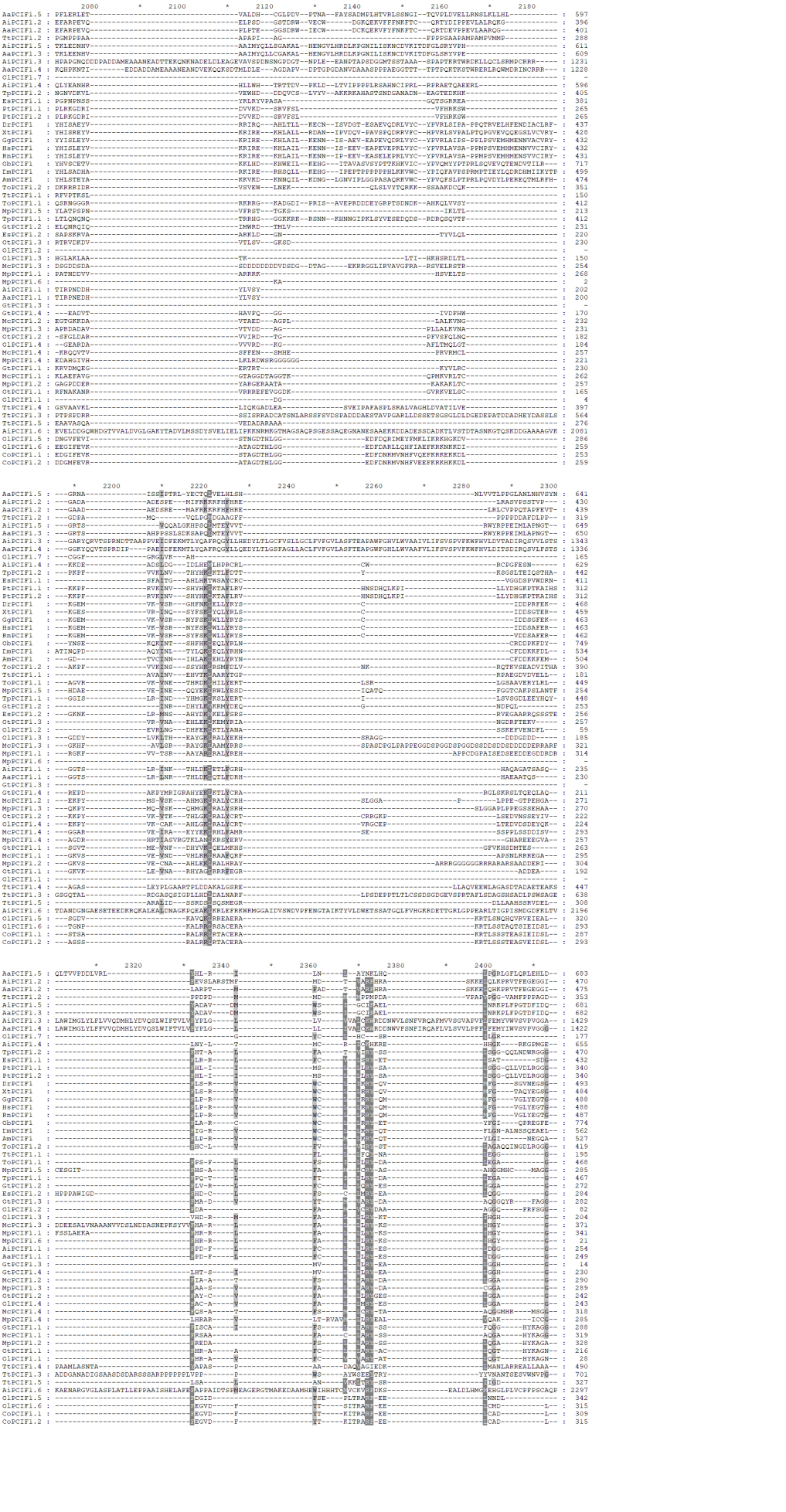


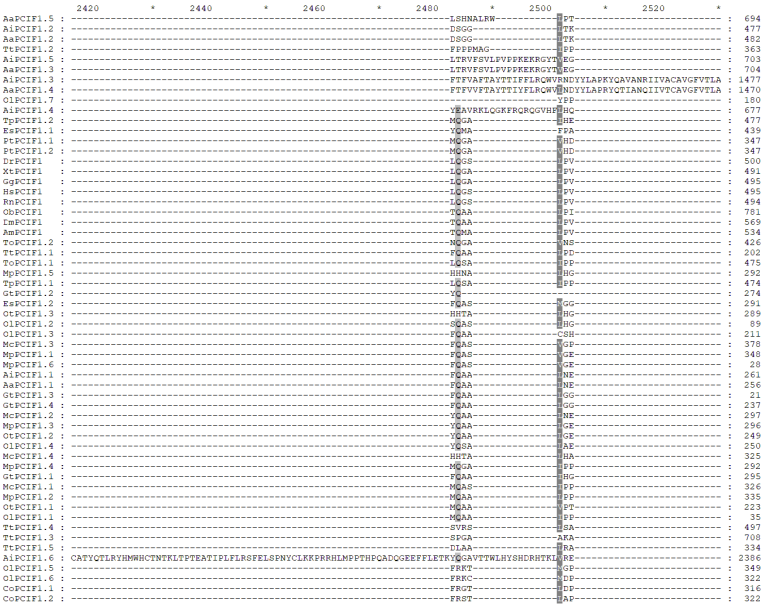


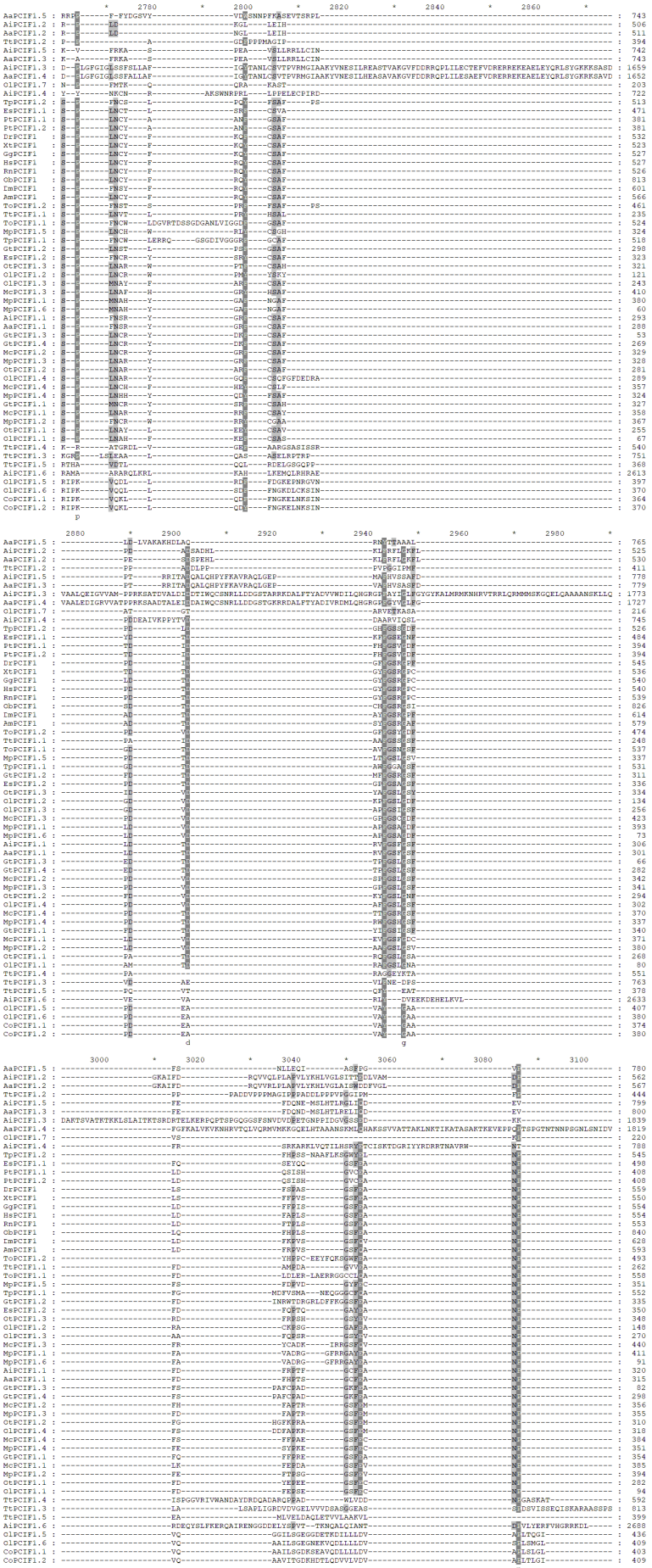


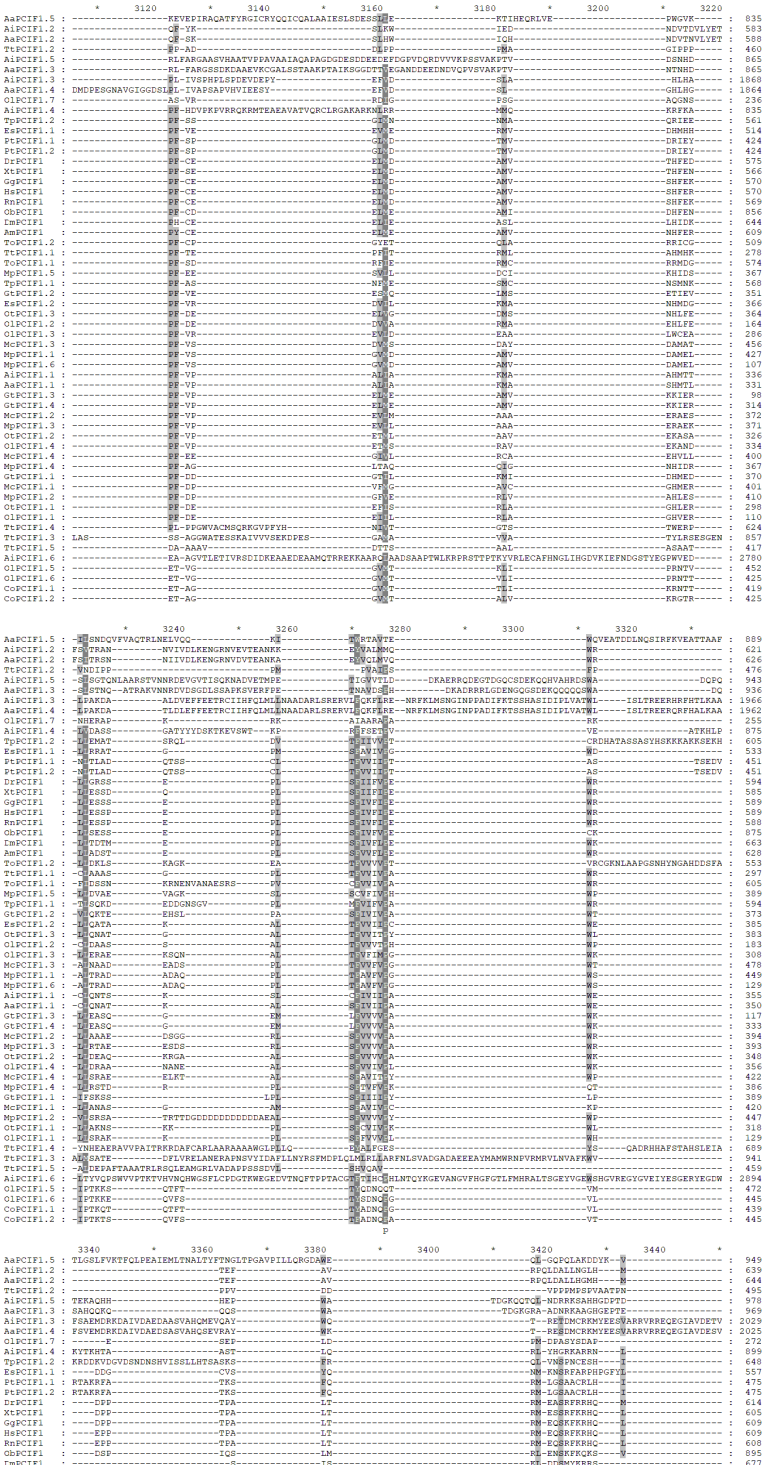


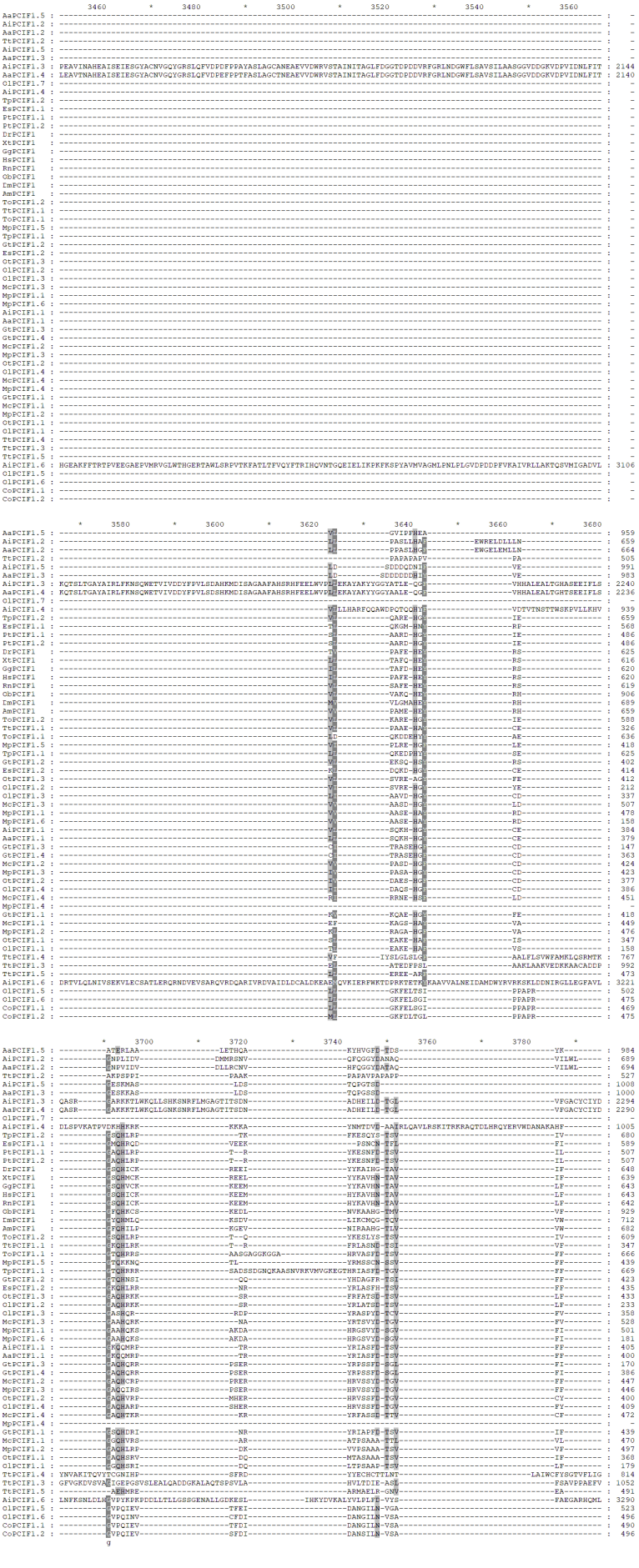


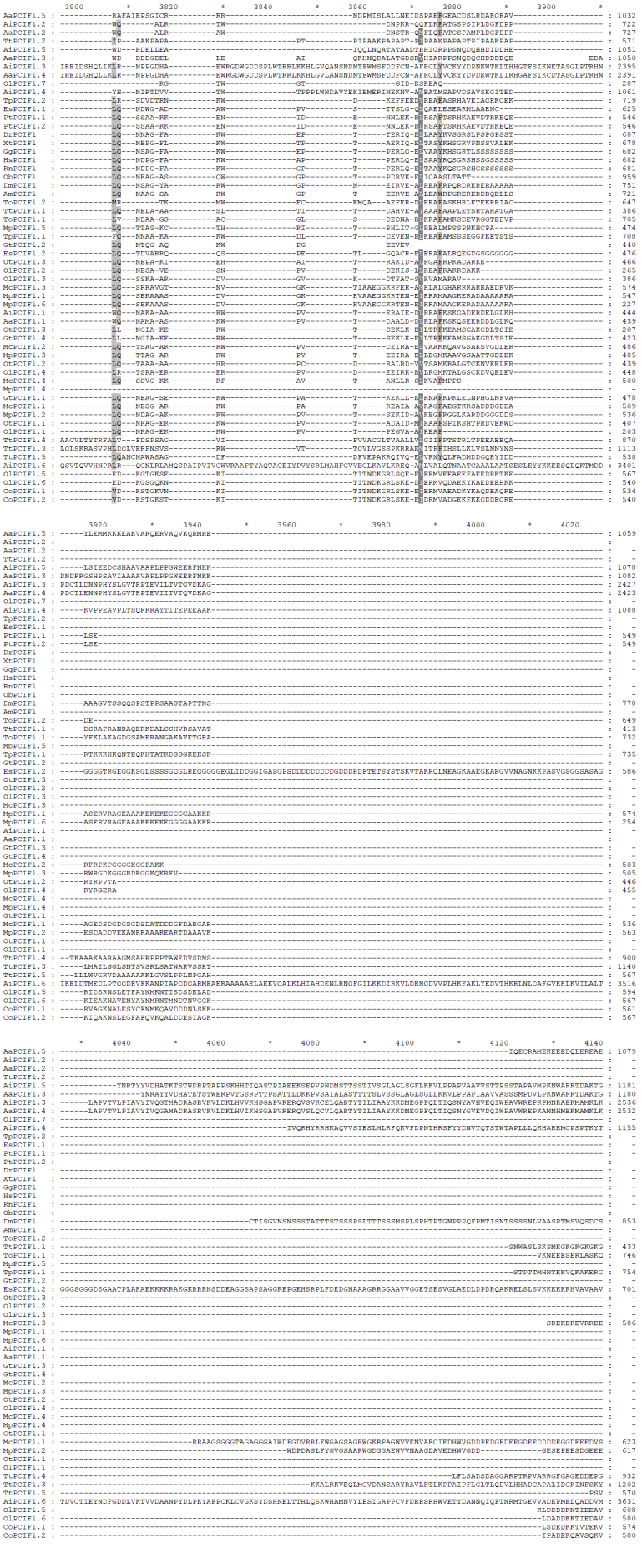






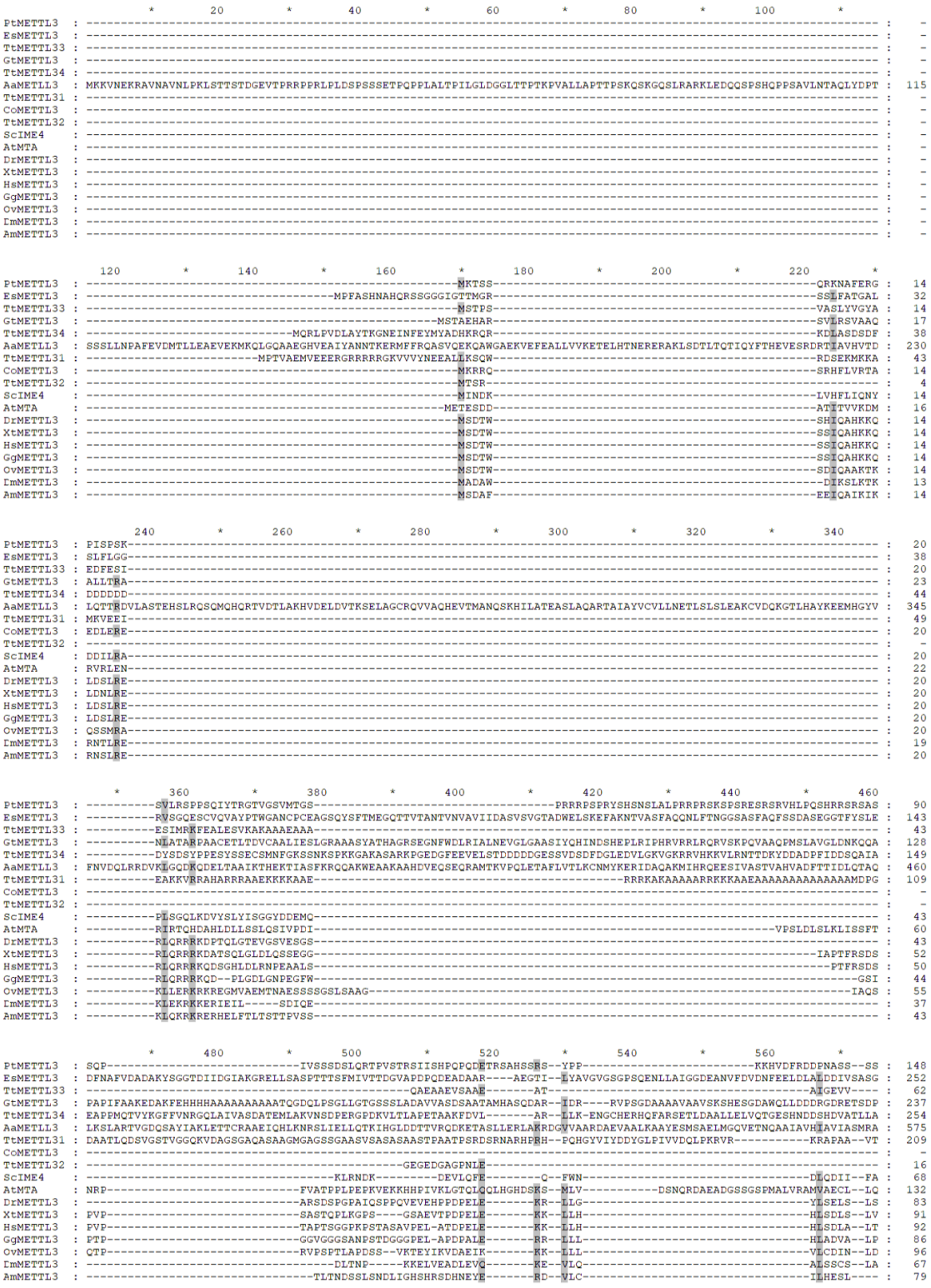


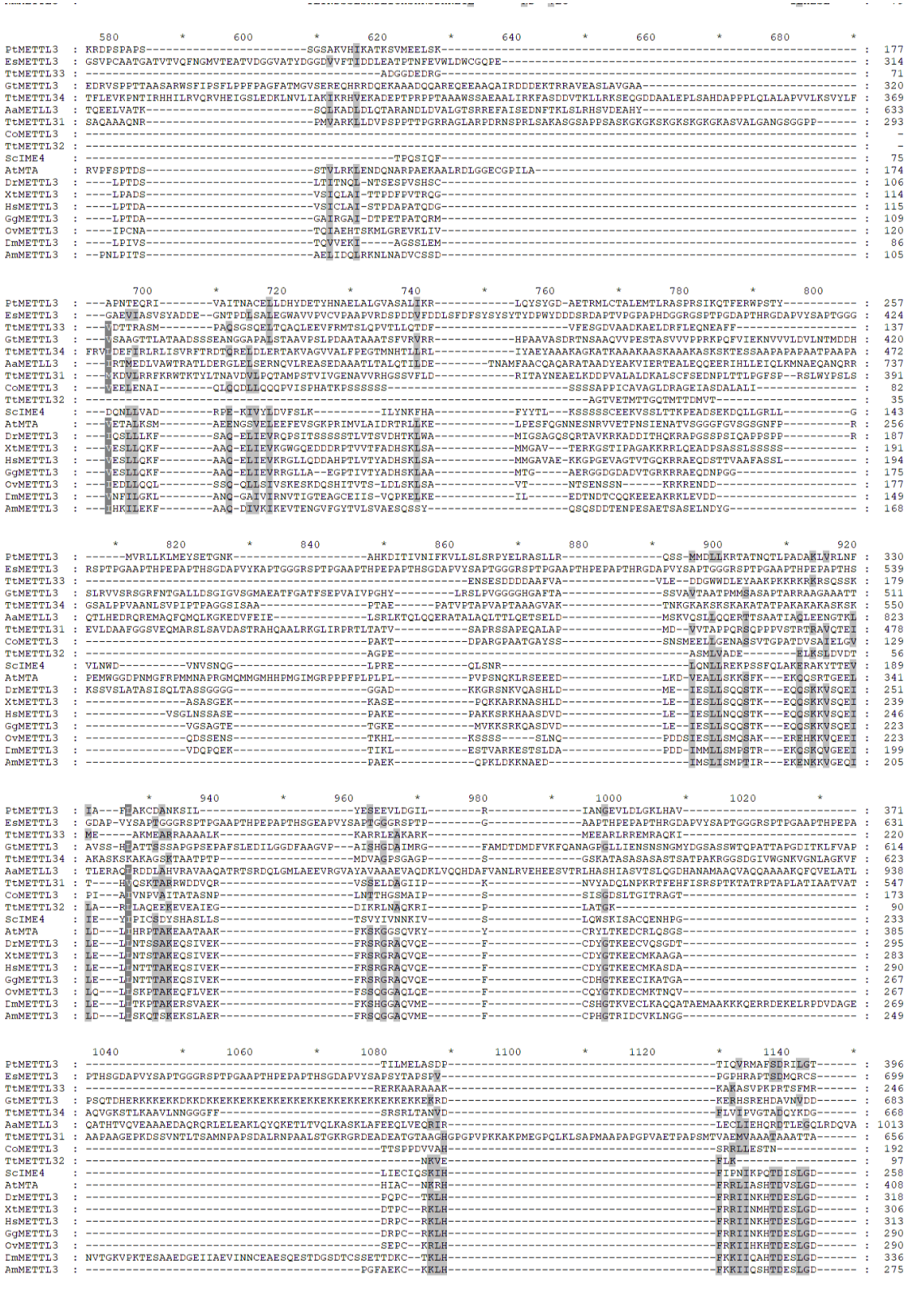


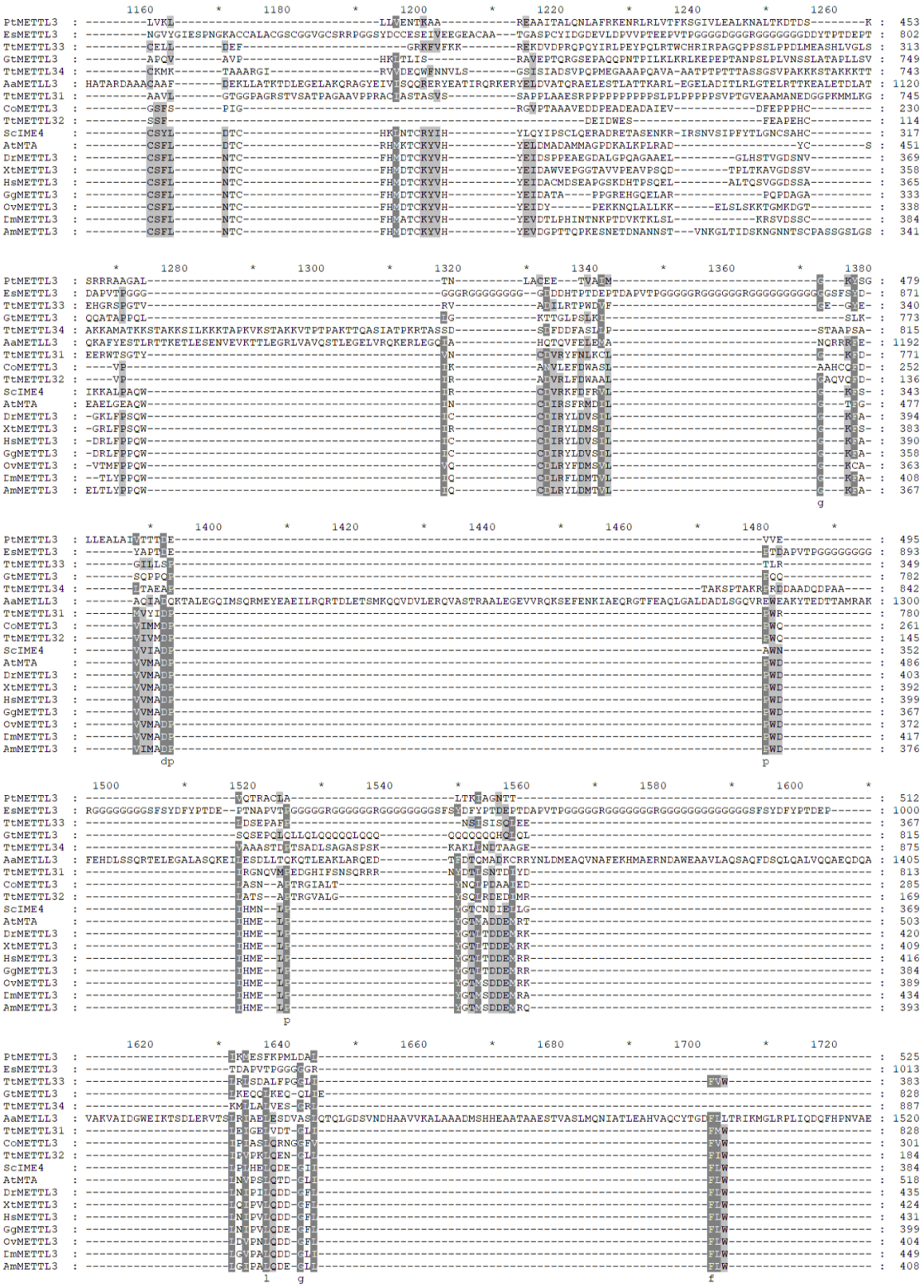


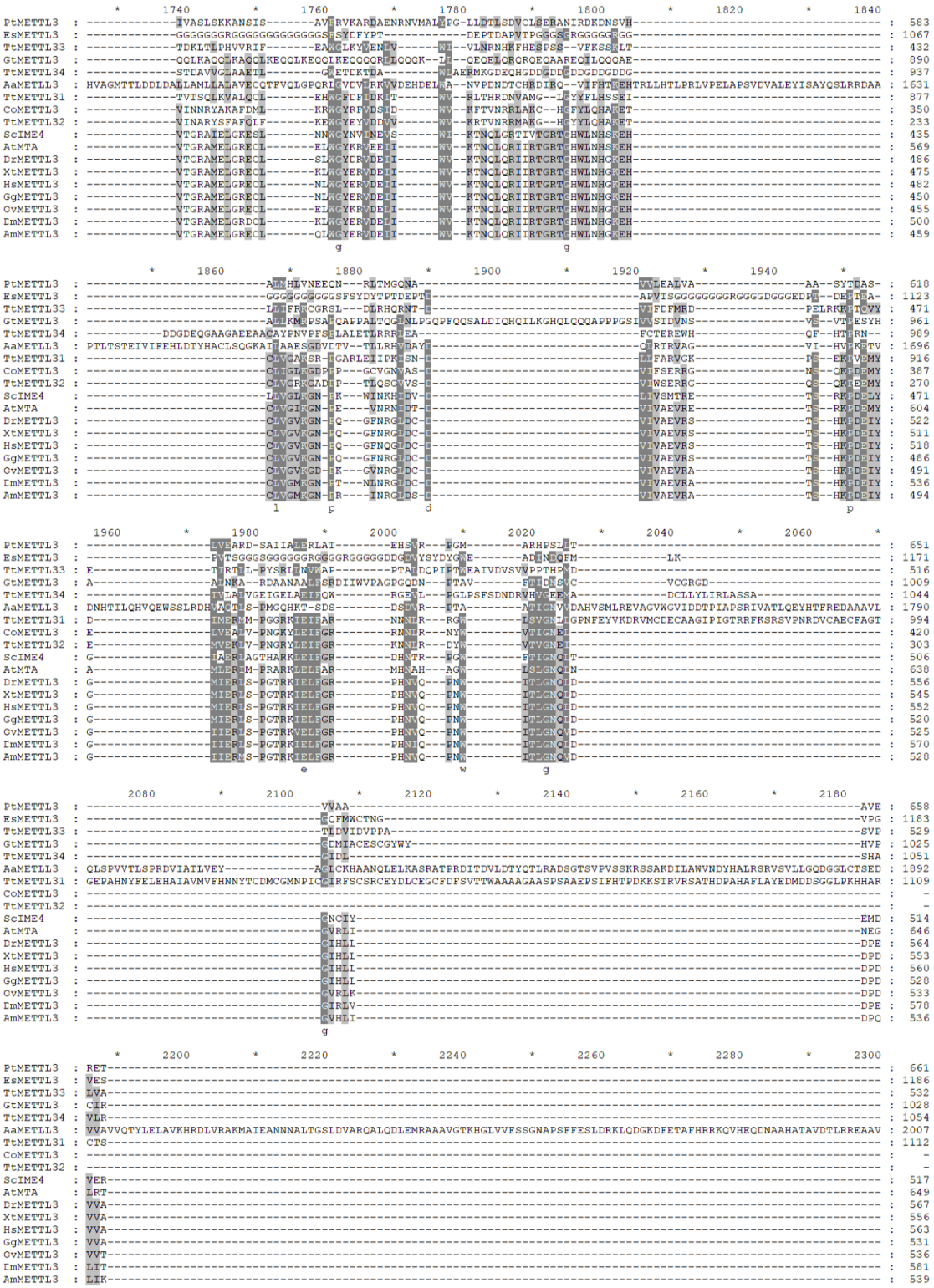


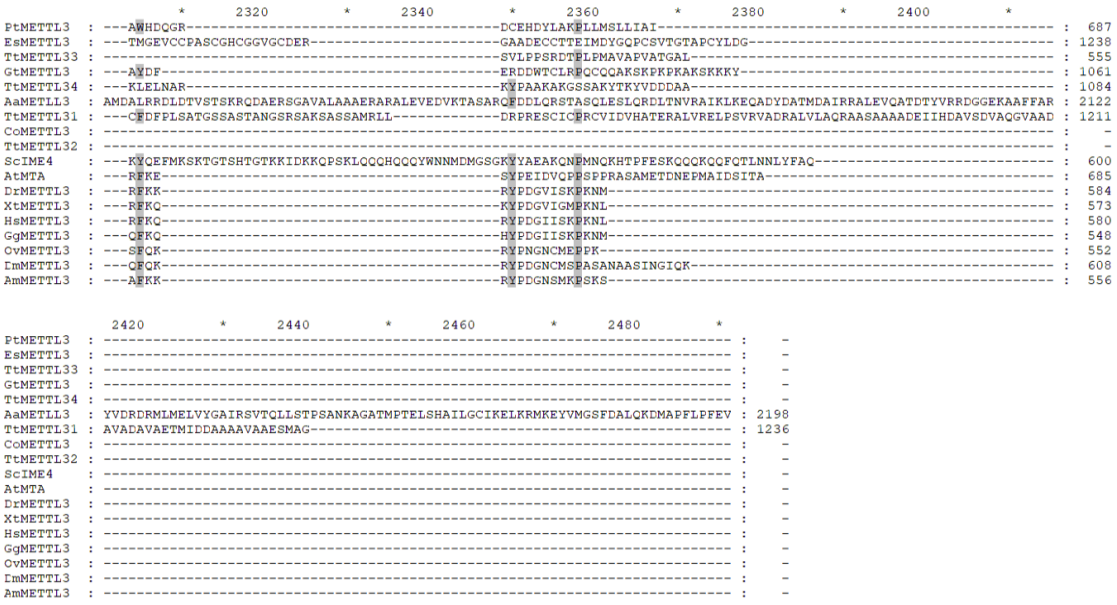
Protein Alignment used in the METTL3 tree of similarity. The alignments were made using muscle and displayed using Genedoc. There are four levels of shading, primary which is 100% match between the amino acids; secondary which is 80% match between the amino acids; tertiary which is 60% match between the amino acids; and quaternary which is less than 60%.



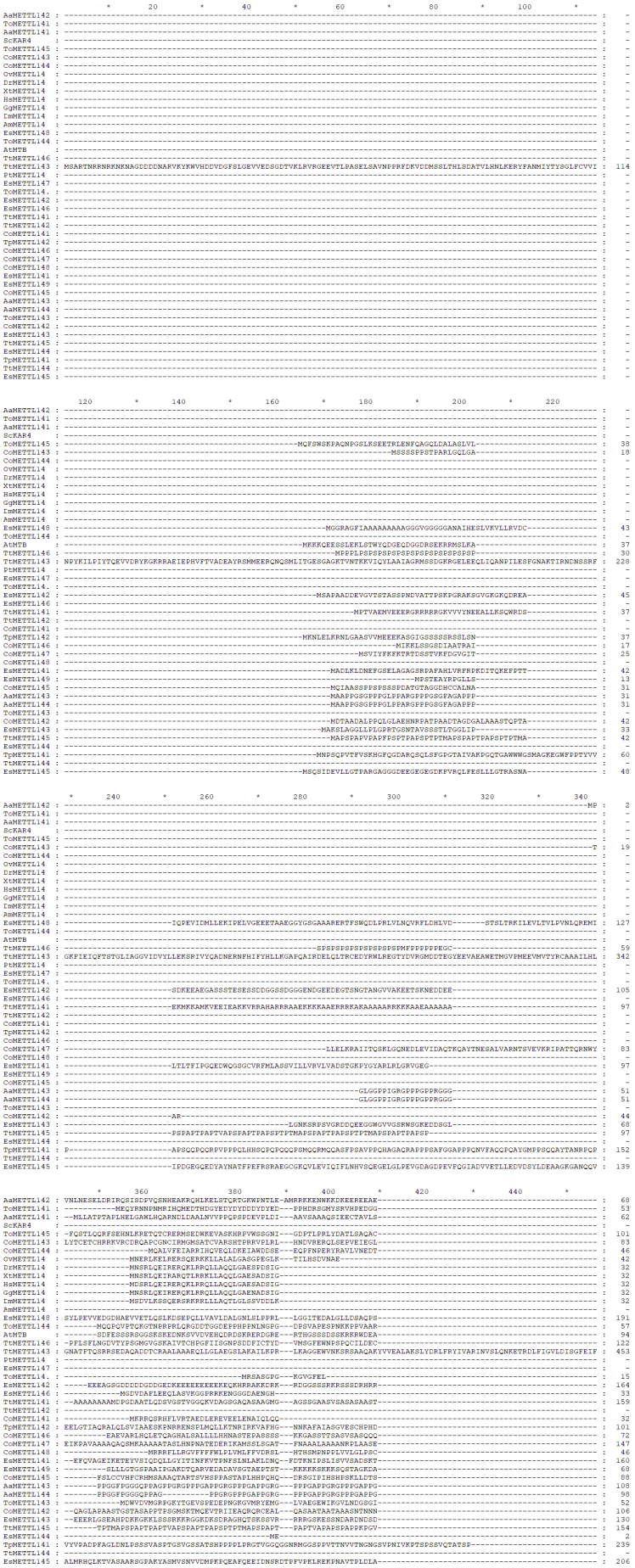


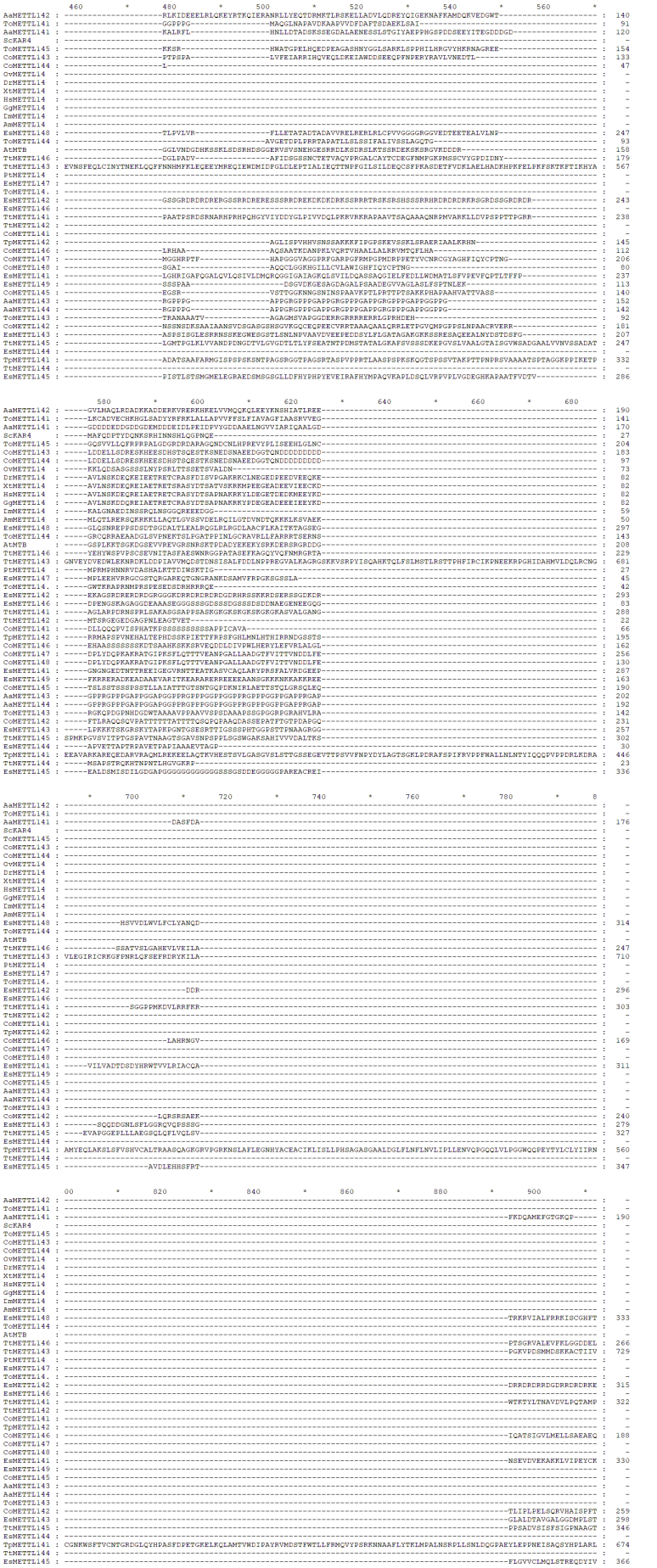


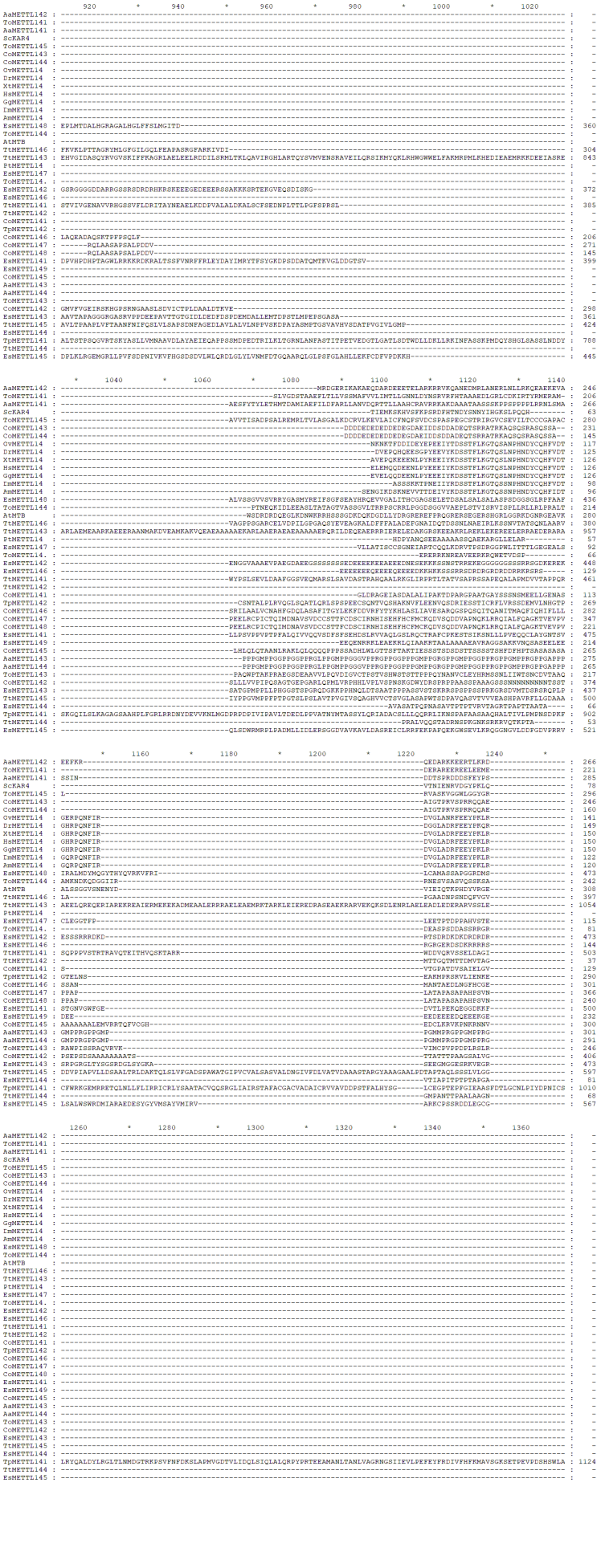

PtMETTL3
EsMETTL3
TtMETTL33
GtMETTL3
TtMETTL34
AaMETTL3
TtMETTL31
CoMETTL3
TtMETTL32
ScIME4
AtMTA
DrMETTL3
XtMETTL3
HsMETTL3
GgMETTL3
OvMETTL3
DmMETTL3
AmMETTL3

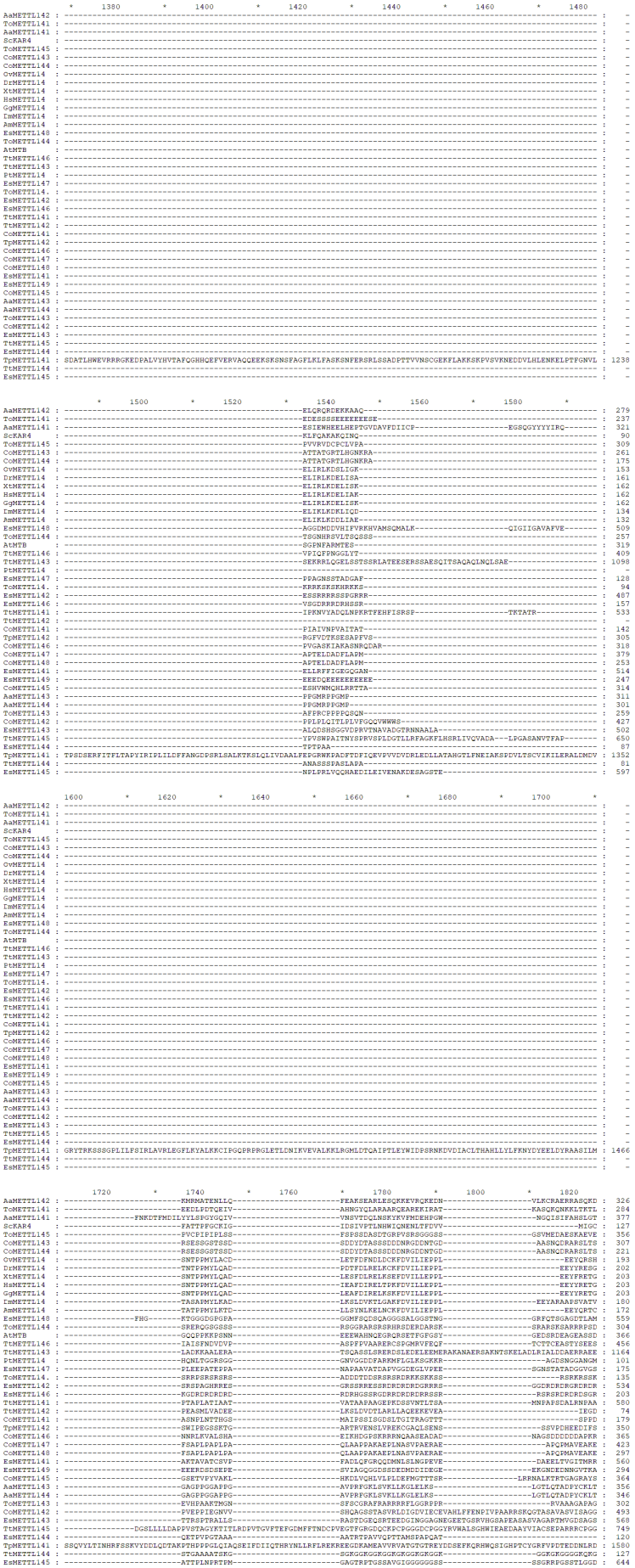

2320
2340
2360
2380
2400
PtMETTL3 : ---AWHDQGR---------- DCEHDYLAKPLLMSLLIAI---------- : 687
EsMETTL3 : ---TMGEVCCPASCGHCGGVGCDER---------- GAADECCTTEIMDYGQPCSVTGTAPCYLDG---------- : 1238
TtMETTL33 : ---------- SVLPPSRDTPLPMAVAPVATGAL---------- : 555
GtMETTL3 : ---AYDF---------- ERDDWTCLRPQCCQAKSKPKPKAKSKKKY---------- : 1061
TtMETTL34 : ---KLELNAR---------- KYPAAKAKGSSAKYTKYVDDDAA---------- : 1084
AaMETLL3 : AMDALRRDLDTVSTSKRQDAERSGAVALAAAERARALEVEDVKTASARQFDDLQRSTASQLESLQRDLTNVRAIKLKEQADYDATMDAIRRALEVQATDTYVRRDGGEKAAFFAR : 2122
TtMETTL31 : ---CFDFPLSATGSSASTANGSRSAKSASSAMRLL---------- DRPRESCICPRCVIDVHATERALVRELPSVRVADRALVLAQRAASAAAADEIIHDAVSDVAQGVAAD : 1211
CoMETTL3 : ---------- : -
TtMETTL32 : ---------- : -
ScIME4 : ---KYQEFMKSKTGTSHTGTKKIDKKQPSKLQQQHQQQYWNNMDMGSGKYYAEAKQNPMNQKHTPFESKQQQKQQFQTLNNLYFAQ---------- : 600
AtMTA : ---RFKE---------- SYPEIDVQPPSPPRASAMETDNEPMAIDSITA---------- : 685
DrMETTL3 : ---RFKK---------- RYPDGVISKPKNM---------- : 584
XtMETTL3 : ---RFKQ---------- KYPDGVIGMPKNL---------- : 573
HsMETTL3 : ---RFKQ---------- RYPDGIISKPKNL---------- : 580
GgMETTL3 : ---CFKQ---------- HYPDGIISKPKNM---------- : 548
OvMETTL3 : ---SFQK---------- RYPNGNCMEPPK---------- : 552
ImMETTL3 : ---CFQK---------- RYPDGNCMSPASANAASINGIQK---------- : 608
AmMETTL3 : ---AFKK---------- RYPDGNSMKPSKS---------- : 556
2420
2440
2460
2480
PtMETTL3 : ---------- : -
EsMETTL3 : ---------- : -
TtMETTL33 : ---------- : -
GtMETTL3 : ---------- : -
TtMETTL34 : ---------- : -
AaMETLL3 : YVDRDRMLMELVYGAIRSVTQLLSTPSANKAGATMPTELSHAILGCIKELKRMKEYVMGSFDALQKDMAPFLPFEV : 2198
TtMETTL31 : AVADAVAETMIDDAAAAVAAESMAG---------- : 1236
CoMETTL3 : ---------- : -
TtMETTL32 : ---------- : -
ScIME4 : ---------- : -
AtMTA : ---------- : -
DrMETTL3 : ---------- : -
XtMETTL3 : ---------- : -
HsMETTL3 : ---------- : -
GgMETTL3 : ---------- : -
OvMETTL3 : ---------- : -
ImMETTL3 : ---------- : -
AmMETTL3 : ---------- : -

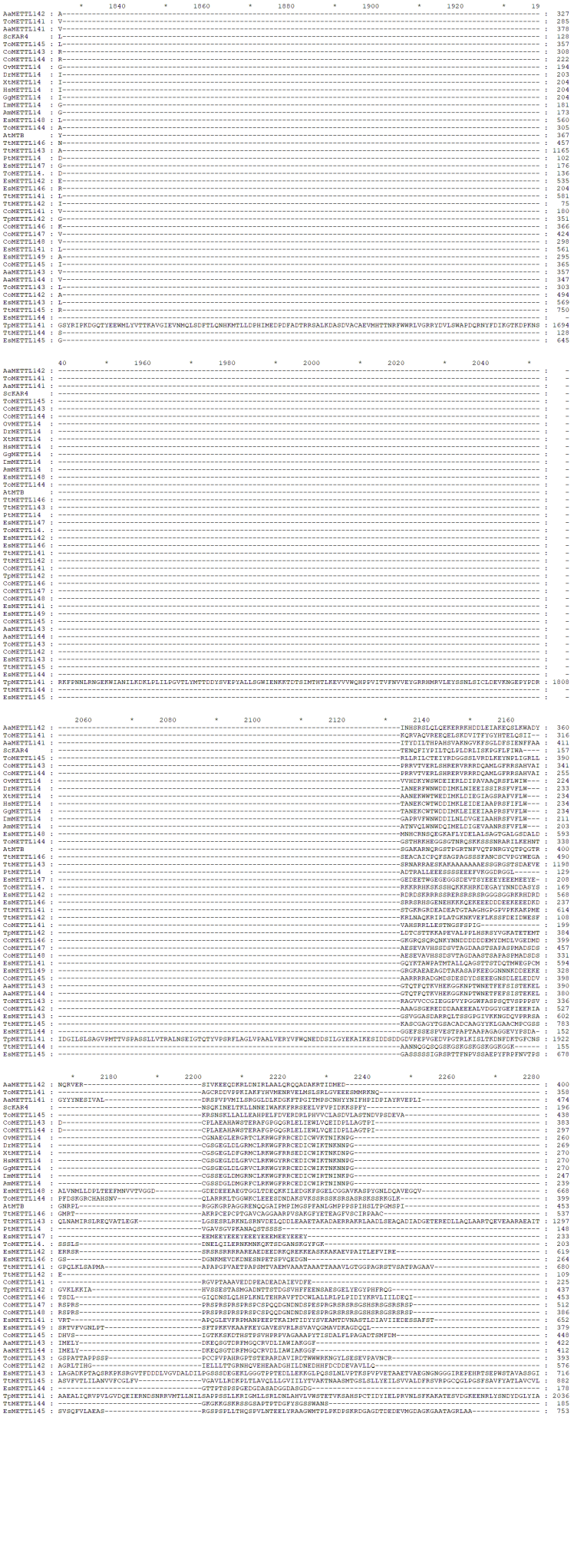
Protein Alignment used in the METTL14 tree of similarity. The alignments were made using muscle and displayed using Genedoc. There are four levels of shading, primary which is 100% match between the amino acids; secondary which is 80% match between the amino acids; tertiary which is 60% match between the amino acids; and quaternary which is less than 60%.

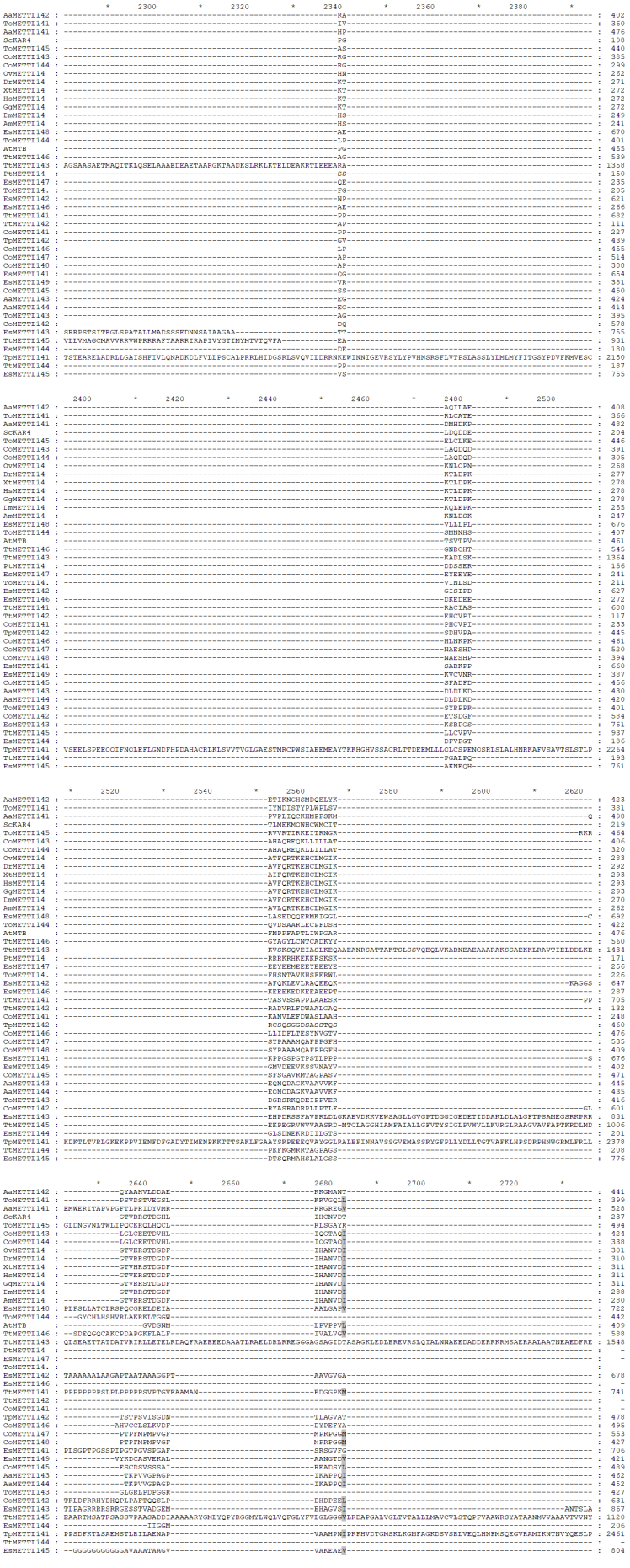


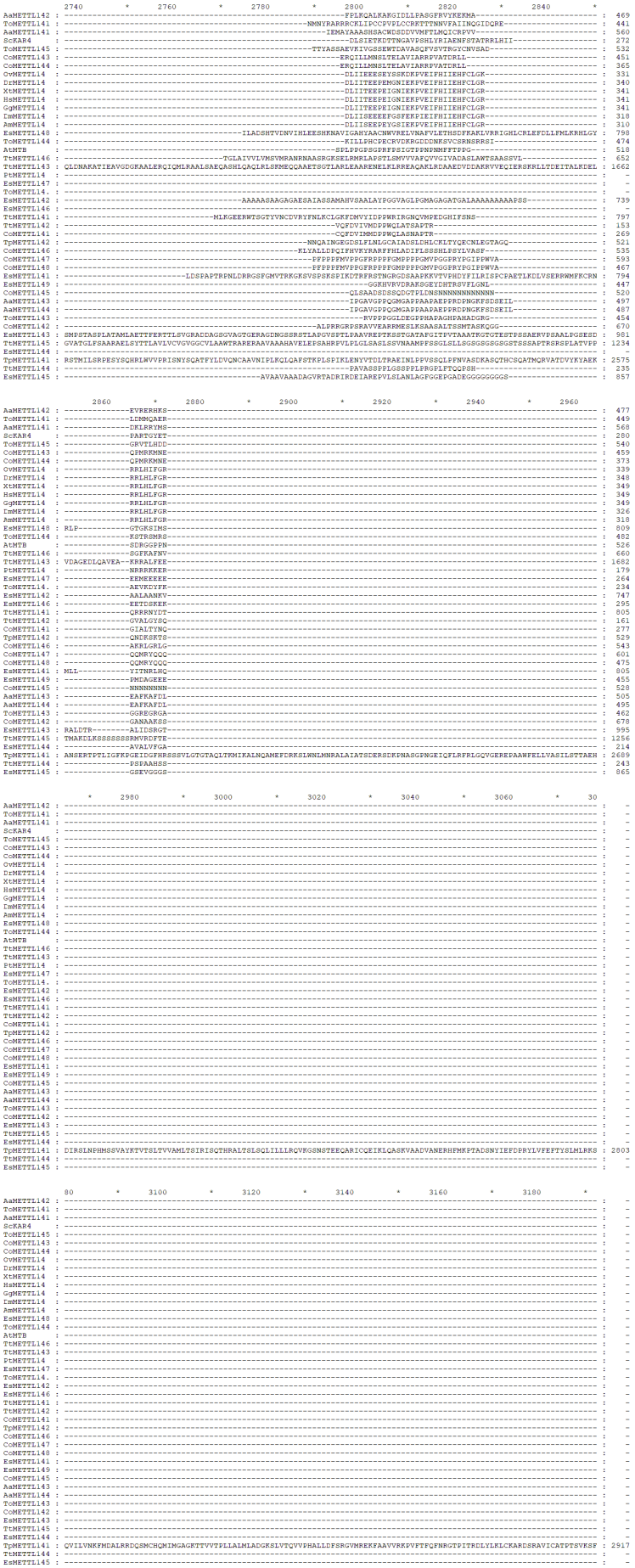


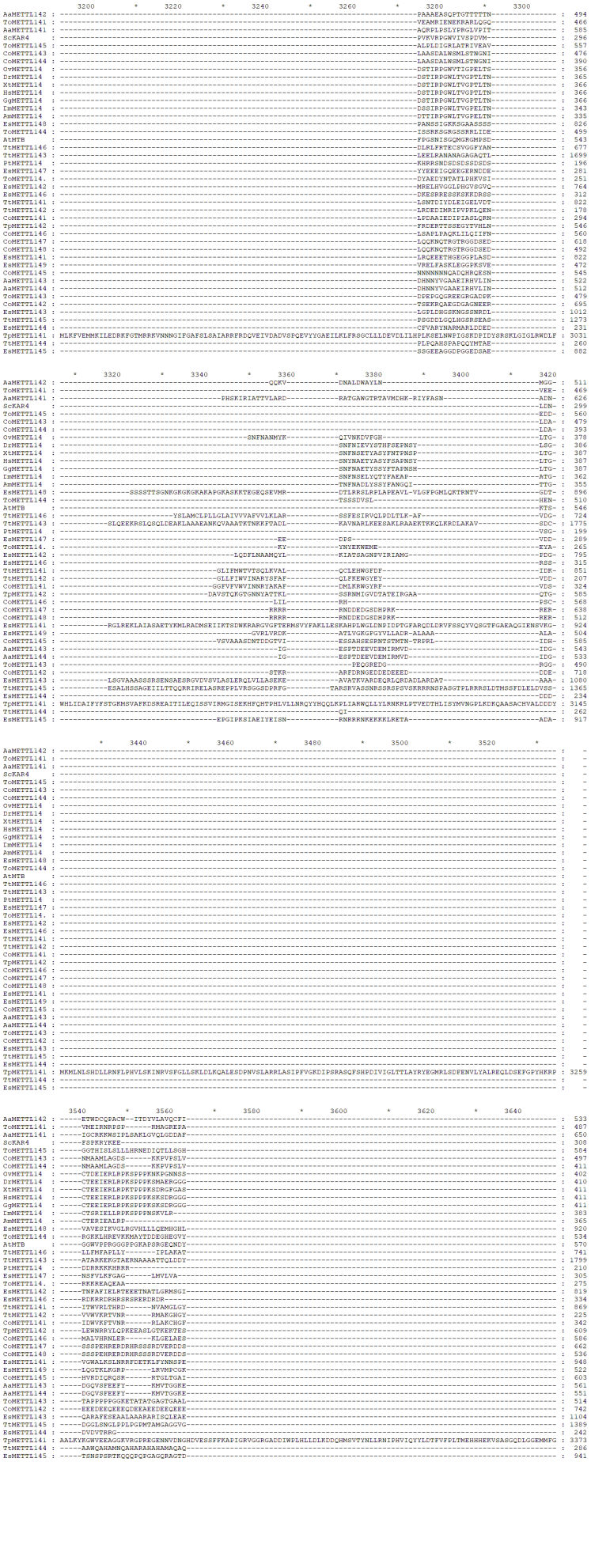


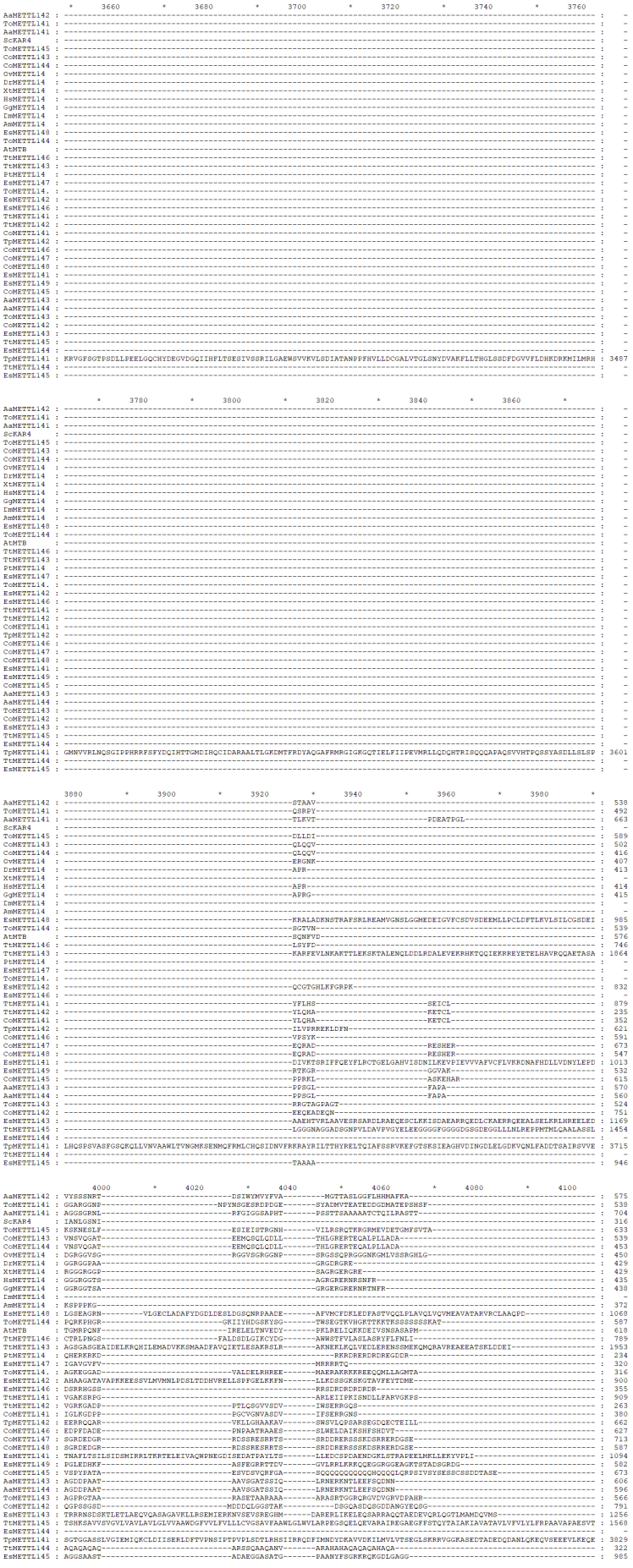


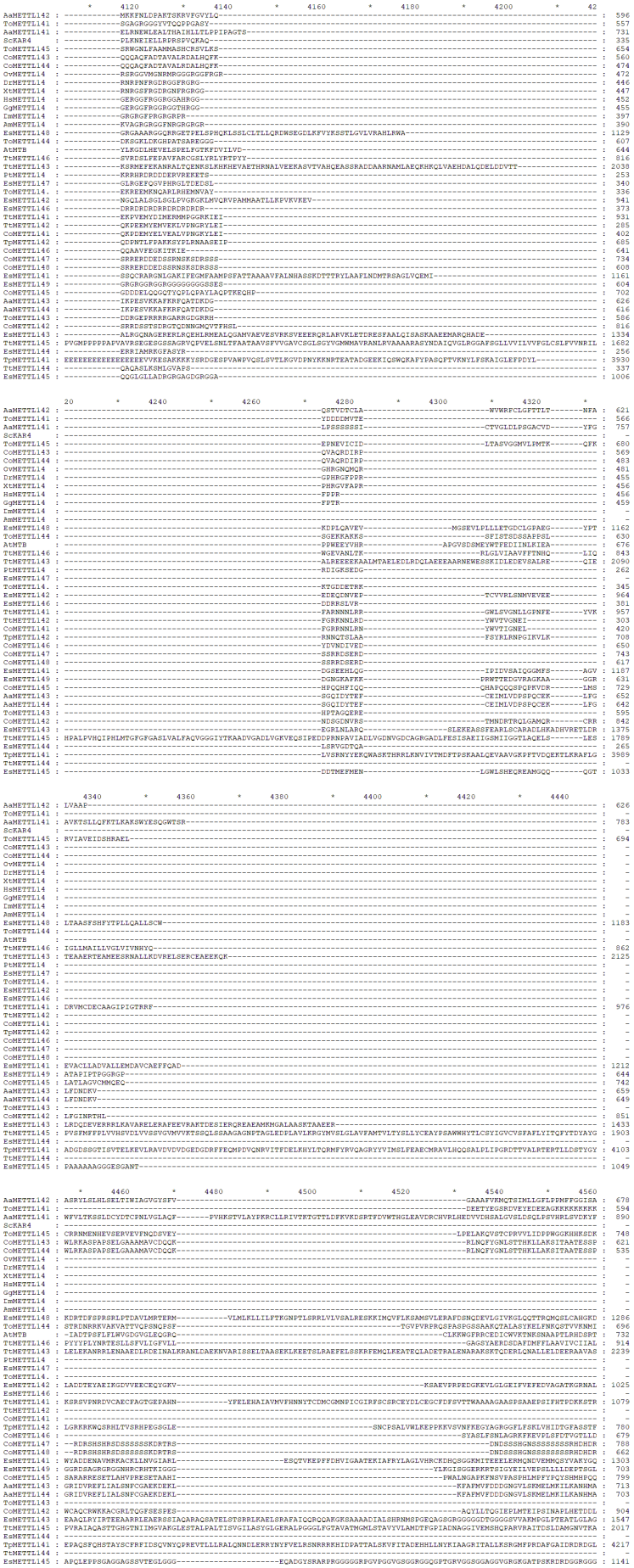


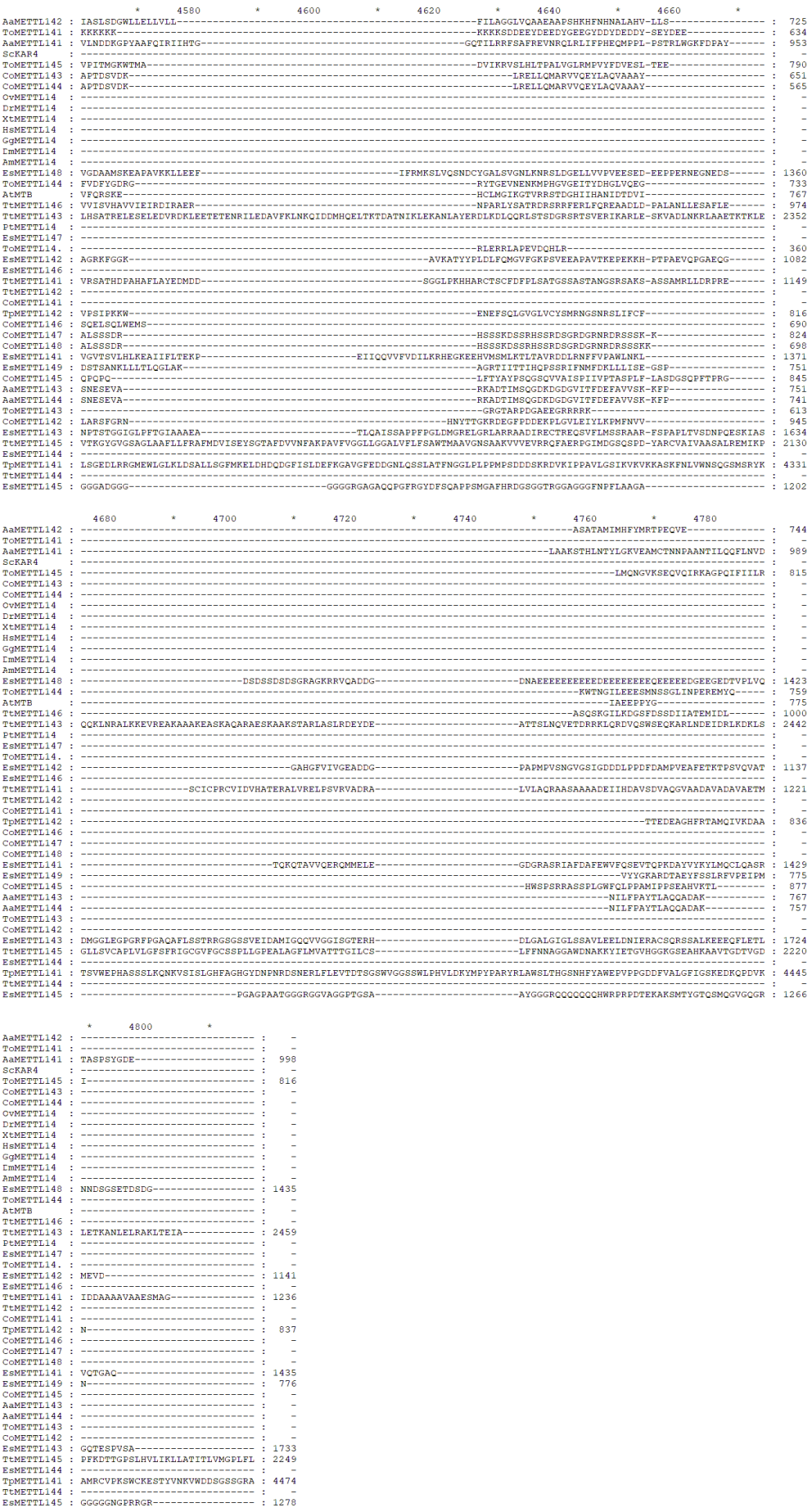




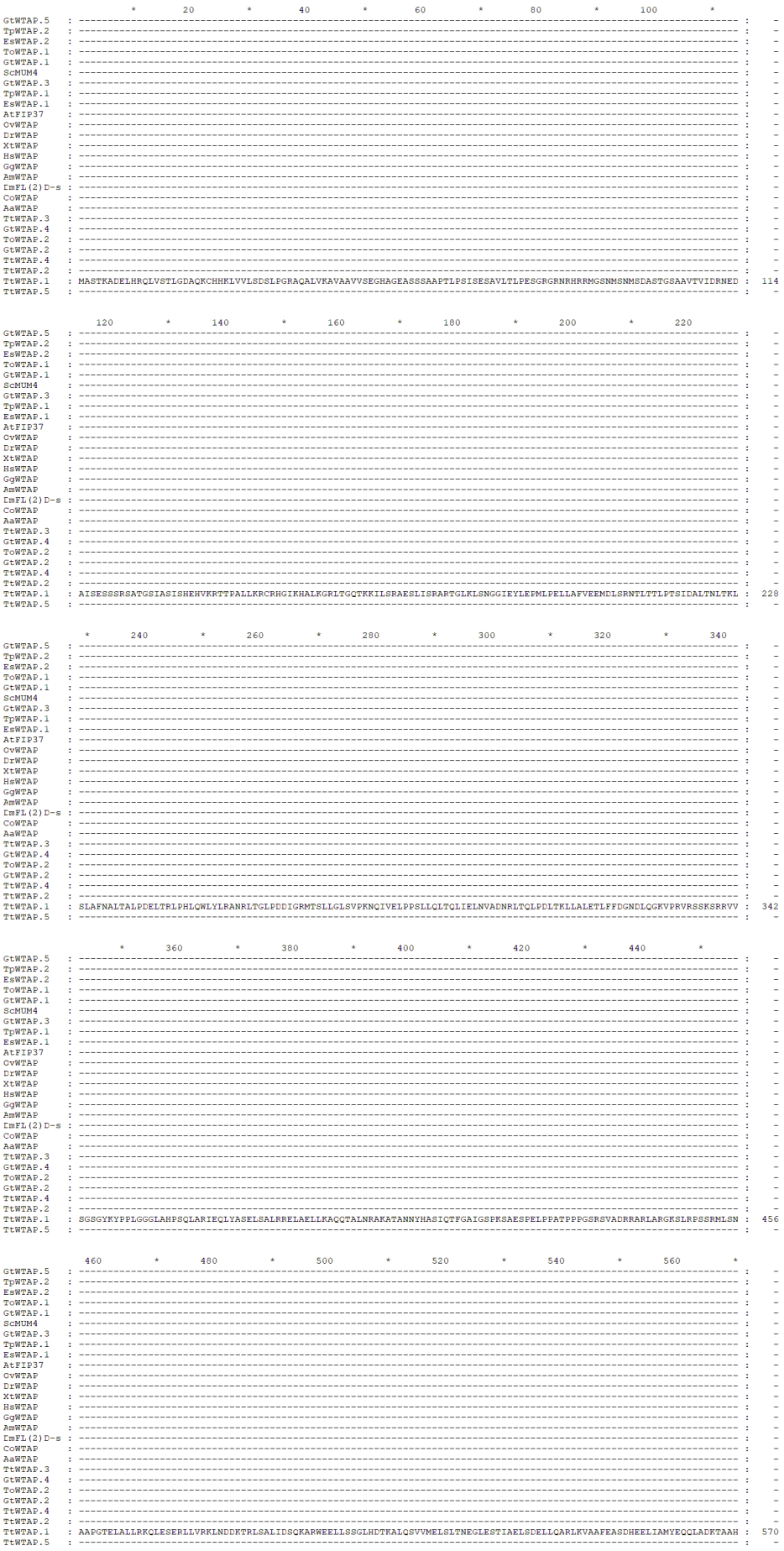


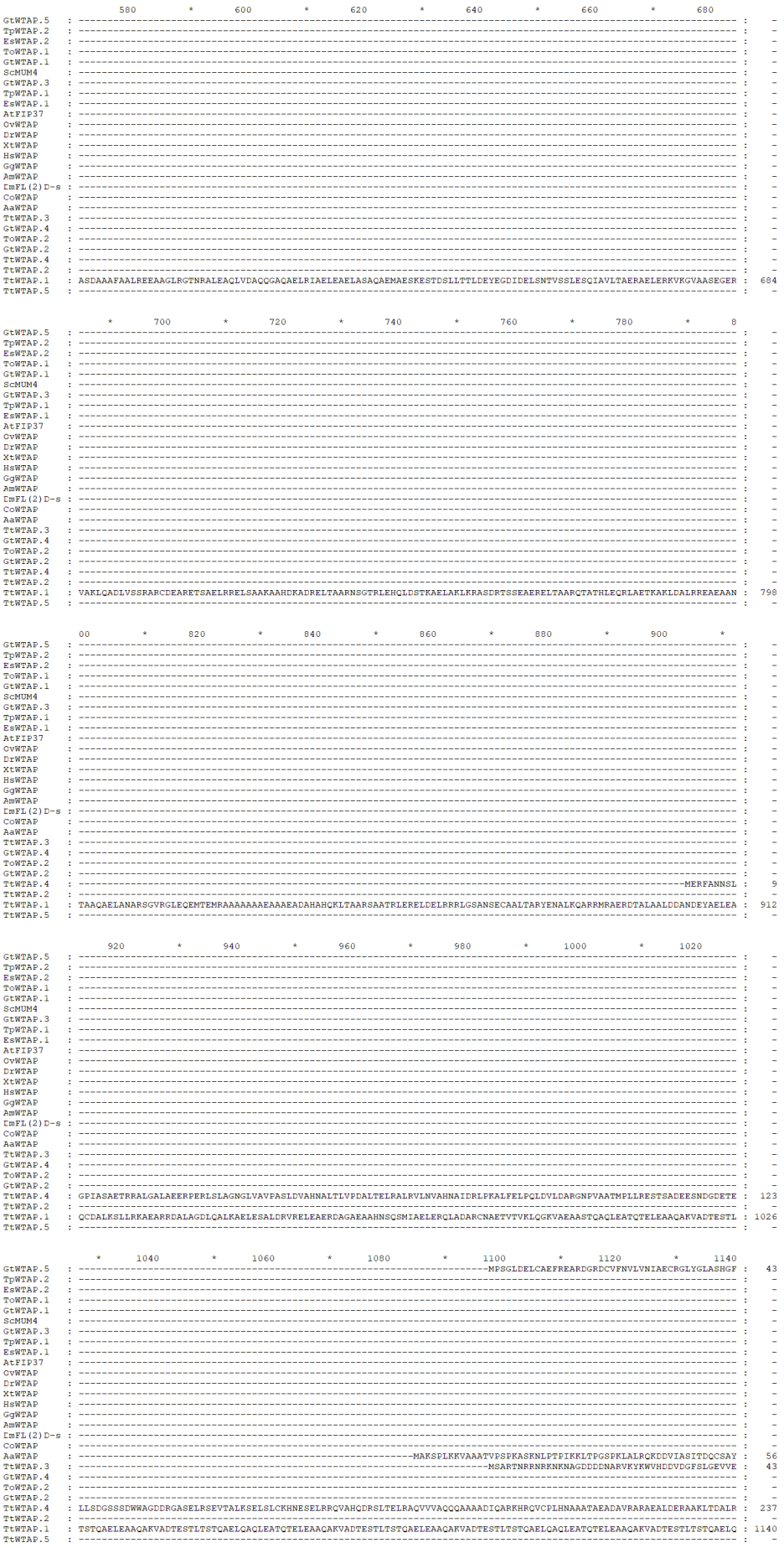




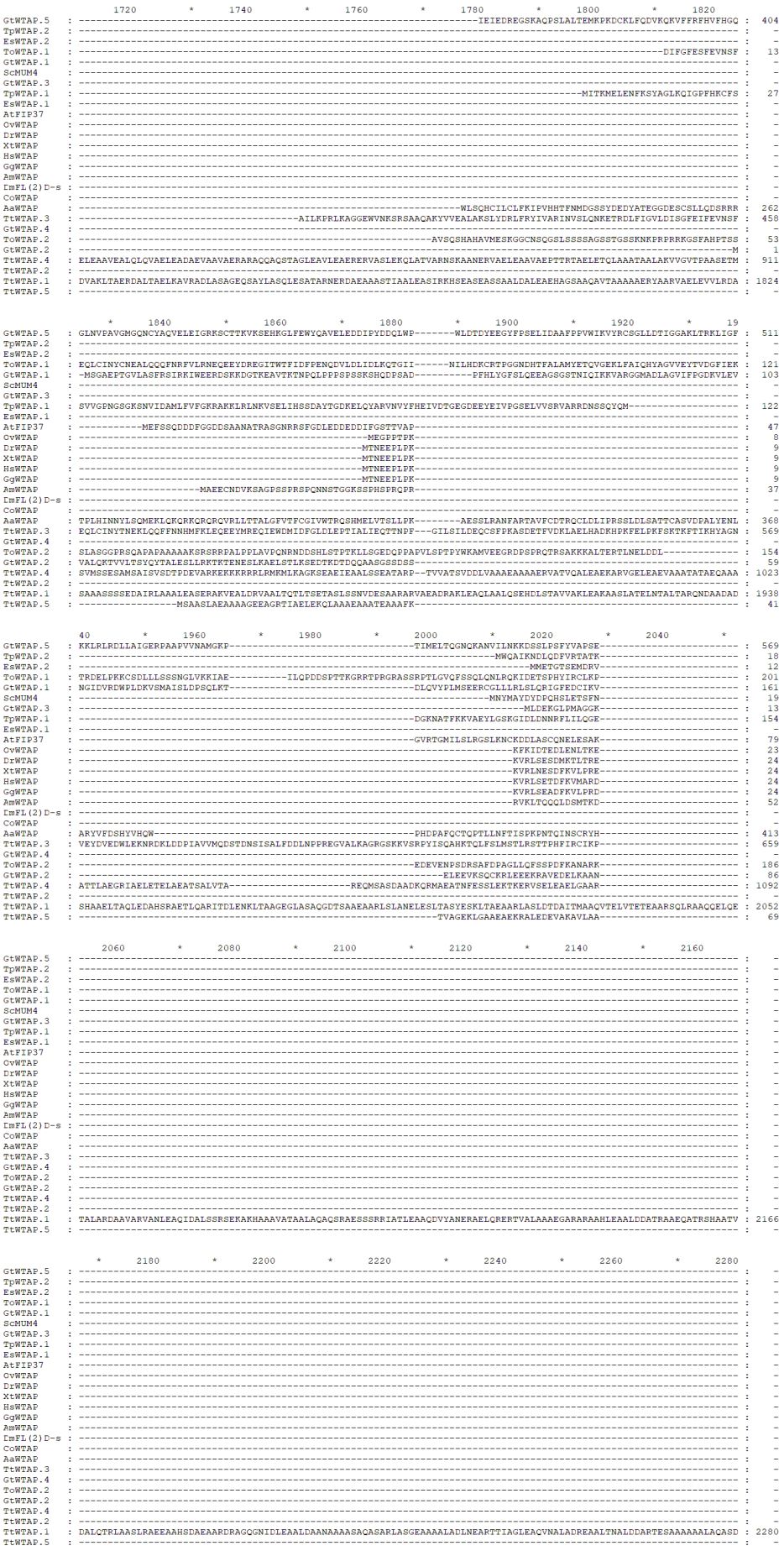


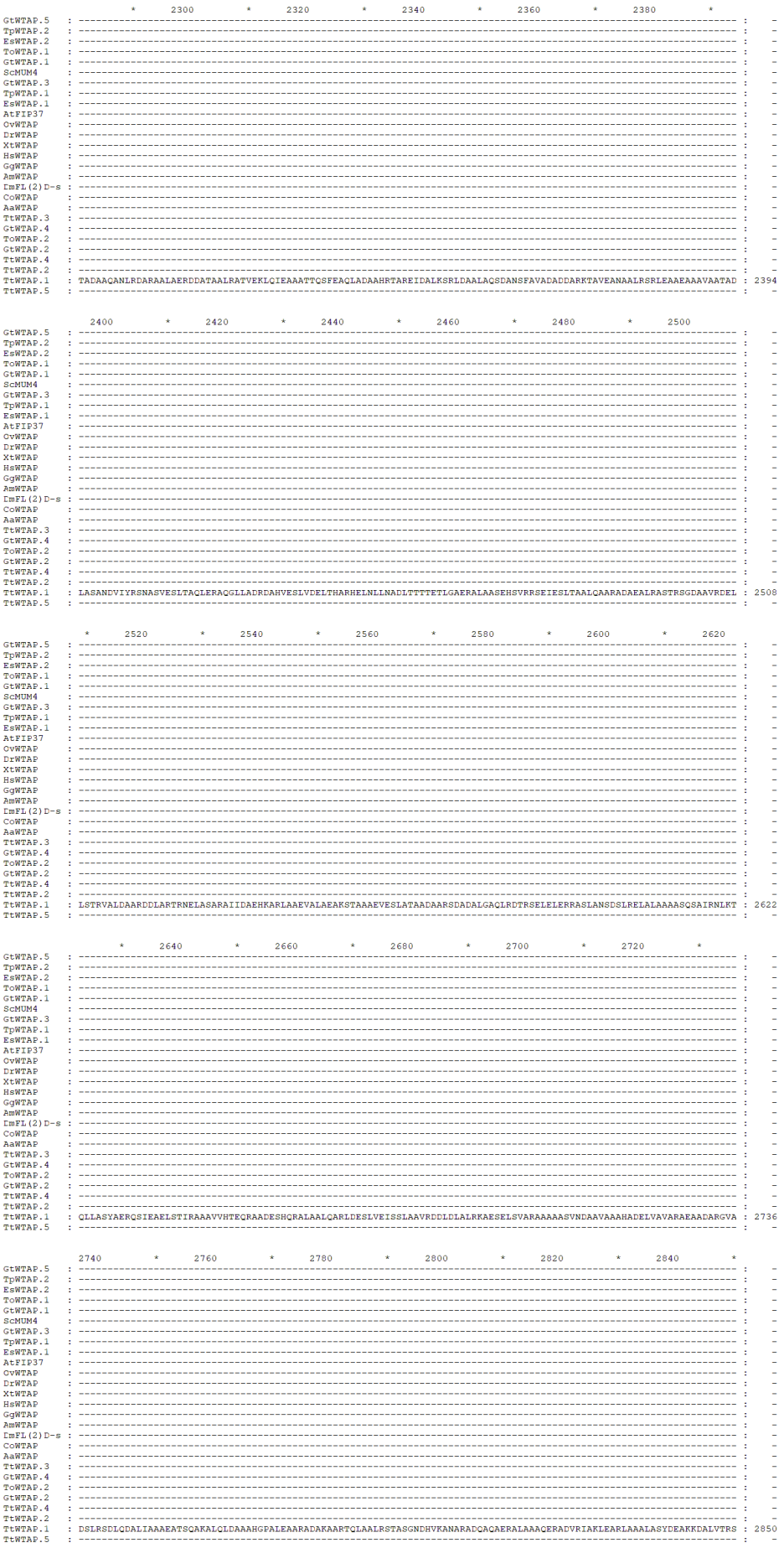
Protein Alignment used in the WTAP tree of similarity. The alignments were made using muscle and displayed using Genedoc. There are four levels of shading, primary which is 100% match between the amino acids; secondary which is 80% match between the amino acids; tertiary which is 60% match between the amino acids; and quaternary which is less than 60%.

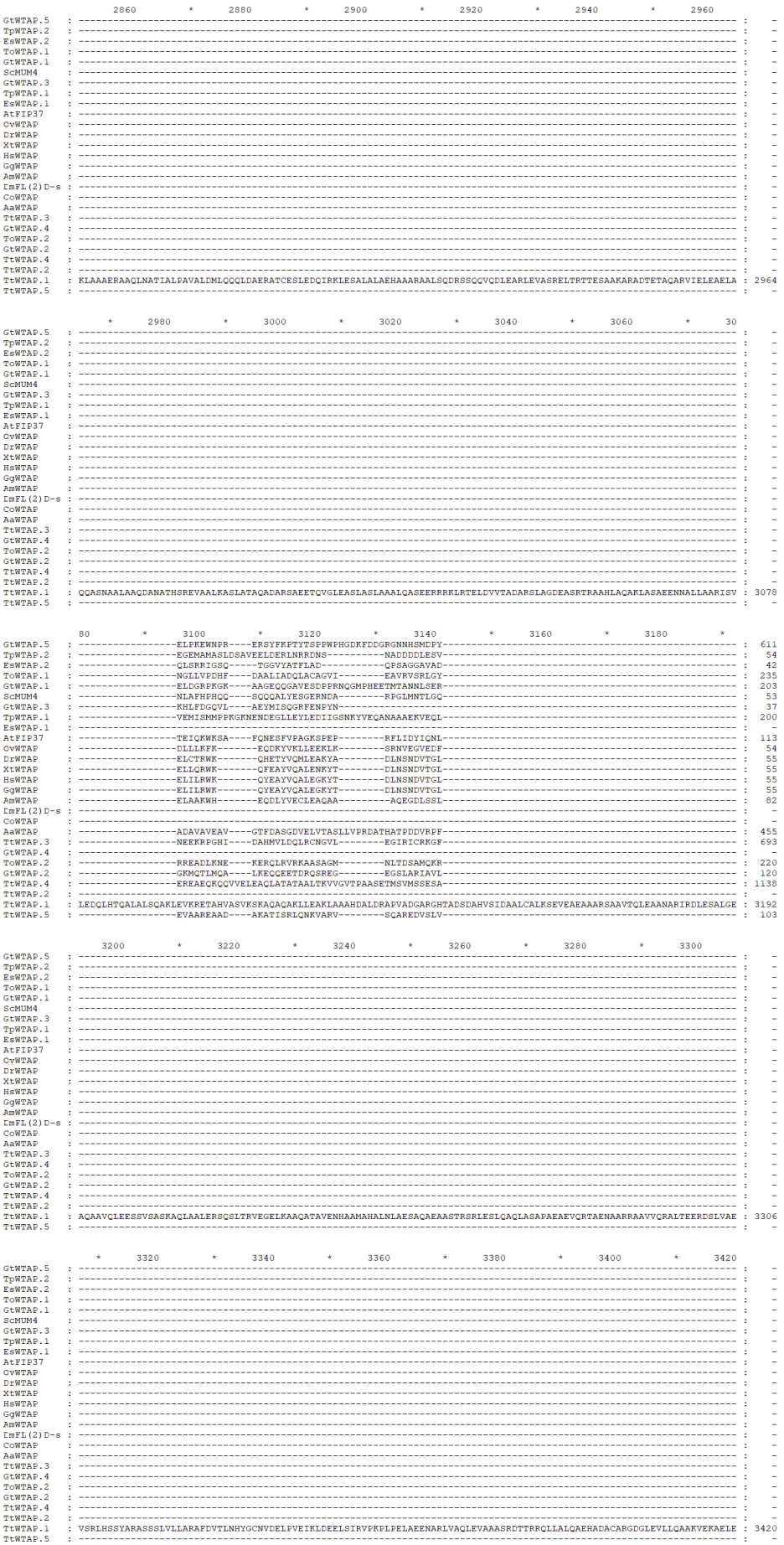

GtWTAP.5
TpWTAP.2
EsWTAP.2
ToWTAP.1
GtWTAP.1
ScMUM4
GtWTAP.3
TpWTAP.1
EsWTAP.1
AtFIP37
OvWTAP
DrWTAP
XtWTAP
HsWTAP
GgWTAP
AmWTAP
DmFL(2)D-s
CoWTAP
AaWTAP
TtWTAP.3
GtWTAP.4
ToWTAP.2
GtWTAP.2
TtWTAP.4
TtWTAP.2
TtWTAP.1
TtWTAP.5
MASTKADELHRQLVSTLGDAQKCHHKLVVLSDSLPGRAQALVKAVAAVVSEGHAGEASSSAAPTLPSISESAVLTLPESGRGRNRHRRMGSNMSNMSDASTGSAAVTVIDRNED
AISESSSSRSATGSIASISHEHVKRTTPALLKRCRHGIKHALKGRLTGQTKKILSRAESLISRARTGLKLSNGGIEYLEPMLPELLAFVEEMDLSRNTLTTLPTSIDALTNLTKL
SLAFNALTALPDELTRLPHLQWLYLRANRLTGLPDDIGRMTSLLGLSVPKNQIVELPPSLLQLTQLIELNVADNRLTQLPDLTKLLALETLFFDGNDLQGKVPRVRSSKSRRVV
SGSGYKYPPLGGGLAHPSQLARIEQLYASELSALRRELAELLKAQQTALNRAKATANNYHASIQTFGAIGSPKSAESPELPPATPPPGSRSVADRRARLARGKSLRPSSRMLSN
AAPGTELALLRKQLESERLLVRKLNDDKTRLSALIDSQKARWEELLSSGLHDTKALQSVVMELSLTNEGLESTIAELSDELLQARLKVAAFEASDHEELIAMYEQQLADKTAAH

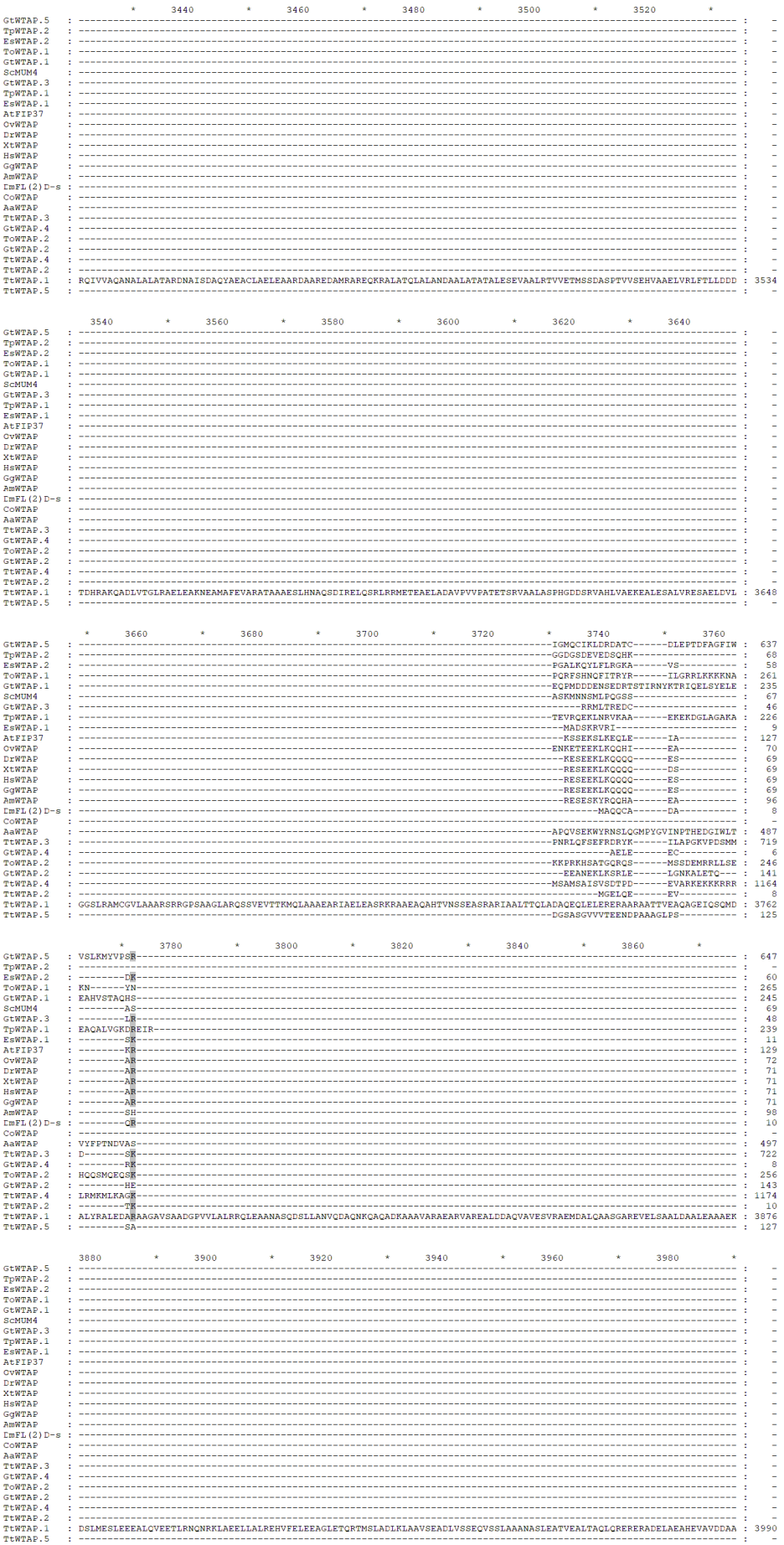




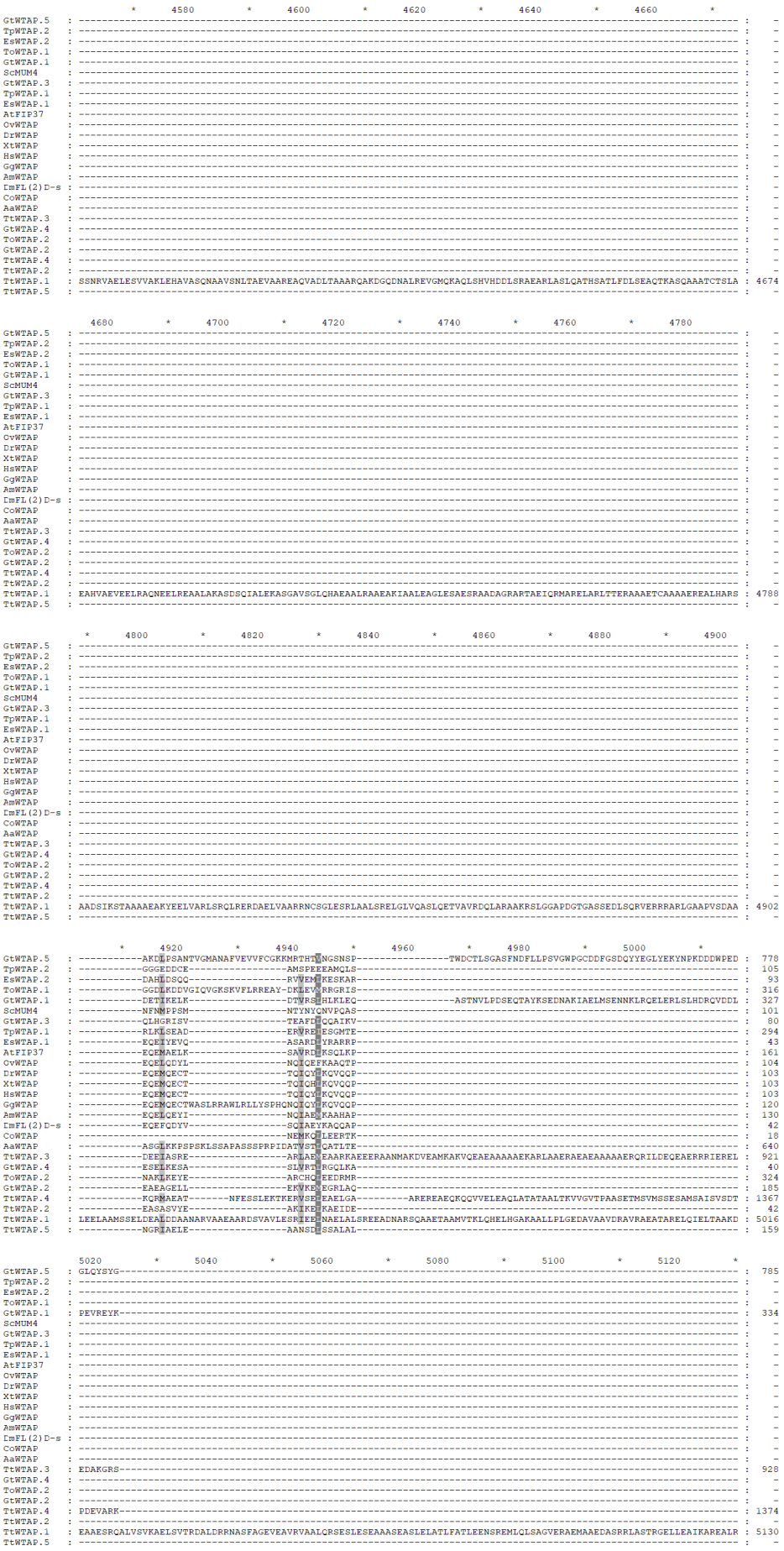


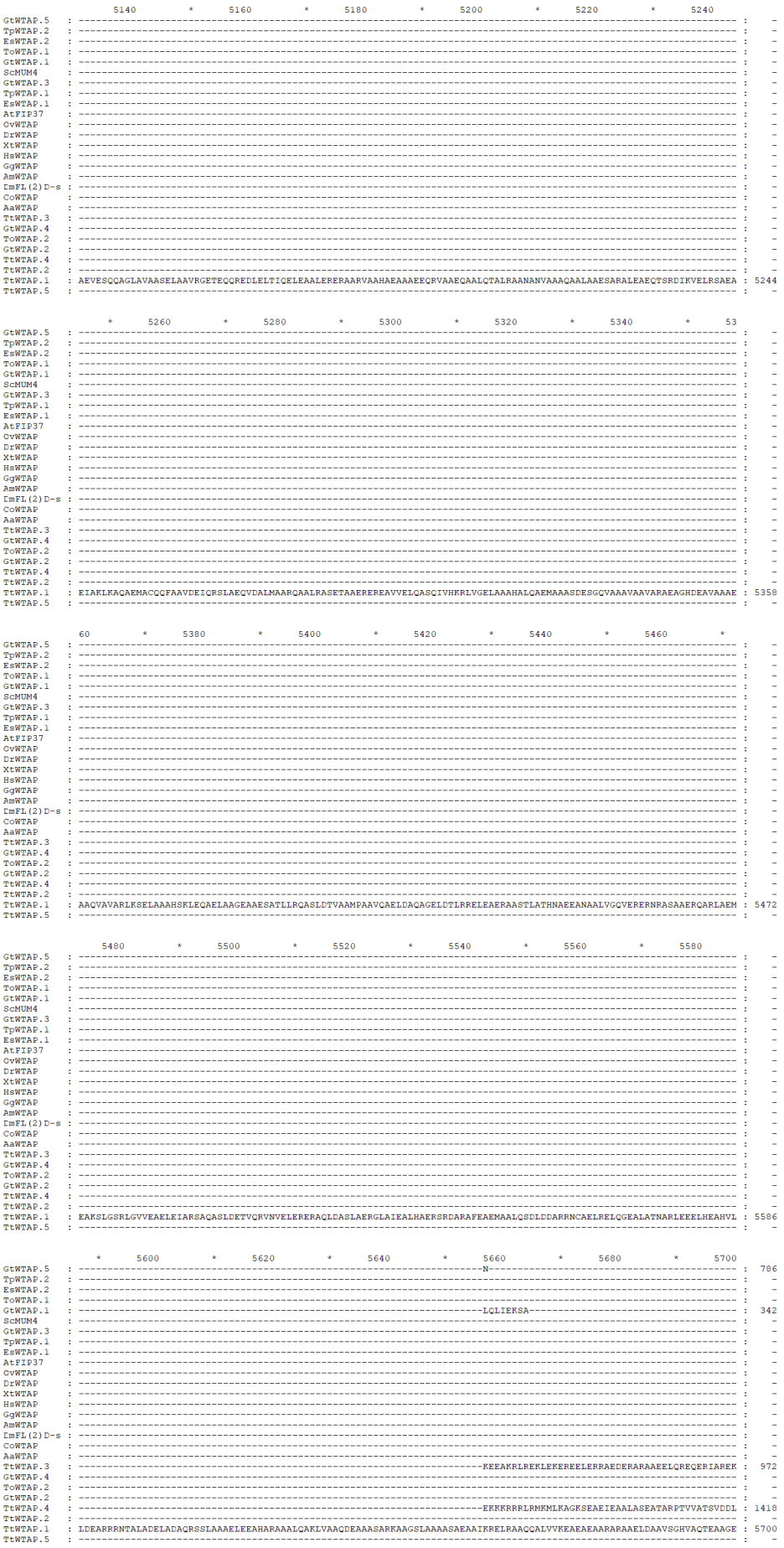

2300
2320
2340
2360
2380
GtWTAP.5
TpWTAP.2
EsWTAP.2
ToWTAP.1
GtWTAP.1
ScMUM4
GtWTAP.3
TpWTAP.1
EsWTAP.1
AtFIP37
OvWTAP
DrWTAP
XtWTAP
HsWTAP
GgWTAP
AmWTAP
DmFL(2)D-s
CoWTAP
AaWTAP
TtWTAP.3
GtWTAP.4
ToWTAP.2
GtWTAP.2
TtWTAP.4
TtWTAP.2
TtWTAP.1
TtWTAP.5
TADAAQANLRDARAALAERDDATAALRATVEKLQIEAAATTQSFEAQLADAAHRTAREIDALKSRLDAALAQSDANSFAVADADDARKTAVEANAALRSRLEAAEAAAVAATAD
2400
2420
2440
2460
2480
2500
LASANDVIYRSNASVESLTAQLERAQGLLADRDAHVESLVDELTHARHELNLLNADLTTTTETLGAERALAASEHSVRRSEIESLTAALQAARADAEALRASTRSGDAAVRDEL
2520
2540
2560
2580
2600
2620
LSTRVALDAARDDLARTRNELASARAIIDAEHKARLAAEVALAEAKSTAAAEVESLATAADAARSDADALGAQLRDTRSELELERRASLANSDSLRELALAAAASQSAIRNLKT
2640
2660
2680
2700
2720
QLLASYAERQSIEAELSTIRAAAVVHTEQRAADESHQRALAALQARLDESLVEISSLAAVRDDLDLALRKAESELSVARAAAAASVNDAAVAAAHADELVAVARAEAADARGVA
2740
2760
2780
2800
2820
2840
DSLRSDLQDALIAAAEATSQAKALQLDAAAHGPALEAARADAKAARTQLAALRSTASGNDHVKANARADQAQAERALAAAQERADVRIAKLEARLAAALASYDEAKKDALVTRS


2860 2880 2900 2920 2940 2960
GtWTAP.5 TpWTAP.2 EsWTAP.2 ToWTAP.1 GtWTAP.1 ScMUM4 GtWTAP.3 TpWTAP.1 EsWTAP.1 AtFIP37 OvWTAP DrWTAP XtWTAP HsWTAP GgWTAP AmWTAP DmFL(2)D-s CoWTAP AaWTAP TtWTAP.3 GtWTAP.4 ToWTAP.2 GtWTAP.2 TtWTAP.4 TtWTAP.2 TtWTAP.1 TtWTAP.5
KLAAAERAAQLNATIALPAVALDMLQQQLDAERATCESLEDQIRKLESALALAEHAAARAALSQDRSSQQVQDLEARLEVASRELTRTTESAAKARADTETAQARVIELEAELA 2964
2980 3000 3020 3040 3060 30
QQASNAALAAQDANATHSREVAALKASLATAQADARSAEETQVGLEASLASLAAALQASEERRRKLRTELDVVTADARSLAGDEASRTRAAHLAQAKLASAEENNALLAARISV 3078
80 3100 3120 3140 3160 3180
ELPKEWNPR----ERSYFKPTYTSPFWPHGDKFDDGRGNNHSMDPY 611
EGEMAMASLDSAVEELDERLNRRDNS---------NADDDDLESV 54
GLSRRIGSQ------TGGYVATFLAD---------QPSAGGAVAD 42
NGLLVPDHF----DAALIADQLACAGVI--------EAVRVSRLGY 235
ELDGRPKGK----AAGEQQGAVESDPPRNQGMPHEETMTANNLSER 203
NLAFHPHQQ----SQQQALYESGERNDA--------RPGLMNTLGQ 53
KHLFDGQVL----AEYMISQGRFENPYN 37
VEMISMMPPKGKNENDEGLLEYLEDIIGSNKYVEQANAAAEKVEQL 200
TEIQKWKSA----FQNESFVPAGKSPEP--------RFLIDYIQNL 113
DLLLKFK------EQDKYVKLLEEKLR--------SRNVEGVEDF 54
ELCTRWK------QHEYVQMLEAKYA--------DLNSNDVTGL 55
ELLCRWK------QFEAYVQALENKYT--------DLNSNDVTGL 55
ELILRWK------QYEAYVQALEGKYT--------DLNSNDVTGL 55
ELILRWK------QYEAYVQALEGKYT--------DLNSNDVTGL 55
ELAAKWH------EQDLYVECLEAQAA--------AQEGDLSSL 82
ADAVAVEAV----GTFDASGDVELVTASLLVPRDATHATPDDVRPF 455
NEEKRPGHI----DAHMVLDQLRCNGVL--------EGIRICRKGF 693
RREADLKNE----KERQLRVRKAASAGM--------NLTDSAMQKR 220
GKMQTLMQA----LKEQQEETDRQSREG--------EGSLARIAVL 120
EREAECKQQVVELEAQLATATAALTKVVGVTPAASETMSVMSSESA 1138
EVAAREAAD----AKATISRLQNKVARV--------SQAREDVSLV 103
3200 3220 3240 3260 3280 3300
AQAAVQLEESSVSASKAQLAALERSQSLTRVEGELKAAQATAVENHAAMAHALNLAESAQAEAASTRSRLESLQAQLASAPAEAEVQRTAENAARRAAVVQRALTEERDSLVAE 3306
3320 3340 3360 3380 3400 3420
VSRLHSSYARASSSLVLLARAFDVTLNHYGCNVDELPVEIKLDEELSIRVPKPLPELAEENARLVAQLEVAAASRDTTRRQLLALQAEHADACARGDGLEVLLQAAKVEKAELE 3420

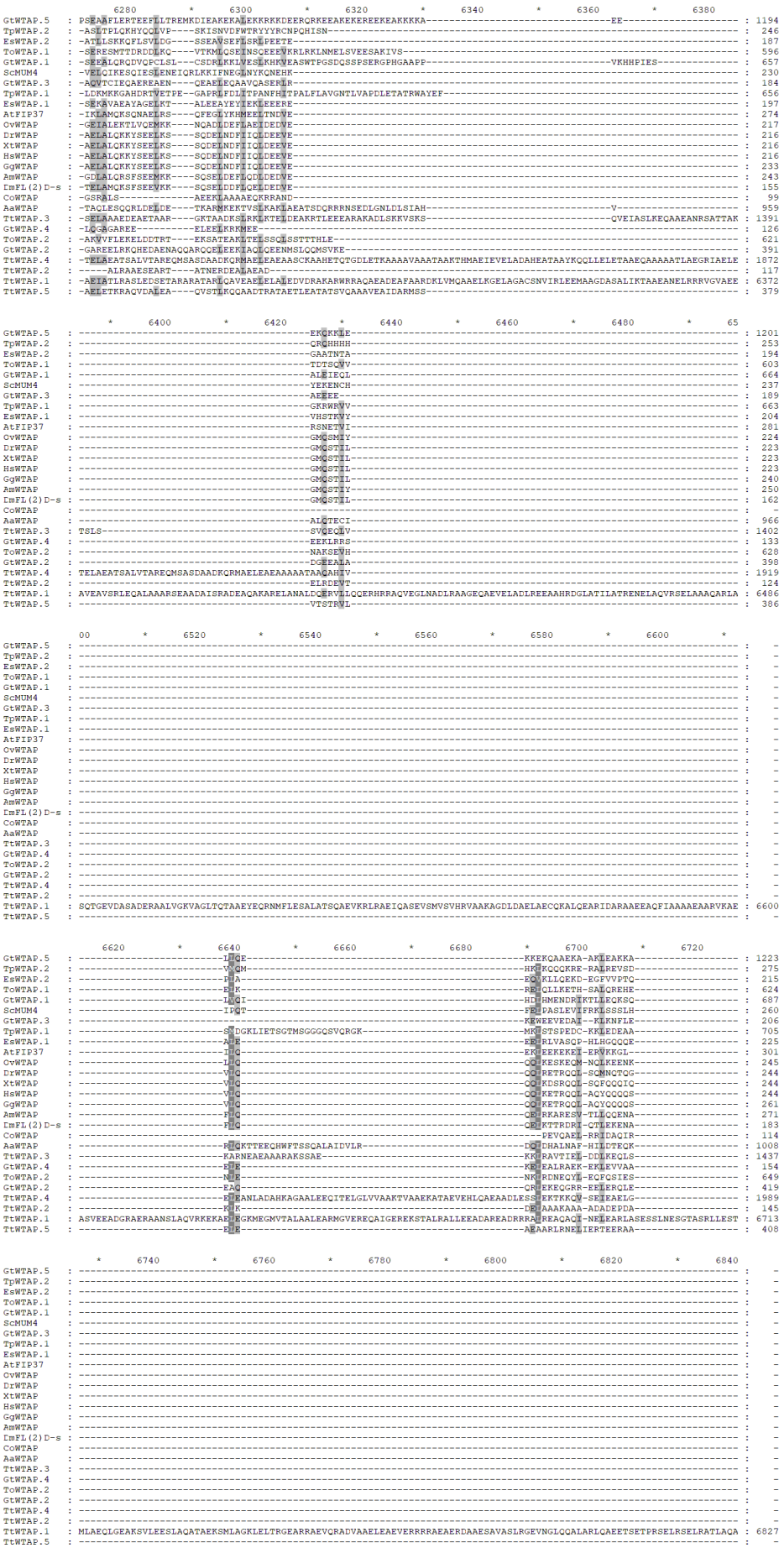


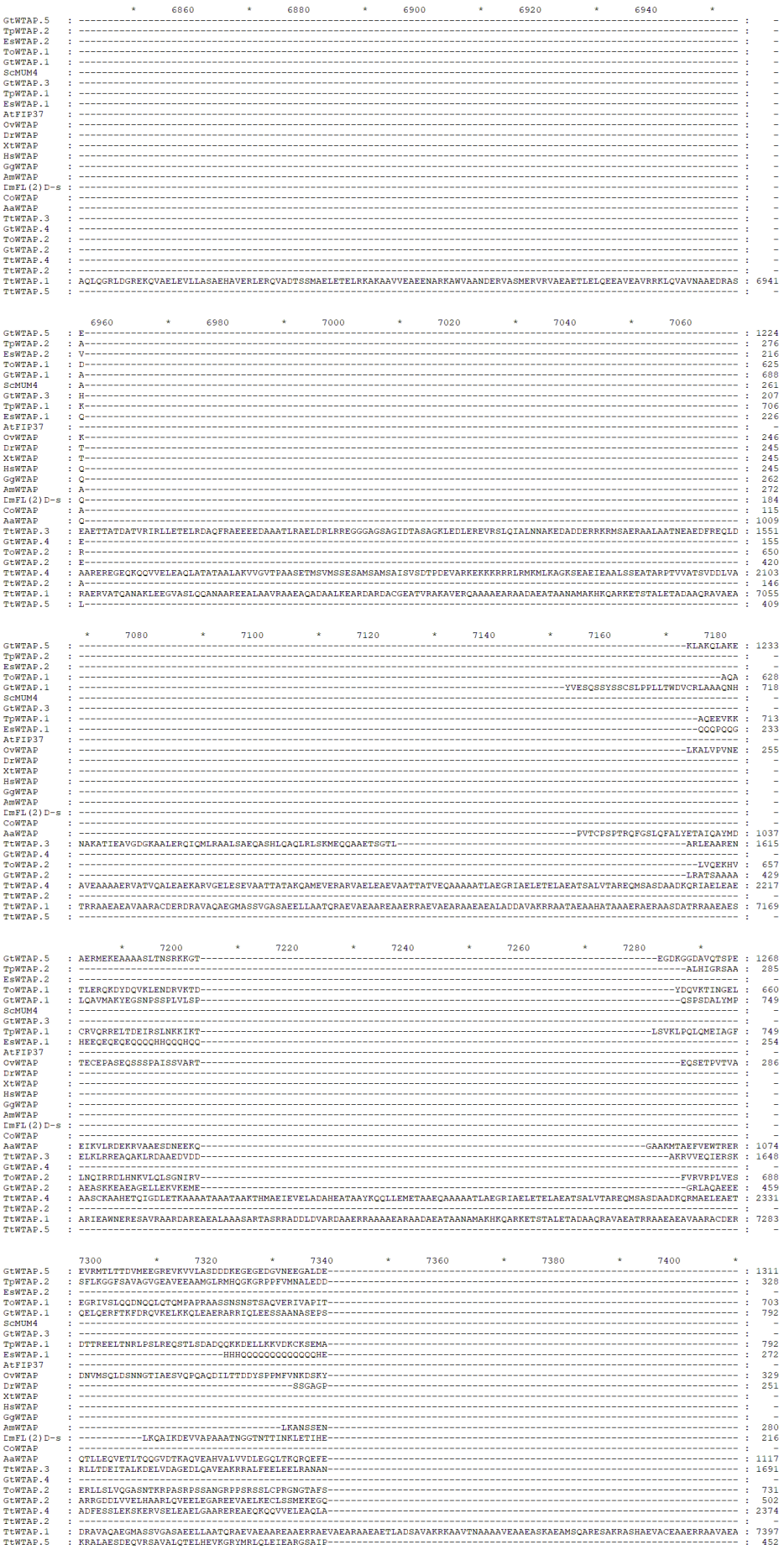


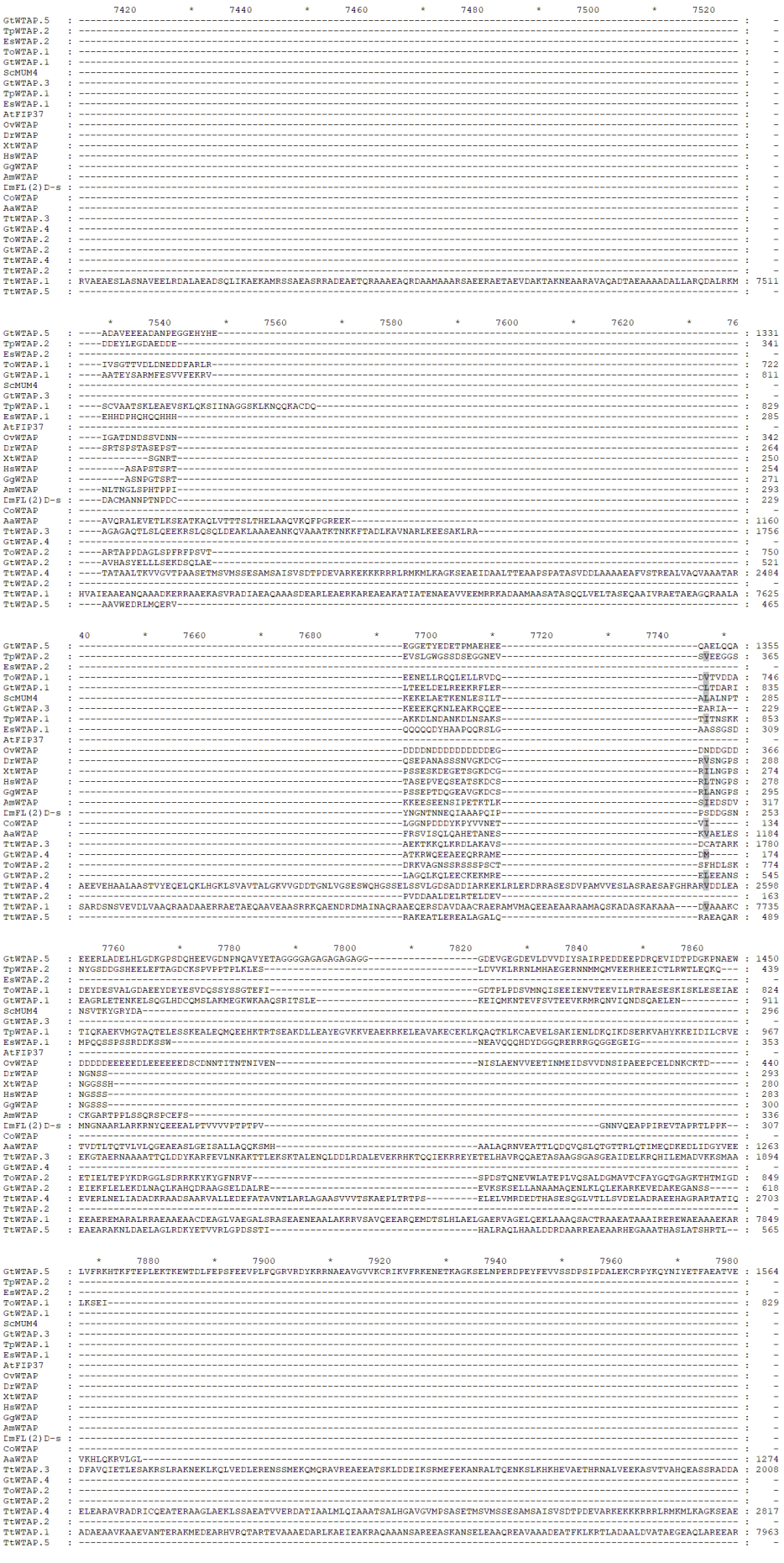


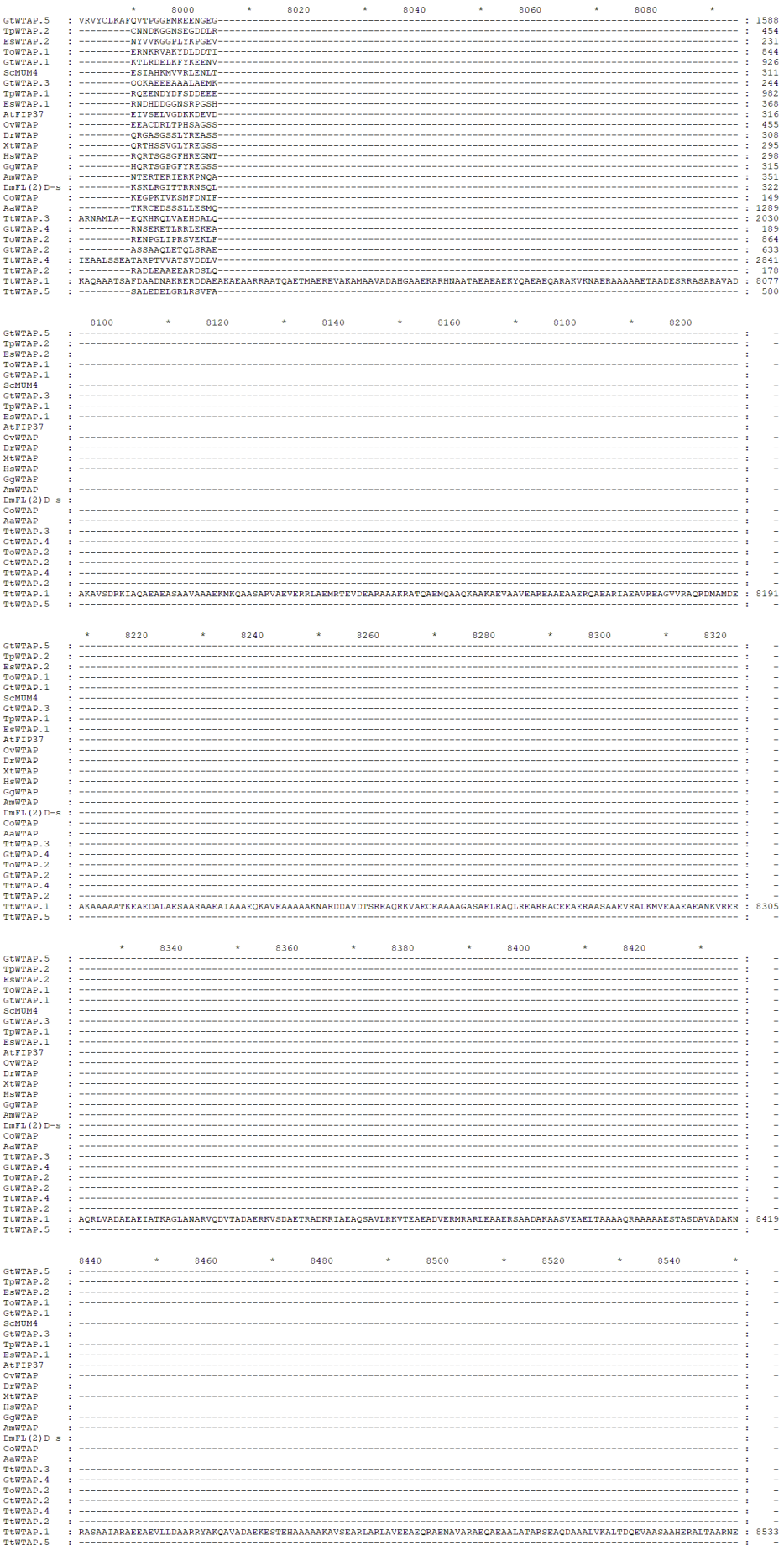

5140 5160 5180 5200 5220 5240
GtWTAP.5
TpWTAP.2
EsWTAP.2
ToWTAP.1
GtWTAP.1
ScMUM4
GtWTAP.3
TpWTAP.1
EsWTAP.1
AtFIP37
OvWTAP
DrWTAP
XtWTAP
HsWTAP
GgWTAP
AmWTAP
DmFL(2)D-s
CoWTAP
AaWTAP
TtWTAP.3
GtWTAP.4
ToWTAP.2
GtWTAP.2
TtWTAP.4
TtWTAP.2
TtWTAP.1 : AEVESQQAGLAVAASELAAVRGETEQQREDLELTIQELEAALERERAARVAAHAEAAAEEQRVAAEQAALQTALRAANANVAAAQAALAAESARALEAEQTSRDIKVELRSAEA : 5244
TtWTAP.5
5260 5280 5300 5320 5340 53
TtWTAP.1 : EIAKLKAQAEMACQQFAAVDEIQRSLAEQVDALMAARQAALRASETAAERERE AVVELQASQIVHKRLVGELAAAHALQAEMAAASDESGQVAAAVAAVARAEAGHDEAVAAAE : 5358
60 5380 5400 5420 5440 5460
TtWTAP.1 : AAQVAVARLKSELAAAHSKLEQAELAAGEAAESATLLRQASLDTVAAMPAAVQAELDAQAGELDTLRRELEAERAASTLATHNAEEANAALVGQVERERNRASAAERQARLAEM : 5472
5480 5500 5520 5540 5560 5580
TtWTAP.1 : EAKSLGSRLGVVEAELEIARSAQASLDETVQRVNVELERERAQLDASLAERGLAIEALHAERSRDARAFEAEMAALQSDLDDARRNCAELRELQGEALATNARLEEELHEAHVL : 5586
5600 5620 5640 5660 5680 5700
GtWTAP.5 : ----------N---------- : 786
GtWTAP.1 : ----------LQLIEKSA---------- : 342
TtWTAP.3 : ----------KEEAKRLREKLEKEREELERRAEDERARAAEELQREQERIAREK : 972
TtWTAP.4 : ----------EKKKRRRLRMKMLKAGKSEAEIEAALASEATARPTVVATSVDDL : 1418
TtWTAP.1 : LDEARRRNTALADELADAQRSSLAAAELEEAHARAAALQAKLVAAQDEAAASARKAAGSLAAAASAEAAIKRELRAAQQALVVKEAEAEAARARAAELDAAVSGHVAQTEAAGE : 5700

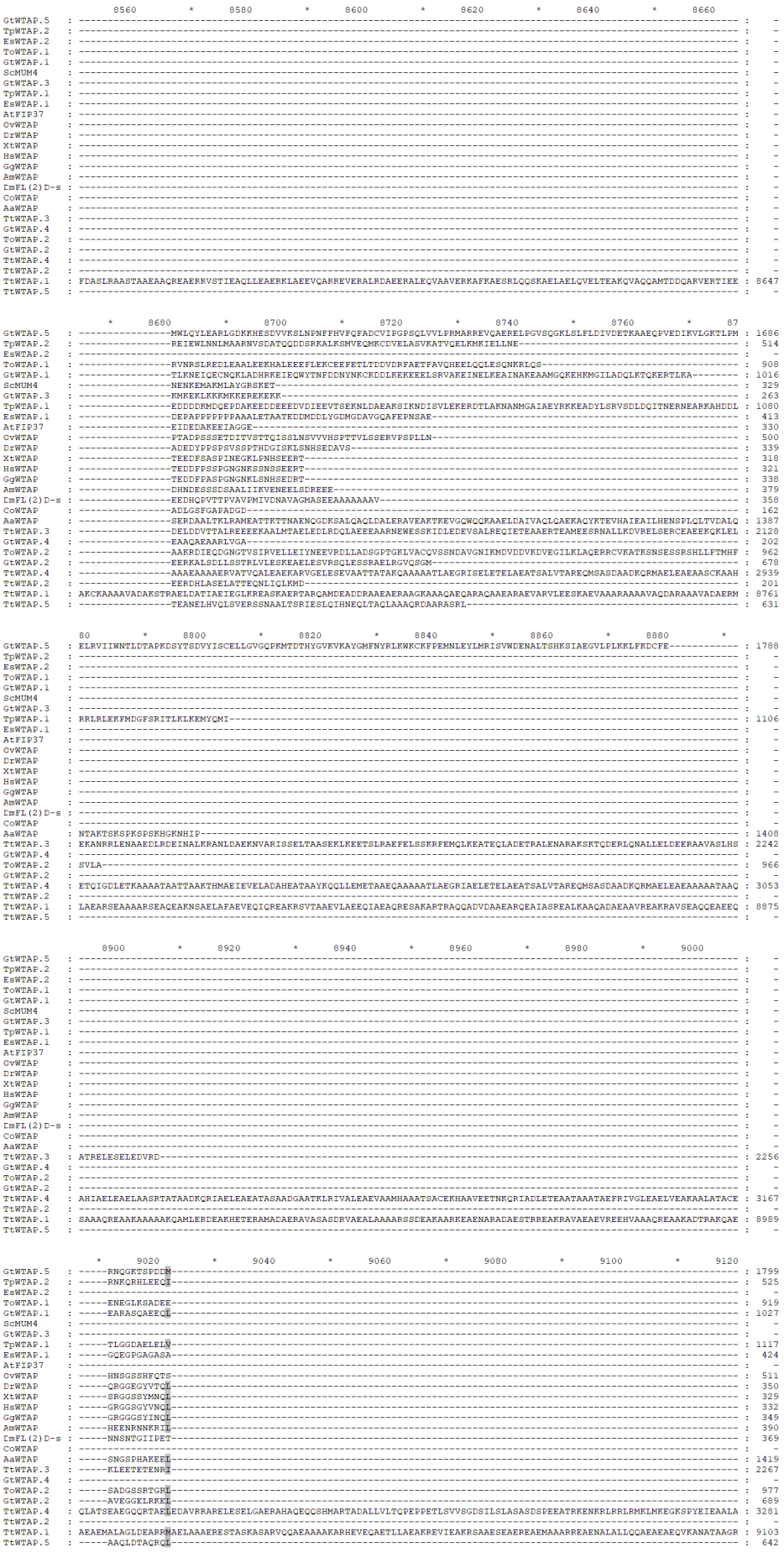


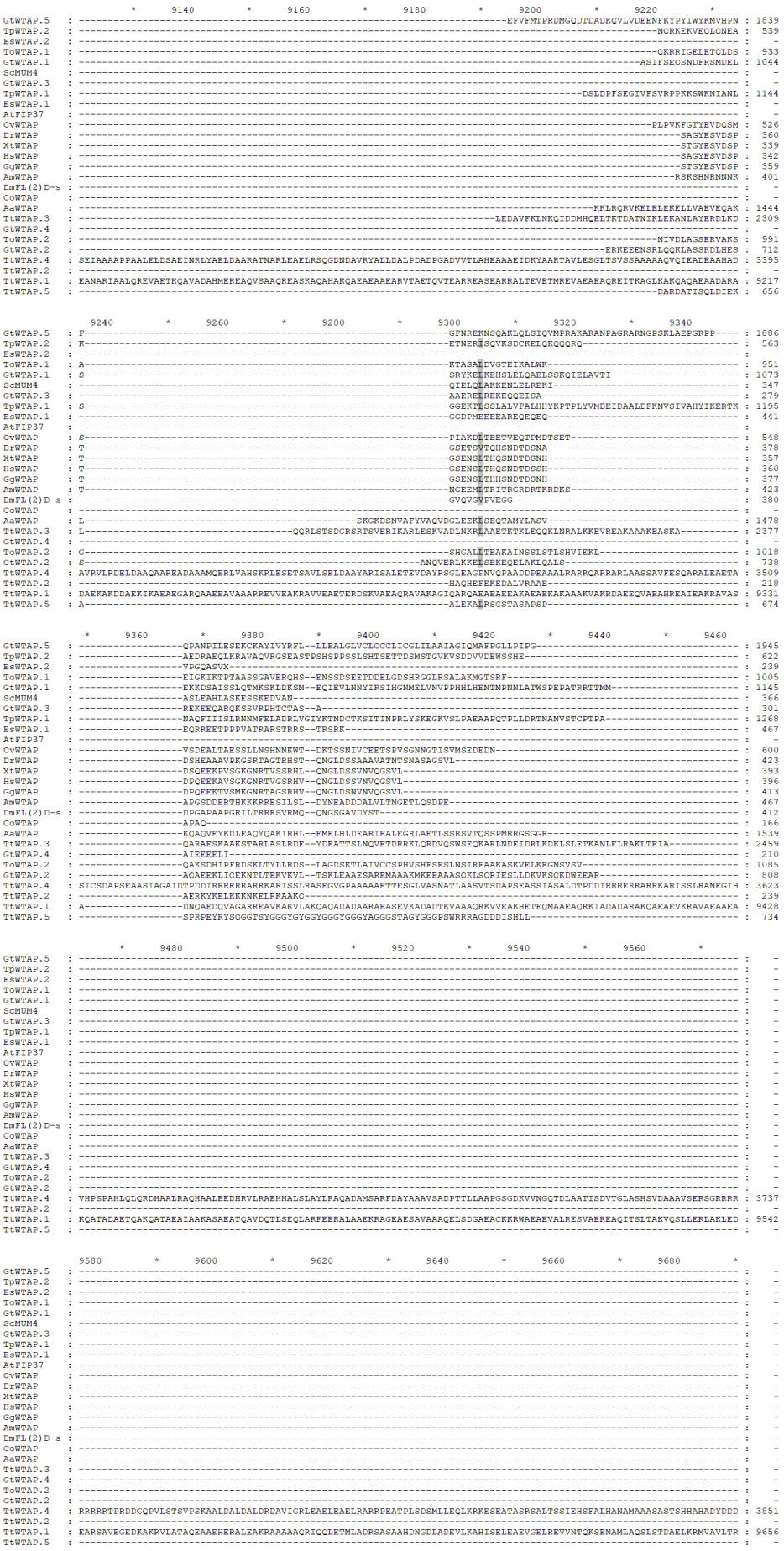













9700 * 9720 * 9740 * 9760 * 9780 * 9800
GtWTAP.5 : - :
TpWTAP.2 : - :
EsWTAP.2 : - :
ToWTAP.1 : - :
GtWTAP.1 : - :
ScMUM4 : - :
GtWTAP.3 : - :
TpWTAP.1 : - :
EsWTAP.1 : - :
AtFIP37 : - :
OvWTAP : - :
DrWTAP : - :
XtWTAP : - :
HsWTAP : - :
GgWTAP : - :
AmWTAP : - :
DmFL(2)D-s : - :
CoWTAP : - :
AaWTAP : - :
TtWTAP.3 : - :
GtWTAP.4 : - :
ToWTAP.2 : - :
GtWTAP.2 : - :
TtWTAP.4 : DNVFGASANVEREWRRDGL- : 3870
TtWTAP.2 : - :
TtWTAP.1 : ALADARAEAKAEAAKAVRLTSALHGLNSRVEVMIADNARLKARNDELAAEVADDSGALKRMAHALSLRLEDALDAKEAAEAKLLAPSTTHVLASSSEAAAALAEAQMEMAALVD : 9770
TtWTAP.5 : - :
* 9820 * 9840 * 9860 * 9880 * 9900 *
GtWTAP.5 : - :
TpWTAP.2 : - :
EsWTAP.2 : - :
ToWTAP.1 : - :
GtWTAP.1 : - :
ScMUM4 : - :
GtWTAP.3 : - :
TpWTAP.1 : - :
EsWTAP.1 : - :
AtFIP37 : - :
OvWTAP : - :
DrWTAP : - :
XtWTAP : - :
HsWTAP : - :
GgWTAP : - :
AmWTAP : - :
DmFL(2)D-s : - :
CoWTAP : - :
AaWTAP : - :
TtWTAP.3 : - :
GtWTAP.4 : - :
ToWTAP.2 : - :
GtWTAP.2 : - :
TtWTAP.4 : - :
TtWTAP.2 : - :
TtWTAP.1 : EREGLLEQLAELEHRQRAGGLVDVLVARLEAIHAATARLAVSEGARAWASEASTLAGEVTGELAALTEQEASLAAAHEAFARRSGIMDDAAREQVLAILQQVSRGV : 9876
TtWTAP.5 : - :

Ensembl IDs for putative m^{6}Am binding proteins

Table of Ensembl IDs with gene names for the putatively m^{6}Am methylated transcripts from the foetal brain m^{6}A sequencing.

Ensembl ID	Gene name	Full Gene Name
ENSG00000139631	CSAD	cysteine sulfinic acid decarboxylase
ENSG00000214447	CCDC103	coiled-coil domain containing 103
ENSG00000134318	ROCK2	Rho associated coiled-coil containing protein kinase 2
ENSG00000196998	WDR45	WD repeat domain 45
ENSG00000136490	LIMD2	LIM domain containing 2
ENSG00000165186	PTCHD1	patched domain containing 1
ENSG00000166974	MAPRE2	microtubule associated protein RP/EB family member 2
ENSG00000134121	CHL1	cell adhesion molecule L1 like
ENSG00000268471	MIR4453	microRNA 4453
ENSG00000140650	PMM2	phosphomannomutase 2
ENSG00000204673	AKT1S1	AKT1 substrate 1
ENSG00000229117	RPL41	ribosomal protein L41
ENSG00000163798	SLC4A1AP	solute carrier family 4 member 1 adaptor protein
ENSG00000151458	ANKRD50	ankyrin repeat domain 50
ENSG00000110651	CD81	CD81 molecule
ENSG00000277957	SENP3	SUMO1/sentrin/SMT3 specific peptidase 3
ENSG00000139990	DCAF5	DDB1 and CUL4 associated factor 5
ENSG00000196290	NIF3L1	NGG1 interacting factor 3 like 1
ENSG00000213719	CLIC1	chloride intracellular channel 1
ENSG00000149582	TMEM25	transmembrane protein 25
ENSG00000108107	MIR6805	microRNA 6805
ENSG00000163975	MELTF	melanotransferrin
ENSG00000151327	LOC101927178	uncharacterized LOC101927178
ENSG00000167967	E4F1	E4F transcription factor 1
ENSG00000149480	MTA2	metastasis associated 1 family member 2
ENSG00000153187	HNRNPU	heterogeneous nuclear ribonucleoprotein U
ENSG00000232838	Pet117	PET117 homolog
ENSG00000198300	PEG3	paternally expressed 3
ENSG00000174939	ASPHD1	aspartate beta-hydroxylase domain containing 1
ENSG00000278970	HEIH	hepatocellular carcinoma up-regulated EZH2-associated long non-coding RNA
ENSG00000128191	MIR1306	microRNA 1306
ENSG00000143420	ENSA	endosulfine alpha
ENSG00000161904	LEMD2	LEM domain containing 2
ENSG00000128596	CCDC136	coiled-coil domain containing 136
ENSG00000128594	LRRC4	leucine rich repeat containing 4

ENSG00000065883	CDK13	cyclin dependent kinase 13
ENSG00000132475	MIR4738	microRNA 4738
ENSG00000055130	CUL1	cullin 1
ENSG00000187555	USP7	ubiquitin specific peptidase 7
ENSG00000179632	MAF1	MAF1 homolog, negative regulator of RNA polymerase III
ENSG00000083290	ULK2	unc-51 like autophagy activating kinase 2
ENSG00000069849	ATP1B3	ATPase Na+/K+ transporting subunit beta 3
ENSG00000132718	SYT11	synaptotagmin 11
ENSG00000161013	MGAT4B	mannosyl -glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme B
ENSG00000154710	RABGEF1	RAB guanine nucleotide exchange factor 1
ENSG00000130338	TULP4	tubby like protein 4
ENSG00000171862	PTEN	phosphatase and tensin homolog
ENSG00000184402	SS18L1	SS18L1, nBAF chromatin remodeling complex subunit
ENSG00000187624	C17orf97	chromosome 17 open reading frame 97
ENSG00000116221	MRPL37	mitochondrial ribosomal protein L37
ENSG00000185278	ZBTB37	zinc finger and BTB domain containing 37
ENSG00000116903	EXOC8	exocyst complex component 8
ENSG00000234857	HNRNPUL2-BSCL2	HNRNPUL2-BSCL2 readthrough
ENSG00000169169	CPT1C	carnitine palmitoyltransferase 1C
ENSG00000090924	PLEKHG2	pleckstrin homology and RhoGEF domain containing G2
ENSG00000115053	NCL	nucleolin
ENSG00000138814	PPP3CA	protein phosphatase 3 catalytic subunit alpha
ENSG00000163939	PBRM1	polybromo 1
ENSG00000176101	SSNA1	SS nuclear autoantigen 1
ENSG00000145016	MIR922	microRNA 922
ENSG00000063587	LOC105373378	uncharacterized LOC105373378
ENSG00000246627	CACNA1C-AS1	CACNA1C antisense RNA 1
ENSG00000137872	SEMA6D	semaphorin 6D
ENSG00000011566	MAP4K3	mitogen-activated protein kinase kinase kinase kinase 3
ENSG00000100697	DICER1	dicer 1, ribonuclease III
ENSG00000264589	MAPT-AS1	MAPT antisense RNA 1
ENSG00000103502	CDIPT	CDP-diacylglycerol--inositol 3-phosphatidyltransferase
ENSG00000092964	DPYSL2	dihydropyrimidinase like 2
ENSG00000170365	SMAD1	SMAD family member 1
ENSG00000172216	CEBPB	CCAAT/enhancer binding protein beta
ENSG00000112473	SLC39A7	solute carrier family 39 member 7
ENSG00000163412	EIF4E3	eukaryotic translation initiation factor 4E family member 3
ENSG00000148516	ZEB1	zinc finger E-box binding homeobox 1
ENSG00000056097	ZFR	zinc finger RNA binding protein

ENSG00000047346	FAM214A	family with sequence similarity 214 member A
ENSG00000249673	NOP14-AS1	NOP14 antisense RNA 1
ENSG00000182004	SNRPE	small nuclear ribonucleoprotein polypeptide E
ENSG00000149256	TENM4	teneurin transmembrane protein 4
ENSG00000176022	B3GALT6	beta-1,3-galactosyltransferase 6
ENSG00000100836	PABPN1	poly binding protein nuclear 1
ENSG00000139211	AMIGO2	adhesion molecule with Ig like domain 2
ENSG00000173457	PPP1R14B	protein phosphatase 1 regulatory inhibitor subunit 14B
ENSG00000124006	OBSL1	obscurin like 1
ENSG00000173452	TMEM196	transmembrane protein 196
ENSG00000213190	MLLT11	myeloid/lymphoid or mixed-lineage leukemia; translocated to, 11
ENSG00000185787	MORF4L1	mortality factor 4 like 1
ENSG00000078328	RBFOX1	RNA binding protein, fox-1 homolog 1
ENSG00000176986	SEC24C	SEC24 homolog C, COPII coat complex component
ENSG00000226137	BAIAP2-AS1	BAIAP2 antisense RNA 1
ENSG00000185129	PURA	purine rich element binding protein A
ENSG00000185127	C6orf120	chromosome 6 open reading frame 120
ENSG00000140199	SLC12A6	solute carrier family 12 member 6
ENSG00000187514	MIR1244-1	microRNA 1244-1
ENSG00000163536	SERPINI1	serpin family I member 1
ENSG00000138650	PCDH10	protocadherin 10
ENSG00000162959	MEMO1	mediator of cell motility 1
ENSG00000125445	MRPS7	mitochondrial ribosomal protein S7
ENSG00000115307	AUP1	ancient ubiquitous protein 1
ENSG00000065609	SNAP91	synaptosome associated protein 91
ENSG00000254102	LOC401463	uncharacterized LOC401463
ENSG00000243232	PCDHAC2	protocadherin alpha subfamily C, 2
ENSG00000156508	EEF1A1	eukaryotic translation elongation factor 1 alpha 1
ENSG00000065427	KARS	lysyl-tRNA synthetase
ENSG00000115839	RAB3GAP1	RAB3 GTPase activating protein catalytic subunit 1
ENSG00000203875	SNHG5	small nucleolar RNA host gene 5
ENSG00000142444	TIMM29	translocase of inner mitochondrial membrane 29
ENSG00000152784	PRDM8	PR/SET domain 8
ENSG00000114062	UBE3A	ubiquitin protein ligase E3A
ENSG00000083812	ZNF324	zinc finger protein 324
ENSG00000112658	SRF	serum response factor
ENSG00000153815	CMIP	c-Maf inducing protein
ENSG00000125352	RNF113A	ring finger protein 113A
ENSG00000175334	BANF1	barrier to autointegration factor 1
ENSG00000148671	ADIRF	adipogenesis regulatory factor
ENSG00000145242	EPHA5	EPH receptor A5
ENSG00000113712	CSNK1A1	casein kinase 1 alpha 1

ENSG00000075856	SART3	squamous cell carcinoma antigen recognized by T-cells 3
ENSG00000185551	MIR1469	microRNA 1469
ENSG00000165516	KLHDC2	kelch domain containing 2
ENSG00000132694	ARHGEF11	Rho guanine nucleotide exchange factor 11
ENSG00000267322	SNHG22	small nucleolar RNA host gene 22
ENSG00000172985	SH3RF3	SH3 domain containing ring finger 3
ENSG00000173011	TADA2B	transcriptional adaptor 2B
ENSG00000169764	UGP2	UDP-glucose pyrophosphorylase 2
ENSG00000243667	WDR92	WD repeat domain 92
ENSG00000110711	AIP	aryl hydrocarbon receptor interacting protein
ENSG00000135045	C9orf40	chromosome 9 open reading frame 40
ENSG00000176463	SLCO3A1	solute carrier organic anion transporter family member 3A1
ENSG00000171791	BCL2	BCL2, apoptosis regulator
ENSG00000142252	GEMIN7	gem nuclear organelle associated protein 7
ENSG00000176055	MBLAC2	metallo-beta-lactamase domain containing 2
ENSG00000235875	ARHGEF7-AS2	ARHGEF7 antisense RNA 2
ENSG00000100911	MIR7703	microRNA 7703
ENSG00000175073	VCPIP1	valosin containing protein interacting protein 1
ENSG00000198356	ASNA1	arsA arsenite transporter, ATP-binding, homolog 1
ENSG00000143222	UFC1	ubiquitin-fold modifier conjugating enzyme 1
ENSG00000160714	UBE2Q1	ubiquitin conjugating enzyme E2 Q1
ENSG00000251562	MALAT1	metastasis associated lung adenocarcinoma transcript 1
ENSG00000106829	TLE4	transducin like enhancer of split 4
ENSG00000187257	RSBN1L	round spermatid basic protein 1 like
ENSG00000109846	CRYAB	crystallin alpha B
ENSG00000236144	TMEM147-AS1	TMEM147 antisense RNA 1
ENSG00000182446	NPLOC4	NPL4 homolog, ubiquitin recognition factor
ENSG00000172260	NEGR1	neuronal growth regulator 1
ENSG00000112333	NR2E1	nuclear receptor subfamily 2 group E member 1
ENSG00000107372	ZFAND5	zinc finger AN1-type containing 5
ENSG00000198783	ZNF830	zinc finger protein 830
ENSG00000185883	ATP6V0C	ATPase H+ transporting V0 subunit c
ENSG00000172432	GTPBP2	GTP binding protein 2
ENSG00000137266	SLC22A23	solute carrier family 22 member 23
ENSG00000169683	LRRC45	leucine rich repeat containing 45
ENSG00000118263	KLF7	Kruppel like factor 7
ENSG00000268439	EMG1	EMG1, N1-specific pseudouridine methyltransferase
ENSG00000182310	SPACA6	sperm acrosome associated 6
ENSG00000148303	RPL7A	ribosomal protein L7a
ENSG00000160216	AGPAT3	1-acylglycerol-3-phosphate O-acyltransferase 3

ENSG00000132466	ANKRD17	ankyrin repeat domain 17
ENSG00000145592	RPL37	ribosomal protein L37
ENSG00000166173	LARP6	La ribonucleoprotein domain family member 6
ENSG00000143303	RRNAD1	ribosomal RNA adenine dimethylase domain containing 1
ENSG00000139154	AEBP2	AE binding protein 2
ENSG00000136930	PSMB7	proteasome subunit beta 7
ENSG00000268592	RAET1E-AS1	RAET1E antisense RNA 1
ENSG00000112972	HMGCS1	3-hydroxy-3-methylglutaryl-CoA synthase 1
ENSG00000170037	CNTROB	centrobin, centriole duplication and spindle assembly protein
ENSG00000146731	CCT6A	chaperonin containing TCP1 subunit 6A
ENSG00000170445	HARS	histidyl-tRNA synthetase
ENSG00000068097	HEATR6	HEAT repeat containing 6
ENSG00000010810	FYN	FYN proto-oncogene, Src family tyrosine kinase
ENSG00000169992	NLGN2	neuroligin 2
ENSG00000132002	DNAJB1	DnaJ heat shock protein family member B1
ENSG00000022267	FHL1	four and a half LIM domains 1
ENSG00000167131	CCDC103	coiled-coil domain containing 103
ENSG00000166483	WEE1	WEE1 G2 checkpoint kinase
ENSG00000033122	LRRC7	leucine rich repeat containing 7
ENSG00000135932	CAB39	calcium binding protein 39
ENSG00000105323	HNRNPUL1	heterogeneous nuclear ribonucleoprotein U like 1
ENSG00000206190	ATP10A	ATPase phospholipid transporting 10A
ENSG00000182087	TMEM259	transmembrane protein 259
ENSG00000173548	snx33	sorting nexin 33
ENSG00000182648	LINC01006	long intergenic non-protein coding RNA 1006
ENSG00000131584	acap3	ArfGAP with coiled-coil, ankyrin repeat and PH domains 3
ENSG00000112183	RBM24	RNA binding motif protein 24
ENSG00000213465	ARL2	ADP ribosylation factor like GTPase 2
ENSG00000129351	ILF3	interleukin enhancer binding factor 3
ENSG00000101746	NOL4	nucleolar protein 4
ENSG00000135930	EIF4E2	eukaryotic translation initiation factor 4E family member 2
ENSG00000063244	U2AF2	U2 small nuclear RNA auxiliary factor 2
ENSG00000140992	PDPK1	3-phosphoinositide dependent protein kinase 1
ENSG00000026025	VIM	vimentin
ENSG00000146457	WTAP	Wilms tumor 1 associated protein
ENSG00000256537	SMIM10L1	small integral membrane protein 10 like 1
ENSG00000225206	MIR2682	microRNA 2682
ENSG00000156103	MMP16	matrix metallopeptidase 16
ENSG00000103253	HAGHL	hydroxyacylglutathione hydrolase-like
ENSG00000081692	JMJD4	jumonji domain containing 4
ENSG00000134278	SPIRE1	spire type actin nucleation factor 1

ENSG00000175928	LRRN1	leucine rich repeat neuronal 1
ENSG00000196367	TRRAP	transformation/transcription domain associated protein
ENSG00000198740	ZNF652	zinc finger protein 652
ENSG00000156650	KAT6B	lysine acetyltransferase 6B
ENSG00000133019	CHRM3	cholinergic receptor muscarinic 3
ENSG00000132635	PCED1A	PC-esterase domain containing 1A
ENSG00000196361	ELAVL3	ELAV like RNA binding protein 3
ENSG00000214226	c17orf67	chromosome 17 open reading frame 67
ENSG00000196365	LONP1	lon peptidase 1, mitochondrial
ENSG00000135298	ADGRB3	adhesion G protein-coupled receptor B3

Table of Ensembl IDs with gene names for the putatively m^6Am methylated transcripts from the grey matter m^6A sequencing.

Ensembl ID	Gene name	Full Gene Name
ENSG00000109971	HSPA8	heat shock protein family A member 8
ENSG00000022567	SLC45A4	solute carrier family 45 member 4
ENSG00000168077	SCARA3	scavenger receptor class A member 3
ENSG00000171126	KCNG3	potassium voltage-gated channel modifier subfamily G member 3
ENSG00000139438	FAM222A	family with sequence similarity 222 member A
ENSG00000214447	CCDC103	coiled-coil domain containing 103
ENSG00000125841	NRSN2	neurensin 2
ENSG00000007202	KIAA0100	KIAA0100
ENSG00000035681	NSMAF	neutral sphingomyelinase activation associated factor
ENSG00000215041	neurl4	neuralized E3 ubiquitin protein ligase 4
ENSG00000163155	LYSMD1	LysM domain containing 1
ENSG00000170185	USP38	ubiquitin specific peptidase 38
ENSG00000168282	mgat2	mannosyl -glycoprotein beta-1,2-N-acetylglucosaminyltransferase
ENSG00000120253	NUP43	nucleoporin 43
ENSG00000219607	PPP1R3G	protein phosphatase 1 regulatory subunit 3G
ENSG00000163704	PRRT3	proline rich transmembrane protein 3
ENSG00000214655	ZSWIM8	zinc finger SWIM-type containing 8
ENSG00000229117	RPL41	ribosomal protein L41
ENSG00000272916	NDST2	N-deacetylase and N-sulfotransferase 2
ENSG00000163798	SLC4A1AP	solute carrier family 4 member 1 adaptor protein
ENSG00000125651	GTF2F1	general transcription factor IIF subunit 1
ENSG00000145242	EPHA5	EPH receptor A5
ENSG00000117505	DR1	down-regulator of transcription 1
ENSG00000173786	CNP	2',3'-cyclic nucleotide 3' phosphodiesterase
ENSG00000065970	FOXJ2	forkhead box J2
ENSG00000117500	TMED5	transmembrane p24 trafficking protein 5
ENSG00000167371	PRRT2	proline rich transmembrane protein 2
ENSG00000165516	KLHDC2	kelch domain containing 2
ENSG00000143153	ATP1B1	ATPase Na+/K+ transporting subunit beta 1
ENSG00000273154	ZGPAT	zinc finger CCCH-type and G-patch domain containing
ENSG00000134323	MYCN	v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog
ENSG00000148948	LRRC4C	leucine rich repeat containing 4C
ENSG00000100906	NFKBIA	NFKB inhibitor alpha
ENSG00000149403	GRIK4	glutamate ionotropic receptor kainate type subunit 4
ENSG00000077782	FGFR1	fibroblast growth factor receptor 1
ENSG00000149480	MTA2	metastasis associated 1 family member 2
ENSG00000131773	KHDRBS3	KH RNA binding domain containing, signal transduction associated 3

ENSG00000177272	KCNA3	potassium voltage-gated channel subfamily A member 3
ENSG00000186212	SOWAHB	sosondowah ankyrin repeat domain family member B
ENSG00000166068	SPRED1	sprouty related EVH1 domain containing 1
ENSG00000204237	oxld1	oxidoreductase like domain containing 1
ENSG00000237187	NR2F1-AS1	NR2F1 antisense RNA 1
ENSG00000177707	NECTIN3	nectin cell adhesion molecule 3
ENSG00000231721	LINC-PINT	long intergenic non-protein coding RNA, p53 induced transcript
ENSG00000127564	pkmyt1	protein kinase, membrane associated tyrosine/threonine 1
ENSG00000070610	GBA2	glucosylceramidase beta 2
ENSG00000103353	UBFD1	ubiquitin family domain containing 1
ENSG00000205808	plpp6	phospholipid phosphatase 6
ENSG00000138080	EMILIN1	elastin microfibril interfacer 1
ENSG00000143420	ENSA	endosulfine alpha
ENSG00000169006	NTSR2	neurotensin receptor 2
ENSG00000100664	EIF5	eukaryotic translation initiation factor 5
ENSG00000084234	APLP2	amyloid beta precursor like protein 2
ENSG00000176788	BASP1	brain abundant membrane attached signal protein 1
ENSG00000155636	RBM45	RNA binding motif protein 45
ENSG00000128594	LRRC4	leucine rich repeat containing 4
ENSG00000122390	NAA60	N-acetyltransferase 60, NatF catalytic subunit
ENSG00000165795	MIR6717	microRNA 6717
ENSG00000179632	MAF1	MAF1 homolog, negative regulator of RNA polymerase III
ENSG00000262304	shpk	sedoheptulokinase
ENSG00000278540	ACACA	acetyl-CoA carboxylase alpha
ENSG00000117152	RGS4	regulator of G-protein signaling 4
ENSG00000069849	ATP1B3	ATPase Na+/K+ transporting subunit beta 3
ENSG00000132718	SYT11	synaptotagmin 11
ENSG00000154710	RABGEF1	RAB guanine nucleotide exchange factor 1
ENSG00000272269	LOC105374952	uncharacterized LOC105374952
ENSG00000077147	TM9SF3	transmembrane 9 superfamily member 3
ENSG00000141750	STAC2	SH3 and cysteine rich domain 2
ENSG00000115241	PPM1G	protein phosphatase, Mg2+/Mn2+ dependent 1G
ENSG00000105679	GAPDHS	glyceraldehyde-3-phosphate dehydrogenase, spermatogenic
ENSG00000148303	RPL7A	ribosomal protein L7a
ENSG00000234745	HLA-B	major histocompatibility complex, class I, B
ENSG00000164620	RELL2	RELT like 2
ENSG00000162592	CCDC27	coiled-coil domain containing 27
ENSG00000108551	RASD1	ras related dexamethasone induced 1
ENSG00000167123	CERCAM	cerebral endothelial cell adhesion molecule
ENSG00000204569	PPP1R10	protein phosphatase 1 regulatory subunit 10
ENSG00000125285	SOX21	SRY-box 21

ENSG00000092421	SEMA6A	semaphorin 6A
ENSG00000240303	ACAD11	acyl-CoA dehydrogenase family member 11
ENSG00000005884	ITGA3	integrin subunit alpha 3
ENSG00000091592	NLRP1	NLR family pyrin domain containing 1
ENSG00000074800	ENO1	enolase 1
ENSG00000059915	PSD	pleckstrin and Sec7 domain containing
ENSG00000176209	SMIM19	small integral membrane protein 19
ENSG00000132002	DNAJB1	DnaJ heat shock protein family member B1
ENSG00000078114	NEBL	nebulette
ENSG00000048540	LMO3	LIM domain only 3
ENSG00000246627	CACNA1C-AS1	CACNA1C antisense RNA 1
ENSG00000179104	TMTC2	transmembrane and tetratricopeptide repeat containing 2
ENSG00000135932	CAB39	calcium binding protein 39
ENSG00000168092	PAFAH1B2	platelet activating factor acetylhydrolase 1b catalytic subunit 2
ENSG00000169242	EFNA1	ephrin A1
ENSG00000105323	HNRNPUL1	heterogeneous nuclear ribonucleoprotein U like 1
ENSG00000121579	NAA50	N-acetyltransferase 50, NatE catalytic subunit
ENSG00000145335	snca	synuclein alpha
ENSG00000172216	CEBPB	CCAAT/enhancer binding protein beta
ENSG00000112473	SLC39A7	solute carrier family 39 member 7
ENSG00000163412	EIF4E3	eukaryotic translation initiation factor 4E family member 3
ENSG00000117385	P3H1	prolyl 3-hydroxylase 1
ENSG00000145743	FBXL17	F-box and leucine rich repeat protein 17
ENSG00000172733	PURG	purine rich element binding protein G
ENSG00000202198	rn7sk	RNA, 7SK small nuclear
ENSG00000178568	ERBB4	erb-b2 receptor tyrosine kinase 4
ENSG00000136854	STXBP1	syntaxin binding protein 1
ENSG00000197114	ZGPAT	zinc finger CCCH-type and G-patch domain containing
ENSG00000280789	PAGR1	PAXIP1 associated glutamate rich protein 1
ENSG00000263753	LINC00667	long intergenic non-protein coding RNA 667
ENSG00000139211	AMIGO2	adhesion molecule with Ig like domain 2
ENSG00000102316	MAGED2	MAGE family member D2
ENSG00000174473	GALNTL6	polypeptide N-acetylgalactosaminyltransferase-like 6
ENSG00000109670	FBXW7	F-box and WD repeat domain containing 7
ENSG00000122557	HERPUD2	HERPUD family member 2
ENSG00000078328	RBFOX1	RNA binding protein, fox-1 homolog 1
ENSG00000156860	FBRS	fibrosin
ENSG00000081692	JMJD4	jumonji domain containing 4
ENSG00000071553	ATP6AP1	ATPase H+ transporting accessory protein 1
ENSG00000158773	USF1	upstream transcription factor 1
ENSG00000132635	PCED1A	PC-esterase domain containing 1A

Table of Ensembl IDs with gene names for the putatively m^6Am methylated transcripts from the white matter m^6A sequencing.

Ensembl ID	Gene name	Full Gene Name
ENSG00000138650	PCDH10	protocadherin 10
ENSG00000109971	HSPA8	heat shock protein family A member 8
ENSG00000022567	SLC45A4	solute carrier family 45 member 4
ENSG00000154710	RABGEF1	RAB guanine nucleotide exchange factor 1
ENSG00000158528	PPP1R9A	protein phosphatase 1 regulatory subunit 9A
ENSG00000130338	TULP4	tubby like protein 4
ENSG00000140043	PTGR2	prostaglandin reductase 2
ENSG00000115307	AUP1	ancient ubiquitous protein 1
ENSG00000184402	SS18L1	SS18L1, nBAF chromatin remodeling complex subunit
ENSG00000136068	FLNB	filamin B
ENSG00000136930	PSMB7	proteasome subunit beta 7
ENSG00000076053	RBM7	RNA binding motif protein 7
ENSG00000204569	PPP1R10	protein phosphatase 1 regulatory subunit 10
ENSG00000007202	KIAA0100	KIAA0100
ENSG00000156976	MIR1248	microRNA 1248
ENSG00000092421	SEMA6A	semaphorin 6A
ENSG00000170185	USP38	ubiquitin specific peptidase 38
ENSG00000100811	MIR6764	microRNA 6764
ENSG00000116903	EXOC8	exocyst complex component 8
ENSG00000264058	krt222	keratin 222
ENSG00000125817	CENPB	centromere protein B
ENSG00000133069	TMCC2	transmembrane and coiled-coil domain family 2
ENSG00000165119	HNRNPK	heterogeneous nuclear ribonucleoprotein K
ENSG00000246731	MGC16275	uncharacterized protein MGC16275
ENSG00000229117	RPL41	ribosomal protein L41
ENSG00000170445	HARS	histidyl-tRNA synthetase
ENSG00000238197	PAXBP1-AS1	PAXBP1 antisense RNA 1
ENSG00000165566	AMER2	APC membrane recruitment protein 2
ENSG00000238741	SCARNA7	small Cajal body-specific RNA 7
ENSG00000144306	SCRN3	secernin 3
ENSG00000145242	EPHA5	EPH receptor A5
ENSG00000115053	NCL	nucleolin
ENSG00000173786	CNP	2',3'-cyclic nucleotide 3' phosphodiesterase
ENSG00000013293	SLC7A14	solute carrier family 7 member 14
ENSG00000146373	RNF217	ring finger protein 217
ENSG00000207955	MIR219A2	microRNA 219a-2
ENSG00000184984	CHRM5	cholinergic receptor muscarinic 5
ENSG00000146966	DENND2A	DENN domain containing 2A
ENSG00000134590	FAM127A	family with sequence similarity 127 member A

ENSG00000048540	LMO3	LIM domain only 3
ENSG00000137871	LOC145783	uncharacterized LOC145783
ENSG00000234456	MAGI2-AS3	MAGI2 antisense RNA 3
ENSG00000172458	IL17D	interleukin 17D
ENSG00000178467	P4HTM	prolyl 4-hydroxylase, transmembrane
ENSG00000275778	PRH1-PRR4	PRH1-PRR4 readthrough
ENSG00000173548	snx33	sorting nexin 33
ENSG00000187951	LOC100288637	OTU deubiquitinase 7A pseudogene
ENSG00000151327	LOC101927178	uncharacterized LOC101927178
ENSG00000104722	NEFM	neurofilament, medium polypeptide
ENSG00000172216	CEBPB	CCAAT/enhancer binding protein beta
ENSG00000250846	EPHA5-AS1	EPHA5 antisense RNA 1
ENSG00000153187	HNRNPU	heterogeneous nuclear ribonucleoprotein U
ENSG00000160752	FDPS	farnesyl diphosphate synthase
ENSG00000187189	TSPYL4	TSPY like 4
ENSG00000256087	znf432	zinc finger protein 432
ENSG00000197114	ZGPAT	zinc finger CCCH-type and G-patch domain containing
ENSG00000139651	ZNF740	zinc finger protein 740
ENSG00000114867	EIF4G1	eukaryotic translation initiation factor 4 gamma 1
ENSG00000198300	PEG3	paternally expressed 3
ENSG00000164161	HHIP	hedgehog interacting protein
ENSG00000280789	PAGR1	PAXIP1 associated glutamate rich protein 1
ENSG00000187391	MAGI2	membrane associated guanylate kinase, WW and PDZ domain containing 2
ENSG00000236104	ZBTB22	zinc finger and BTB domain containing 22
ENSG00000162944	RFTN2	raftlin family member 2
ENSG00000070610	GBA2	glucosylceramidase beta 2
ENSG00000251562	MALAT1	metastasis associated lung adenocarcinoma transcript 1
ENSG00000143753	DEGS1	delta 4-desaturase, sphingolipid 1
ENSG00000128656	CHN1	chimerin 1
ENSG00000185787	MORF4L1	mortality factor 4 like 1
ENSG00000233705	SLC26A4-AS1	SLC26A4 antisense RNA 1
ENSG00000246067	RAB30-AS1	RAB30 antisense RNA 1
ENSG00000184719	RNLS	renalase, FAD dependent amine oxidase
ENSG00000065883	CDK13	cyclin dependent kinase 13
ENSG00000103111	MON1B	MON1 homolog B, secretory trafficking associated
ENSG00000122390	NAA60	N-acetyltransferase 60, NatF catalytic subunit
ENSG00000137497	NUMA1	nuclear mitotic apparatus protein 1
ENSG00000100167	SEPT3	septin 3
ENSG00000152942	RAD17	RAD17 checkpoint clamp loader component
ENSG00000163104	SMARCAD1	SWI/SNF-related, matrix-associated actin-dependent regulator of chromatin, subfamily a, containing DEAD/H box 1
ENSG00000278540	ACACA	acetyl-CoA carboxylase alpha

ENSG00000196220	SRGAP3	SLIT-ROBO Rho GTPase activating protein 3
ENSG00000080603	SRCAP	Snf2 related CREBBP activator protein
ENSG00000187514	MIR1244-1	microRNA 1244-1
ENSG00000172534	HCFC1	host cell factor C1
ENSG00000173141	MRPL57	mitochondrial ribosomal protein L57
ENSG00000090238	YPEL3	yippee like 3
ENSG00000132718	SYT11	synaptotagmin 11
ENSG00000166862	CACNG2	calcium voltage-gated channel auxiliary subunit gamma 2
ENSG00000166012	MIR1304	microRNA 1304
ENSG00000185883	ATP6V0C	ATPase H+ transporting V0 subunit c