Investigation of the role of tau gene transcriptional regulation in neurodegeneration

Reta Lila Weston Institute of Neurological Studies
Institute of Neurology
University College London (UCL)
University of London

December 2011

Juan Fidel Anaya
PhD

I, Juan Fidel Anaya, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Although the tau gene (*MAPT)* is not mutated in the majority of tauopathies, there are pathological disturbances in tau-isoform homeostasis. Investigation into *MAPT* linkage disequilibrium and haplotype structure, and the common variation of *MAPT* associated with increased risk of progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) identified a possible basis for the disturbances in tau-isoform homeostasis. Of multiple haplotypes in the H1 clade, the H1c sub-haplotype drives the association with PSP and CBD, suggesting that it carries the pathogenic variation leading to increased risk of these largely sporadic disorders. The aim of this project was to investigate the regulation of the transcription of *MAPT* by first identifying transcription factors (TFs) and TF complexes that bind to the *MAPT* core promoter and its conserved regulatory domains. The specific aim was to determine if differential binding of TFs and TF complexes to the allelic variants of the SNPs in the *MAPT* promoter region forms the basis of the allele-specific differences in *MAPT* transcription and splicing that we have observed *in vitro* and *in vivo*.

Results from electrophoretic mobility shift assays, pull-down experiments and mass spectrometry had shown potentially novel proteins binding to the disease-associated SNP under investigation. Bioinformatic analyses helped to stratify these proteins according to possible relevance. While immunoblotting alongside co-transfection of siRNA and luciferase promoter construct experiments had shown initial evidence of a complicated relationship between *cis* and *trans* factors within the tau promoter region. These findings were novel and they provide new insight into the regulation of one of the most important genes in neurodegeneration. The identification of *MAPT* transcriptional machinery would provide further insight into the role of *MAPT* in neurodegeneration and the basis for therapeutic intervention.

(280 words)

Contents

i. Figures

ii. Acknowledgements

I would like to thank my supervisor Dr. Rohan de Silva and the Reta Lila Weston Institute of Neurology, UCL for taking me onboard to undertake this PhD. It would not have been possible without the generosity of the Alzheimer's Research UK and UCL Overseas Research Student (ORS) award for their funding towards making this research project possible. All the staff at the Reta Lila Weston Institute of Neurology has been very supportive and patient with me and their advice has been invaluable. Special mention also goes to the staff at the Institute of Neurology, Queen Square and the core proteomics unit of the UCL Institute of Child Health for allowing me access to their facility which was invaluable.

iii. List of publications

1. **Anaya F.**, Simone R., Kay V., Hondhamuni G., Lees A., de Silva R. (2011) Allele-specific regulation of tau gene transcription by promoter-binding regulatory proteins (in preparation)
2. **Anaya F.**, Lees A. and de Silva R. (2011) Tau gene promoter rs242557 and allele-specific protein binding. Translational Neuroscience **2(2)** 176-205
3. Vandrovcova J.*, **Anaya F.***, Kay V., Lees A., Hardy J., de Silva R. (2010) Disentangling the role of tau gene locus in sporadic tauopathies. Current Alzheimer Research **7(8)** 726-734

iv. Abbreviations

2DGE	two-dimensional gel electrophoresis
AAO	age at symptomatic onset
ACE	A+C rich element
AD	Alzheimer's disease
ADPRT1	NAD(+) ADP-ribosyltransferase 1
AgD	argyrophillic grain disease
AICD	amyloid precursor protein intracellular domain
AIF	apoptotic inducing factor
ALS	amyotrophic lateral sclerosis
ALY	Ally of AML-1 and LEF-1
AMPK	AMP-activated kinase
APBB1	amyloid beta A4 precursor protein-binding family B member 1
APLP2	amyloid precursor like protein 2
ARE	A+U-rich element
ARHGEF10	guanine nucleotide exchange factor 10
BAF	Brg or hBrm associated factors
BDNF	brain derived neurotrophic factor
BP	branch point
BPS	branch point sequence
BRCT	BRCA1 carboxy-terminal repeat motif
BSA	bovine serum albumin
BTB/POZ	Broad complex, Tramtrack and Bric a brac/Poxviruses and Zinc Finger
bZIP	basic region leucine-zipper
CBD	corticobasal degeneration
CBS	corticobasal syndrome
CCD	charge-coupled device
CHCA	α-cyano-4-hydroxycinnamic acid
ChIP	chromatin immunoprecipitation

CID	collision-induced dissociation
CLK	CDC20like kinases
CNS	central nervous system
CtBP	C-terminal binding protein
CTD	C-terminal domain
Cyt	cytoplasmic
dATP	deoxyadenosinetriphosphate
DBD	DNA-binding domain
dCTP	deoxycytidinetriphosphate
DDX21	nucleolar RNA helicase II
dGTP	deoxyguanidinetriphosphate
DM1	myotonic dystrophy type 1
DMSO	Dimethyl sulphoxide
DNMT1	DNA methyltransferase 1
dNTPs	deoxynucleoside triphosphates
DOPE	L-dioleoyl phosphstidylethanolamine
DTE	1, 4-Dithioerythritol
dTTP	deoxythymidinetriphosphate
ECACC	European Collection of Cell Cultures
ECL	enhanced chemiluminescence
EDTA	Ethylenediamine tetraacetic acid
EJC	exon junction complex
EKLF	erythroid krueppel like factor
EMSA	electrophoretic mobility shift assay
ERK1/2	extracellular signal-regulated kinases 1/2
ESE	exonic splicing enhancer
ESI	electrospray ionization
ESS	exonic splicing silencer
FTD	frontotemporal dementia
FTDP-17	frontotemporal dementia with Parkinsonism linked to chromosome 17
FTLD	frontotemporal lobar degeneration
GAPDH	glyceraldehyde-3-phosphate

GFP	green fluorescent protein
GO	gene ontology
GRK3	G protein receptor kinase 3
GSK3 β	glycogen synthase kinase 3β
GWAS	genome wide association study
HAT	histone acetyltransferase
HDAC1	histone deacetylase 1
h-IBM	hereditary inclusion body myositis
HIC1	Hypermethylated in cancer 1
HMG	High mobility group
hnRP	heterogeneous nuclear ribonucleoprotein
HPLC	high pressure liquid chromatography
HRP	horseradish peroxidase
ICD	intracellular domain
ISM	intronic splicing modulator
ISS	intronic splicing silencer
LBP	leader-binding protein
LD	linkage disequilibrium
LOAD	late onset AD
LTP	long term potentiation
M17	BE(2)-M17
mH2A1.1	macroH2A1.1
MALDI	matrix assisted laser desorption ionization
MAPs	microtubule-associated proteins
MAPT	microtubule associated protein, tau
MC	minichromosomal
MDM	monocyte-derived macrophages
MHC	major histocompatibility complex
MES	2-(N-morpholino)-ethanesulfonic acid
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
MOPS	3-(N-morpholino)-propanesulfonic acid
MOWSE	Molecular Weight Search
mRNP	messenger ribonucleoprotein

MS	mass spectrometry
NAD^+	nicotinamide adenine dinucleotide
NCAM	neural cell adhesion molecule
NCBI	National Centre for Biotechnology Information
NFL	low molecular weight neurofilament
NFTs	neurofibrillary tangles
NLS	nuclear localization signal
NO	nitric oxide
NPC	Niemann-Pick type C
NTF	neurogenic element binding transcription factor
Nuc	nuclear
NUCL	nucleolin
NXF	nuclear export factor
OPRD1	δ-opoid receptor
P(number)	postnatal day (number)
PAR	poly-(ADP)-ribose
PARP1	Poly [ADP-ribose] polymerase 1
p54nrb	54 kDa nuclear RNA and DNA binding protein
PBS	phosphate buffered saline
P/CAF	p300/CBP-associated factor
PCR	polymerase chain reaction
PD	Parkinson's disease
PEG	polyethylene glycol
PHFs	paired, helically wound filaments
PIC	preinitiation complex
PiD	Pick's disease
PIN1	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1
PKAN	panthothenate kinase-associated neurodegeneration
PKC	protein kinase C
PLGS	ProteinLynx Global Server
PNS	peripheral nervous system
Poly (dI.dC)	Poly(deoxyinosinic-deoxycytidylic) acid
PP	protein phosphatase

PPE	polypurine enhancer
PPT	polypyrimidine tract
PSP	progressive supranuclear palsy
PTB	polypyrimidine tract binding
PTBP1/hnRNPI	Polypyrimidine tract binding protein
QTOF	quadrupole time-of-flight
RA	retinoic acid
REF	RNA and export factor
ROD	heterogeneous ribonucleoprotein D0
ROS	reactive oxygen species
RPA70	replication protein 70 kDa
RNAPII	RNA polymerase II
RNP	ribonucleoprotein
SAF	scaffolding attachment factor
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SF	straight filaments
s-IBM	sporadic inclusion body myositis
SKY5	SH-SY5Y
S/MAR	scaffold / matrix attachment regions
snRNP	small nuclear ribonucleoprotein
SNPs	single-nucleotide polymorphisms
SR	serine-arginine rich
Staf	Selenocysteine tRNA gene transcription activating factor
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TAE	Tris-Acetate buffer
TAR	transactive response
TBE	Tris-Borate buffer
TBP	TATA-binding protein
TCA	trichloroacetic acid
TDP-43	Transactive Response DNA-binding protein-43
TERT	telomerase catalytic subunit
TFs	transcription factors
TFA	trifluoroacetic acid

TFCP2	transcription factor CP2
T_m	melting temperature
TPH2	tryptophan hydroxylase 2
TPM	tropomyosin
TONEBP/OREBP	tonicity-responsive enhancer/osmotic response element-binding protein
TOP1	DNA topoisomerase 1
TR	telomerase RNA component
U2AF65	U2 auxiliary factor 65 kDa
UCSC	University of California Santa Cruz
UTR	untranslated region
VIP	Vasoactive intestinal polypeptide
WHO	world health organization
WSTF	Williams syndrome transcription factor
WT1	Wilms' tumour
WT	wild-type
ZNF143	zinc finger protein 143

1 Introduction

1.1 Overview

In 1975, Kirschner's group suggested either phosphorylation of tubulins or the presence of an exchangeable factor that might be responsible for microtubule elongation and dynamics. In search of this factor, tubulin was purified and proteins bound to tubulin were separated.[1] One of the first proteins eluted was a factor termed tau for its involvement in tubulin elongation. When added to purified tubulin under non-polymerizing conditions, tau was sufficient for both nucleation and elongation of microtubules from purified tubulin as seen by electron microscopy.[2] This formed the basis of the assay for which to monitor the purification of tau's activity. Research into microtubule-associated proteins (MAPs) and their importance as regulators of microtubule dynamics due to their direct and intimate binding with microtubules began to intensify.[3, 4, 5]

As brain was used due to its significantly greater amounts of tubulin than other tissues, the discovery and function of tau has had been intimately placed in a neuronal context. This became more apparent when neurofibrillary tangles (NFTs) such as those seen in Alzheimer's disease (AD) were found to contain tau. NFTs are intra-neuronal cytoplasmic inclusions consisting of non-membrane bound bundles of paired, helically wound filaments (PHFs) which are 10 nm in diameter. The sub-unit of PHFs is the MAP, tau which is hyperphosphorylated. These have been shown to exist in other tauopathies such as frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17), progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD). However, little is known about the molecular and cellular biology of tau to fully appreciate the significance of PHFs in the disease.

1.2 Molecular properties of tau

Human tau is encoded by a single gene, *MAPT*; approximately 150 kb large on chromosome 17q21.3 consisting of 1 non-coding and 14 coding exons.[6] Exon -1 is transcribed but not translated and comprises the core promoter region. Exons 2, 3 which encode N-terminal inserts and exon 10 which encode 1 of 4 microtubule binding domains are alternative spliced in the adult human central nervous system (CNS), leading to 6 tau isoforms (0N/3R, 1N/3R, 2N/3R, 0N/4R, 1N/4R, 2N/4R where N is an N-terminal insert and R is a microtubule-binding repeat). Exons 4a, 6 and 8 are expressed in neuronal subtypes such as in the peripheral nervous system (PNS) (Figure 1.1).[6, 7] Complimentary DNA cloning has yielded tau sequences from mouse, rat and cow.[8] As examples, orthologues have been found in *Caenorhabditis elegans*, *Drosophila melanogaster* and *Xenopus laevis*.[5, 9]

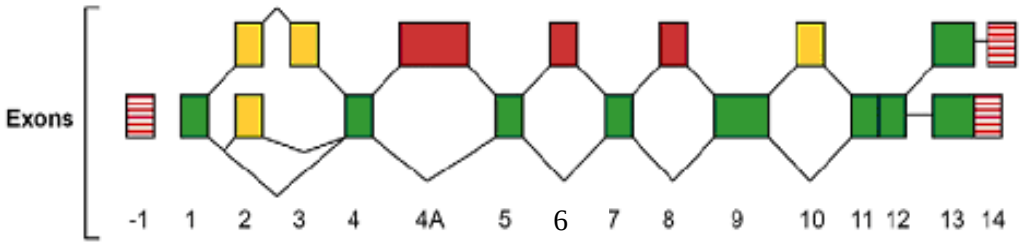


Figure 1.2.1: Tau gene, MAPT structure. Exon -1 is located in the 5' untranslated region. Exons 2, 3 and 10 in yellow are alternatively spliced in the CNS. Exons 4A, 6 and 8 in red are absent in the CNS but exon 4A is present mostly in the PNS, very few in the CNS. Modified from reference [6].

The 6 tau isoforms produced in the adult CNS have apparent molecular weights between 48 and 67 kDa due to the alternative mRNA splicing of exons 2, 3 and 10 (Figure 1.2.1). Alternative splicing of exons 2 or 3 leads to either 0, 1 or 2 N-terminal insert (termed 0, 1 or 2 N tau). Alternative splicing of exon 10 leads to an addition of the repeat binding domain either 3- or 4-R tau. In foetal brain only the shortest isoform, 0N, 3R is expressed. Circular dichroism was used to establish the secondary structure of 2N, 4R tau protein in solution. A 12% α-helical and 20% β-sheet composition was estimated, although this is likely to be an overestimate and tau's structure is probably best described as a random coil.[10] This hinders structural studies of tau's interaction with microtubules and other proteins.

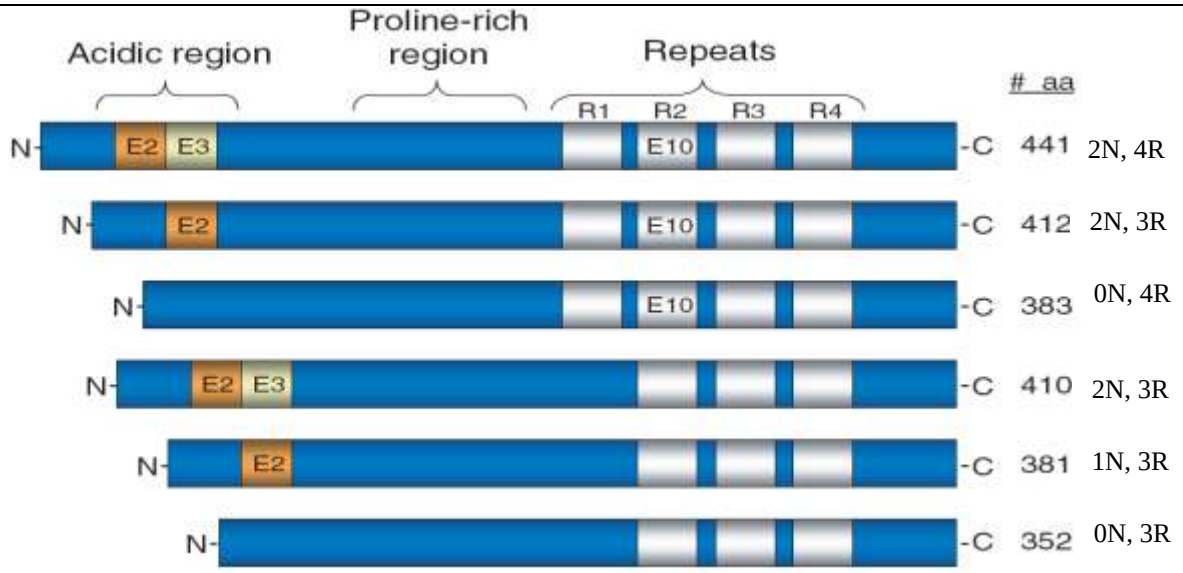


Figure 1.2.2: 6 Tau isoforms in adult human CNS. This shows the tau isoforms in adult CNS. The number of amino acids are indicated at the right (# aa). The protein is broadly divided into 3 domains, the N-terminal acidic region with either exon 2 or both exons 2 and 3 or neither, the middle proline-rich domain and the C-terminal microtubule binding domain with or without the 2nd microtubule binding repeat (R2) to yield either 4 or 3 microtubule binding domains. Modified from reference[11].

Each N-terminal insert has 29 amino acids and is rich in proline and positively charged residues. These are flanked by 2 acidic domains followed by a basic stretch before a proline-rich domain that separates the projection domain from the assembly domain (Figure 1.2.2). The C-terminus is rich in basic residues as well as the microtubule binding repeats (31-32 amino acids) each containing a characteristic PGGG motif as well as a conserved KXGS motif that can be phosphorylated.[5, 8] Each repeat can be divided to an 18 residue minimal tubulin binding region and a less conserved inter repeat of 13/14 residues.[12] Furthermore, there is a pseudo repeat (Figure 1.2.3). The regions that flank the microtubule binding regions also influence its microtubule interactions.

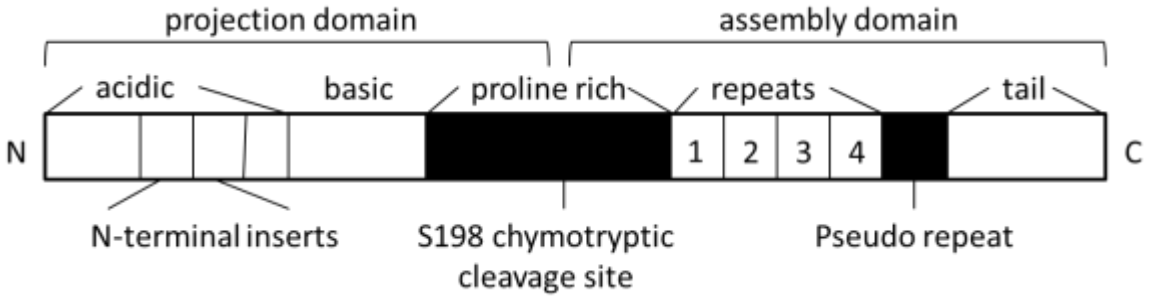


Figure 1.2.3: Protein domains of hTau[441]. Chymotryptic cleavage separates regions that bind (C-terminal half) or do not bind (N-terminal half) to microtubules. The 29-residue inserts near the N-terminal are sandwiched between 2 acidic regions followed by a basic region, a middle proline rich region, 4 microtubule binding repeats and a pseudo repeat. Modified from references[13,14].

1.3 Tauopathies

Tauopathies are diseases that are characterized by tau pathology. These include AD, PSP, CBD, FTDP-17, Pick's disease (PiD), argyrophillic grain disease (AgD), amyotrophic lateral sclerosis (ALS), Niemann-Pick type C (NPC), dementia pugilistica, tangle-only dementia, subacute sclerosing panencephalitis, Wernicke's encephalopathy, Parkinson's dementia complex of Guam, pantothenate kinase-associated neurodegeneration (PKAN), progressive nonfluent aphasia, Down's syndrome, several variants of Prion diseases, myotonic dystrophy, sporadic inclusion body myositis (s-IBM), hereditary inclusion body myopathy (h-IBM) and glaucoma.[15, 16, 17, 18, 19, 20, 21, 22, 23, 24] Until a consensus in 1994, frontotemporal dementia (FTD) was often classified as PiD.[25] FTD is the second most common form of primary degenerative dementia after AD and accounts for 20% of presenile dementia cases.[26] Mutations in *MAPT* accounts for 18% of all FTD and 41% of familial FTD cases.[27] With PSP being the second most common cause of parkinsonism, the deregulation of tau has far and wide implications in neurology. According to the World Health Organization (WHO), almost one billion people worldwide are expected to be 60 years or older by 2025 and 42 million people are expected to have dementia by 2020, therefore the need for greater understanding for the basis of these diseases cannot be overstated (http://www.who.int/ageing/en/).

1.4 Tau mutations

Since the cloning of human *MAPT* and the establishment of linkage between FTD and 17q21.31, FTDP-17, originally defined in 1996, has been described as a collection of autosomal dominant inherited tauopathies.[24,28,29] Intensive research was carried out on characterizing the various mutations in *MAPT* and how these enlighten our understanding of the pathologies mentioned above. More than 100 families with FTDP-17 were described worldwide with 42 different mutations in *MAPT*.[30] These include 26 missense, 19 intronic, 3 silent and 2 amino acid deletion mutations. A pictorial representation of these mutations along the tau protein can be seen in Figure 1.4.1. Approximately half of the mutations had their primary effect at the RNA level while the other half had their primary effect at the protein function level. While these mutations might exert different effects at the molecular level, there was extensive clinical overlap between them.

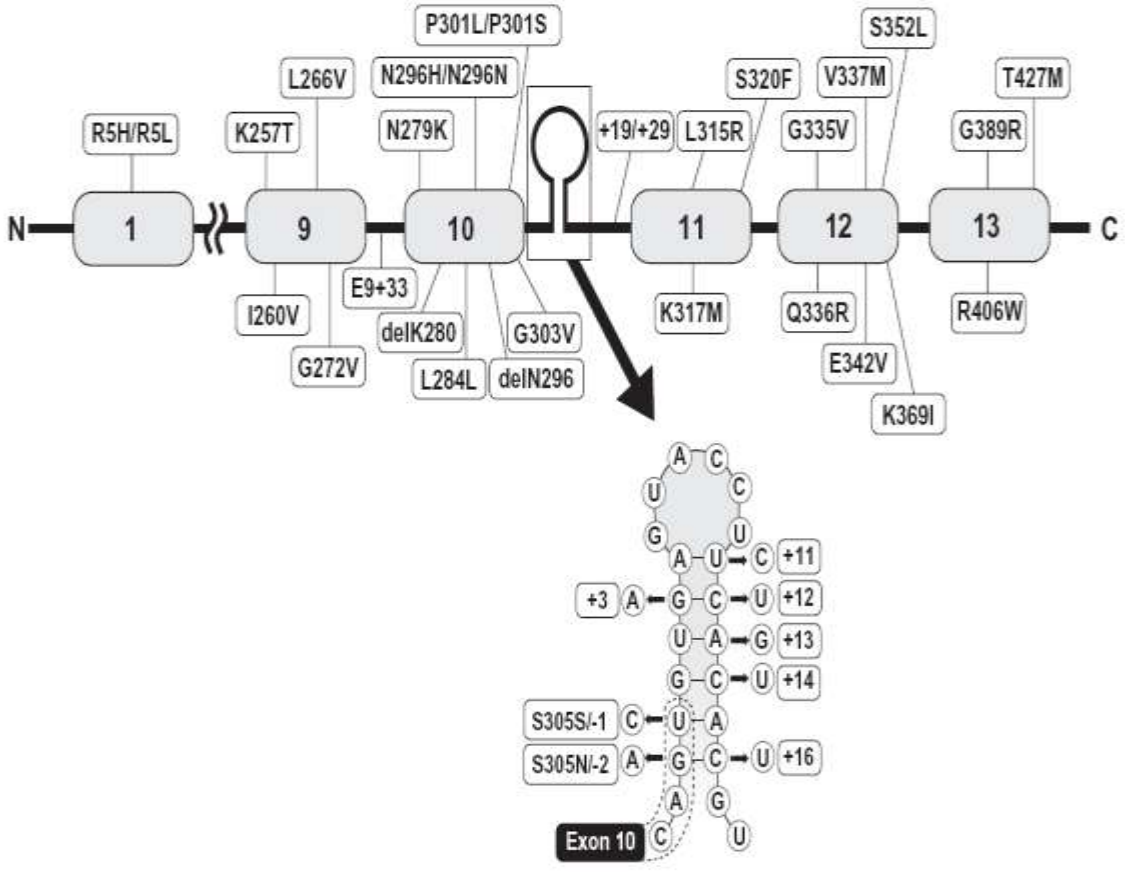


Figure 1.4.1: List of most FTDP-17 mutations along the tau protein. Modified from[31].

1.5 Clinicopathology of tau mutations

Generally, dementia predominant FTDP-17 was usually seen in families with mutations in exons 1, 9, 10, 11, 12 and 13. Mutant tau can be mostly found in both neuronal and glial tissues as narrow twisted ribbons.[31] However, parkinsonism-plus predominant FTDP-17 was usually seen in families with intronic and exonic mutations affecting exon 10. Most of these led to neuronal tau pathology and very little glial tau pathology except for R5L/H, I260V, L266V and L315R which had more glial tau pathology. Intronic mutations and those in exons 1 and 10 were mostly associated with neuronal and glial tau deposition which looked like CBD while those in exons 9, 11, 12 and 13 were predominantly neuronal. Three mutations (K257T, L266V, G272V) in exon 9, one in exon 10, two (L315R, S320F) in exon 11, three (Q336R, E342V, K369I) in exon 12 and two (G389R, T427M) in exon 13 were associated with Pick-body-like inclusions. Other mutations in exons 12 (V337M) and 13 (R406W) led to tangles similar to those seen in AD.[17] The R5L, N279K, ΔN296, S305S and +16 mutations had phenotypes similar to PSP.

The R5L mutation in exon 1 generated a PSP-like phenotype.[32] Unlike the R5H mutation which showed prominent glial tau deposits, the R5L mutation led to both neuronal and glial pathology.[33] G272V, K257T or L266V mutation in exon 9 exhibited phenotypic resemblance to PiD with Pick-body-like inclusions consisting of only 3R tau but without motor dysfunction.[34,35,36] However, mutations and pathology are not clear cut. Clinical diagnoses of patients with P301L mutation in exon 10 ranged from PSP to CBD and PiD with narrow, twisted-ribbon structures of 4R-tau and little 3R-tau.[37] Both P301L and P301S mutations led to neuronal and glial pathology.[38] The N279K mutation typically led to PSP with dementia.[39] ΔK280 in exon 10 displayed pathological features largely in the form of Pick-body-like bodies with some 4R-tau found in them.[40] However, it is not clear if ΔK280 is pathogenic. ΔN296, however, generated atypical PSP-like symptoms.[41] K317M in exon 11 led to degeneration of frontotemporal, corticospinal tract and substantia nigra without Lewy bodies causing Parkinsonism with motor neuron disease and FTD. Abnormal tau deposition was observed in both neuronal and glial tissues but no Pick-body-like inclusions were observed.[42] Other mutations in exon 11, however, such as L315R led to memory loss and behavioral changes with neuronal and astrocytic tau inclusions while S320F mutation led to memory decline with neuronal inclusions and only rare coiled bodies found in

oligodendroglial cells.[43,44] Unlike K317M, Pick-body-like inclusions were observed with the other two mutations in exon 11. Although three mutations in exon 12 (Q336R, E342V and K369I) all resulted in the development of Pick-like bodies, V337M produced widespread neurofibrillary tangles, both PHFs and straight filaments (SF), containing all six isoforms similar to AD.[45,46,47,48] Recently, another mutation in exon 12, V363I, had been associated with progressive nonfluent aphasia.[49] The G389R in exon 13 showed phenotypic variability resulting in FTD, corticobasal syndrome (CBS) or PiD without motor dysfunction with large number of Pick-body-like and axonal inclusions.[50]

1.6 Molecular implications of tau protein mutations

In P301L brains, there was selective aggregation of mutant tau in the deposits which were partly resistant to degradation as confirmed by pulse-chase experiments.[51] In R406W brains, unlike P301L, there were equal amounts of wild-type (WT) and mutant tau present in the aggregated form. Hence, there was no selective aggregation of R406W tau isoforms.[52] Although the same level of phosphorylation was seen in WT and R406W aggregated tau, soluble R406W was less phosphorylated than soluble WT tau at Ser-396 and 404. This was probably due to a steric restriction imposed by the substitution. R406W mutant tau also led to the potential loss of phosphorylation at Thr-231 and Ser-235 by GSK3β compared to WT tau, implying long-range conformational effects.[53] No differences in phosphorylation were seen between P301L and V337M compared to WT tau. In neuroblastoma cell culture experiments, P301L and V337M mutations led to decreased phosphorylation at the pathologically relevant Ser-202/Thr-205 site but R406W did not.[54] For Cdk5-dependent phophorylation, R406W mutant tau was less phosphorylated at Ser-404 whereas no difference was seen with K257T, P301L/S mutant tau compared to WT tau.[55] The reduction in Cdk5 phosphorylation was only seen with full-length R406W tau suggesting again long-range conformational changes. Upon binding to microtubules, more P301L, P301S and R406W tau were phosphorylated compared to WT tau, implying that the conformation change brought about by the mutations was negated by microtubule binding. JNK, activated in the mitotic phase, directly hyperphosphorylated R406W mutant tau at the PHF-1 site suggesting a functional role of R406W tau in mitotic cell.[56]

Effects on phosphorylation of tau are only half the picture. It had been shown that G272V, ΔK280, P301L, P301S, S305N, V337M, G389R and R406W mutations inhibited the binding of recombinant 3R- and 4R-tau to protein phosphatase 2A (PP2A) by 20-95%.[57] The reduction in binding was greatest with G272V, ΔK280 and V337M mutations. R406W and G389R mutations, however, showed isoform-specific inhibition of 3R-tau to PP2A compared to 4R-tau.

Both P301L and P301S mutant recombinant tau were able to respond strongest to the stimulation of heparin-induced filament assembly in 4R-tau only.[58] G272V mutation in exon 9 produced an increase in tau filaments in both 3R- and 4R-tau. However, V337M in exon 12 had a small stimulatory effect in 3R-tau while the ΔK280 mutation had no obvious effect on tau filament assembly. G272V and P301L mutations increased the rates of both filament nucleation and extension reactions whereas R5L and V337M increased the nucleation phase.[59] Using tryptophan scanning mutagenesis, it was shown that both ΔK280 and P301L led to enhanced propensity for β-structure where packing is tightest in filaments.[60] Sf9 insect cells overexpressing G272V, P301L and R406W mutant tau were able to form filaments resembling PHFs.[61] G272V, in both 3R- and 4R-tau, ΔK280 and P301L mutant tau, all exhibited significant reduction in their ability to regulate microtubule dynamic instability compared to WT tau.[62] This meant they were less able to increase the amount of time microtubules were attenuated. R406W, however, had little effect on regulating microtubule dynamics, suggesting indirect biological effects. These mutational effects were reduced when cells were treated with lithium implicating a role for GSK3β. Largest effects on aggregation were observed with G272V and P301L tau mutations.[45] They were predicted to be highly sensitive to changes in free cytosolic tau concentration and their ability to aggregate at submicromolar concentrations. Both G272V and P301L mutations lie in segments adjacent to hexapeptide motifs involved in inducing cross-β-sheet structure in the repeat region. In hippocampal neurons, N279K potentiated the apoptotic effect of tau 151-421, the cleavage product of caspase-3.[63] This probably was related to the N279K, and in a similar way, S305N tau's ability to bind to microtubules with equal efficacy but compromised their abilities to regulate microtubule growth rate and catastrophe frequency.[64] Microtubule effects of these mutations are summarized in Figure 1.6.1.

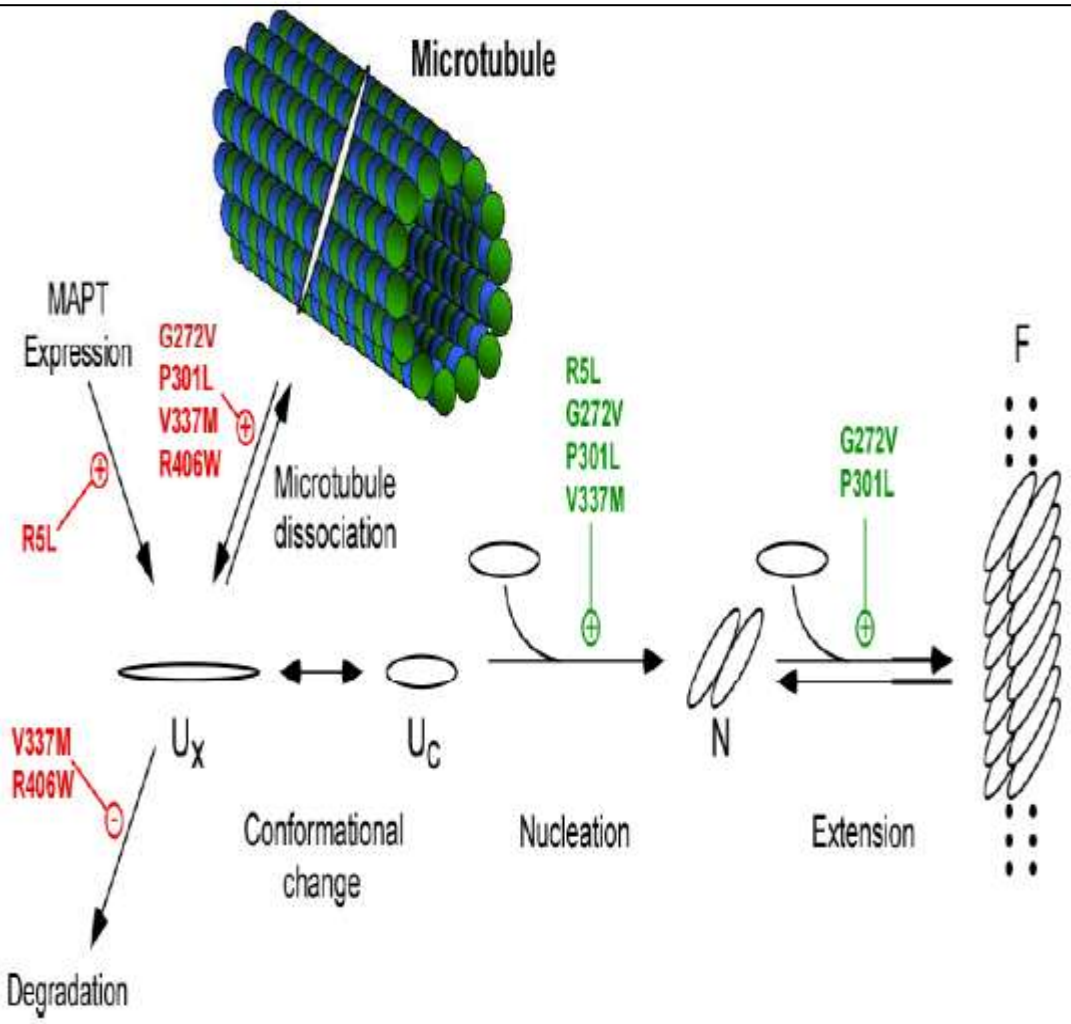


Figure 1.6.1: Effects of FTDP-17 mutations on the tau fibrillization pathway. U_x represents tau in an assembly incompetent state while U_c represents an assembly competent tau. N represents the nucleus of fibrillization with tau as a dimer. Modified from[60].

1.7 Molecular implications of tau splicing mutations

Besides affecting tau at the protein level, mutations in and around intron and exon 10 contribute to changes in 4R:3R tau expression levels, most of which lead to an increase in 4R-tau. The 5' and the 3' exon 10 splice sites differ from the canonical sites by 4 and 1 nucleotides respectively, making them weaker binding sites.[15] Furthermore, exon 10 is flanked by two reasonably large introns, intron 9 (13.6 kb) and intron 10 (3.8 kb). The 5'

splice site sequence contains three components: an invariant dinucleotide at the 5' end of the intron separated from the branchpoint sequence (BPS) by a central polypyrimidine tract (PPT). The BPS consensus sequence is complementary to the U2 snRNA recognition sequence. At the 5' end of exon 10, three exonic splicing enhancers (ESEs) have been identified: a SC35-like enhancer, a polypurine enhancer (PPE) and an AC-rich element (ACE).[65] An exon splicing silencer (ESS) in exon 10 is sandwiched between these ESEs upstream and another ESE downstream, but still within the same exon.

The 3' splice site is located at the boundary between exon 10 and intron 10. This sequence is complimentary to and base-pairs with the 5' end of U1 snRNA.[66] The sequence located further down had been proposed to form a stable, folded stem-loop structure which acts as an intron splicing silencer (ISS).[67] Two different structures have been proposed (Figure 1.7.1B).[68, 69] Downstream of this ISS, an intron splicing modulator (ISM) from +19 to +26 had been described. It had also been proposed that the +29 intron 10 mutation which led to decreased exon 10 (more 3R-tau) splicing might be due to a smaller double-stem loop structure (Figure 1.7.1C).[70]

During the initial splicing pathway, the 65 kDa U2 auxiliary factor (U2AF65) interacts loosely with the PPT which is stabilized by U2AF35 binding to the dinucleotide of the 3' splice site.[71, 72] U1 snRNP then associates with the 3' splice site and further stabilizes U2AF65 at the PPT. This leads to the strengthening of the recruitment of U2 snRNP to the BPS. This and all subsequent steps are energy dependent. SR factors bound to ESEs are important for the recruitment and stabilization of U2AF, U1 and U2 snRNPs (Figure 1.7.1A). Exon 10 inclusion into mRNA is regulated by CDC2-like kinases (CLK1, 2, 3 and 4) which phosphorylate serine-arginine-rich (SR) proteins which regulate pre-mRNA splicing.[73] CLKs achieve this through releasing specific proteins from nuclear storage sites. Other proteins thought to bind to the intron 10 and contribute to splicing include RBM4, CELF3 and CELF4.[74,75]

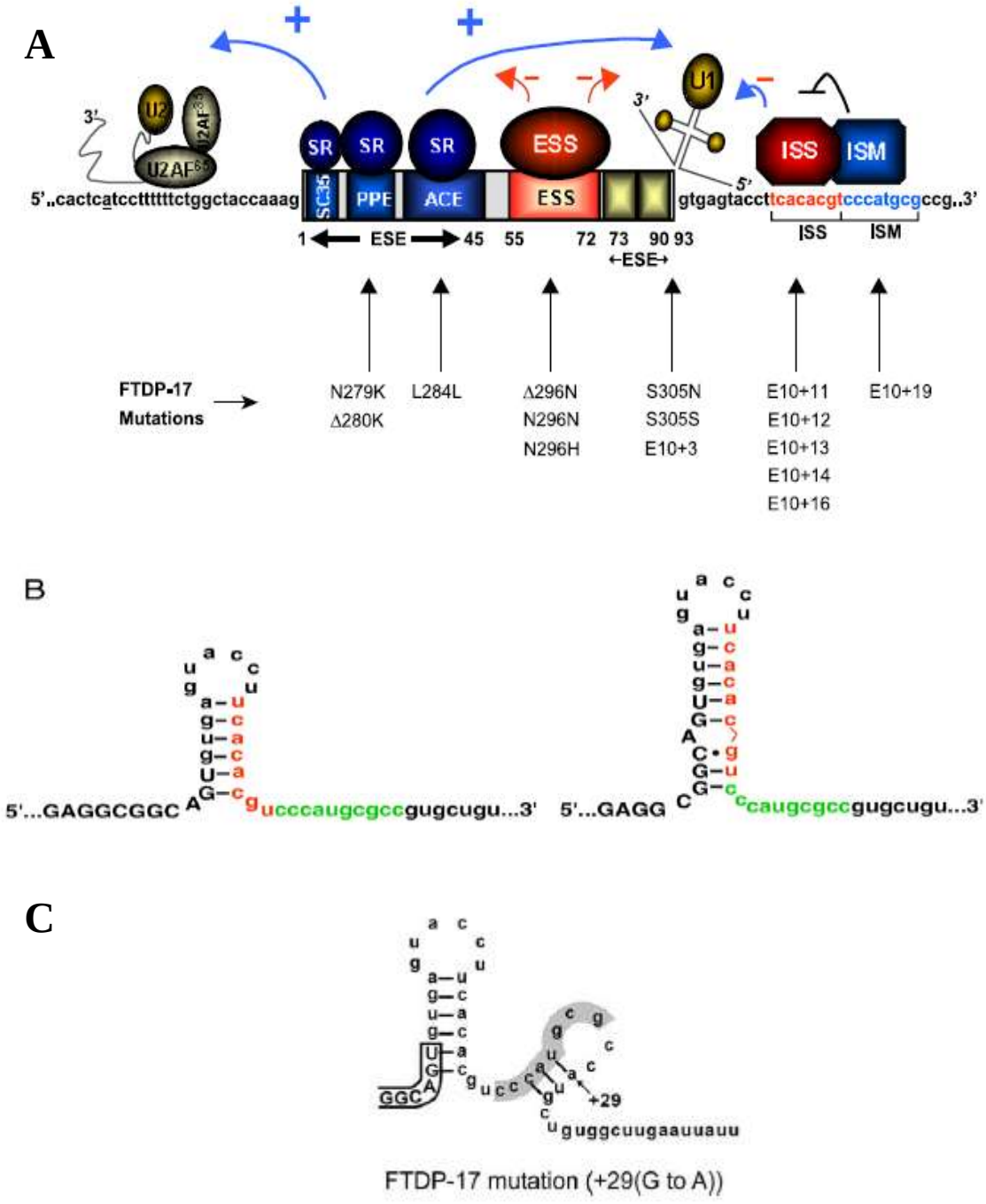


Figure 1.7.1: Model of tau exon 10 splicing. A. Position of *cis*-elements and related trans-factors along *MAPT* that enhance or reduce exon 10 splicing. FTDP-17 mutations affecting exon 10 splicing are listed below. B. Two of the variant stem-loop structures proposed to form at the ISS. C. The secondary structure of +29 mutation predicted to form with the ISM highlighted in grey. Modified from[71,76,89].

A mutation in intron 9, -10, added an additional pyrimidine, strengthening the polypyrimidine tract by having more consecutive pyrimidines. This had been shown to strengthen the splice acceptor site and to promote the inclusion of exon 10.[77] The N279K mutation (AAT to AAG) not only affected the peptide sequence but its location in the PPE created a purine-rich stretch (AAGAAGAAG) which made it more resemble an ESE consensus sequence.[78] The ΔK280 mutation, however, abolished exon 10 splicing by removing an AAG.[79] The PPE acts as a binding site for trans-acting factors such as SF2/ASF, Tra2β, SRp30c and SRp54.[80] The silent mutation L284L increased the AC content of ACE which made it more effective as an ESE.[61] Three mutations (N296H, N296N and ΔN296) within the ESS, disrupted this 18 nucleotide sequence which enhanced exon 10 splicing.[81] The S305N mutation (AGT to AAT) and the +3 mutation occurred in the 5' splice site and strengthened its binding to U1 snRNA.[55,56] This mutation had been associated with Pick-body-like inclusions while S305S mutation was reminiscent of PSP. S305S disrupted the stem-loop structure without the predicted effect on U1 snRNA.[82] A new mutation, S305I, had a morphological phenotype resembling AgD. It had a strong effect on exon 10 inclusion by substituting the complimentary binding partner of +16 and weakening the stem-loop structure.[83] The +11, +12, +13, +14 and +16 mutations within the stem-loop structure of the ISS destabilized this structure and prevented it from acting as an ISS, leading to more 4R-tau. The +19 mutation, 3' of the stem-loop, disrupted the ISM which acted to destabilize the effects of the ISS, resulting in more 3R-tau.[84]

In myotonic dystrophy type 1 (DM1), patients suffer from degeneration of cortical neurons with reduced levels of exon 2-containing tau.[85] Splicing of exon 2 in tau pre-mRNA is regulated by CELF protein ETR-3.[86] Inclusion of exon 6c in tau decreased in DM1 brains compared to control brains.[87] This isoform of tau might be needed in neuronal differentiation and neurite extension and exon 6 splicing is aided by CELF-5 and -6. Figure 1.7.2 illustrates the sequences involved in exon 2 and 3 splicing.

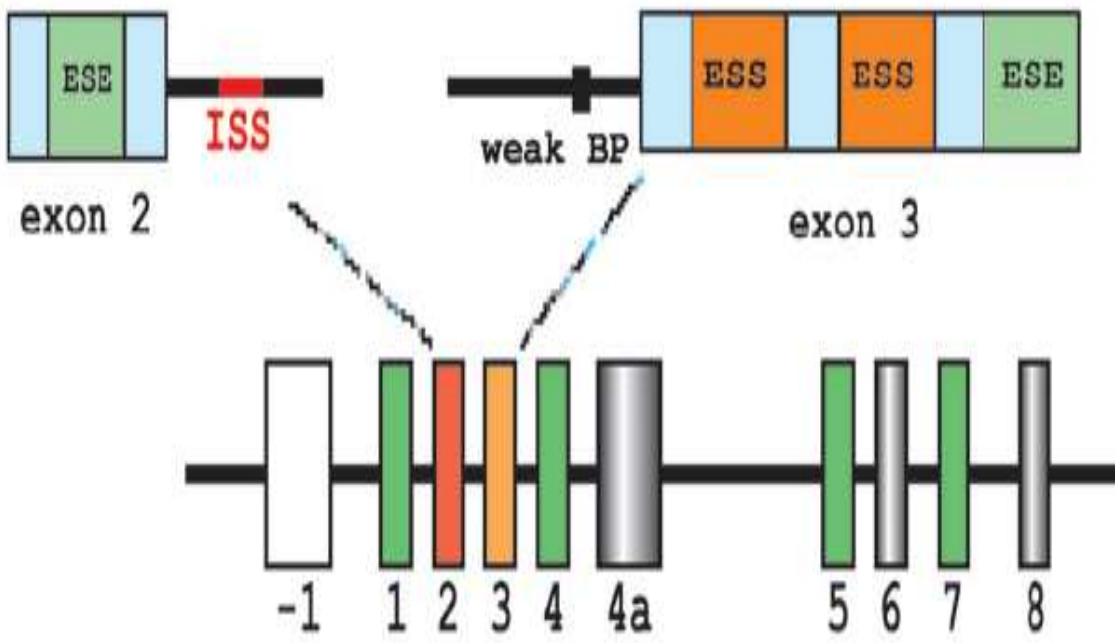


Figure 1.7.2: Model of tau exon 2/3 splicing. The exon splicing enhancers (ESEs), exon splicing silencers (ESSs), intron splicing silencer (ISS) and weak branch point (BP) for exons 2 and 3 are shown. Unspliced exons are coloured green, spliced exons are coloured orange and exons expressed in the peripheral nervous system are coloured grey. Modified from[88].

From all these studies, two models have been proposed for the regulatory elements in tau pre-mRNA: linear *cis*-element model and the stem-loop model.[5] The stem-loop model, however, does not support the *in vivo* modulatory effect of the ISM. In the linear *cis*-element model, multiple weak interactions between factors binding to sequences inside intron 9, exon 10 and intron 10 act in a linear fashion to modulate the splicing of exon 10.

Recently, a partial deletion of the *MAPT* gene (17.3 kb from exon 6 to 9) was detected in a patient with FTD.[89] This meant that the first microtubule binding domain was absent. *In vivo* binding assays showed that this dramatically reduced binding to microtubules.

1.8 Complex architecture of the human tau gene locus

The FTDP-17 candidate region had been studied and mapped extensively.[90] The human *MAPT* locus spans 130,987 bp on chromosome 17 and has approximately one SNP for every 200 bp and an insertion or deletion every 1 kb. Two *MAPT* haplotypes with 8 SNPs and the TG dinucleotide repeat have been described.[91] H2 was characterized by a 238 bp deletion in intron 9. Chromosome 17 is prone to recurrent inversions because the inversion breakpoints correspond to core duplications within the leucine-rich repeat containing (LRRC37) gene family.[92] LCR A is located 250 kb centromeric to *MAPT* and inverted relative to LCRs B (which is 180kb telomeric to *MAPT*) and C (Figure 1.8.1). LCR D is located in the same orientation as LCR A but 19 Mb telomeric of *MAPT*. LCR A and B flank the extended H1 and H2 haplotypes that resulted from a historical inversion. H1 was assumed to be the ancestral sequence as most of sub-Saharan African haplotypes were of the H1 clade. However, H2 is the most likely great ape and human ancestral state. The divergence was estimated to be about 3 million years ago. The H2 is the minor haplotype in Caucasian populations but still accounted for 20% of the Caucasian compared to 10% of the Russian, 6% of the Guamanian Chamorros and 0% of the Japanese population.[93] The distribution of the haplotype and analysis of the slippage of dinucleotide repeat markers within the haplotype suggested that it entered *Homo sapiens* populations between approximately 10,000 and 30,000 years ago when *Homo neanderthalensis* co-existed.[94]

Looking at the detailed BAC-based sequence assembly of the approximately 970 kb inverted human H2 haplotype and comparing it to the sequence structure and genetic variation of the corresponding region for the H1 human haplotype and that of chimpanzee and orangutan, provided evidence that the inversions occurred independently in chimpanzee and humans.[95]

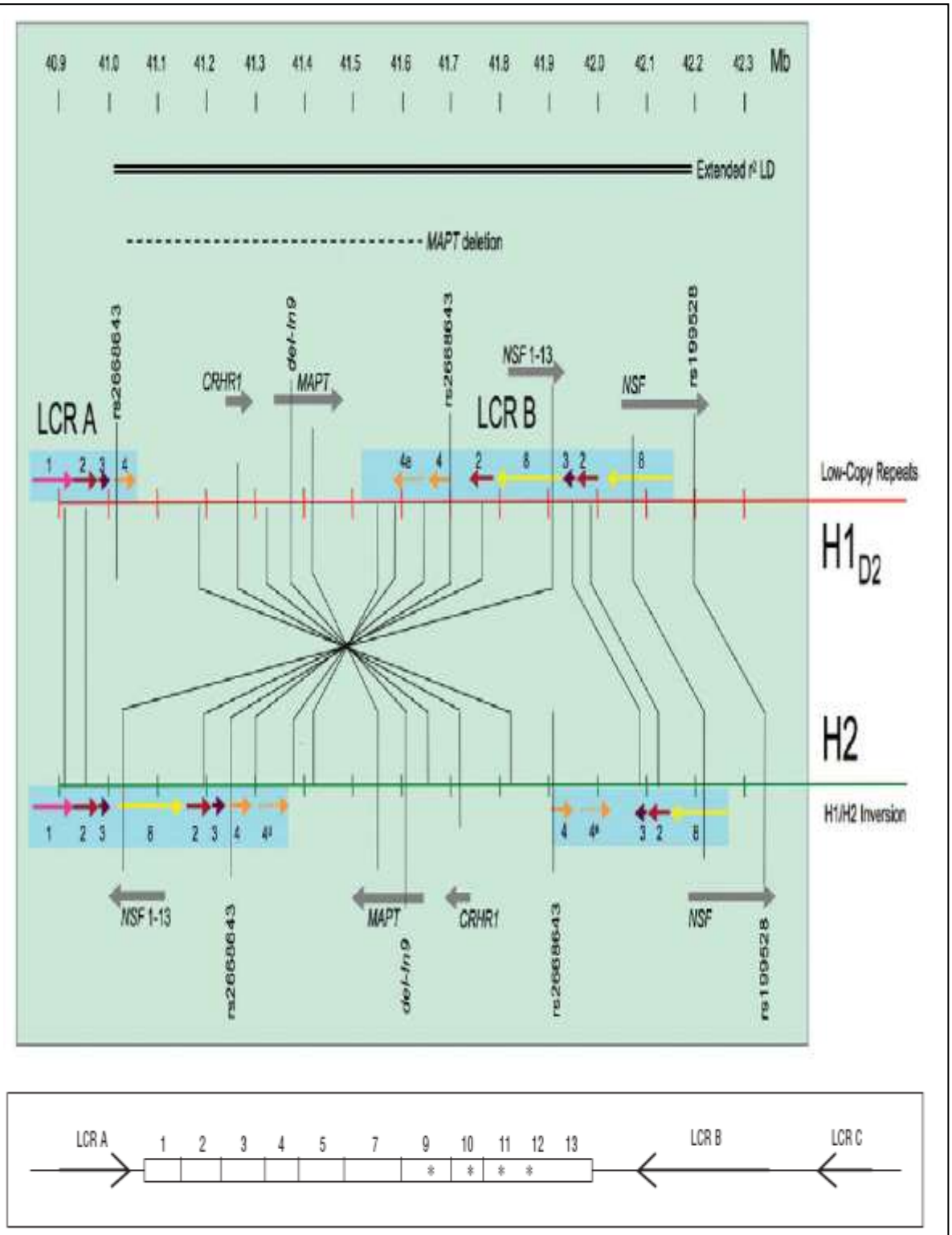


Figure 1.8.1: Inverted *MAPT* haplotypes at 17q21.31. Besides the inversion between H1 and H2, there are also segmental duplications in both directions. Modified from[96].

1.9 Tau gene haplotypes and disease

Since 1997, the first reported genetic predisposition to sporadic PSP of the A0 allele of *MAPT* of a repeat polymorphism in intron 9 was made.[97] The entire, fully extended H1 haplotype was associated with PSP and CBD as a pathogenic locus.[98,99,100] This locus was defined to at least 56 kb from rs242557 to rs2471738. The sole H2 derived haplotype was underrepresented in PSP and CBD. The H1/H1 genotype correlated the parkinsonism phenotype of FTDP-17 while the H1/H2 genotype correlated with the dementia phenotype.[101] H2/H2 was found to be overrepresented in FFTD patients especially in APOE ε4 negative FFTD patients.[102] This was also associated with reduced tau expression. From linear regression analysis, the H1/H1 genotype was associated with significantly earlier age at symptomatic onset (AAO) in patients with PSP and may contribute to earlier clinical expression of PSP-related pathology.[103] In another study, H1 and H2 haplotypes were not found to associate with a significant difference in the risk of developing FTD.[104]

The H1c subhaplotype was shown to be associated with increased risk of AD with the most likely region for a pathogenic variant lying between rs242557 (5' of exon 1) and rs2471738 (in intron 9) of *MAPT*.[105] However, there was no evidence of association between the H1c haplotype or rs242557 with late onset AD (LOAD).[106] When gender and *APOE4* status were taken into account, there was a suggestive association between rs242557 and LOAD.[107] There was a slight increase in H1/H1 genotype in AD cases with AgD, suggesting a possibility of *MAPT* H1 as a risk factor for AgD.[108] People with Down's syndrome (DS) and the H1/H2 haplotype had a greater risk of developing dementia before 45 years old compared to H1/H1.[109] *MAPT* H1 subhaplotypes had been shown to be preferentially associated with PD even though there was no consistent association between the H1 haplotype and PD in three ethnically diverse populations.[110,111] In the Serbian sample, however, there was significant evidence for an association between H1 and PD suggesting that this role might be ethnically dependent.[112] More recently, H1-clade but not H1c was associated with increased risk of sporadic PD in single-locus associations.[113,114] However, the linkage disequilibrium did not extend beyond the *MAPT* haplotype to neighbouring genes.[115] This indicated the probability of a different mechanism involved in PD compared to other pathologies.

H2 carriers, however, were associated with significant reductions of glucose utilization in the superior and medial frontal gyri in both hemispheres. Compared to H2/H2 patients, H1 carriers showed up to 19% smaller grey matter volume in all the regions compared, frontal gyrus, insula, temporal and cerebellar lobes.[116] These results might shed some light on pathology as FTD, PSP and CBD have some form of cerebral atrophy and reduced glucose metabolism. There was a detection of a relative decrease in *MAPT* H1 expression with increased age in human brain. H1c was also associated with both Richardson's syndrome and PSP-Parkinsonism compared to controls.[117]

The H1 haplotype was also shown to express significantly more 4R-tau mRNA and protein in post-mortem tissue compared to H2 while the H2 haplotype was shown to express more 2N-tau compared to H1.[118,119] This might suggest that the inclusion of exon 3 might contribute to the protective effects of H2.

1.10 Molecular function of *MAPT* SNPs

The rs242557 SNP falls within a 182 bp transcriptionally active region that is conserved in human, chimpanzee, mouse, dog and rat. It is 47 kb downstream from the core promoter and19 kb upstream of the first coding exon.[79] Genotypes at rs2258689 interact with rs242557 to confer risk for Guam ALS/PDC and dementia.[120] There was a significant association of the rs242557-A allele with AD.[121] The H1 risk in young PSP patients could be explained by one SNP, rs242557 within the 22 kb regulatory region of intron 0.[122] In two different human cell lines, H1 haplotype was more efficient at driving gene expression compared to H2.[92] The reporter assay was consistent with the view that rs242557 was a key part of promoter variability and may underlie some of the association.[96, 123] In the Chinese Han population, the 347 C/C promoter allele within the H1/H1 genotype was over-represented in 252 sporadic AD patients compared to control.[124] Promoter transcriptional activity of the 347 C/C genotype was significantly higher compared to 347 G/G genotype. Survival analysis of rs1880753, 160 kb and 50 kb upstream of tau and CRHR1 genes respectively, was consistently associated with disease risk and with AAO in PSP and CBD.[125] It might act as a *cis*-element that regulates gene expression. Another polymorphism within the tau promoter region such as del-568Tins (a deletion at position -568 or an insertion of a T allele) was found to correlate with AAO in PD, suggesting variants within *MAPT* might modify the pathogenesis process of PD.[126] A novel H1 haplotype with rs1800547 was significantly associated with the risk of PD.[127]

The general aim of this PhD was to gain an understanding of the regulation of the transcription of *MAPT* by identifying transcription factors (TF) and TF complexes that bind to the *MAPT* core promoter and its conserved regulatory domains. The specific aim was to test the hypothesis that the differential binding of TFs and TF complexes to the allelic variants of the SNPs in the *MAPT* promoter region forms the basis of the allele-specific differences in *MAPT* transcription that Dr. de Silva's group has observed *in vitro* and *in vivo*. The identification of *MAPT* transcriptional machinery would provide the basis for therapeutic intervention and the individual TFs would be plausible candidates for analysis of genetic variants that exacerbate or reduce these differences.

2 Methods

2.1 DNA and protein quantification

DNA or protein quantity and quality was monitored using the Nanodrop™ ND-1000 Spectrophotometer (Thermo Scientific). It is a full spectrum (220 – 750 nm) spectrophotometer that measures 1 µl samples with high accuracy and reproducibility without the use of cuvettes or the need for dilutions (between 2 – 3700 ng/µl for double-stranded DNA or 0.1 – 100 mg/ml for protein) (Figure 2.1.1).[128]

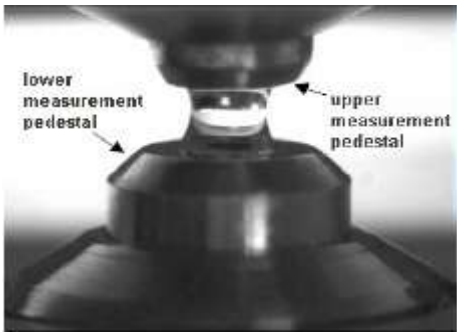


Figure 2.1.1: Upper and lower sampling arms of ND-1000. Adapted from[129].

The absorption at 260 nm indicated the concentration of nucleic acid in a sample. The approximate concentration of sample protein mixtures is calculated using the absorption at 280 nm while using the mass extinction coefficient of bovine serum albumin (BSA) which is 6.7 $L \cdot g^{-1} cm^{-1}$ at 280 nm for a 10 mg/ml BSA solution.

The ratio of absorption at 260 nm to 280 nm provides an indication of the purity of DNA in relation to protein contamination. A ratio of 1.8 is generally accepted as pure for DNA samples. The ratio of absorption at 260 nm to 230 nm provides an indication of DNA purity in relation to contamination by carbohydrates, peptides, urea, phenols and other aromatic compounds. A ratio of 2.0 is generally accepted as pure. For proteins, the ratio of absorption at 260 nm to 280 nm of 0.6 is generally accepted as pure for protein samples.

2.2 Polymerase chain reaction (PCR), Generation of concatamers & Annealing complimentary pairs of oligonucleotides

Developed in 1984 by Kary Mullis, PCR is an *in vitro* method for the exponential amplification of a target DNA sequence within a DNA sample.[130] This was aided by the use of temperature cycling. Knowing the target DNA sequence, designing the primers was done using the computer program Primer3 v. 0.4.0 (http://frodo.wi.mit.edu/). Having primers, heat-stable DNA polymerase and an excess of the four deoxynucleoside triphosphates (dNTPs) : deoxyadenosinetriphosphate (dATP), deoxycytidinetriphosphate (dCTP), deoxyguanidinetriphosphate (dGTP) and deoxythymidinetriphosphate (dTTP), and in permissive buffer conditions, the DNA polymerase will catalyze the synthesis of new complimentary DNA strands containing the target DNA sequence. With every temperature cycling, newly synthesized DNA strands from previous cycles will act as DNA templates for subsequent cycles. Thus, this chain reaction will continue until DNA strands containing the target DNA sequence predominate. After about 25 cycles of DNA synthesis, there will be about 10^5 copies of PCR product with the target DNA sequence.

The generation of concatamers of the region of interest through PCR was carried out using Platinum® *Pfx* DNA polymerase (Invitrogen). This uses a recombinant DNA polymerase from *Thermococcus* species strain KOD with proofreading (3' to 5' exonuclease) activity. Typical 50 µl reactions contained 1x *Pfx* amplification buffer, 1x PCR enhancer solution, 0.15 mM dNTPs, 1 mM magnesium sulfate, 7.5 nM of each forward and reverse primer, and 1 unit of Platinum® *Pfx* DNA polymerase. Designing primers that were tandem repeat and complimentary to each other negated the need for a separate DNA template.[131] These primers form a partial duplex which act as the substrate for DNA polymerase (Figure 2.2.1). This was carried out using either an Eppendorf Mastercycler or Techne Touchgene. After an initial denaturing step at 94°C for 2 minutes, there were 25 cycles of denaturation at 94°C for 30 s, annealing step at 55°C for 30 s and extension step at 68°C for 1 minute. The PCR products were held at 4°C until collection.

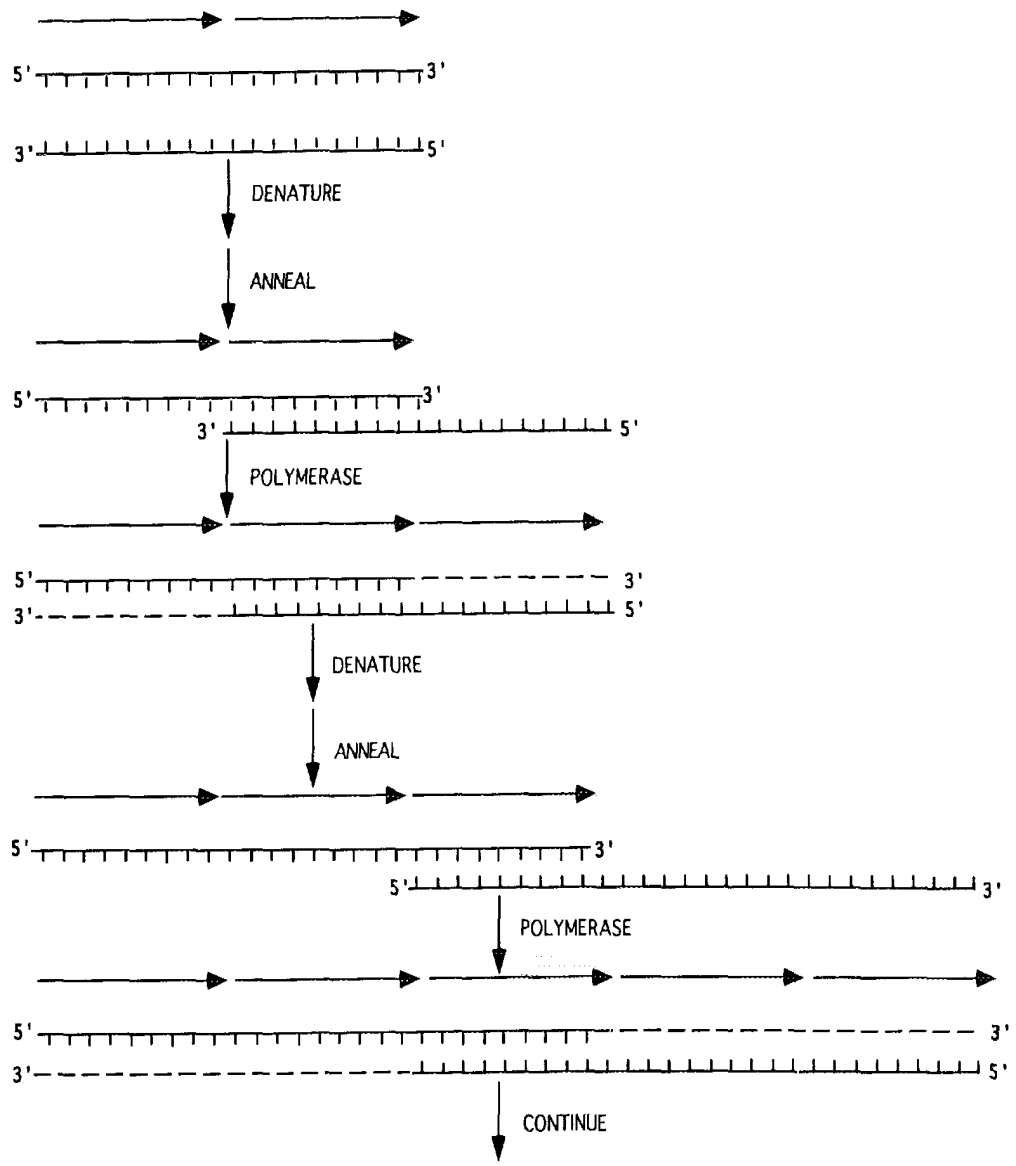


Figure 2.2.1: Synthesis of concatamers using PCR and partially duplex oligonucletoides. Adapted from[132].

The process of annealing complimentary oligonucleotides was achieved using the thermocycling function of the PCR machine. After an initial denaturing step at 95°C for 5 minutes, the temperature was reduced by 1°C per minute until the melting temperature (T_m) was reached. The reaction was held for 30 minutes at this temperature before reducing the temperature by 1°C per minute until the temperature reached 25°C. The annealed oligonucleotides were then held at 4°C until collection.

2.3 Agarose gel electrophoresis

Agarose gels were used to analyze the PCR product sizes and quality. 1% w/v agarose gels were routinely used. Agarose powder (Sigma-Aldrich) was melted in Tris-Acetate buffer (TAE) in a microwave oven for 2 minutes. Once the gel solution was hand-warm, 5 µl SafeView nucleic acid stain (NBS Biologicals) was added to every 100 ml gel solution. This was then poured into a cast and a comb was used to form wells once the gel had set. After setting, the gel was submerged in an electrophoresis tank filled with TAE buffer. Samples were mixed with 5x gel loading dye before being loaded into the wells. Samples were electrophoresed for 20 minutes at 80 mV. The DNA bands were then visualized on a UV transilluminator and captured digitally using the EDAS Geldoc system (Labtech International).

2.4 Immunoblot analysis of proteins

Immunoblot analysis of proteins is a separation of proteins according to their molecular weight using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), followed by immobilization onto a membrane before detection of specific proteins by enzyme conjugated antibodies.

Equal masses of protein samples were mixed with 10x NuPAGE reducing agent (Invitrogen) to reduce any disulfide bonds, thus removing tertiary and quaternary structures. 4x NuPAGE LDS sample buffer (Invitrogen) was then added to envelope the proteins in negative charges to denature the proteins as well as to weigh them down and act as a tracking dye. Samples were heated to 70°C for 10 minutes. To run SDS-PAGE gels, the NuPAGE (Invitrogen) system was used. This comprised of pre-cast polyacrylamide gels of defined concentrations. Samples were loaded into the wells and SeeBlue (Invitrogen) pre-stained molecular mass markers were added to each end of the gel. Electrophoresis was carried out in either 2-(N-morpholino)-ethanesulfonic acid (MES) or 3-(N-morpholino)-propanesulfonic acid (MOPS) buffer with NuPAGE antioxidant (Invitrogen) added. The gels were allowed to run at 200V until the dye front reached the foot of the gel.

Gels were then removed from their casts and transferred to Hybond P membranes (Amersham Biosciences) using the electrophoresis blot method. The gels and membranes were sandwiched between filter paper (BioRad) and sponges (Invitrogen) and held together between two electrodes. 1x transfer buffer (Invitrogen) containing 20% methanol (VWR) is then added to complete the circuit. The apparatus was left to run at 30V for 1 hour. Successful blotting can be observed by the presence of pre-stained molecular mass markers on the membrane.

Excess protein binding sites on the membranes were blocked by incubating them in 5% fat-free milk powder (Premier Brands) dissolved in phosphate buffered saline (PBS, Roche) with 0.1% Tween-20 (Sigma-Aldrich) rocking for 1 hour. Primary antibodies against the proteins of interest were diluted at defined ratios in PBS with 0.1% Tween-20 (PBS tween) and 5% fat-free milk and incubated with the membranes rocking for a defined period at 4°C. Membranes were then washed in PBS tween, once quickly followed by three

10 minute washes at room temperature. A secondary antibody (Santa Cruz Biotechnologies, supplied by Autogen Bioclear) against the primary antibody, conjugated to horseradish peroxidase (HRP), was then incubated with the membrane rocking for 1 hour at room temperature. They were diluted at defined ratios in PBS tween with 5% fat-free milk. This was followed by a similar set of washes with PBS tween.

Presence of HRP was detected using the Super Signal West Pico kit (Perbio) which utilizes enhanced chemiluminescence (ECL). HRP reacts with luminol in the presence of hydrogen peroxide and an azine enhancer to emit light. ECL reagents were left to react with the blots for 5 minutes before exposing them to Kodak Biomax film (Sigma-Aldrich) for 1 second or up to 1 hour, depending on the strength of the signal.

2.5 Nuclear and cytoplasmic fractionation

Nuclear extracts from SH-SY5Y and M17 neuroblastoma cell lines were obtained using the NE-PER® nuclear and cytoplasmic extraction reagents (Pierce). These enabled stepwise separation of cytoplasmic and nuclear extracts. The reagents were derived from work done by Dignam and colleagues where optimum salt concentration of low ionic strength and alkaline pH was used to extract the nuclei.[132] Before any extraction, fresh protease inhibitor pellets (Roche) were dissolved accordingly into the cytoplasmic and nuclear extraction reagents and kept on ice. Empty Eppendorfs were weighed so that weight of packed cells can be determined. Knowing this, it was then possible to scale-up or –down the extraction process. Addition of the first two reagents to the cell pellet results in the disruption of cell membranes and release of cytoplasmic contents. After recovering the intact nuclei through centrifugation at 16,000 x g for 5 minutes, the cytoplasmic supernatant was aliquoted immediately into prechilled Eppendorfs on dry ice before storing at -80°C. The remaining nuclei were lysed with a third reagent to yield the nuclear extract. After centrifugation, the nuclear extract was aliquoted into prechilled Eppendorfs on dry ice before storing at -80°C. Nuclear extracts obtained in this manner were compatible with a variety of downstream processes such as western blotting and gel shift assays.

2.6 Electrophoretic mobility shift assay (EMSA)

EMSA (gel shift or gel retardation assay) has been used extensively for studying the interaction of DNA-binding proteins with their cognate DNA recognition sequence.[133] This technique operates on the basis that DNA-protein complexes migrate slower compared to unbound DNA in a native polyacrylamide or agarose gel. This results in a shift in migration of the labeled DNA band (Figure 2.6.1).[134] The LightShift® chemiluminescent EMSA kit (Pierce) uses a nonisotopic method to detect DNA-protein interactions.[135] Biotin end-labeled DNA containing the binding site of interest was incubated with nuclear extract before being subjected to gel electrophoresis on a native polyacrylamide gel and then transferred to a positively-charged nylon membrane (Sigma-Aldrich). The biotin labeled DNA was detected using streptavidin-HRP conjugate and a chemiluminescent substrate. Specificity of DNA binding was established using competition experiments with an unlabelled DNA of the same sequence.

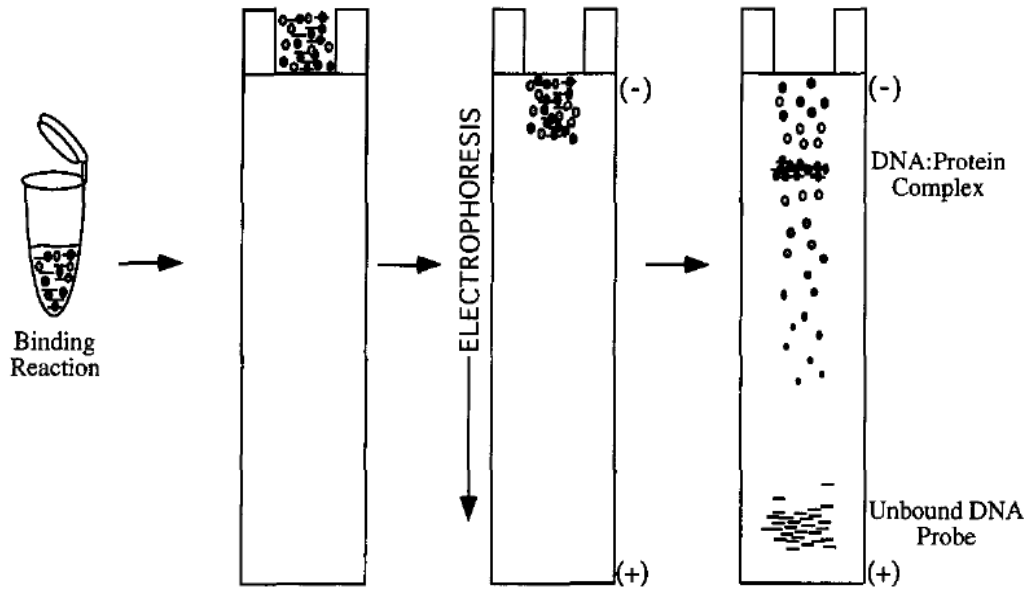


Figure 2.6.1: EMSA.

The protein extract (•) and DNA probe (-) are incubated before loading them onto a native polyacrylamide gel. This separates the bound DNA probe from the unbound DNA probe. Only complexes bound to the DNA probe and unbound DNA are visualized. Adapted from[135].

A typical 20 µl reaction volume consisted of 1x protein-binding buffer (Roche), 50 ng/µl Poly(deoxyinosinic-deoxycytidylic) acid [Poly (dI.dC)] (Roche or Sigma Aldrich) to inhibit non-specific binding of proteins present in crude nuclear extracts, 5 ng/µl Poly L-lysine (Roche or Sigma Aldrich) to stabilize protein-protein and protein-DNA interactions[136], 8 pmol of unlabelled target DNA (Invitrogen, for competition reactions), H1c-A (5'-TTCGCCCAGGGTACACCAGGACA-3'), H1b-G (5'-TTCGCCCAGGGTGCACCAGGACA-3'), 2.5 µl of nuclear extract from 2.5 and 20 fmol of biotin end-labeled target DNA (Invitrogen), H1c-A (5'-biotin-TTCGCCCAGGGTACACCAGGACA-3'), H1b-G (5'-biotin-TTCGCCCAGGGTGCACCAGGACA-3'). The reagents have been listed in the order they should be added. Before the crude nuclear extract was added, the buffer, non-specific DNA and stabilizer were allowed to equilibrate for 15 minutes at room temperature. After the nuclear extract was added, the reaction was allowed to equilibrate on ice for 15 minutes. After the biotin end-labeled target DNA was added, the reaction was left on ice for 30 minutes to allow for the formation of protein-DNA complexes.

Loading buffer was then added before the reaction was loaded onto native 6% Tris-Borate buffer (TBE) gels (Invitrogen). These gel matrices are effective in separating proteins from 15-500 kDa and 30-200 bp of DNA fragments. Current was set to 100V using Invitrogen's NuPAGE® mini-gel system until the bromophenol dye migrated approximately ¾ of the gel. Sequence-specific interactions are transient and they are stabilized by the low ionic strength of 0.5x TBE used.

Gels were then removed from their casts and transferred to BioBond™-Plus Nylon membrane (Sigma Aldrich) using the electrophoresis blot method. The gels and membranes were sandwiched between filter paper (BioRad) and sponges (Invitrogen) and held together between two electrodes. 0.5x TBE cooled with ice is then added to complete the circuit. The apparatus was left to run at 30V for 2 hours on ice. When the transfer was completed, transferred DNA was cross-linked to the membrane on a transilluminator with 312 nm bulb for 15 minutes.

Excess protein binding sites on the membranes were blocked by incubating them in Blocking buffer (Pierce) warmed to 45°C for 15 minutes with gentle shaking. 1:300

dilution of stabilized streptavidin-HRP conjugate (Pierce) in blocking buffer was then incubated with the membranes rocking for 15 minutes. Membranes were then washed in 1x wash solution warmed to 45°C, once quickly followed by four 5 minute washes. Membranes were then transferred to a new container and incubated in substrate equilibration buffer (Pierce) cooled to 4°C for 5 minutes with gentle shaking.

Presence of HRP was detected using the chemiluminescent nucleic acid detection module (Pierce) which utilizes enhanced chemiluminescence (ECL). HRP reacts with luminol in the presence of hydrogen peroxide and an azine enhancer to emit light. ECL reagents were left to react with the blots for 5 minutes before exposing them to Kodak Biomax film (Sigma-Aldrich) for 1 second or up to 1 hour, depending on the strength of the signal.

2.7 Cell culture

The SH-SY5Y (SKY5) and BE(2)-M17 (M17) human neuroblastoma cells were obtained from the European Collection of Cell Cultures (ECACC). The SKY5 cell line was a thrice-cloned sub-line originally derived from the SK-N-SH cell line from a bone marrow biopsy of a neuroblastoma patient in December 1970 (Figure 2.7.1).[137] It has dopamine-β-hydroxylase activity and converts glutamate to GABA.[138] It also has a substrate-adherent phenotype and naturally grows processes in culture. The M17 cell line was twice-cloned sub-line isolated from the SK-N-BE(2) cell line which was from a bone marrow biopsy of another neuroblastoma patient in November 1972.[139] These cells have tyrosine hydroxylase and dopamine-β-hydroxylase activity as well as specific uptake of noradrenalin. Intermediate filaments and vimentin were also expressed in M17 cells.

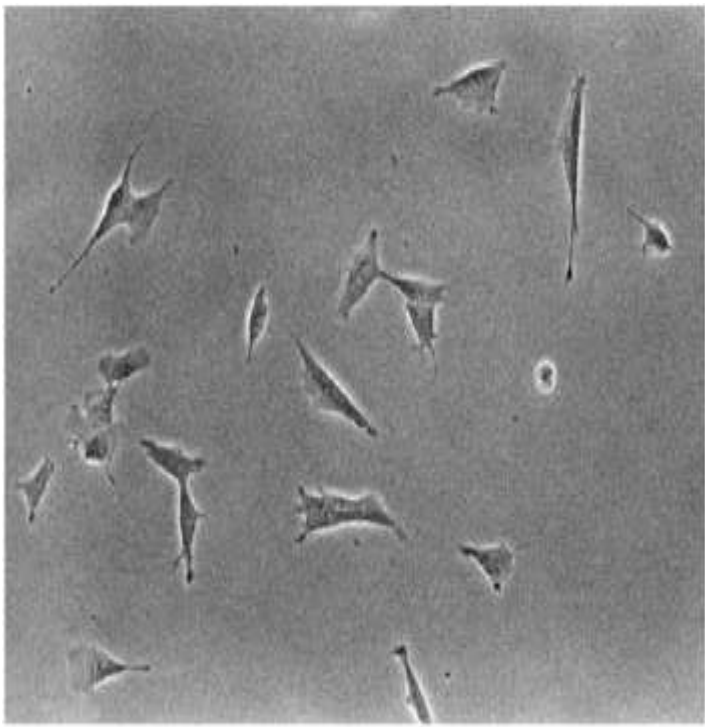


Figure 2.7.1: Undifferentiated SKY5 cells.
Roundish features with few short processes. Adapted from[140].

Both SKY5 and M17 cells were cultured in SKY5 culture medium (42% v/v Ham's F12 nutrient mixture, 42% v/v Eagle's minimum essential medium, 15% v/v fetal calf serum, 2 mM L-glutamine, 1% v/v non essential amino acids, 20 U/ml penicillin, 20 mg/ml

streptomycin, 250 ng/ml amphotericin B) maintained at 37°C in 100% relative humidity and 5% carbon dioxide. The growth medium was changed every two to three days.

Cells were subcultured by diluting (splitting) at a ratio between 1:4 and 1:100 after they reach 80% confluency, approximately seven days after subculture. Old culture medium was aspirated before rinsing the adherent cells with warmed 1x PBS solution (Sigma Aldrich) to remove any remaining culture medium. Cells are then detached by incubating them with a thin layer of trypsin-EDTA solution (Invitrogen) at 37°C for 5 minutes. Fresh culture medium was then added to quench the reaction. Cells were harvested to remove the supernatant containing trypsin. Cell pellets were then resuspended in warmed fresh culture medium and diluted accordingly to further seed culture flasks.

For medium term storage of cells, SKY5 and M17 cells were harvested by trypsinisation and pelleted by centrifugation at 800xg for 5 minutes and resuspending the cells in 1 ml of freezing medium. They were then transferred to polypropylene cryotubes and slowly frozen in a bath of isopropanol in a -80°C freezer. Long term storage of cells were placed in liquid nitrogen.

2.8 DNA-binding protein purification

DNA-binding proteins were purified using the DNA-binding protein purification kit (Roche) (Figure 2.8.1). Streptavidin magnetic particles provided had been coupled to double-stranded 16mer (5'-GGAATTCGGATCCGCG-3') oligonucleotide via a 3'-end labeled biotin at one site. The DNA region of interest, H1c-A (5'-TTCGCCCAGGGTACACCAGGACATTCGCCCAGGGTACACCAGGACA-3'), H1b-G (5'-TTCGCCCAGGGTGCACCAGGACATTCGCCCAGGGTGCACCAGGACA-3'), control oligonucleotides, R2 (5'-ATCAGTGGTAACCGTTAAATACCATCAGTGGTAACCGTTAAATACC-3') and R3 (5'-ATGCCTAGCAAGCAGTCGTTATA ATGCCTAGCAAGCAGTCGTTATA -3') was concatamarized using a self-primed PCR technique described in section 2.2. The generated DNA with multiple DNA-binding protein binding sites of interest was then ligated to the tethered oligonucleotide on the magnetic particles using DNA ligase. Nuclear extracts was then added to the magnetic particles to allow the DNA-binding proteins of interest to be captured. Unspecific DNA-binding proteins were removed by the addition of poly (dI•dC). Specific DNA-binding proteins were eluted using a magnetic separator and elution buffer of high ionic strength. Next, the elution buffer was removed from the proteins by microdialysis and the proteins analyzed by SDS PAGE.

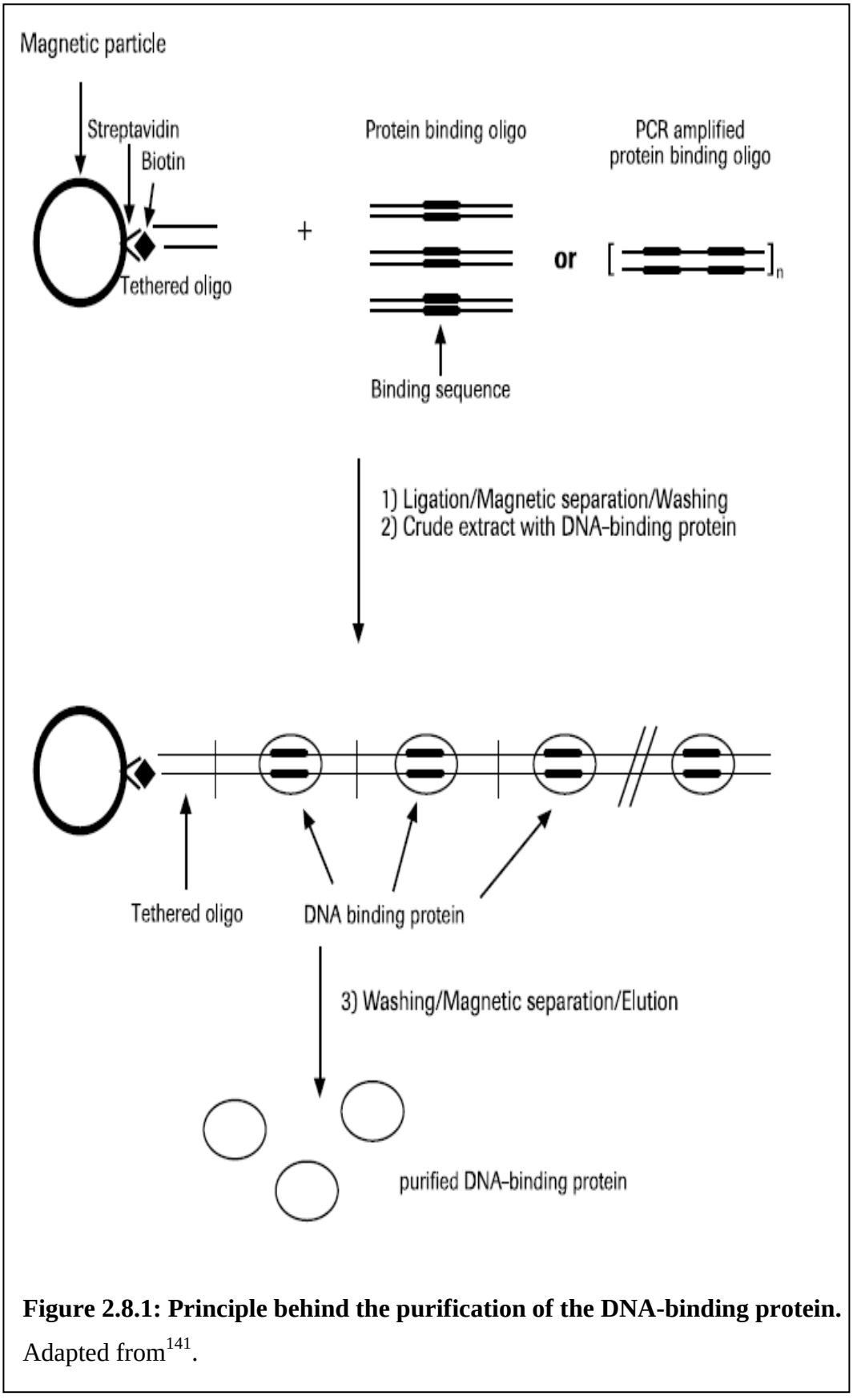


Figure 2.8.1: Principle behind the purification of the DNA-binding protein. Adapted from[141].

2.9 Mass spectrometry

As a general rule, protein bands that are visible with Coomassie stain have sufficient amount of protein to be analyzed by mass spectrometry (Figure 2.9.1). Proteins of interest purified using the method described in 2.8 were subjected to SDS-PAGE described in the beginning of 2.4.

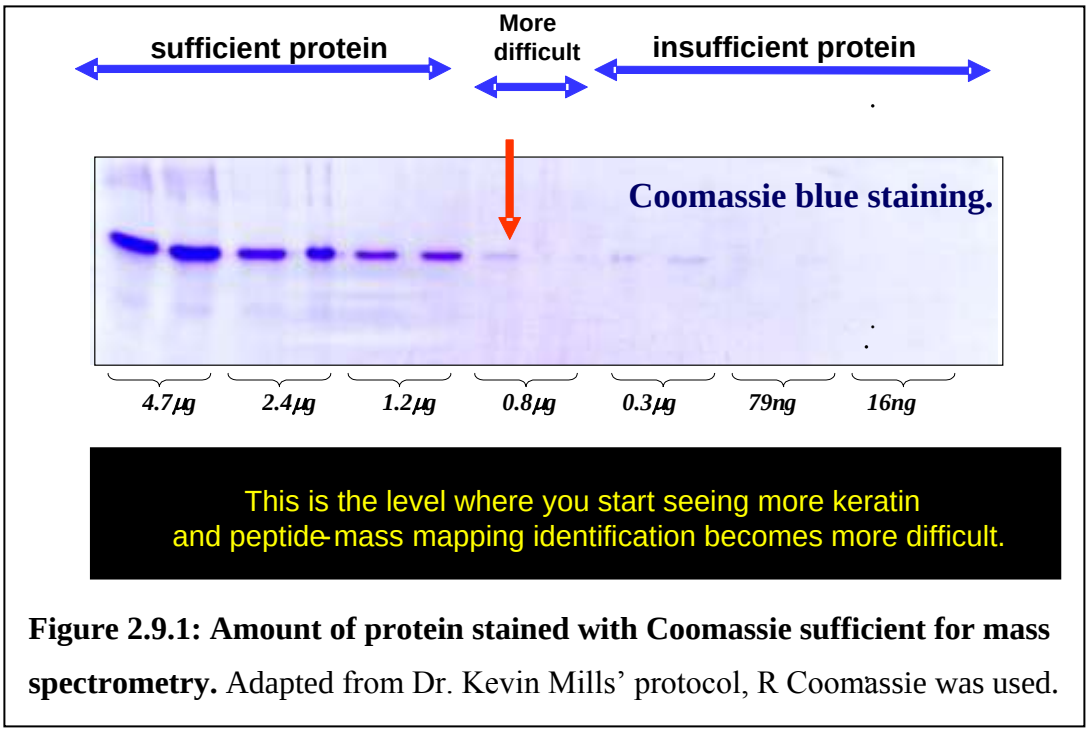


Figure 2.9.1: Amount of protein stained with Coomassie sufficient for mass spectrometry. Adapted from Dr. Kevin Mills' protocol, R Coomassie was used.

A clean scalpel was used to cut bands 2 mm long and 1 mm thick. The excised gel pieces were placed into silanised Eppendorfs (Sigma-Aldrich) and destained using 50% methanol (Sigma-Aldrich) containing 0.1% acetic acid (VWR) in a 37°C water bath. The gel pieces were then washed three times with 50 mM ammonium bicarbonate buffer, pH 7.8 (Sigma-Aldrich).

Proteins needed to be carboamidomethylated to prevent proteins from re-folding and to allow trypsin to digest the proteins efficiently. Before carboamidomethylation, any disulfide bridges needed to be broken using 1, 4-Dithioerythritol (DTE) (Sigma-Aldrich). First, the gel pieces were quickly dehydrated by shaking them in LC-MS grade acetonitrile (Sigma-Aldrich). For complete dehydration, they were placed in a centrifugal evaporator until gel pieces become brittle. Second, the gel pieces were then incubated in a 37°C water

bath with 50 mM ammonium bicarbonate buffer, pH 7.8 with 10 mM DTE for 1 hour. After washing them, 50 mM ammonium bicarbonate, pH 7.8 containing 55 mM iodoacetamide (Sigma-Aldrich) was then added and the gel pieces were left to incubate in the dark for 45 minutes.

$$-CH_2-SH + I-CH_2-CONH_2 \rightarrow -CH_2-S-CH_2-CONH_2 + HI$$

Cystine iodoacetamide S-carboamidomethyl cystine hydrogen iodide

Trypsin (Sigma-Aldrich) extensively purified from porcine pancreas was used to digest the protein samples. Trypsin is a highly specific serine endoprotease that cleaves on the carboxyl side of arginine or lysine residues. The rate of hydrolysis is reduced when an acidic residue is next to the cleavage site and mis-cleavage might occur when a proline is on the carboxyl side. Reductive methylation of the lysine residues made trypsin resistant to autolysis while TPCK treatment removed the chymotryptic activity of the enzyme. Trypsin solution was added to the gel pieces and incubated overnight in a 37°C water bath.

After digestion, the trypsin solutions were removed into new tubes. Hydrophilic peptides in the gel were extracted by the addition of 1% formic acid shaking for 20 minutes. Tryptic peptides were then extracted using 50% acetonitrile containing 1% formic acid shaking for 20 minutes. This was repeated once. All the extraction were then pooled together with the trypsin solution and freeze-dried.

Peptide samples were resuspended in 0.1% trifluoroacetic acid (TFA) (Sigma-Aldrich) to prevent peptides from associating with the tube walls. 1.5 µl of the sample was added to 1.5 µl of a U.V. absorbing co-matrix of α-cyano-4-hydroxycinnamic acid (CHCA) (Sigma-Aldrich) and 50 mM fucose (Sigma-Aldrich). Two 1.5 µl spots for each sample were then added to a clean mass spectrometry (MS) 96-well sample plate (Figure 2.9.2) to form a distinct drop.

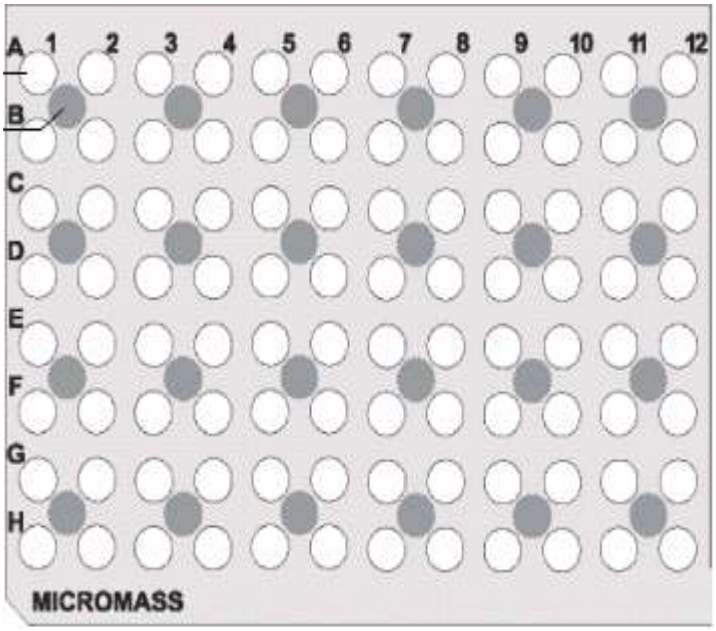


Figure 2.9.2: MS 96-well sample plate.

Samples are loaded onto the clear circles. Standards are loaded onto the shaded circles. Adapted from[142].

10 pmol/µl of protein standards such as glu-fibrinogen, 1570.6 Da (Sigma Aldrich) and bovine serum albumin (BSA) (Sigma-Aldrich), 66430 Da were used to calibrate the machine as well as polyethylene glycol (PEG) (VWR). The spots were allowed to dry before loading them into the MS machine.

The biological MS centre houses a quadrupole time-of-flight (QTOF) mass spectrometer (Waters® Micromass® MALDI Q-TOF Premier™). It is fitted with an interchangeable matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) sources.

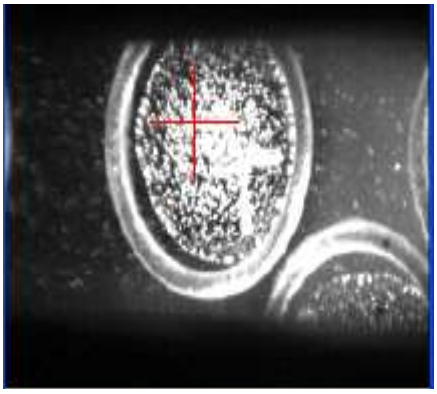


Figure 2.9.3: Camera picture of a spot of matrix and sample peptides. Adapted from[146].

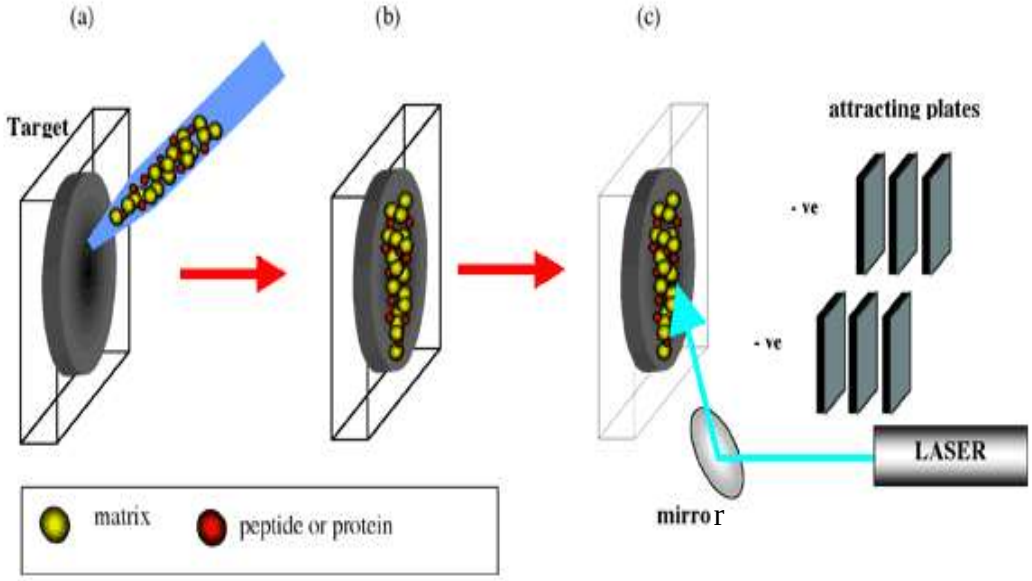


Figure 2.9.4: Principle of MALDI ionization. (a) peptide and matrix spotted on a grid (b) Mixture allowed to dry (c) UV laser fired onto mixture to ionize. Adapted from Dr. Kevin Mills' protocol.

For MALDI, once the spots were housed in the ionization chamber (Figure 2.9.3), the energy of the laser was absorbed by the co-matrix and resulted in the ionization of the sample peptides (Figure 2.9.4). This was a soft ionization technique where matrix material absorbed the UV laser beam and became ionized. The ionized matrix then charges the protein molecules.

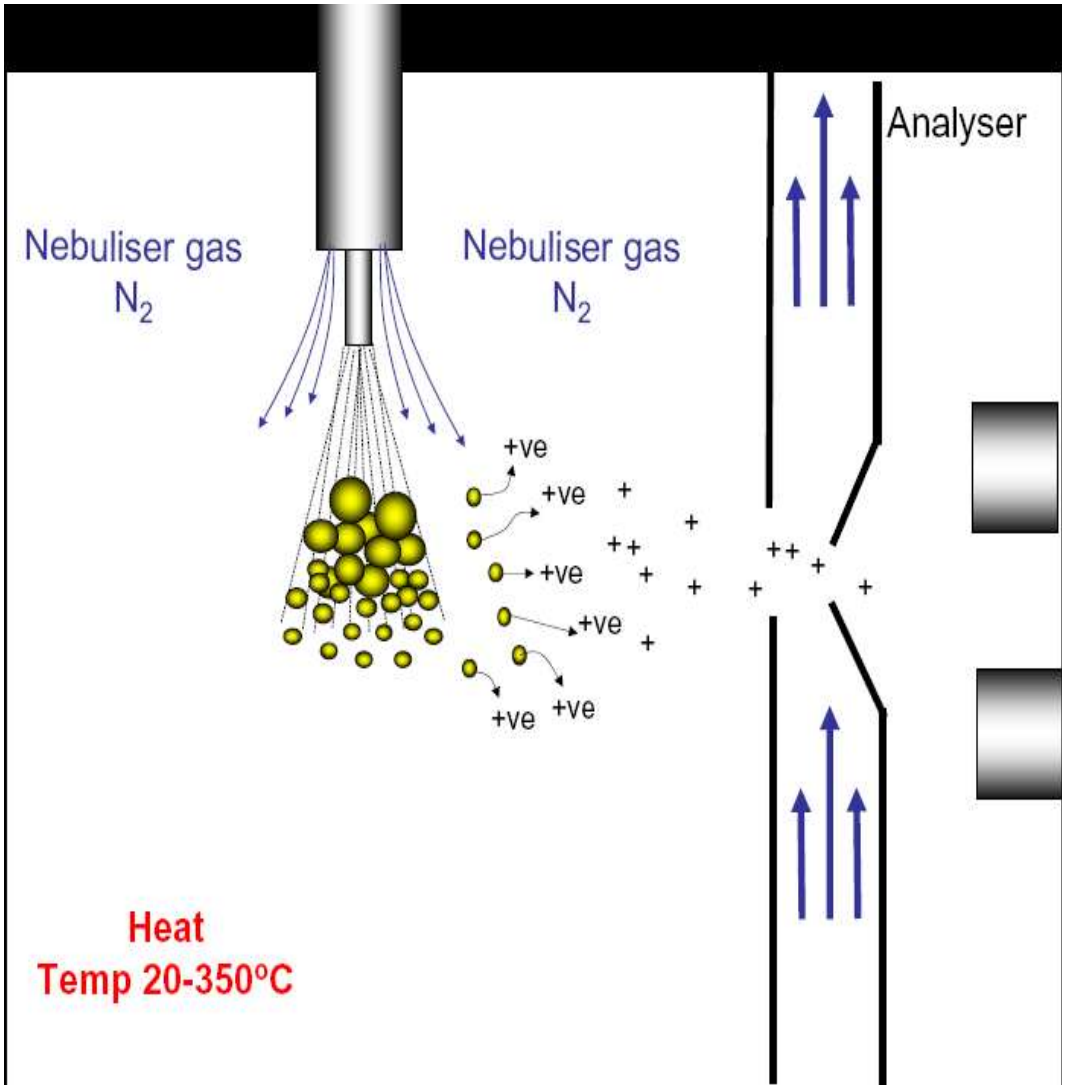


Figure 2.9.5: Principle of ESI ionization. Peptides are mixed with a volatile organic compound. The aerosol is sampled into the chamber through a capillary. The charge droplets shrink and become unstable until they emit charged jets. Adapted from Dr. Kevin Mills' protocol.

Nanoflow high pressure liquid chromatography (HPLC) (Waters® nanoACQUITY) combined with ESI-QTOF allowed for greater separation of peptides as they enter the mass spectrometer and hence a greater number of peptides that will be sequenced resulting in a larger sequence coverage of the unknown protein. Charged peptide micelles were pumped through a small aperture at high pressure resulting in multiply charged peptides (Figure 2.9.5).

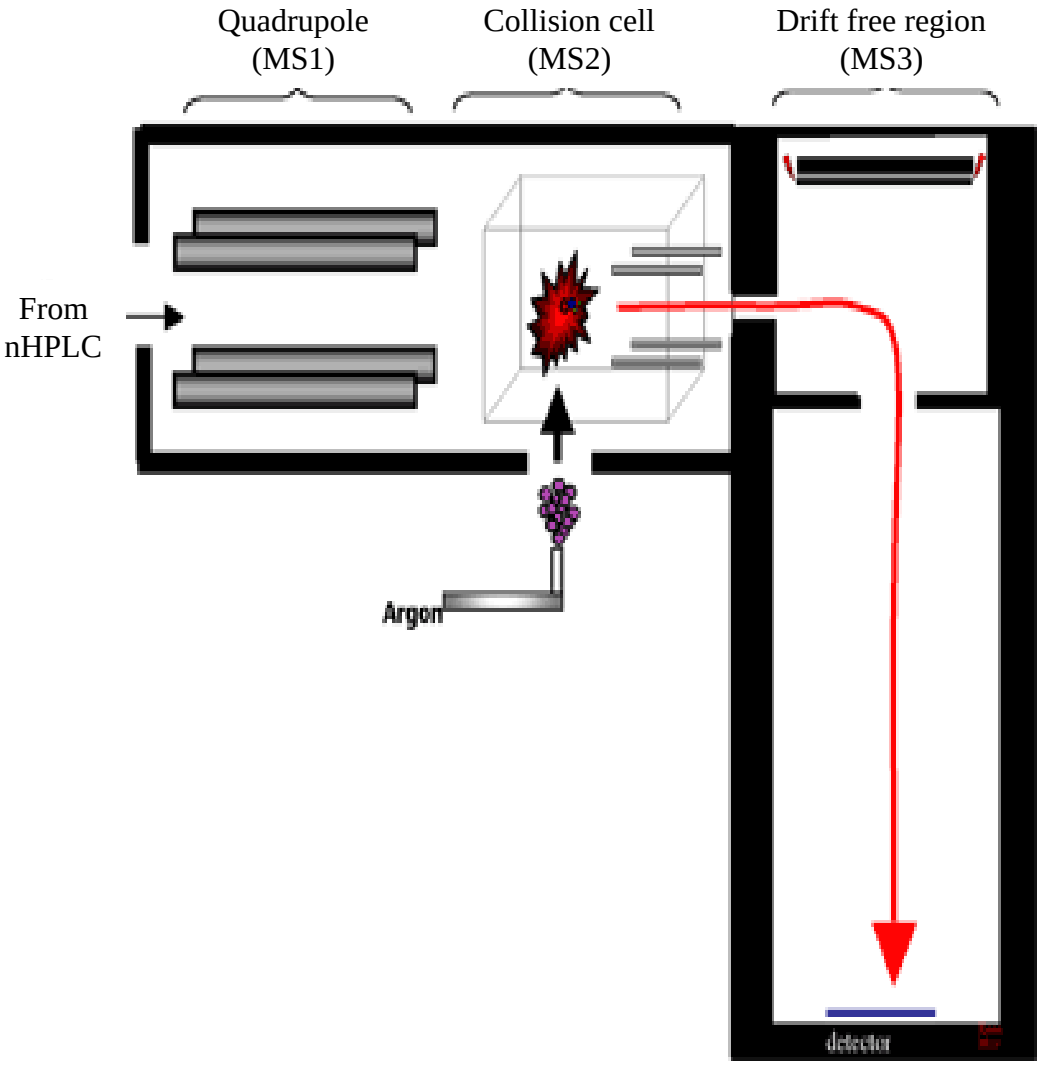


Figure 2.9.6: Principle of a QTOF analyzer.
Adapted from Dr. Kevin Mills' protocol.

Whether peptides were ionized through MALDI or ESI, they were then sequenced in the QTOF (Figure 2.9.6). A parent ion of a particular m/z value was selected in the first quadrupole section of the mass analyzer. Only this peptide was allowed into the collision cell where it was collided with inert argon gas resulting in several fragments of the peptide through collision-induced dissociation (CID). The masses of these fragment ions were then measured according to their time of flight.

2.10 Bioinformatics

The National Centre for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/) is a resource for molecular biology and genetics consisting of publicly available databases for nucleotide sequence data and polymorphism frequency data.

The University of California Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu/) is a useful resource for the visualization of an assembled reference human genome annotated with gene positions, polymorphic variation, repeats, cross-species conservation and structural variation. It also has programs for the identification of the location of nucleotide sequences on the genome (BLAT) and *in silico* PCR for the identification or PCR products generated of primer-pairs when using genomic DNA as a template.

The Japanese DNA transcription factor binding site prediction program (TFSEARCH) (www.cbrc.jp/research/db/TFSEARCH.html) was written by Yutaka Akiyama in 1995. It searches sequence fragments against TFMATRIX transcription factor binding site profile database in TRANSFAC databases developed at GBF-Braunschweig, Germany.[143]

MatInspector (http://www.genomatix.de/matinspector) is another software tool that uses a large library of matrix descriptions for transcription factor binding sites to locate matches in DNA sequences. It also assigns quality rating to the matches and allows for filtering and selection of matches.[144] This avoids redundant and false-positive matches.[145]

Mascot (www.matrixscience.com/mascot) is a search engine which uses mass spectrometry data to identify proteins from primary sequence databases. It carries out probability-based scoring based on Molecular Weight Search (MOWSE) score. Searches can be carried out using peptide mass fingerprint, sequence query or MS/MS ion search. Experimental mass values are compared with calculated peptide mass or fragment ion mass values from applying cleavage rules to the entries in a comprehensive primary sequence database. Such databases include MSDB, NCBInr, Swissprot and dbEST. Mascot is also able to consider multiple chemical modifications and missed cleavages.[146]

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) v8.0 (http://string.embl.de) is a metaresource that pools most of the available information on protein-protein associations. There are more than 50 million stored interactions from 2.5 million proteins from 630 organisms.[147] It weighs and integrates information from experimental repositories, public text collections and computational prediction methods.[148] These resources include SMART, PDB, COG, Ensembl, Intact, RefSeq, Medline, Reactome, DIP, BioGRID, MINT, KEGG, SGD, FlyBase, SwissProt/UniProt, HUGO, OMIM, NCI/Nature PID, The Interactive Fly, HPRD, BIND, EcoCyc, NCI-Nature Pathway Interaction Database, Gene Ontology (GO) protein complexes and all abstracts from PubMed.[149]

Reactome (http://www.reactome.org) is a manually curated resource for human pathway data. Every reaction comes from evidence in the biomedical literature with appropriate citations.[150] This could either come from direct assays on human cells or indirectly from experiments or observations in non-human species. It has grown into a collaboration between the Ontario Institute for Cancer Research, Cold Spring Harbor Laboratory, New York University School of Medicine and The European Bioinformatics Institute.[151] The SkyPainter, a tool used to map a user's list of genes or proteins of interest into reaction pathways has now been updated and replaced by the Pathway Analysis tool.[152] The overrepresentation analysis allows for clustering of the data into pathways, the higher the degree of clustering, the higher the probability of the proteins involved.

2.11 Native Western blotting of EMSA

Native Western blotting of proteins is the separation of protein complexes according to their native sizes and conformations using native polyacrylamide gel electrophoresis (native-PAGE), followed by immobilization onto a membrane before detection of specific proteins by enzyme conjugated antibodies.

A typical 20 µl reaction volume consisted of 1x protein-binding buffer (Roche), 50 ng/µl Poly(deoxyinosinic-deoxycytidylic) acid [Poly (dI.dC)] (Roche or Sigma Aldrich) to inhibit non-specific binding of proteins present in crude nuclear extracts, 5 ng/µl Poly L-lysine (Roche or Sigma Aldrich) to stabilize protein-protein and protein-DNA interactions[137], 2.5 µl of nuclear extract from 2.1.5 and 20 fmol of biotin end-labeled target DNA (Invitrogen). As controls, biotin end-labeled random DNA (Sigma) R2 (5'-biotin-ATCAGTGGTAACCGTTAAATACC-3') and R3 (5'-biotin-ATGCCTAGCAAGCAGTCGTTATA-3') were used instead. The reagents have been listed in the order they should be added. Before the crude nuclear extract was added, the buffer, non-specific DNA and stabilizer were allowed to equilibrate for 15 minutes at room temperature. After the nuclear extract was added, the reaction was allowed to equilibrate on ice for 15 minutes. After the biotin end-labeled target DNA was added, the reaction was left on ice for 30 minutes to allow for the formation of protein-DNA complexes.

Loading buffer was then added before the reaction was loaded onto native 6% Tris-Borate buffer (TBE) gels (Invitrogen). Current was set to 100V until the bromophenol dye migrated approximately ¾ of the gel. Sequence-specific interactions are transient and they are stabilized by the low ionic strength of 0.5x TBE used.

Gels were then removed from their casts and soaked in 2x transfer buffer (Invitrogen) for 10 minutes to aid in the transfer of protein complexes onto Hybond P membranes (Amersham Biosciences) using the electrophoresis blot method. The gels and membranes were sandwiched between filter paper (BioRad) and sponges (Invitrogen) and held together between two electrodes. 1x transfer buffer containing 20% methanol (VWR) is then added to complete the circuit. The apparatus was left to run at 30V for 2 hours.

Native protein complexes were fixed onto the membrane by incubating the blots in 8% acetic acid (VWR) for 15 minutes, rinsing in double distilled water and air-drying overnight. Before blocking, the membrane had to be re-wet with methanol and rinsed with double-distilled water. Excess protein binding sites on the membrane were blocked by incubating them in 5% fat-free milk powder (Premier Brands) dissolved in phosphate buffered saline (PBS, Roche) with 0.1% Tween-20 (Sigma-Aldrich) rocking for 1 hour. Primary antibodies against the proteins of interest were diluted at defined ratios in PBS with 0.1% Tween-20 (PBS tween) and 5% fat-free milk and incubated with the membranes rocking for a defined period at 4°C. Membranes were then washed in PBS tween, once quickly followed by three 10 minute washes at room temperature. A secondary antibody (Santa Cruz Biotechnologies, supplied by Autogen Bioclear) against the primary antibody, conjugated to horseradish peroxidase (HRP), was then incubated with the membrane rocking for 1 hour at room temperature. They were diluted at defined ratios in PBS tween with 5% fat-free milk. This was followed by a similar set of washes with PBS tween.

Presence of HRP was detected using the Super Signal West Pico kit (Perbio) which utilizes enhanced chemiluminescence (ECL). HRP reacts with luminol in the presence of hydrogen peroxide and an azine enhancer to emit light. ECL reagents were left to react with the blots for 5 minutes before exposing them to Kodak Biomax film (Sigma-Aldrich) for 1 second or up to 1 hour, depending on the strength of the signal.

2.12 Transfection & Luciferase assay

All transfections of SH-SY5Y cell line were carried out using TransFast™ transfection reagent (Promega). It is comprised of a synthetic cationic lipid (+)-N,N-[bis(2-hydroxyethyl)-N-methyl-N-[2,3-di(tetradecanoyloxy)propyl] ammonium iodide (Figure 2.12.1) and the neutral lipid L-dioleoyl phosphstidylethanolamine (DOPE) (Figure 2.12.2).

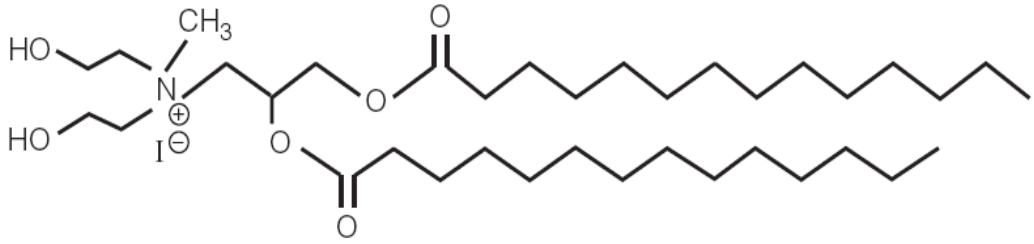


Figure 2.12.1: Structure of synthetic cationic lipid. Adapted from[153].

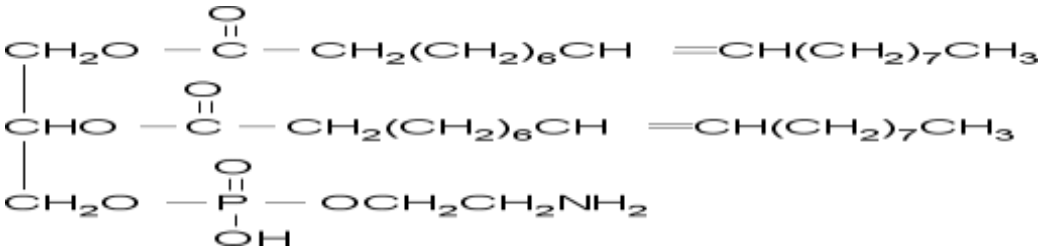


Figure 2.12.2: Structure of DOPE. From Sigma-Aldrich website.

These lipids form liposomes with water and engulf nucleic acids probably via electrostatic interactions. These neutralized micelles are then able to approach closer to the negatively-charged cell membrane. Entry into the cell may either be via endocytosis or fusion with the plasma membrane.

Before transfection, $1x10^4$ cells are seeded onto wells and grown overnight in SKY5 media. On the day of transfection, plasmids (contructed and prepared by another PhD student) only transfection was carried out by adding 200 ng of endotoxin-free plasmid DNA to every 1.5 µl of Transfast™ transfection reagent with 50 ng of Renilla plasmid in SKY5 media without fetal bovine serum. The mixture was then incubated at room temperature for 10 minutes. The cell culture medium was removed and replaced with this mixture. The cells were then incubated at 37°C for 1 hour. The wells were then overlaid

with SKY5 culture medium and incubated for at least 24 hours. Co-transfections were carried out similarly by including 1µl of 2 µM siRNA desired into the mix.

The Dual-Glo™ luciferase assay system (Promega) was used to quantify the action of specific DNA *cis*-elements on promoter transcriptional activity. A change in the transcription and translation of a reporter molecule can be tracked in a biological system as a marker for promoter activity. The promoter and regulatory sequences to be studied were inserted into the multiple-cloning site upstream of the reporter gene in the pGL4.10 reporter gene plasmid (Promega). This was then transfected into cell cultures. A gene from the firefly (*Photinus pyralis*) when translated, gives rise to the monomeric 61 kDa firefly luciferase enzyme. This acts as a reporter by catalyzing the oxidation of luciferin with ATP•Mg^{2+} as a cosubstrate which produces light in a linear relationship with enzyme activity (Figure 2.12.3).

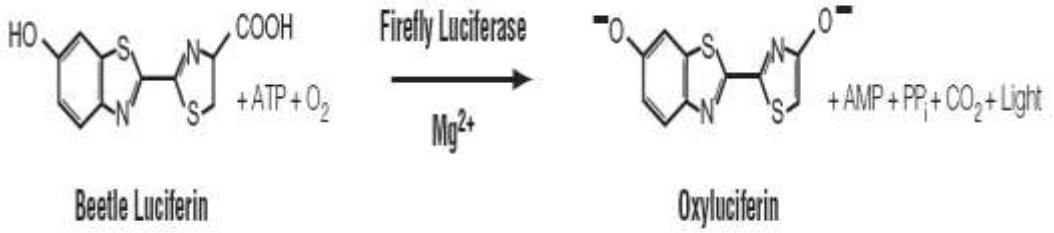


Figure 2.12.3: Bioluminescent reaction catalysed by firefly luciferase. Adapted from[154].

Readings were normalized by co-transfection with the pRL-CMV internal control plasmid which expresses the sea pansy (*Renilla reniformis*) luciferase. This monomeric 36 kDa enzyme catalyzes the oxidation of coelenterate-luciferin (coelentrazine) to produce light (Figure 2.12.4). This allowed control for varying transfection efficiencies.

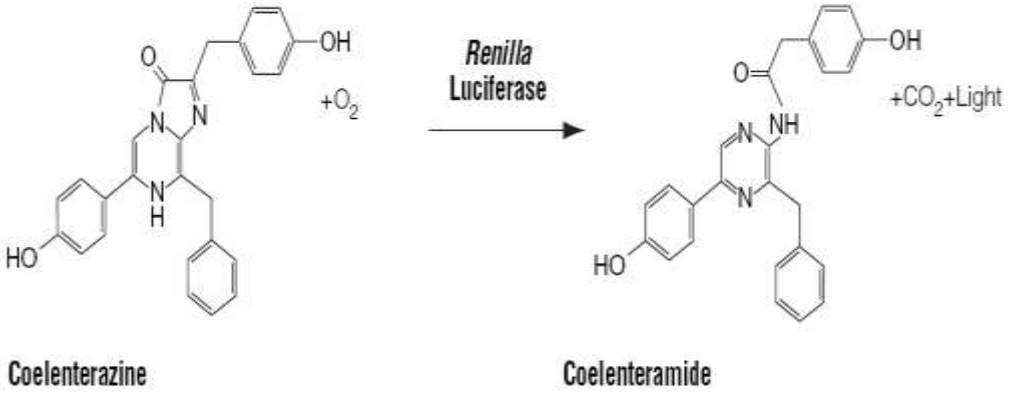


Figure 2.12.4: Bioluminescent reaction by Renilla luciferase.
Adapted from[155].

Luciferase assays were carried out on mammalian cell cultures seeded onto opaque 96-well plates (Greiner) in 200 µl of SKY5 media. 125 µl of the media was removed before 75 µl of the Dual-Glo™ luciferase reagent was added. It lysed the cells and acted as a substrate for firefly luciferase which produced a luminescent signal that was stable for over 2 hours. Luminescence was measured using a 96-well reader luminometer (Tecan GENios). The process was then repeated using 75 µl of the Dual-Glo™ Stop and Glo reagent to quench the firefly reaction and provide the substrate for the *Renilla* luciferase.

2.13 Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) is a method that is performed over several days and prone to variability in results. It is powerful for analyzing the direct or indirect protein interaction with DNA. The technique is based on formaldehyde's ability to form reversible Schiff's base cross-linkages between protein and DNA or protein and protein or both (Figure 2.13.1 A and B). This can be done either by acid or by increased temperature. Variability is introduced through the specificity of the antibody, the crosslinking conditions and digestion conditions. Antibodies used must be able to recognize formalin-fixed epitopes. These variabilities need to be empirically determined.

The first step of ChIP is to crosslink DNA to associated proteins using formaldehyde (Figure 2.13.1 C). Next, the DNA-protein complexes are extracted from the cells. Thirdly, these complexes are fragmented with micrococcal nuclease. The fourth step involves immunoprecipitation of DNA fragments through the binding of antibodies to the DNA-associated proteins. The immunoprecipitated DNA-protein complexes are then de-crosslinked in the fifth step. Lastly, the purified DNA fragments are then subjected to PCR with primers specific to the DNA regions of interest. It is important to use a large number of cells to improve the signal-to-noise ratio.

Sufficient quantity of 36.5% formaldehyde (Sigma F8775) was added to a flask of cells to obtain a final concentration of 1% formaldehyde. After mixing, the cells are left to incubate in the formaldehyde for ten minutes. 10x glycine solution (Pierce Agarose ChIP kit, cat no. 26156) was then added and left to incubate for five minutes to stop the reaction. After the media had been removed, the cells were washed twice with ice-cold PBS solution. The cells were then detached by scraping with ice-cold PBS along with protease and phosphatase inhibitor (Pierce Agarose ChIP kit). The cells were pelleted before proceeding to cell lysis and micrococcal nuclease digestion.

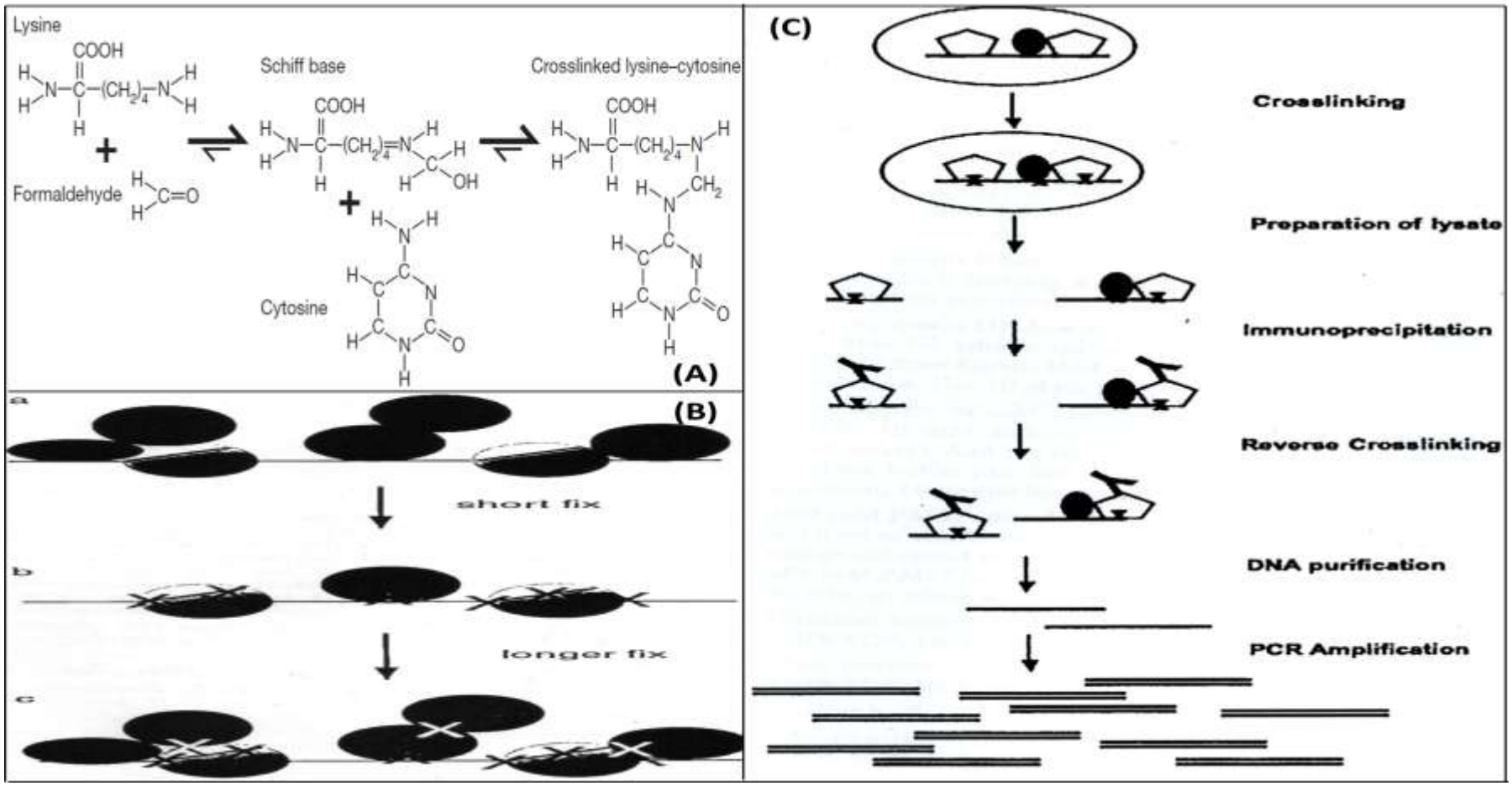


Figure 2.13.1: Schematic of formaldehyde crosslinking step of chromatin immunoprecipitation (ChIP) and ChIP workflow.
(A) Formaldehyde crosslinking chemistry. (B) Depending on the length of formaldehyde crosslinking, only direct DNA-protein interaction could be crosslinked (short fix – line: DNA and circles: proteins) or indirectly associated proteins could also be crosslinked (longer fix). (C) Six steps of a ChIP workflow. Adapted from[156,157,158].

The cells were lysed according to the manufacturer's instructions while the micrococcal nuclease digestion had been optimized previously. The digested chromatin was then immunoprecipitated with either the positive control antibody (anti-RNA polymerase II), or the negative control antibody (rabbit IgG) or the target specific antibody overnight. ChIP grade protein A/G plus agarose beads were then used to capture the immunoprecipitates. After washing, the beads were washed before elution. De-crosslinking was achieved by incubation in a 65°C water bath for 40 minutes before digesting the proteins using proteinase K. DNA fragments were then recovered using a DNA clean-up column before proceeding with PCR.

3 Chapter 1 Results

3.1 Subcellular fractionation of nuclear and cytoplasmic extracts

As a lot of the work involved DNA-binding proteins and SNP rs242557, a reproducible and effective means of isolating nuclear extracts from human neuroblastoma cell lines was needed. Obtaining nuclear extracts was carried out using a commercially available kit from Pierce (NE-PER® nuclear and cytoplasmic extraction kit). This ensured that there was little variation between the different batches of nuclear extracts. SH-SY5Y (SKY5) and BE3-M17 (M17) human neuroblastoma cell lines were used for their neuron-like properties. To validate the purity of cytoplasmic and nuclear extracts derived from SKY5 and M17 cells, antibodies against glyceraldehyde-3-phosphate (GAPDH) and histone deacetylase 1 (HDAC1) were used, respectively. All western blots were carried out in independent biological triplicate experiments as shown and the bands quantified using Bio-Rad's Quantity One software.

GAPDH is primarily located in the cytoplasm where it catalyses the reversible oxidative phosphorylation of glyceraldehyde-3-phosphate. During apoptosis, however, GAPDH translocates into the nucleus. Hence, as can be seen in Figure 3.1.1, even though there was more GAPDH in the cytoplasmic extract of both cell lines, this was approximately four (M17) to six (SKY5) as much as those seen in the nuclear extracts which was a significant ($p<0.05$) cytoplasmic protein enrichment.

HDAC1 is a nuclear protein which catalyses the deacetylation of histones. This is associated with transcriptional silencing. In Figure 3.1.1, the nuclear extracts were enriched with three (M17) to nine (SKY5) times more HDAC1 as compared to the cytoplasmic extracts. Although there was no significant ($p=0.08$) enrichment of nuclear proteins with M17, SKY5 did show significant ($p<0.05$) and robust enrichment. This shows that the nuclear extraction kit is effective in fractionating and enriching the nuclear proteins from the cell line SH-SY5Y.

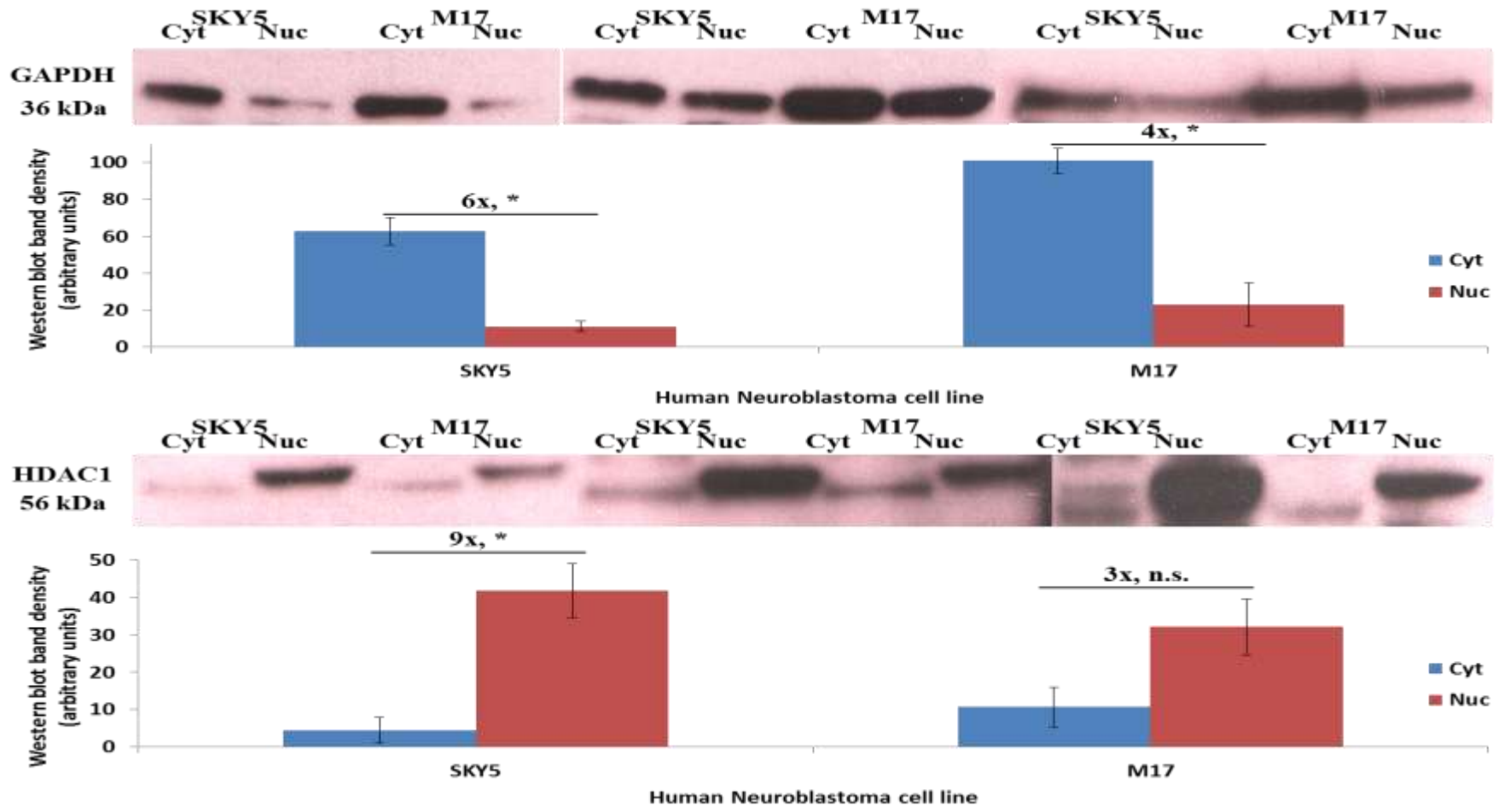


Figure 3.1.1: Western blot of undifferentiated SKY5 & M17 cytoplasmic (Cyt) and nuclear (Nuc) extracts with polyclonal cytoplasmic- (anti-GAPDH) and nuclear- (anti-HDAC) specific antibodies. These were done in triplicate, *: $p<0.05$, n.s.: non-significant (Two-tailed T-test).

3.2 Chemiluminescent electrophoretic mobility shift assay (EMSA)

Biotin-labelled 23-mer oligonucleotide probes containing the H1c-A and H1b/H2-G allelotypes of rs242557 were incubated with nuclear extracts of undifferentiated SH-SY5Y. Complexes containing the probe and DNA-binding proteins migrated slower through the native gel compared to the free probe. Non-specific binding was blocked using poly-(dI.dC) oligonucleotides. Specificity of binding between DNA-binding proteins and the probe containing rs242557 was demonstrated using unlabelled probe, which competed out binding to the labelled probe.

The results in Figure 3.2.1 showed that the H1b/H2-G allelotype of rs242557 was able to recruit more DNA-binding proteins as compared to the A allelotype. On an average of five independent experiments, the H1b/H2-G allelotype pulled down four fold more nuclear proteins than the H1c-A allelotype normalised against the amount of free biotin probe ($p<0.01$).

This result was promising as it demonstrated a potential biological significance of a SNP that has been associated with increased risk of sporadic PSP, CBD and AD. The next step was to characterize these DNA-binding proteins by identifying them and to investigate if there were any nuclear proteins that differentially bind to the H1c-A or H1b/H2-G allelotypes.

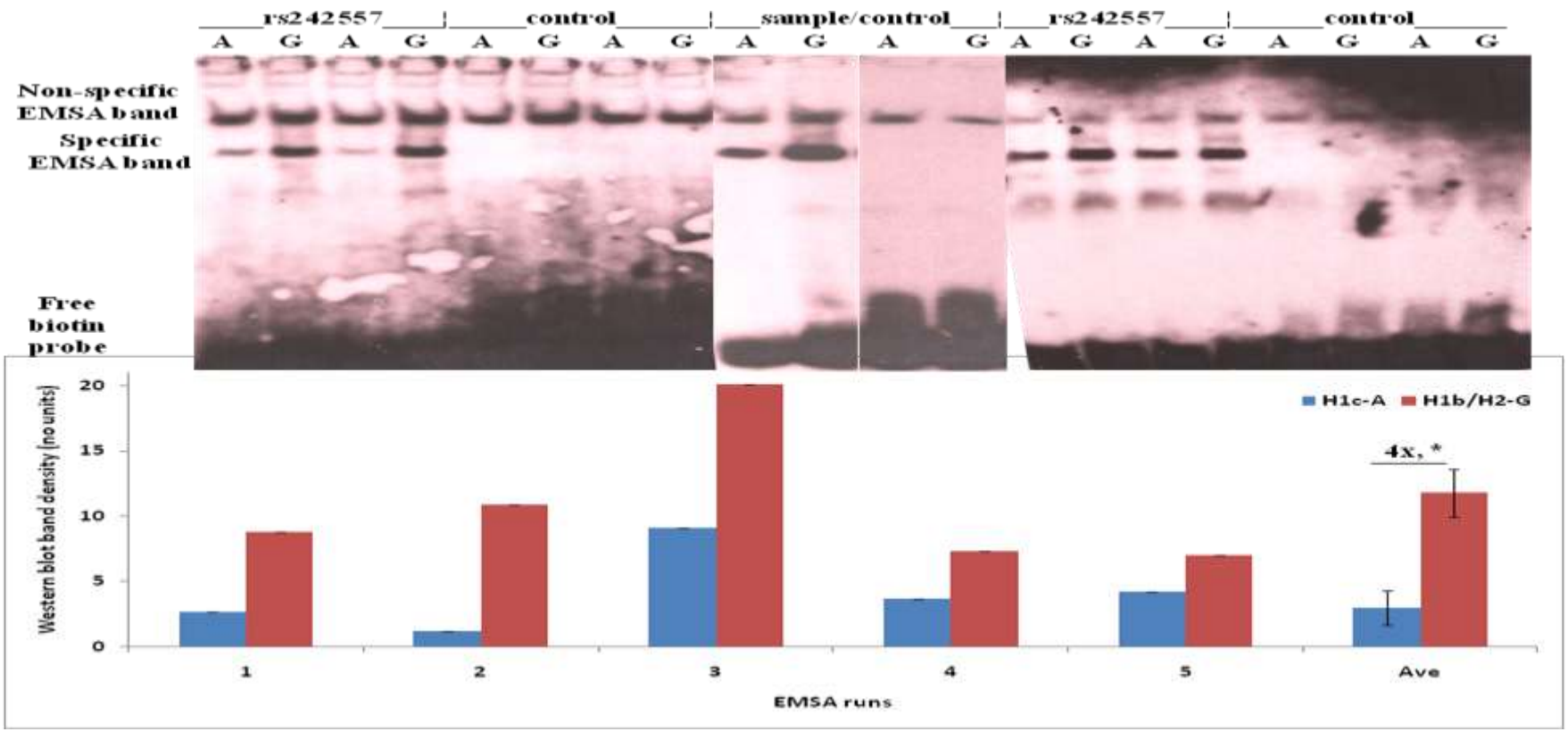


Figure 3.2.1: EMSA of rs242557 H1c-A or H1b/H2-G variant.

Biotin-labelled 23-mer oligonucleotides containing the A (H1c) and G (H1b & H2) allelotypes of rs242557 were incubated with nuclear extracts from undifferentiated SH-SY5Y cells. Incubation with unlabelled probe confirmed the specific interaction between probe and nuclear proteins. On average, the H1b/H2-G allelotype had 4-fold more blot density than the H1c-A allelotype. Five replicates were done. As replicate 3 was an outlier, only repliactes 1-2 and 4-5 were used for the average and standard deviation. *: $p<0.01$, two-tailed T-test.

3.3 Affinity purification of DNA-binding proteins

To purify these DNA-binding proteins, much larger quantities of nuclear extracts were used. In order to maximize capture of the nuclear proteins of interests, concatamers of the 23-mer oligonucleotide binding site containing either the H1c-A or H1b/H2-G allelotype of rs242557 were created using PCR (see 2.1.2). These were ligated to DNA linkers that were biotin-streptavidin linked to magnetic beads using T4 DNA ligase and incubated with the nuclear extracts. The beads were washed thoroughly to remove unbound proteins and eluted with a 3.5 M potassium chloride salt elution.

As illustrated in Figure 3.3.1, there was not much difference between the washes and the supernatants indicating that only a few specific proteins were bound to the probe. When the SDS-containing sample buffer was added to the elution, however, a SDS-soluble fraction (Figure 3.3.1) and a less SDS-soluble fraction formed. More SDS-containing sample buffer, heating and vortexing were needed to solubilise the less SDS-soluble fraction. Elution bands from both gels that were of sufficient intensity and/or of different intensity between H1c-A and H1b/H2-G were sliced with a clean scalpel and analysed by mass spectrometry (see 3.4).

We were now able to observe in greater detail the allele-specific differences of discrete protein bands of promoter trapped DNA-binding proteins. Although the eluted proteins were spread across two gels, looking from the protein identification, only two proteins were common to both at the same gel band, poly ADP ribose polymerase 1 (PARP1) and actin. Actin, however, had more of it bound to the H1b/H2-G allelotype of rs242557 for both gels and was identified as the only protein within those gel bands. Bands 34 and 31 kDa (Histone H1.5 and fumarate hydratase as identified later) were similar to actin in being the only protein present in a band that could be identified and bound more strongly to H1b/H2-G. Other gel bands that had more than one protein identified made it difficult at this stage to decipher which protein was responsible for such allelic differences. Bands 95, 89, 64, 44, 42, 34, 32, 28 kDa, however, were bands like actin that had one protein confidently present in each band but unlike actin interacted with the H1c-A allelotype of rs242557 more strongly. These bands were subsequently identified (see 3.4) as DNA topoisomerase 1 (TOP1), nucleolin (NUCL), replication protein (RPA70),

heterogeneous ribonucleoprotein D0 (ROD), nucleolar RNA helicase II (DDX21), H1 histone (H1.2 or H1.4), high mobility group protein 2 (HMGB2).

Thus, one SNP seems to have the ability to vary protein-DNA interaction subtlely as in the case of the HMG proteins or more pronouncedly as in the case of DDX21. How these proteins were identified and the significance of their identifications will be discussed in the following section.

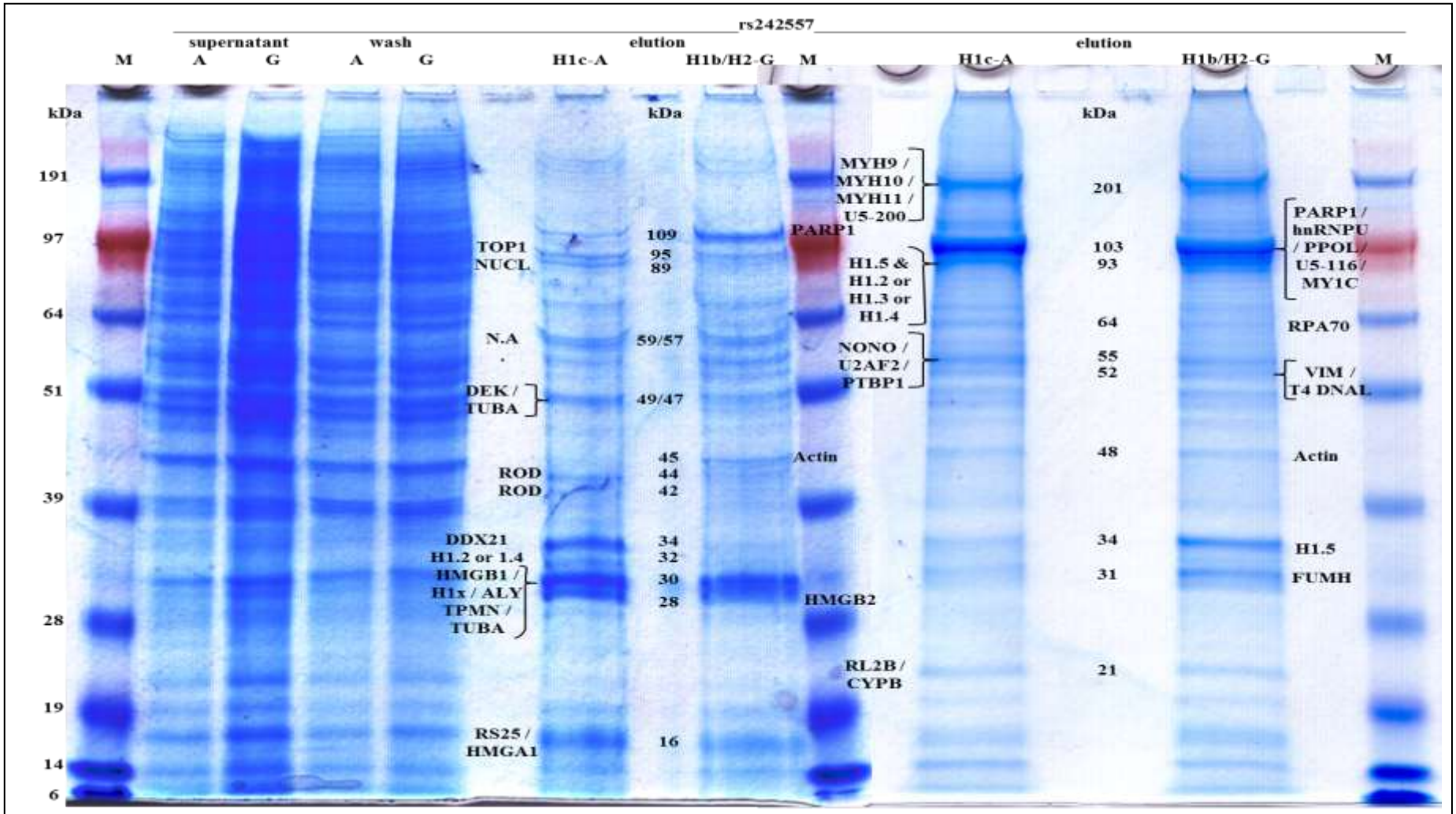


Figure 3.3.1: SDS-PAGE gel of DNA-binding proteins purified using concatamerised probes of the region containing the H1c-A and H1b/H2-G allelotypes of rs242557 incubated with nuclear extracts of undifferentiated SH-SY5Y, less SDS-soluble fractions on the right.

3.4 Identification of DNA-binding proteins

To identify the allele-specific binding of proteins to rs242557, as many discrete bands that were visible to the eye were cut carefully using a clean scalpel. A general guide for proteins with sufficient yield for MS sequencing was that they were visible on Coomassie stained protein gels. As these proteins had been separated on a 1-D gel, each protein band might include more than one protein of interest. Performing MALDI-QTOF on the bands characterized the most abundant protein within a band based on the molecular weight of the digests. Hence it was necessary to carry out ESI-QTOF/TOF sequencing the digested peptide fragments and was sensitive enough to characterize other proteins within the same protein gel band. MS data generated from the runs were translated into protein hits using the manufacturer's software. In total, there were 43 protein hits from 22 protein bands, 37 of which were unique (Table 3.4.1). Ten protein bands yielded single protein identifications while 11 protein bands yielded multiple protein identifications between two to five protein matches. One protein band did not have sufficient peptides for any positive identification. It was more difficult to positively identify some of the histone linker H1 proteins as there were different isoforms and the peptides sequenced did not match to the unique regions of any isoforms (H1.2 – 1.4). The known molecular weights of most of the proteins identified matched well with the observed molecular weights of the gel bands from which they were identified from. The exceptions were the 20 kDa H1 proteins that were the only proteins identified in the 93 kDa band, the 80 kDa nucleolar RNA helicase II in the 34.1 kDa band and the 55 kDa fumarate hydratase in the 30.7 kDa band. The smaller molecular weight proteins found in a larger gel band might be due to the different isoforms of H1 interacting with each other or multimers of the same isoform or they might be interacting with another protein that was not identified. A larger protein might have been truncated or fragmented or partially degraded that might have resulted in it being identified in a smaller molecular weight gel band.

A positive identification of a protein depends on various factors: the abundance of protein within a gel band, the efficiency of the trypsin digestion, the condition of the peptide, the size and composition of the peptide. Waters analysis and scoring system, ProteinLynx Global Server (PLGS) used the selectivity of the exact mass measured data, the specificity of the analysis coupled with filtering and scoring routines to generate the

PLGS score. These were also manually verified to prevent mis-assignment of spectra. A score above 10 was usually considered as highly confident identification. For example, hnRNPU was identified from a one peptide that represented 2.2% of the protein while myosin heavy chain was identified from 55 peptides that accounted for 30.5% of the protein. Both were assigned a score of 11.6. The specificity and uniqueness of a peptide to protein match was significant regardless of protein coverage. Even 13 peptides that represented 44.5% coverage for actin was assigned a score of 8.4, possibly due to the poor fragmentation data for some of the peptides or the lack of specificity of the peptide match.

A positive control was found with the identification of bacteriophage T4 DNA ligase. It was used to ligate the concatamers of the rs242557-containing oligonucleotides to the linker DNA attached to the magnetic beads. Twenty four peptides covering 46% of the protein with a PLGS score of 11.6 gave a confident identification for this protein. However, the fact that the ligase carried over could have led to spurious non-specific protein-protein interactions with some of the other proteins in the pull-down assay.

No	Band ID (kDa)	Accession no.	Protein ID	Gel bands were from / Abb.	Theoretical MW (Da)	Peptide sequence or % coverage	PLGS score
1	201.5	P35580	Myosin Heavy Chain	2 / MYH10	228796	30.5% coverage	11.6
		O75643	U5 snRNP 200 kDa helicase	2 / U5-200	194354	3.4% coverage	11.6
		P35579	Myosin heavy chain nonmuscle A	2 / MYH9	226390	25.4% coverage	11.6
		P35749	Smooth muscle myosin heavy chain	2 / MYH11	227339	4.0% coverage	11.2
2	109.4	P09874	Poly ADP ribose polymerase	1 / PARP1	112881	33.5% coverage	11.6
3	103.5	P09874	Poly ADP ribose polymerase	2 / PARP1	112881	39.5% coverage	11.6
		O00159	Myosin 1c	2 / MY1C	117965	3.7% coverage	11.6
		Q00839	hnRNP U	2 / hnRNPU	90422	2.2% coverage	11.6
		Q9H9Y6	DNA directed RNA polymerase I	2 / PPOL	122058	2.4% coverage	11.6
		Q8IXJ3	U5 snRNP 116 kDa	2 / U5-116	95372	6.2% coverage	10.6
4	95.4	P11387	DNA topoisomerase I	1 / TOP1	90668	See Figure 3.4.5	11.6
		P19338	Nucleolin Protein C23	1 / NUCL	76167	GLSEDTTEETLK GFGFVDFNSEEDAK	10.2
5	93	P16403	Histone H1.2 /H1d	2 / H1.2	21220	ASGPPVSELITK SGVSLAALK	7
		P10412	Histone H1.4 /H1b	2 / H1.4	21720		
							6.9
		P16402	Histone H1.3 /H1c	2 / H1.3	22205		
							6.9

		P16401	Histone H1.5 /H1a	2 / H1.5	22435	ATGPPVSELITK	10.8
6	64	P27694	Replication protein A 70kDa	2 / RPA70	68095	25 % coverage	11.6
7	59/57	N.A.	Not enough peptides	1 / -	N.A.	N.A.	N.A.
8	55	Q15233	NONO	2 / NONO	54066	44.5 % coverage	11.6
		P26368	U2AF2	2 / U2AF2	53467	16.2% coverage	10.9
		P26599	PTBP1	2 / PTBP1	57185	4.3% coverage	10.2
9	52.5	P08670	Vimentin	2 / VIM	53522	7.3% coverage	11.6
		P00970	DNA Ligase from Bacteriophage T4	2 / T4DNAL	55257	46 % coverage	11.6
10	49/47	P35659	DEK protein	1 / DEK	42647	45.9 % coverage	11.6
		P04687	Tubulin a1 chain brain specific	1 / TUBA	50125	4.4% coverage	8.9
11	48.3	P02570	Actin	2 / Actin	41709	44.5 % coverage	8.4
12	44.5	P02570	Actin	1 / Actin	41709	36.3% coverage	10.0
13	43.9	Q14103	hnRNP D0	1 / ROD	38410	DLKDYFSK FGEVVDCTLK IFVGGLSPDTPEEK	11.6
14	42.2	Q14103	hnRNP D0	1 / ROD	38410	IFVGGLSPDTPEEK	9.3
15	34.3	P16401	Histone H1.5 /H1a	2 / H1.5	22435	ATGPPVSELITK	10.8
16	34.1	Q9NR30	Nucleolar RNA helicase II	1 / DDX21	79607	APQVLVLAPTR	11.0
17	31.6	P16403	Histone H1.2 /H1d	1 / H1.2	21220	SGVSLAALK	8.3
		P10412	Histone H1.4 /H1b	1 / H1.4	21720		8.3

18	30.7	P07954	Fumarate Hydratase	2 / FUMH	54602	YAMT RIKAAMPR	11.6
19	30.0	P09429	High mobility group protein 1	1 / HMGB1	24747	48.1% coverage	10.2
		Q92522	Histone H1x	1 / H1x	22473	5.6% coverage	11.6
		P12324	Tropomyosin cytoskeletal type 3	1 / TPMN	29014	13.3% coverage	10.9
		Q86V81	Thoc4/ALY	1 / ALY	26871	24.9% coverage	11.2
		Q8WU19	Tubulin 1B	1 / TUBA	37194	11.3% coverage	10.7
20	28.2	P26583	High mobility group protein 2	1 / HMGB2	23887	48.1% coverage	10.5
21	21.6	P29316	60S ribosomal protein L23a	2 / RL2B	17684	LAPDYDALDVANK	11.6
		P23284	Peptidyl prolyl cis trans isomerase E	2 / CYPB	22728	TVDNFVALATGEK	10.2
22	16	P25111	40S ribosomal protein S25	1 / RS25	13733	LITPAVVSER AALQELLSK	11.6
		P17096	High mobility group protein A1	1 / HMGA1	11538	KQPPVSPGTALVGSQK	10.9

Confirmed in band Possible proteins

Table 3.4.1: List of protein hits from MS data.

Proteins highlighted in green had been positively identified within the protein bands excised using PLGS. Those in yellow had the possibility of being present in the bands. It was possible for some of the 1-D protein bands to yield more than 1 protein identification. ProteinLynx Global Server (PLGS) was Waters analytical platform which matched MS spectra against those generated from proteins in the Swissprot and Uniprot databases and generated the PLGS score. The higher the score, the higher the confidence in the match.

A closer look at the analysis of the peptide sequencing (Appendix, figures 3.4.2 – 3.4.19) revealed that some of the matches were to exact sequences (coloured blue), partial match (coloured red), modified match (coloured green), partial modified match (coloured yellow). For those proteins that were identified that had extensive peptide coverage such as myosin (Appendix, figure 3.4.2), PARP1 (Appendix, figures 3.4.3 & 3.4.4), RPA70 (Appendix, figure 3.4.8), Non-POU domain-containing octamer-binding protein (NONO) protein (Appendix, figure 3.4.9), actin (Appendix, figure 3.4.11), HMGB2 (Appendix, figure 3.4.17), the coverage for each protein was fairly evenly distributed across the protein sequence and not skewed.

For the rest of the proteins without extensive peptide coverage, a closer look into the fragmentation patterns (b and y ions) was done. The b or y ions were formed from the most common cleavage site along the peptide backbone (CO-NH). Such a cleavage resulted in a neutral and a charged species. When the charge remained on the N-terminal fragment, b ions were generated while y ions were the result of the charge being on the C-terminal fragment. For most of these proteins such as DNA topoisomerase I (Appendix, figure 3.4.5), nucleolin (Appendix, figure 3.4.5 continued), hnRNPD0 (Appendix, figure 3.4.13), nucleolar RNA helicase II (Appendix, figure 3.4.14), 60S ribosomal protein L23a (Appendix, figure 3.4.18), peptidyl prolyl cis trans isomerise E (Appendix, figure 3.4.18 continued) and 40S ribosomal protein S25 (Appendix, figure 3.4.19), they had well-defined y fragmentation pattern for each of the amino acids but their b fragmentation patterns were not as defined. H1.5 (Appendix, figure 3.4.6), H1.2/1.3/1.4 (Appendix, figure 3.4.7), T4 DNA ligase (Appendix, figure 3.4.10), hnRNPD0 (Appendix, figure 3.4.12) and HMGB1 (Appendix, figure 3.4.16), however, had both clearly defined b and y fragmentation patterns which yield more confidence in their protein prediction. Fumarate hydratase (Appendix, figure 3.4.15) and HMGA1 (Appendix, figure 3.4.19 continued) neither had good b and y fragmentation patterns and since fumarate hydratase was of a different molecular weight to the gel band, its protein identification might be less significant.

3.5 Bioinformatics on DNA-binding proteins

To stratify the 36 unique proteins that had been previously identified through mass spectrometry into a hierarchy of manageable numbers of relevant proteins, several approaches were adopted. Besides the crude method of visually discriminating protein bands from Figures 3.3.1 and 3.3.1 (continued) for allelic effects, different bioinformatic tools such as Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) v 8.0 which discerned proteins according to their known or predicted protein-protein interactions, Reactome which grouped proteins by the reaction pathways they were known to be involved in. Lastly, transcription factor prediction software was used to highlight the likelihood of transcription factors recognizing and binding to the rs242557-containing oligonucleotides used.

In order to understand the functional interactions between these proteins, STRING was used to trawl through databases, text-mining, experiments and co-expression studies to identify functional networks amongst the protein hits. In doing so, it was possible to rank the hits according to how likely they are to form complexes as opposed to the false positives.

Figure 3.5.1 is an example of how STRING v 8.0 displays its results, using the highest confidence set of proteins as an example. At such confidence, a 30% coverage of the proteins identified or 15 proteins fell into this category. Hence, in terms of validating these proteins, it was logical to validate the proteins of the highest confidence first. Most of these proteins formed a complex (U2AF2, PTBP1, hnRNPU, hnRNPD0, ALY/Thoc4, U5 SNRNP 200). This complex is involved in splicing which might be important especially in light of the complex splicing pattern of tau. The rest of the proteins (nucleolin, DNA topoisomerase 1 and NONO) were important in DNA binding and transcription activation or silencing. Thus, these proteins suggested a possible coordination between chromatin remodeling and splicing. The rest were cytoskeletal proteins such as actin, tubulin, myosin, tropomyosin and vimentin. These could be contamination from the cytoplasm or relevant architectural proteins.

A slight reduction in confidence level to high confidence, Figure 3.5.2, (21 proteins, 4% coverage) showed the possibility of the non-splicing proteins to interact with each other and with the aid of PARP1. Another HMG protein, HMGA1 and the association between the HMG proteins were also revealed. DEK, an oncoprotein, was also shown to interact with the splicing complex.

At the medium confidence level, Figure 3.5.3 (29 proteins, 58% coverage) although some of the proteins included such as nucleolar RNA helicase II, RNA-dependent helicase, myosin 1C, and other members of the H1 family proved to be more of satellite interactors, not adding much more to the picture, it is interesting to note that H1c was shown to interact with both PARP1 and DNA topoisomerase 1.

The result of the STRING analysis has been summarized in Table 3.5.5 have been colour-coded in such a way that the first set of proteins include those that have highest confidence in associating with each other.

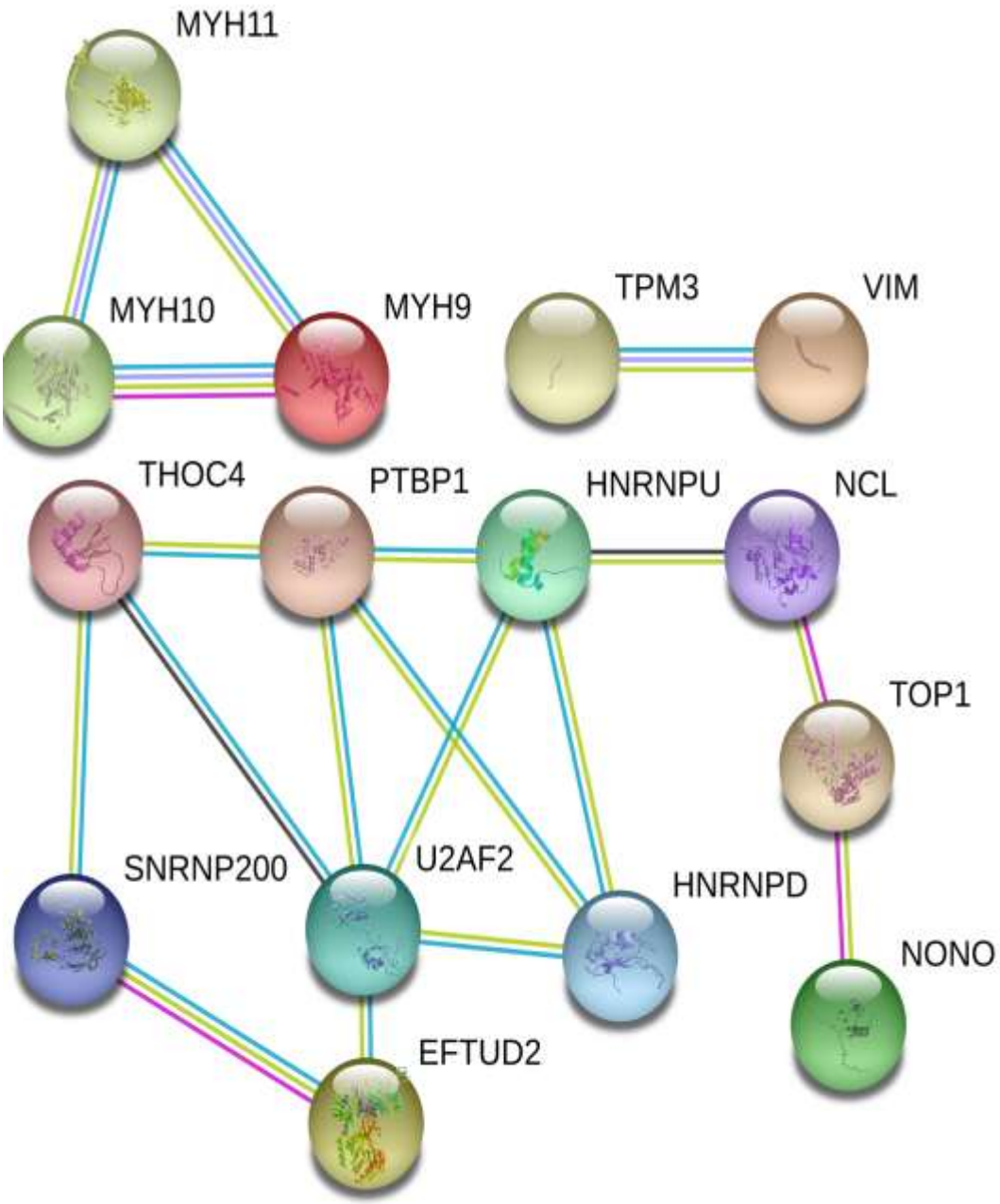


Figure 3.5.1: Highest confidence (0.9=P=1) protein-protein interactions amongst proteins identified through MS.

Blue lines represent associations based on databases. Yellow lines represent associations based on text-mining. Grey lines represent associations based on co-expression studies. Purple lines represent associations from co-immunoprecipitation experiments. Light blue lines represent putative homologs interacting in other species. Obtained using String v. 8.0 (string.embl.de). Size of nodes and length of lines are arbitrary.

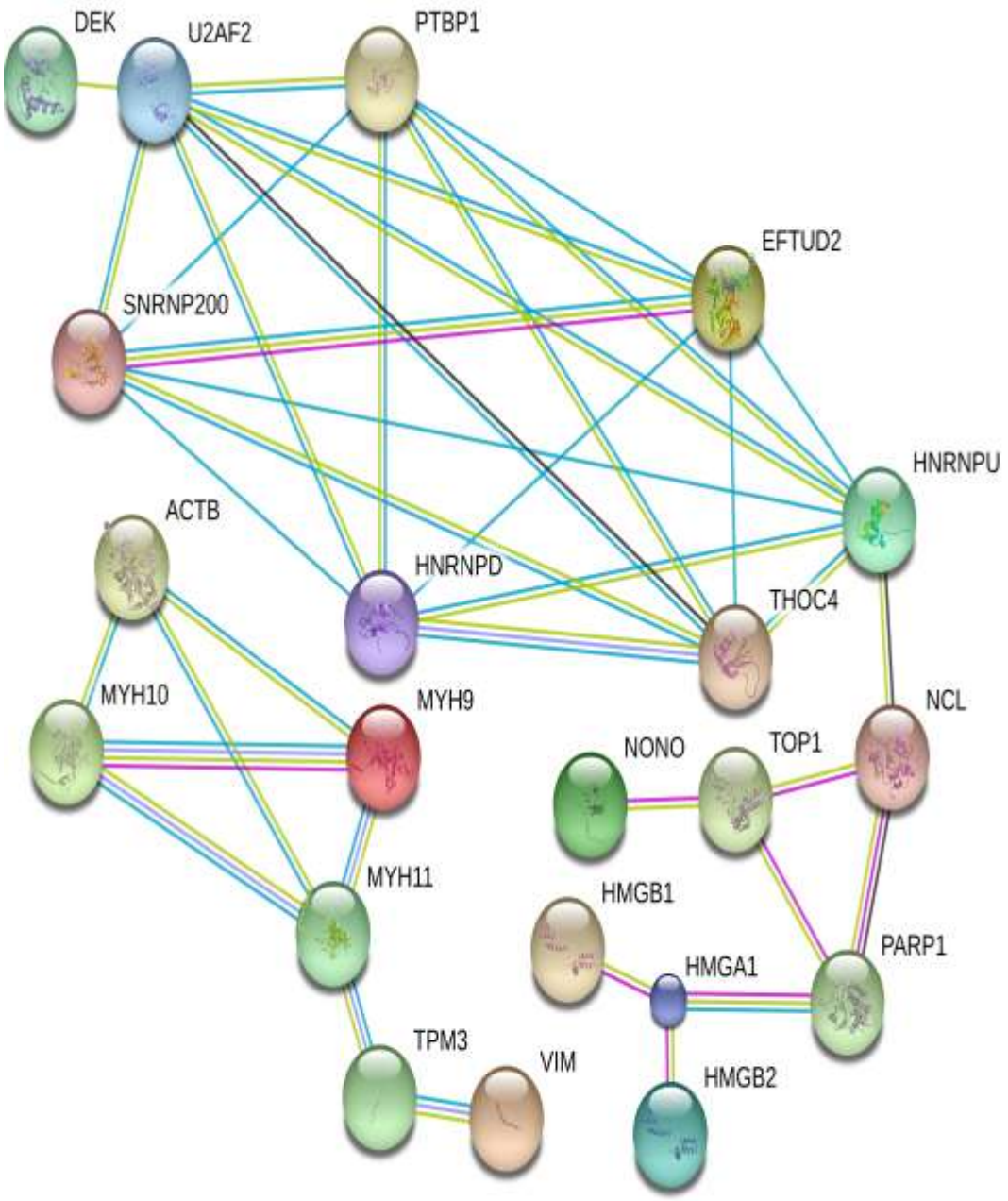


Figure 3.5.2: High confidence (0.7=P=0.9) protein-protein interactions amongst proteins identified through MS.

Blue lines represent associations based on databases. Yellow lines represent associations based on text-mining. Grey lines represent associations based on co-expression studies. Purple lines represent associations from co-immunoprecipitation experiments. Light blue lines represent putative homologs interacting in other species. Obtained using String v. 8.0 (string.embl.de). Size of nodes and length of lines are arbitrary.

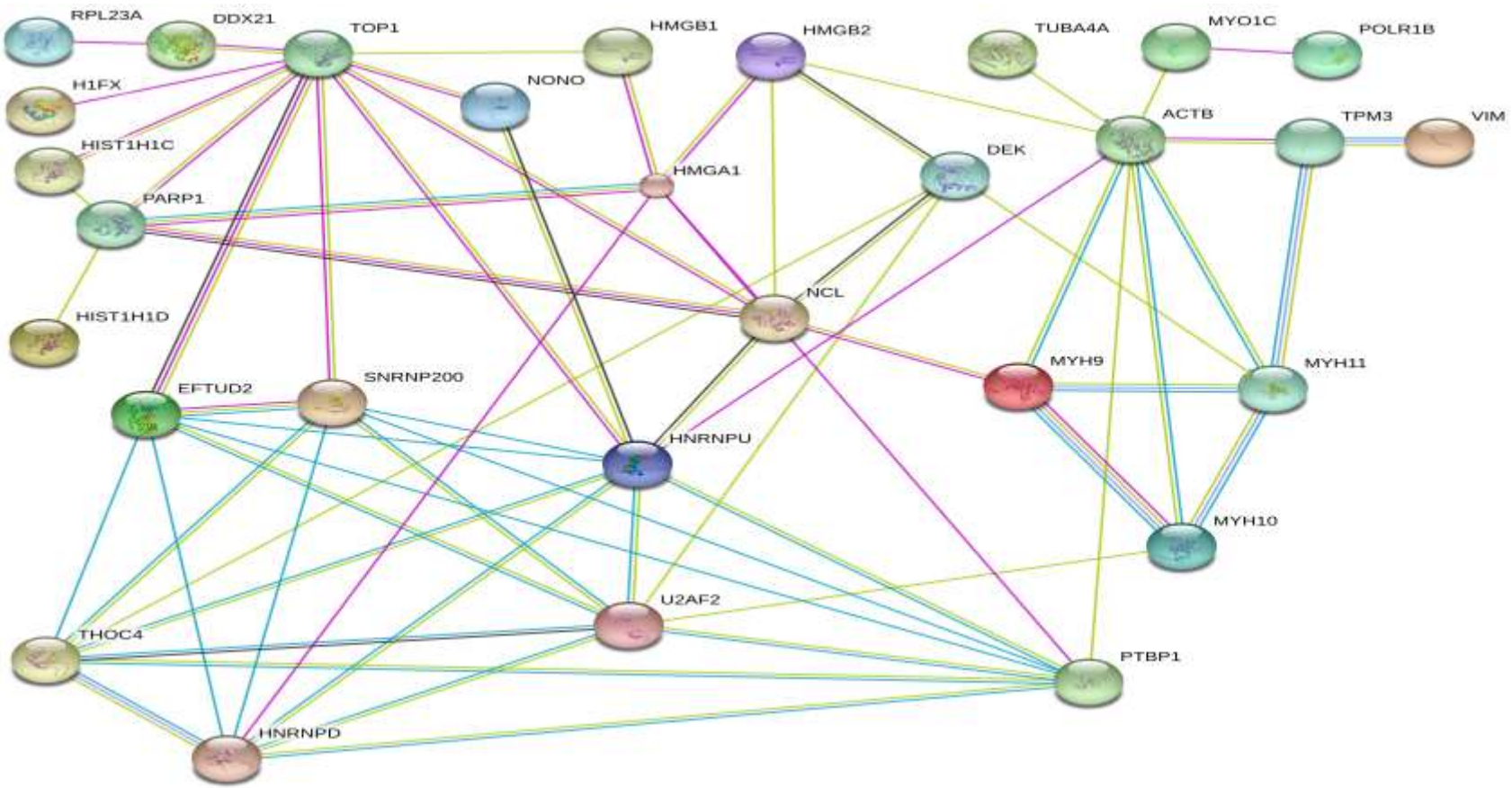


Figure 3.5.3: Medium confidence (0.4=P=0.7) protein-protein interactions amongst proteins identified through MS.

Blue lines represent associations based on databases. Yellow lines represent associations based on text-mining. Grey lines represent associations based on co-expression studies. Purple lines represent associations from co-immunoprecipitation experiments. Light blue lines represent putative homologs interacting in other species. Obtained using String v. 8.0 (string.embl.de). Size of nodes and length of lines are arbitrary.

No	Accession	UniProt Name	Full Name, Gene Name, Chromosome Number	(kDa)
1	P26599	PTB_HUMAN	Polypyrimidine tract binding protein, PTBP1/PTB, Chr 19	57.2
2	P26368	U2AF_HUMAN	Splicing factor U2AF 65 kDa subunit U2 auxiliary, U2AF2, Chr 19	53.5
3	Q00839	ROU_HUMAN	Heterogeneous nuclear ribonucleoprotein U, HNRNPU/SAFA, Chr 1	90.5
4	Q14103	ROD_HUMAN	Heterogeneous nuclear ribonucleoprotein D0, HNRNPD/AUF1, Chr 4	38.4
5	Q86V81	THOC4_HUMAN	THO complex subunit 4, THOC4/ALY/BEF, Chr 17	26.9
6	O75643	U520_HUMAN	U5small nuclear ribonucleoprotein 200kDa helicase, SNRNP 200, Chr2	244.5
7	Q8IXJ3	Q8IXJ3_HUMAN	small nuclear ribonucleoprotein component, SNRP 116, Chr 17	95.4
8	P19338	NUCL_HUMAN	Nucleolin Protein C23, NCL, Chr 2	76.6
9	Q15233	NR54_HUMAN	54 kDa nuclear RNA and DNA binding protein, NONO/NRB54, Chr X	54.2
10	P08670	VIME_HUMAN	Vimentin, VIM, Chr 10	53.7
11	P11387	TOP1_HUMAN	DNA Topoisomerase I, TOP1, Chr 20	90.7
12	P12324	TPMN_HUMAN	Tropomyosin cytoskeletal type 3, TPM3, Chr 1	32.8
13	P35579	MYH9_HUMAN	Myosin heavy chain nonmuscle type A, MYH 9, Chr 22	226.5
14	P35580	MYHA_HUMAN	Myosin Heavy Chain nonmuscle type B, MYH10, Chr 17	229.0
15	P35749	MYH11_HUMAN	Smooth muscle myosin heavy chain, MYH11, Chr 16	121.9
16	P35659	DEK_HUMAN	DEK protein, DEK, Chr 6	42.7
17	P60709	ACTB_HUMAN	Actin, ACTB, Chr 7	41.7

18	P09429	HMG1_HUMAN	High mobility group protein, HMGB1/HMG1, Chr 13	24.9
19	P26583	HMG2_HUMAN	High mobility group protein HMG2, HMGB2, Chr 4	24.0
20	P09874	PPOL_HUMAN	Poly ADP ribose polymerase, PARP1/ADPRT/PPOL, Chr 1	113.1
21	P17096	HMGI_HUMAN	High mobility group protein, HMGA1/HMGIY, Chr 6	11.7
22	Q92522	H1X_HUMAN	Histone H1x, H1FX, Chr 3	22.5
23	P62750	RL2B_HUMAN	60S ribosomal protein L23a, RPL23A, Chr 17	17.7
24	Q8WU19	Q8WU19_HUMAN	TUBA1B, TUBA1B, Chr 12	37.2
25	P16403	H12_HUMAN	Histone H1.2 Histone 1d, HIST1H1C, Chr 6	21.4
26	P16402	H13_HUMAN	Histone H1.3 Histone H1c, HIST1H1D, Chr 6	22.4
27	O00159	MY1C_HUMAN	Myosin 1C, MYO1C, Chr 17	117.9
28	Q9H9Y6	RPA2_HUMAN	DNA-directed RNA polymerase I 135kDa polypeptide, POLR1B, Chr 2	128.2
29	Q9NR30	DDX21_HUMAN	Nucleolar RNA helicase II, DDX21, Chr 10	87.4
30	P27694	RFA1_HUMAN	Replication protein A 70 kDa binding subunit, RPA1/RPA70, Chr 17	68.1
31	P62851	RS25_HUMAN	40S ribosomal protein S25, RPS25, Chr 11	13.7
32	P07954	FUMH_HUMAN	Fumarate Hydratase, FH, Chr 1	54.6
33	P10412	H14_HUMAN	Histone H1.4 Histone 1b, HIST1H1E, Chr 6	21.9
34	P16401	H15_HUMAN	Histone H1.5 Histone 1a, HIST1H1B, Chr 6	22.6
35	Q71U36	TBA1_HUMAN	Tubulin alpha 1 chain brain specific, TUBA1A/TUBA3, Chr 12	50.1

36	P23284	CYPB_HUMAN	Peptidyl prolyl *cis* trans isomerase E, PP1B/CYPB, Chr 15	23.7

Table 3.5.4: List of protein hits from MS data generated after STRING analysis. Proteins highlighted in green are of highest confidence of interacting with each other. Those in orange are of high confidence, red for medium confidence, purple for low confidence and uncoloured for those with no known interactions as indicated by STRING v. 8.0.

3.6 Further bioinformatics on DNA-binding proteins

While STRING was useful in shedding some light on the proteins identified, their interactions with each other and providing a rough framework to choose candidate proteins to carry through, it is still one bioinformatics tool rooted in protein-protein interactions.

Another bioinformatics tool, Reactome, took the list of submitted proteins and allocated them based on the reaction event(s) they are known to be involved in according to a database of curated reaction pathways. Each reaction event could consist of a few molecules to hundreds of molecules. For example, in Figure 3.6.1 (A) – (D), pathways involved in gene expression are highlighted by a series of arrows, each representing a particular event within the pathway. More specific events such as formation and maturation of mRNA transcript, processing of capped intron-containing pre-mRNA and mRNA splicing constitute smaller pathways within gene expression.

Once a list of protein identifiers was submitted, the skypainter tool displayed all the curated reactions and 'painted' or allocated the proteins to their respective reaction event. The more proteins that cluster around the same reaction event, the warmer the colour. As can be seen from Figure 3.6.1, most of the proteins (7 proteins) have functions within the mRNA splicing pathway. As it was a manually curated database, Reactome's strength lies on the extensiveness of their database. Hence, 20 proteins (40% coverage) compared to STRING's 88% coverage of proteins identified was to be reasonably expected.

A breakdown of each reaction event within mRNA splicing in which the proteins were involved in is illustrated in Table 3.6.2. The proteins first involved at the very beginning of the mRNA splicing pathway are hnRNPU, hnRNPD0 and PTBP1. This could possibly suggest that of all the proteins identified, these were the most likely direct interactors with DNA and/or mRNA. However, the reaction event that had the highest probability was the formation of the spliceosomal complex B. This was the earliest event within the mRNA splicing pathway that incorporated seven of the proteins identified. Subsequent downstream events were singled out as well since they involved the same seven proteins. However, it was harder to differentiate when a protein was recruited into a

complex from when it executed its function within a complex from remaining in the complex after its execution.

Other less represented pathways included the U12-dependent splicing pathway as well as the cleavage of mRNA at the 3’ end and polyadenylation, Table 3.6.2 continued.

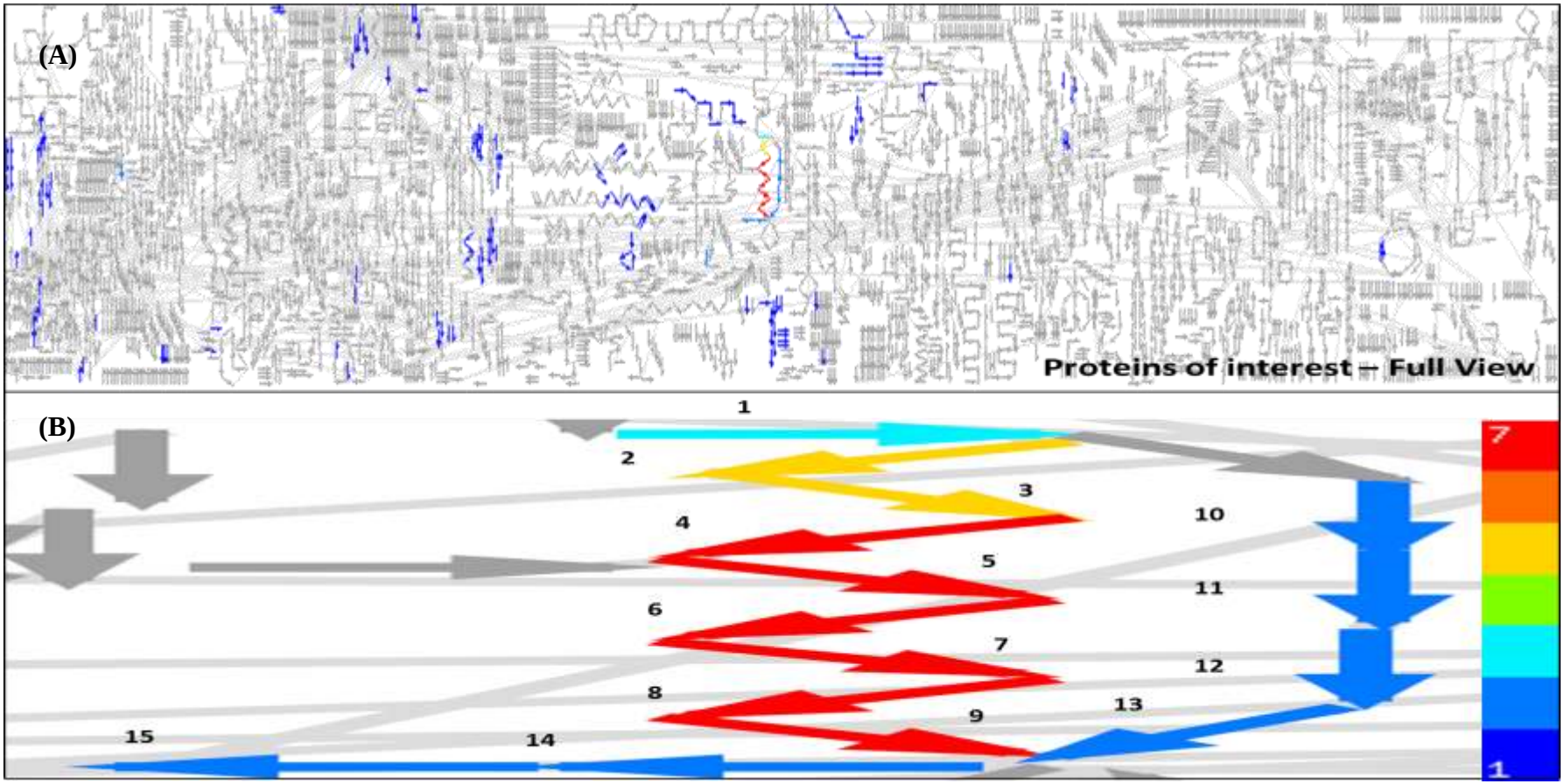


Figure 3.6.1: Reactome over-representation analysis using Skypainter.

All annotated reaction pathways are indicated as arrows and laid out. (A) Reactome clusters the proteins of interests into the reaction pathways they act upon. (B) Specifically, most of the proteins are involved in mRNA splicing with the warmer colours indicating more of the submitted proteins being involved in a reaction pathway. Numbered reactions elaborated in Figure 3.9. Obtained from www.reactome.org.

No.	False discovery rate	Un-adjusted probability of seeing N or more genes in this Event by chance	Number of genes in your query which map to this Event	Total number of genes involved in this Event	Identifier of this Event	Name of this Event	Submitted identifiers mapping to this Event
1	-	0.000391869	3	34	REACT_1877	Formation of pre-mRNPs	hnRNPU, hnRNPD0, PTBP1
2	-	7.12E-06	5	66	REACT_222	Formation of the Spliceosomal E complex	hnRNPU, THOC4, hnRNPD0, U2AF2, PTBP1
3	-	2.34E-05	5	84	REACT_788	Formation of the Spliceosomal A Complex	hnRNPU, THOC4, hnRNPD0, U2AF2, PTBP1
4	-	9.97E-08	7	97	REACT_48	Formation of the Spliceosomal B Complex	hnRNPU, U2AF2, U5 SNRNP200, PTBP1, hnRNPD0, THOC4, U5 SNRNP116
5	-	2.24E-07	7	109	REACT_625	Formation of an intermediate Spliceosomal C complex	hnRNPU, U2AF2, U5 SNRNP200, PTBP1, hnRNPD0, THOC4, U5 SNRNP116
6	-	1.73E-07	7	105	REACT_1554	Formation of the active Spliceosomal C complex	hnRNPU, U2AF2, U5 SNRNP200, PTBP1, hnRNPD0, THOC4, U5 SNRNP116
7	-	1.85E-07	7	106	REACT_1935	Lariat Formation and 5'-Splice Site Cleavage	hnRNPU, U2AF2, U5 SNRNP200, PTBP1, hnRNPD0, THOC4, U5 SNRNP116
8	-	2.10E-07	7	108	REACT_774	Formation of Exon Junction Complex	hnRNPU, U2AF2, U5 SNRNP200, PTBP1, hnRNPD0, THOC4, U5 SNRNP116
9	-	2.10E-07	7	108	REACT_1331	Cleavage at the 3'-Splice Site and Exon Ligation	hnRNPU, U2AF2, U5 SNRNP200, PTBP1, hnRNPD0, THOC4, U5 SNRNP116

Table 3.6.2: List of reactome events over-represented by proteins of interest.

Proteins of interest were matched to the reaction events they participated in. The events have been ordered according to the reaction pathway they follow, refer to Figure 3.8 (G). The warmer the colour, the higher the probability of the proteins involved in the reaction event. Un-adjusted probability refers to no correction for multiple testing. Probability was carried out using the hypergeometric test.

No.	False discovery rate	Un-adjusted probability of seeing N or more genes in this Event by chance	Number of genes in your query which map to this Event	Total number of genes involved in this Event	Identifier of this Event	Name of this Event	Submitted identifiers mapping to this Event
10	-	0.014680417	2	43	REACT_1253	Formation of AT-AC B Complex	U5 SNRNP116, U5 SNRNP200
11	-	0.014680417	2	43	REACT_1615	Formation of AT-AC C complex	U5 SNRNP116, U5 SNRNP200
12	-	0.014680417	2	43	REACT_2241	ATAC spliceosome mediated Lariat formation,5' splice site cleavage	U5 SNRNP116, U5 SNRNP200
13	-	0.014680417	2	43	REACT_1494	ATAC spliceosome mediated 3' splice site cleavage, exon ligation	U5 SNRNP116, U5 SNRNP200
14	-	0.009337753	2	34	REACT_1914	Cleavage of mRNA at the 3'-end	THOC4, U2AF2
15	-	0.005950512	2	27	REACT_1162	mRNA polyadenylation	THOC4, U2AF2

Table 3.6.2 continued: List of reactome events over-represented by proteins of interest.

Proteins of interest were matched to the reaction events they participated in. The events have been ordered according to the reaction pathway they follow, refer to Figure 3.8 (G). The warmer the colour, the higher the probability of the proteins involved in the reaction event. Un-adjusted probability refers to no correction for multiple testing. Probability was carried out using the hypergeometric test.

3.7 *In silico* prediction of transcription factors binding to rs242557 region

Apart from the *de novo* identification of DNA-binding proteins that interact with the sequences for H1c-A and H1b-G flanking the rs242557 polymorphism – the same sequence used for the EMSA experiment (Figure 3.2.1) and pull-down / mass spectrometry experiment (Figure 3.3.1), was analysed by software such as MatInspector and Transfac to predict potential transcription factor binding sites and this yielded potentially promising leads.

First, MatInspector (Table 3.7.1) predicted less transcription factors compared to Transfac (Table 3.7.2) with differential binding to H1c-A and H1c-G. Only one protein, the erythroid krueppel like factor (EKLF) protein, was predicted to bind to the positive strand of H1c-A, stretching across 74% of the sequence. However, EKLF is also predicted to bind to H1b-G suggesting that this binding is not dependent on alleles. Selenocysteine tRNA gene transcription activating factor (Staf) protein is predicted to bind to H1b-G but not H1c-A. MatInspector, had Staf with the lowest confidence score even though it had a 100% coverage of the input sequence. Transfac had also predicted that Staf would bind to H1b-G more than it would bind to H1c-A, however unlike MatInspector, it had assigned Staf with the highest score for H1b-G.

The next transcription factor suggested binding to H1b-G more than H1c-A by both MatInspector and Transfac was CP2. It had an 83% sequence coverage. What made CP2 different to the other proteins was that it was predicted to bind to the negative strand of the sequence used. This could possibly explain why Transfac had lower confidence in predicting CP2 over Staf whereas MatInspector had more confidence in CP2 over Staf.

Lastly, MatInspector predicted Hypermethylated in cancer 1 (HIC1) with the highest score to bind to 57% of the input sequence. Although it was the smallest coverage, the motif which the protein recognized was where rs242557 resided. Since multiple binding sites are required for optimal binding, it made it harder to ascertain the relevance of this prediction. It was also probably why HIC1 did not feature in Transfac's prediction.

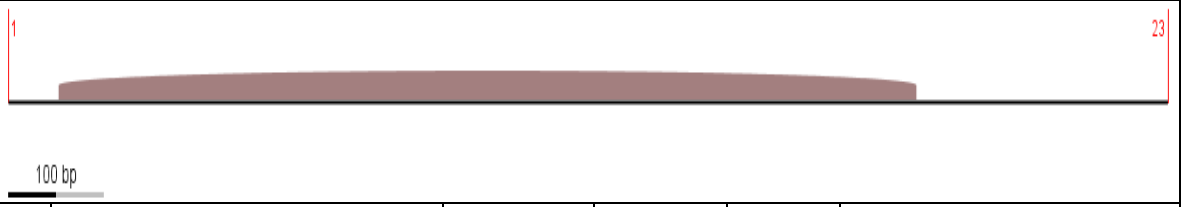
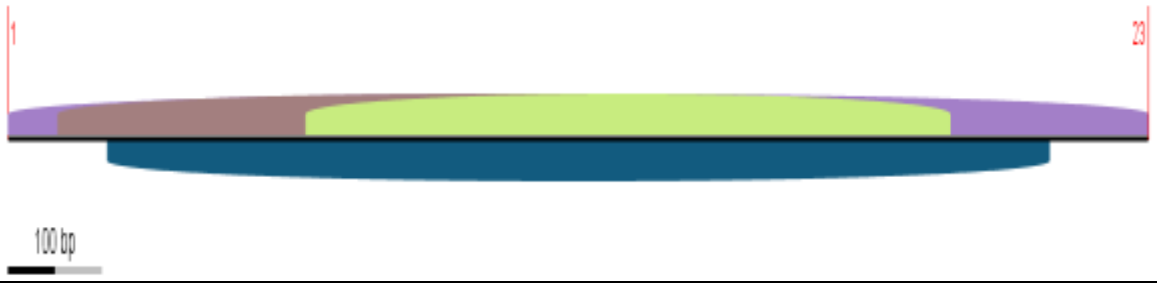
No.	H1c-A	Position from	Strand	Score	Sequence
1	EKLF (Erythroid krueppel like factor)	2-18	+	0.896	tcgcccaGGGTacacca
					
No.	**H1b-G**	**Position from**	**Strand**	**Score**	**Sequence**
1	STAF (Se-Cys tRNA gene transcription activating factor)	1-23	+	0.803	ttcgCCCAgggtgcaccaggaca
2	EKLF (Erythroid krueppel like factor)	2-18	+	0.925	tcgcccaGGGTgcacca
3	CP2 [LBP-1c (leader-binding protein-1c), LSF (late SV40 factor), CP2, SEF (SAA3 enhancer factor)]	3-21	-	0.843	tCCTGgtgcaccctgggcg
4	HIC1 (Hypermethylated in cancer 1, transcriptional repressor containing five Krüppel-like C2H2 zinc fingers, for optimal binding multiple binding sites are required.)	7-19	+	0.937	cagggTGCAccag
					

Table 3.7.1: List of predicted transcription factors to bind to H1c-A and H1b-G according to MatInspector. 23 base pair sequences for H1c-A (TTCGCCCAGGGTACACCAGGACA) and H1b-G (TTCGCCCAGGGTGCACCAGGACA) were submitted to MatInspector and a list of potential transcription factors that could bind was then generated.

No.	H1c-A	Score	H1b-G	Score	No.	H1c-A	Score	H1b-G	Score
1	Ik-3	74.9	Staf	75.0	**21**	GATA-2	65.2	C/EBPa	63.3
2	deltaE	71.9	deltaE	71.9	**22**	c-Rel	64.5	Egr-1	63.0
3	E47	71.2	c-Ets-	71.6	**23**	NF-kap	64.2	CdxA	62.9
4	Lyf-1	70.1	USF	71.1	**24**	GATA-1	64.1	ZID	62.5
5	Ik-1	69.9	C/EBPb	70.4	**25**	XFD-3	63.9	CP2	62.5
6	Staf	69.6	Lyf-1	70.1	**26**	v-Myb	63.6	COUP-T	62.1
7	USF	69.1	AhR/Ar	69.8	**27**	Evi-1	63.0	NGFI-C	61.9
8	N-Myc	68.9	MZF1	68.7	**28**	Arnt	63.0	STATx	61.5
9	AML-1a	68.8	Egr-3	68.6	**29**	C/EBPb	62.9	CHOP-C	61.1
10	MZF1	68.7	N-Myc	67.5	**30**	CdxA	62.9	GR	60.6
11	VBP	68.6	AML-1a	67.4	**31**	MyoD	62.8	NF-kap	60.5
12	HLF	68.4	ARP-1	66.7	**32**	E4BP4	62.3	MyoD	60.5
13	Egr-2	67.4	GATA-1	66.1	**33**	COUP-T	62.1	p300	60.4
14	Egr-3	67.3	Egr-2	64.8	**34**	STATx	61.5	HLF	60.3
15	Ik-2	66.7	E47	64.4	**35**	CREB	61.4	Arnt	60.1
16	ARP-1	66.7	Ik-3	64.3	**36**	RREB-1	61.2	GATA-2	60.1
17	Egr-1	66.6	NF-E2	64.2	**37**	CP2	60.4		
18	c-Ets-	65.7	AP-1	63.8	**38**	GATA-X	60.2		
19	NGFI-C	65.5	v-Myb	63.6					
20	p300	65.3	Elk-1	63.4					

Table 3.7.2: List of predicted transcription factors to bind to H1c-A and H1b-G according to Transfac. 23 base pair sequences for H1c-A (TTCGCCCAGGGTACACCAGGACA) and H1b-G (TTCGCCCAGGGTGCACCAGGACA) were submitted to Transfac and a list of potential transcription factors that could bind was then generated.

To probe CP2, HIC1 and Staf further, a high confidence STRING analysis was done to ascertain their most probable protein-protein interactions. CP2 was shown to interact with other transcription factors such YY1, GATA1 as well as HDAC1 and amyloid beta A4 precursor protein-binding family B member 1 (APBB1) or Fe65 protein. HIC1 like CP2 interacted with HDAC1 and as its name suggested, proteins involved in cell cycle and cancer such as p53, the cell cycle dependent transcription factor E2F-1, the NAD-dependent protein deacetylase sirtuin 1 as well as the DNA-helicase/exonuclease WRN. Staf or the zinc finger protein 143 (ZNF143) was shown to interact with the RNAPII/III transcription factor complex SNAPc, RNAPIII transcription factor

BRF2/TFIIIB50, POU2F1/Oct-1. Besides playing a role in the transcription of RNAPII and III, Staf was also able to interact with RNAPII/III as well as the TATA-binding protein (TBP).

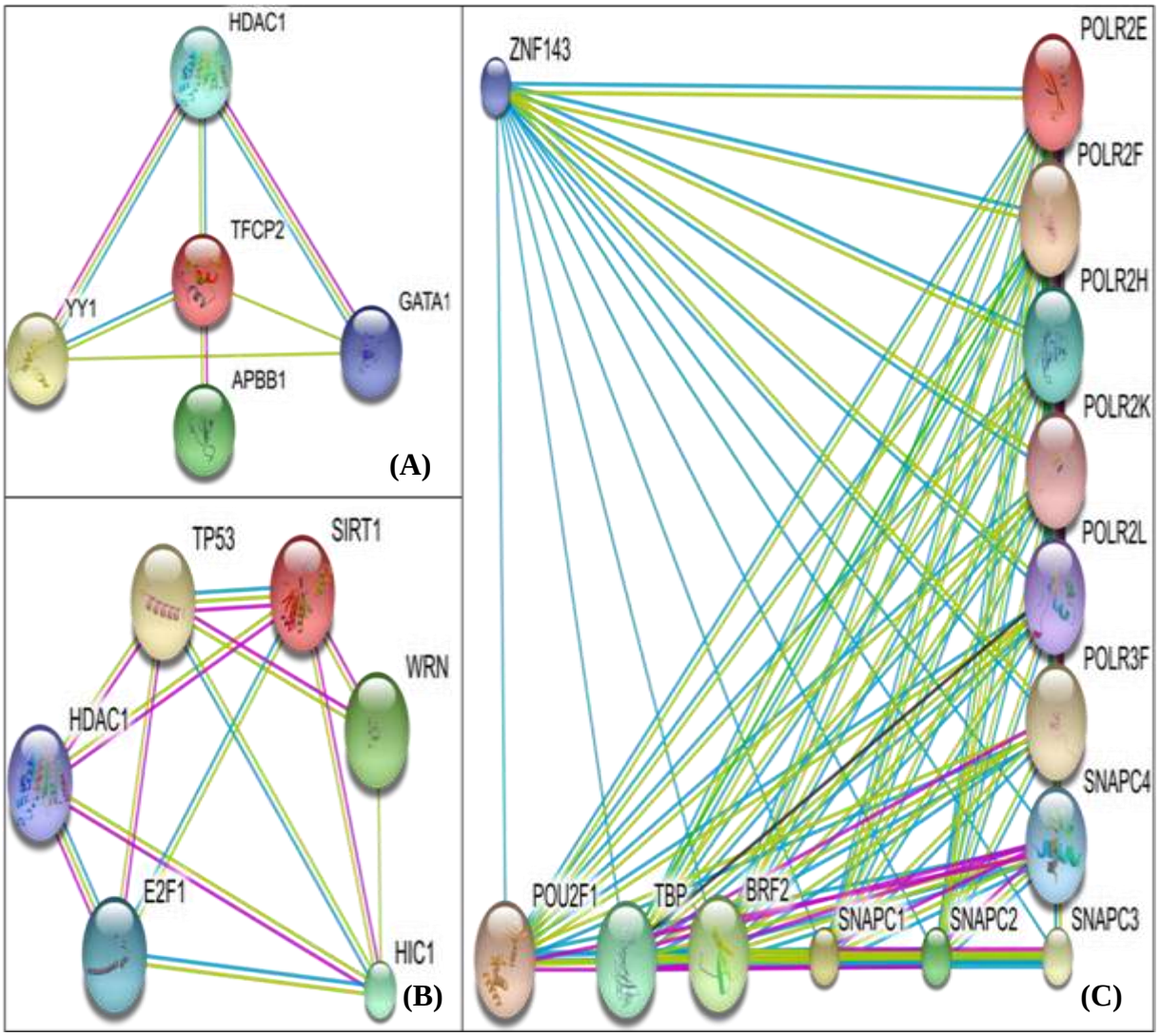


Figure 3.7.3: High confidence (0.7<P<0.9) protein-protein interactions for CP2, Staf and HIC1 proteins.

(A) Protein-protein interactions for transcription factor CP2 (TFCP2) (B) Protein-protein interactions for HIC1 (C) Protein-protein interactions for Staf or Zinc Finger protein 143 (ZNF143). Blue lines represent associations based on databases. Yellow lines represent associations based on textmining. Red lines represent associations based on gene fusion. Grey lines represent associations based on coexpression studies. Purple lines represent associations from experiments. Light blue lines represent homology based association. Obtained using String v. 8.0 (string.embl.de).

3.8 Discussion on DNA-binding proteins identified and bioinformatics analyses

Since the advent of genome-wide association studies (GWAS), this approach has helped create a deeper and more subtle understanding of complex diseases such as tauopathies, Alzheimer's disease and Parkinson's disease. The results of such extensive studies have led to genes previously not immediately associated with disease to be highlighted as well as single nucleotide polymorphisms (SNPs) adding a more complex picture in well-studied genes such as *MAPT*. In one of the more recent PSP GWAS, the conditional analysis still supported the relevance of rs242557 with disease.[159] As rs242557 is located within a conserved region within the *MAPT* promoter region (Region C), downstream of the core promoter, its location would suggest that it might exert its effect on the transcription or splicing machinery or even epigenetically by influencing chromatin dynamics. It is important to remember that in physiological conditions, genes are not linear strands of DNA but highly regulated conformational DNA. Furthermore, there exists a robust cross-talk between the transcription and splicing machinery as well as with their epigenetic environment that surrounds them. Most studies have focused on the carboxy-terminal heptad repeat domain (CTD) of RNA polymerase II (RNAPII) being a key platform for gene expression. When the CTD was truncated *in vivo*, splicing 3' end processing and termination were all inhibited.[160] More specifically, phosphorylation at serine sites within a repeated heptameric sequence within the CTD couples processes within gene expression.[161]

The approach taken in this work to elucidate the functionality of a disease-associated SNP had been applied to various other SNPs within neurology to decipher their contribution to disease. The -258 T/G SNP located within the *parkin* gene was shown, through EMSA and dual-luciferase assay, to be the first functional evidence that it affected gene transcription.[162] In delayed onset of Alzheimer's disease, the *PIN1* (Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1) SNP rs2287839 was shown using EMSA, promoter luciferase assay and chromatin immunoprecipitation (ChIP) to affect the binding of AP4 to its promoter.[163]

In sporadic Alzheimer's disease, the -1314C/T polymorphism was again investigated using a luciferase reporter assay and EMSA as to its functional relevance.[164] This methodological trend was seen across the breadth of neurology such as the G protein receptor kinase 3 (*GRK3*) gene in bipolar disorder, the tryptophan hydroxylase 2 (*TPH2*) gene in neuropsychiatric conditions, *NOS1AP* gene in schizophrenia, *Dlx* homeobox genes in neurodevelopmental disorders, *DYX1C1* gene in dyslexia, the δ-opoid receptor (*OPRD1*) gene in addiction and guanine nucleotide exchange factor 10 (*ARHGEF10*) gene in atherothrombotic stroke.[165,166,167,168,169,170,171]

The model that we worked on throughout this work were the undifferentiated human neuroblastoma cell lines. These cells, in their undifferentiated form, still had neuronal characteristics such as dopamine-β-hydroxylase activity and converting glutamate to GABA. Between the SH-SY5Y and BE(3)-M17 neuroblastoma cell lines, however, SH-SY5Y was the only cell line that showed significant nuclear extract enrichment using HDAC1 as a nuclear extract marker. The HDAC1 nuclear protein was enriched approximately ten-fold in the SH-SY5Y nuclear extracts compared to SH-SY5Y cytoplasmic extracts. Although genotyped human brain samples were available, it would have been premature to work on nuclear protein extracts from these as limited patient sample were available for use that would also be needed for other projects, the protein integrity would vary and it would not be a practical system to probe for protein-DNA interaction, protein identification or protein manipulation.

Once a reproducible method of enriching nuclear proteins was established, it was then possible to test the hypothesis that the effect previously reported that rs242557 had in dual-luciferase promoter assays was due to changes in protein- or protein complex-DNA interactions. EMSA is a native method which allows any differences in binding between alleles and nuclear proteins to be investigated, resulting in the labelled DNA probe to be retarded due to the proteins binding to it. To probe the protein binding affinity of rs242557, a minimal stretch of 23 base-pairs containing rs242557 as the only SNP was used. Repetitive stretches of DNA were used to sequester any non-specific protein-DNA binding while excess unlabelled DNA of interest was used to demonstrate the specificity of the interaction through competition. The result was the first evidence of rs242557 as having a functional allelic control on DNA-protein interaction. While previous dual-luciferase

promoter assays had shown that the risk allele, rs242557-A (H1c), increased the promoter strength, EMSA however showed less protein binding to rs242557-A. This was about four-fold less protein binding compared to the protective allele, rs242557-G (H1b/H2), Figure 3.2.1. This suggested that altered DNA-binding protein binding may account for the differences in allelic expression. It could be that the proteins recruited to rs242557 act primarily to repress the *MAPT* promoter. Since EMSA is a relatively crude method, it was not possible to infer how similar or different the proteins that made up the protein complexes around rs242557-A or –G. What was unexpected from the experiment was another larger complex that seemed to bind in the same allele-specific manner albeit less pronounced. This could have been an artefact as we were not able to compete it out even with a 400-fold excess of unlabelled competition probe.

We used mass spectrometry (MS) to identify the DNA-binding proteins involved with rs242557 region. Nuclear proteins that were affinity purified using allelic probes were separated on a 1-D SDS gel. Since we did not have our own mass spectrometry set up, we collaborated with the core proteomics unit of the UCL Institute of Child Health. We decided that it would not have been cost effective to divide the gel lanes into regular intervals and analyse each gel band. Instead, our focused approach was to target specific pairs of bands that from visual inspection showed differences in band intensity between rs242557-A and –G. Also, as a rule-of-thumb, the mass spectrometer was only likely to give a much more confident identification of a protein band that was visible to the naked eye by Coomassie staining. Hence, a targeted approach was pursued rather than a permissive approach of cutting the whole lane into equal segments and analysing each band.

Although 37 unique proteins were identified, 21 of these showed considerable differences in band intensity in the 1-D gels. However, as these were 1-D gels, there could have been other proteins with similar molecular mass within the same band. Even though the mass spectrometer in most cases could only positively identify one protein in a band, usually the most abundant or most amenable to digestion and mass spectrometry, various non-abundant proteins could still accumulate to give the impression that one band is more intense than another. Hence, different bioinformatics approaches were employed to further stratify these results. One was based on protein-protein interactions (STRING) while another was based on proteins that were involved in the same biological events (Reactome).

STRING is a metasource that weighs and integrates protein-protein information from experimental repositories, public text collections and computational prediction methods. This is different to Reactome which is a manually curated resource for human pathway data. While the results of STRING demonstrated that 23 of the 37 proteins could potentially interact with one or more proteins identified with high confidence, the results of Reactome showed that 20 such proteins were involved in manually curated biological events. These have been summarised in Table 3.8.2.

Consolidating all these different analyses revealed that ten proteins were repeatedly highlighted, eight proteins were common to both STRING and Reactome, five were common to both visual inspection and STRING while there was only one protein common to both visual inspection and Reactome. Looking at Table 3.8.2, eleven of the shortlisted proteins were mostly structural proteins such as myosins, ribosomal proteins, tubulin, vimentin, small nuclear ribonucleoproteins and high mobility group proteins. These were not highlighted as compared to the rest of the shortlisted proteins that had more dynamic roles in splicing and chromatin remodelling. Most of the splicing and chromatin remodelling proteins were common to all of the different analyses (seven out of the ten proteins), making them the more probable candidate binding proteins. Following this trend, six other possible candidate proteins have been highlighted amongst the shortlisted proteins. In total, about half (13 out of 24) of the shortlisted proteins were pursued further. These were PTBP1, U2AF2, hnRNPD0, hnRNPU, HMGB2, histone H1, actin, THOC4/ALY, DEK, PARP1, NONO, nucleolin and RPA70.

A brief overview of what is currently known about each of these proteins follows. The scope of this PhD not only revolves around the tau gene but also transcriptional regulation. Hence, these identified proteins will be discussed in the context of transcriptional regulation and where possible a neuronal or neurological context as well. This brief literature research for each protein allows for not only a better understanding of these proteins but also an independent validation for the relatedness of these proteins either through direct interaction or similar function. Besides the binding proteins we experimentally identified, three proteins (CP2, HIC1 and ZNF43) were predicted *in silico* to bind to the rs242557 region. Hence, these proteins will be discussed in brief too.

No.	Visual differences from gel bands	STRING	Reactome	Proteins of interest shortlist
1	PTBP1	PTBP1	PTBP1	PTBP1
2	U2AF2	U2AF2	U2AF2	U2AF2
3	hnRNPD0	hnRNPD0	hnRNPD0	hnRNPD0
4	hnRNPU	hnRNPU	hnRNPU	hnRNPU
5	HMGB2	HMGB2	HMGB2	HMGB2
6	Histone H1	Histone H1	Histone H1	Histone H1
7	U5 snRNP 116	U5 snRNP 116	U5 snRNP 116	U5 snRNP 116
8	40S Ribosomal protein S25	40S Ribosomal protein S25	40S Ribosomal protein S25	40S Ribosomal protein S25
9	Actin	Actin	Actin	Actin
10	Tubulin α1 brain-specific	Tubulin α1 brain-specific	Tubulin α1 brain-specific	Tubulin α1 brain-specific
11	-	HMGB1	HMBG1	HMBG1
12	-	THOC4/ALY	THOC4/ALY	THOC4/ALY
13	-	U5 snRNP 200	U5 snRNP 200	U5 snRNP 200
14	DNA topoisomerase I	Vimentin	Vimentin	Vimentin
15	Myosin 1c	MYH 9	MYH 9	MYH 9
16	Nucleolar RNA Helicase II	MYH10	MYH10	MYH10
17	Fumarate hydratase	MYH11	MYH11	MYH11
18	PPOL	60S ribosomal protein L23a	60S ribosomal protein L23a	60S ribosomal protein L23a
19	DEK	DEK	Elongation factor 1 beta	DEK
20	PARP1	PARP1	-	PARP1
21	NONO	NONO	-	NONO
22	HMGA1	HMGA1	-	HMGA1
23	Nucleolin	Nucleolin	-	Nucleolin
24	RPA70	-	RPA70	RPA70

Table 3.8.2: Summary of bioinformatics analyses of identified proteins. A visual inspection of band intensity from the gel bands used to identify the proteins was compared against the results of protein-protein interactions generated through STRING as well as protein function through Reactome. Orange – proteins common to all three analyses, green – proteins common to String / Reactome, yellow – proteins common to visual / String, blue – proteins common to visual / Reactome, purple – proteins shortlisted to follow through.

3.8.1 Polypyrimidine tract binding protein (PTBP1 / hnRNPI)

PTBP1 was first discovered as a protein that specifically bound to the polypyrimidine tract of introns.[172] It was shown that PTBP1 represses neuron-specific splicing events in rat cerebellum.[173, 174] A brain-specific isoform of PTBP1 was discovered that inhibits Nova activation of exon selection in neurons in addition to many others.[175, 176, 177] There are many ways that PTBP1 repress splicing. One way was that it competed and sterically impeded the binding of U2AF65 to polypyrimidine tracts.[178, 179] However, it was also possible that PTBP1 promoted exon inclusion by occluding U2AF binding as was shown in calcitonin mRNA.[180] Other ways PTBP1 hindered splicing was through looping via multimerization or interfering with U1 snRNP interactions (Figure 3.8.1.1).[181] Ma, S. et. al. also showed that the tau mRNA also contained the conserved consensus PTBP1 binding site in its neurite localization signal.[182]

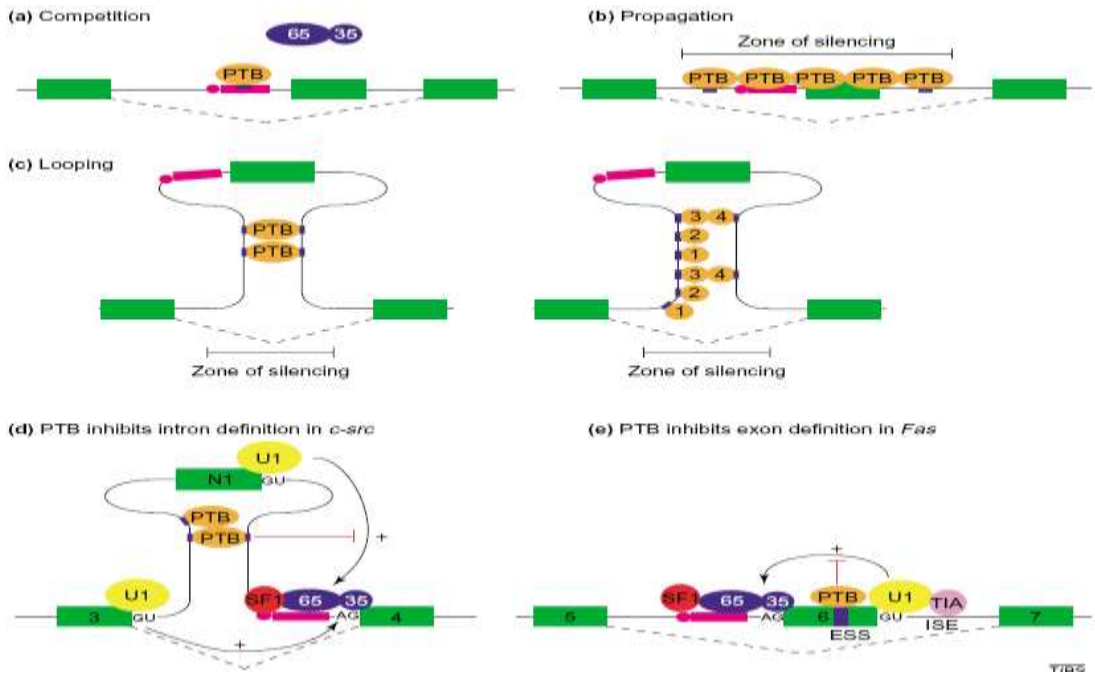


Figure 3.8.1.1: Models of PTB repression. (a) PTB competes with U2AF2 for binding. (b) A series of PTB proteins mask an exon. (c) Looping as a result of binding sites flanking an exon. (d) PTB prevents U1 at N1 to form productive binding with U2AF2. (e) PTB protein binds to an exon-splicing silencer (ESS) and prevents U1 from interacting with U2AF2. Adapted from[170].

3.8.2 Splicing factor U2AF 65 kDa subunit U2 auxiliary (U2AF65)

The large subunit of splicing factor U2AF (U2AF65) binds to the polypyrimidine tract (PPT) and directs contacts with the branch site.[183,184] It is also essential for the assembly of the pre-splicing complex A by forming a short helix between the U2 snRNA and the branch site.[185]

3.8.3 High Mobility Group 2 (HMG2) protein

HMG-2 belongs to a family of non-histone chromatin proteins that regulate transcription, replication, recombination and DNA repair.[186,187] A mechanism for their action in priming the nucleosome is illustrated in Figures 3.8.3.1.[188] The low level expression of HMG-2 in normal human brain and significant upregulation in glioblastoma tissue has been utilized to develop a glioblastoma gene therapy.[189] Genome-wide linkage analysis and genome-wide expression analysis pointed to HMG-2 being differentially expressed in schizophrenia in postmortem dorsolateral prefrontal cortex.[190] HMG-2 was shown to interact with huntingtin and ataxin 1 protein and repressing the genotoxic stress signals induced in these polyglutamine diseases.

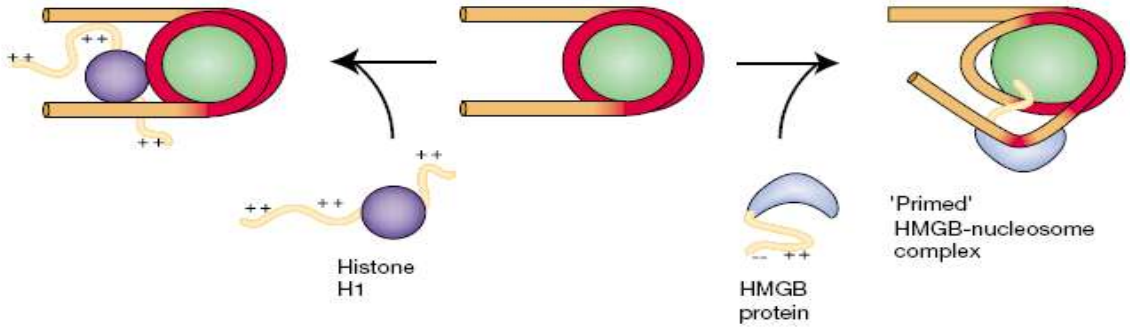


Figure 3.8.3.1: General model for HMG-2 function. HMG-2 protein bound at the entry/exit site induces increased DNA access while less accessible conformation of the nucleosome is stabilized by H1 proteins. Modified from[183].

3.8.4 Histone proteins (H1.2 – 1.5, H1x)

The ten human H1 histones are small basic proteins, consisting of a highly conserved globular domain with variable hydrophilic N- and C-terminal tails.[191] These

proteins are also extensively post-translationally modified.[192] They not only have an essential architectural role of sealing off nucleosomes but also play a role in transcriptional regulation.[193]

3.8.5 Heterogeneous nuclear ribonucleoprotein D0 (hnRNPD / AUF1)

Splicing of AUF1 results in four isoforms: a 37 kDa core protein (p37), a 40 kDa protein (p40) with an N-terminal insertion of exon 2, a 42 kDa protein (p42) with a C-terminal insertion of exon 7 and a 45 kDa protein (p45) containing both insertions.[194] They differ in stabilities, ubiquination and A+U-rich element (ARE)-RNA binding characteristics. ARE in the mRNA 3'-untranslated region (UTR) aids in its degradation. It was found that K_d of ARE binding of AUF1 correlated with the efficiency of an ARE to direct degradation of mRNA.[195]

3.8.6 Heterogeneous nuclear ribonucleoprotein U (hnRNPU / scaffold attachment factor / SAF-1)

hnRNPU binds actin through its C-terminal domain and also associates with the phosphorylated C-terminal domain (CTD) of RNA polymerase II (RNAPII).[196] It also could inhibit CTD phosphorylation by binding to TFIIH and repress transcription elongation.[197] It is also a component of ribonucleoprotein (RNP) particles besides being associated with the transcription activator CBP/p300 and scaffold / matrix attachment regions (S/MAR) on chromosomes coupled to histone acetylation.[198, 199]

3.8.7 THO complex 4 (Ally of AML-1 and LEF-1 proteins / RNA and export factor binding-I / bZIP enhancing factor / Aly)

The human Aly protein is ubiquitously expressed in the nucleus. It was originally associated with a role in the activation of the T-cell receptor α gene.[200] It could also act as a chaperone by interacting with the basic region-leucine zipper (bZIP) TFs and enhancing their DNA-binding.[201] It was also part of the exon junction complex (EJC) which aided in nuclear export and the robustness of mRNA.[202] Lastly, Aly is able to bind directly to some exons but also introns, 5' end of genes near promoters.[203]

3.8.8 Actin

In 1969, the first suggestion of the possibility of nuclear actin was made.[204] Since then, nuclear actin had been implicated in chromatin re-modeling (Figures 3.8.8.1), transcriptional regulation, RNA processing and nuclear export.[205,206,207,208] Actin is also required for all three classes of eukaryotic RNAPs.[209, 210, 211] During transcription, actin interacts with RNAPII including the CTD.[196, 212] It was also found to participate in transcriptional elongation.[213]

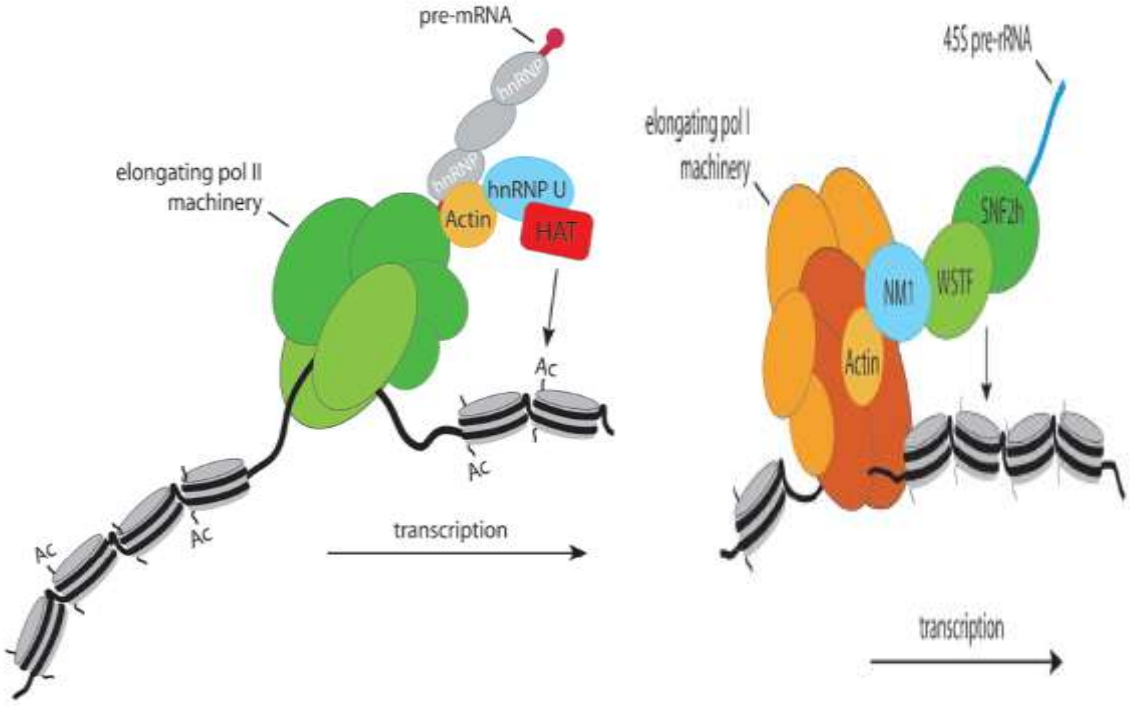


Figure 3.8.8.1: Models of actin involved in RNAPI and II transcription. Left, actin and hnRNPU may recruit the co-activators, histone acetyl transferases (HATs). Right, actin and nuclear myosin recruits the chromatin re-modeling complex. Modified from [214]

3.8.9 DEK

In a specific subtype of acute myeloid leukemia, the chromosome translocation between chromosome 6 and chromosome 9 leads to the fusion of the DEK gene (chromosome 6p22.3) and the CAN (NUP214) gene (chromosome 9q34).[215] This was the first identification of the 375 amino acids, 43 kDa, largely acidic DEK protein.[216]

A scan of brain tumours also showed that DEK was associated with tumour formation.[217] From DNase, RNase and micrococcal nuclease treatment, most of the DEK in the cell was found to be associated with chromatin with only about 10% associated with RNA.[218] Along with DNA topoisomerase I, DEK was able to induce positive supercoils in minichromosomes.[219]

Besides DNA-binding, DEK also has an intimate role in splicing. The spliceosome deposits a 335 kDa complex which includes DEK, ALY and the splicing coactivator SRm160, 20-24 nucleotides upstream of exon-exon junctions, protecting eight nucleotides of mRNA .[220,221]

3.8.10 Poly [ADP-ribose] polymerase 1 (PARP1 / NAD(+) ADP-ribosyltransferase 1 or ADPRT1 / Poly[ADP-ribose] synthase 1)

PARP1 is an abundant nuclear enzyme (1 per 50 nucleosomes) that transforms nicotinamide adenine dinucleotide (NAD^+) and ATP into poly-(ADP)-ribose (PAR), up to 200 units, and adds it to various acceptor proteins via glutamic acid residues. It binds to DNA through its DNA-binding domain consisting of two zinc fingers while catalyzing the PAR reaction through its C-terminal catalytical domain. Between these domains lie an auto-modification domain where phosphorylation by extracellular signal-regulated kinases 1/2 (ERK1/2) and AMP-activated kinase (AMPK) result in modification of PARP1 activation.[222,223] Although as previously mentioned, PARP1 is able to repair DNA under mild or moderate stress, severe stress, however, can lead to the overactivation of PARP1. This would then lead to cell death, both necrosis and apoptosis.[224] Knowing PARP1's involvement in cell death is especially important in the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) model of Parkinson's disease. The PARP1 inhibitor, benzamide, was able to reduce the MPTP-induced neurotoxicity.[225,226] Furthermore, PARP knocked out mice showed that the dopamine containing neurons of the subtantia nigra were spared from the MPTP toxicity.[227]

3.8.11 54 kDa nuclear RNA and DNA binding protein (p54nrb / Nono)

NONO is an abundant nuclear factor. NONO was implicated in the formation of paraspeckles and was able to bind to DNA and regulate gene transcription.[228, 229] It was implicated in transcriptional regulation, splicing, DNA unwinding, viral RNA processing, cell proliferation and circadian rhythm.[230, 231] NONO was also shown to be one of the major interacting-proteins with DJ-1 in dopaminergic neuronal cells and blocked oxidative stress as well as α-synuclein-induced cell death.[232]

3.8.12 Transcription factor CP2

CP2 also known as leader-binding protein (LBP)-1c or late simian virus 40 factor is a transcription factor that was first identified to bind to the promoter of the murine α-globin and that the murine CP2 was 96.4% identical to the human CP2 protein.[233] Analysis of the CP2 protein domains highlighted that it shares certain motifs with that of the histone linker protein H1 amongst other DNA-binding proteins and that it recognises a fragmented sequence motif spanning one DNA helix turn.[234] It falls under the neurogenic element binding transcription factor (NTF)-like family of transcription factors including LBP1a. When CP2 knockout mice were generated, however, very little effect was seen suggesting that LBP1a or other compensatory mechanisms might have taken over.[235]

In British and French populations with sporadic Alzheimer's disease, an association with a non-coding polymorphism in the 3' untranslated region of CP2 was found with a trend seen with the north American population.[236] This might be due to a decrease in binding to other nuclear proteins. This association was confirmed in another British cohort, two different north American cohorts as well as an Italian cohort.[237,238,239,240] Extensive *in vitro* work had subsequently shown that the phosphorylation of the amyloid precursor protein intracellular domain (AICD) led to the formation of a ternary complex with Fe65 and CP2 which induced the expression of glycogen synthase kinase-3β.[241,242] Furthermore, CP2 regulates the expression of transferrin by binding to its promoter region and that this regulation is modulated by $A\beta_{1-42}$ and $A\beta_{25-35}$.[243] Hence, because of its association of with Alzheimer's disease both on the genetic and protein level, this makes CP2 a highly promising predicted transcription factor. In Figure 3.8.12.1, besides interacting with

GATA-1 and Fe65 (APBB1) that had been highlighted, STRING also suggested that CP2 can associate with the shortlisted proteins through HDAC1 and the transcription factor YY1. The HDAC1-CP2-YY1 interaction was obtained through the National Cancer Institute's Pathway Interaction Database.

3.8.13 Transcription factor Staf or ZNF143

Staf or selenocysteine tRNA gene transcription activation factor, was first identified in *Xenopus laevis* as a transcription factor which binds to and activates the selenocysteine tRNA gene using its zinc fingers.[244,245] The human ortholog zinc finger protein 143 (ZNF143) was chromosomally fine-mapped to 11p15.3-p15.4.[246] ZNF143 was predicted by STRING to associate with the list of shortlisted proteins through protein subunits common to the RNA polymerase complexes. This was derived from Reactome's curated pathway which highlighted ZNF143's involvement in transcription via RNA polymerase III (Figure 3.8.13.1).

3.8.14 Transcription factor HIC1

Hypermethylated in cancer 1 (HIC1) was first identified as a candidate tumour suppressor gene activated by p53.[247] From its name, HIC1 was hypermethylated in multiple tumour types as well as gliomas.[248] Its genomic structure consists of alternative promoters that lead to various isoforms, the two main isoforms termed 1a and 1b. This gene encodes a sequence-specific transcriptional factor with an N-terminal BTB/POZ (Broad complex, Tramtrack and Bric a brac/Poxviruses and Zinc Finger) protein dimerization domain as well as a C-terminal sequence specific DNA-binding domain consisting of five Krüppel-like C_2H_2 zinc fingers.[249,250] This protein also has a central region which interacts with the co-repressor protein CtBP (C-terminal binding protein) and a SUMOylation/acetylation switch motif controlled by the deacetylases HDAC4 and SIRT1 (Figure 3.8.14.1 B). HIC1 was also known to be post-translationally modified by O-GlcNAc glycosylation.[251] Although this modification did not have any effect on the full-length protein, it did have an effect on the N-terminal truncated protein. It was plausible that HIC1 could associate with the list of proteins identified through HDAC1 (same as CP2) and p53 as shown in Figure 3.8.14.2.

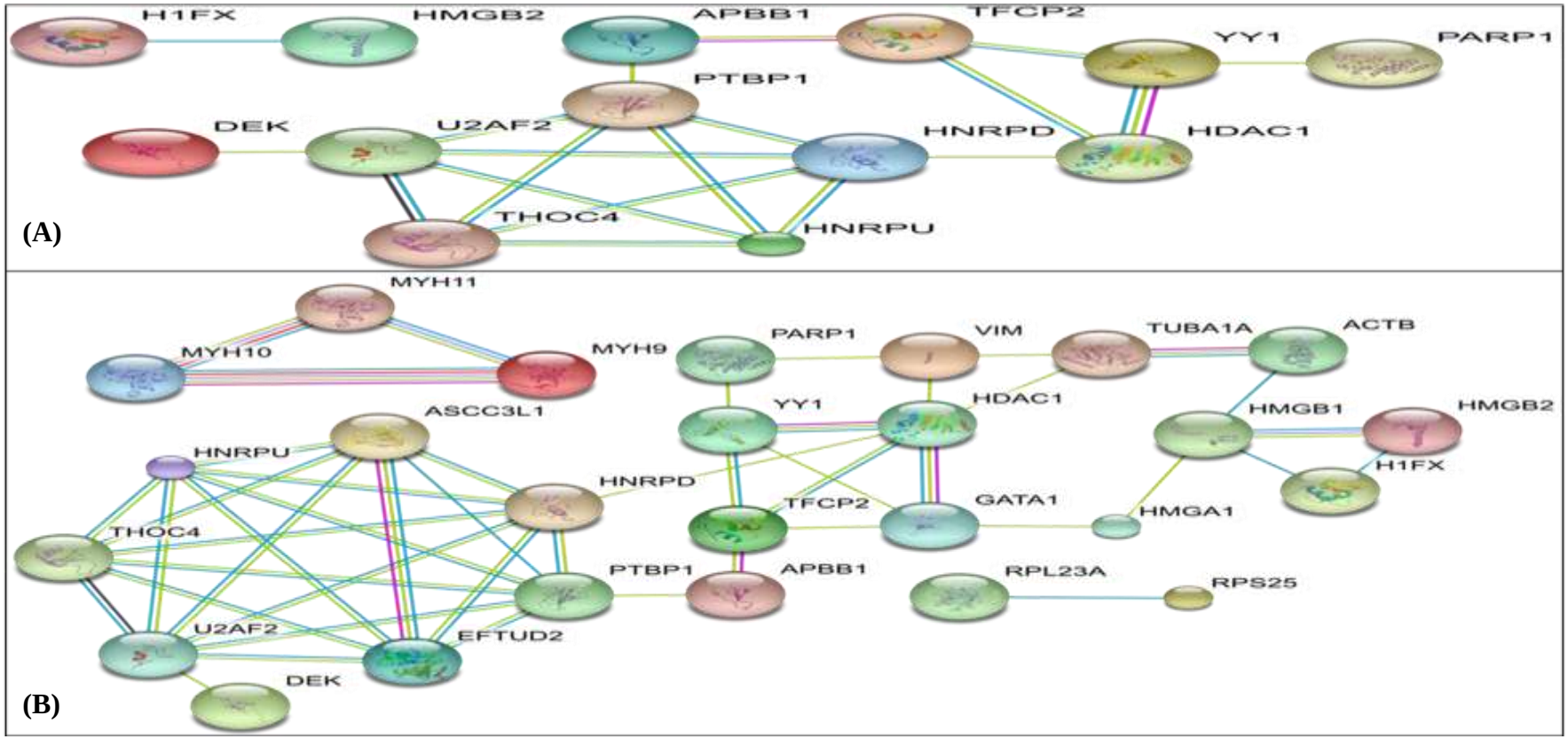


Figure 3.8.12.1: High confidence (0.7<P<0.9) protein-protein interactions for CP2 protein with shortlisted proteins of interest.

(A) STRING analysis of CP2 and its interactors with shortlisted proteins highlighted in pink from Figure 3.8.1. (B) STRING analysis of CP2 and its interactors with all of the shortlisted proteins from Figure 3.8.1. Blue lines represent associations based on databases. Yellow lines represent associations based on textmining. Red lines represent associations based on gene fusion. Grey lines represent associations based on coexpression studies. Purple lines represent associations from experiments. Light blue lines represent homology based association. Obtained using String v. 8.0 (string.embl.de).

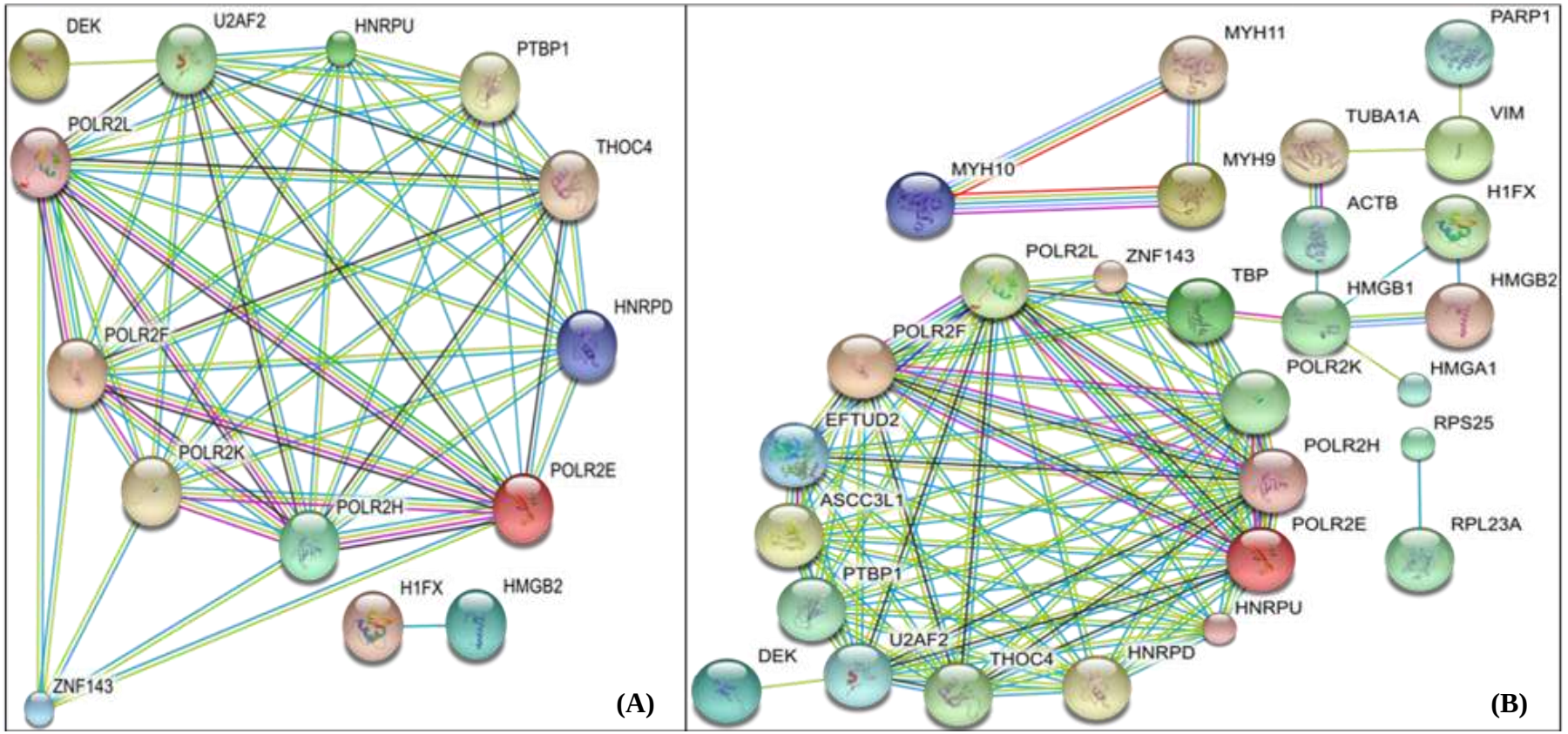


Figure 3.8.13.1: High confidence (0.7<P<0.9) protein-protein interactions for Staf (ZNF143) protein with shortlisted proteins of interest. (A) STRING analysis of Staf and its interactors with shortlisted proteins highlighted in pink from Figure 3.8.1. (B) STRING analysis of Staf and its interactors with all of the shortlisted proteins from Figure 3.8.1. Blue lines represent associations based on databases. Yellow lines represent associations based on textmining. Red lines represent associations based on gene fusion. Grey lines represent associations based on coexpression studies. Purple lines represent associations from experiments. Light blue lines represent homology based association. Obtained using String v. 8.0 (string.embl.de).

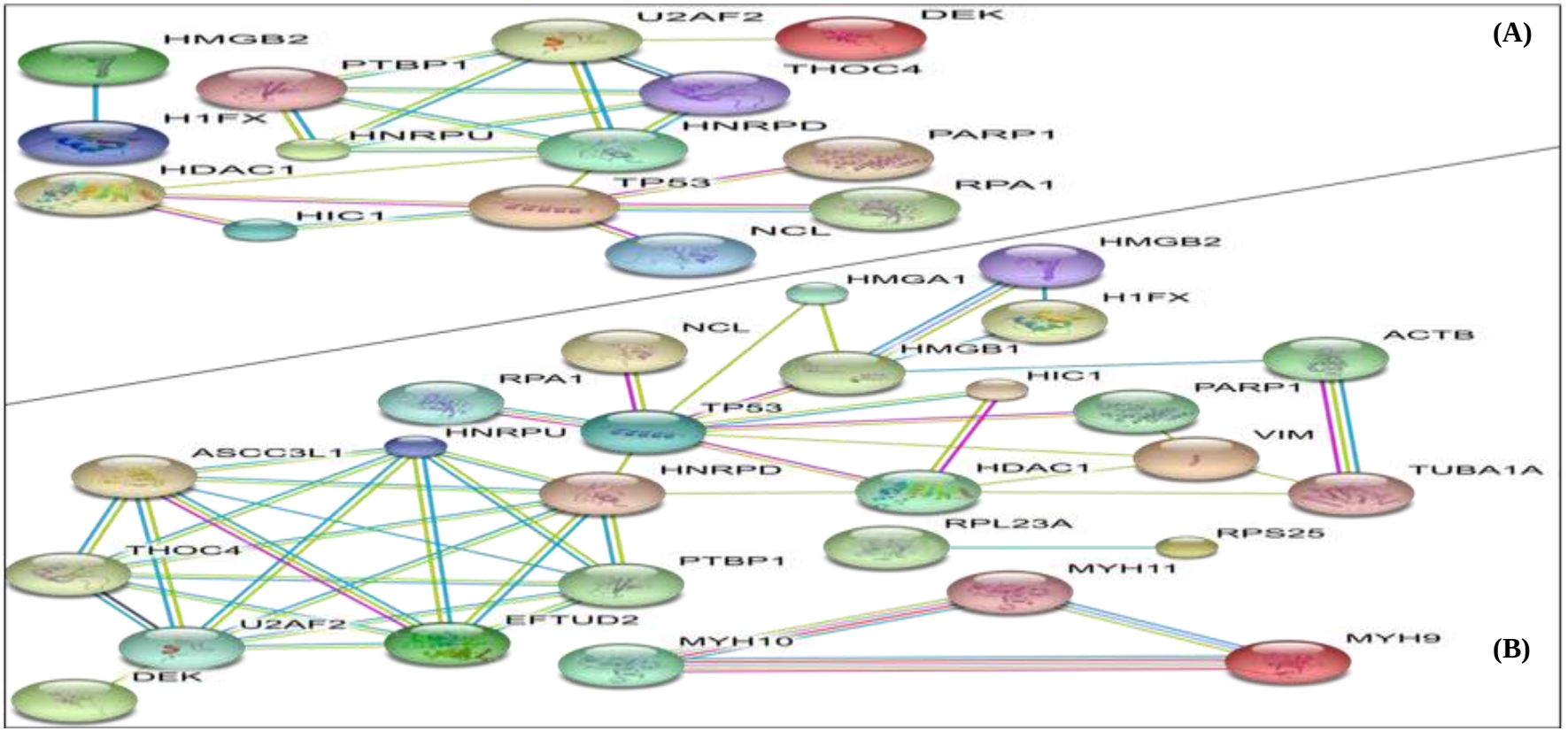


Figure 3.8.14.1: High confidence (0.7<P<0.9) protein-protein interactions for HIC1 protein with shortlisted proteins of interest. (A) STRING analysis of HIC1 and its interactors with shortlisted proteins highlighted in pink from Figure 3.8.1. (B) STRING analysis of HIC1 and its interactors with all of the shortlisted proteins from Figure 3.8.1. Blue lines represent associations based on databases. Yellow lines represent associations based on textmining. Red lines represent associations based on gene fusion. Grey lines represent associations based on coexpression studies. Purple lines represent associations from experiments. Light blue lines represent homology based association. Obtained using String v. 8.0 (string.embl.de).

3.9 Conclusions from analyses of proteins identified

Understanding how genes function requires an appreciation of the dynamic nuclear architecture. The results discussed so far are promising in highlighting the relationship between the risk associations of a SNP with neurological diseases and how a SNP influences the biology and subsequently the pathology of a neurological disease. The literature and evidence on the most likely proteins involved with rs242557 substantiate the most confident interactions highlighted by the STRING and Reactome programs. The hypothesis now seems to be that the region containing rs242557 is either occupied by histone proteins probably during a transcriptionally repressed state or is bound to a range of factors involved in the regulation of transcription, splicing and post-transcriptional activities. Also, of all the predicted transcription factor binding proteins, CP2 seems like a very promising candidate protein that should be included with the other experimentally shortlisted proteins and pursued further.

The effects of the individual DNA-binding proteins will have to be rigorously studied. First, the proteins of interest highlighted need to be verified using antibodies and applied to the pull-down experiment. Other more stringent experiments to illustrate proteins binding to the region of interest such as chromatin immunoprecipitation (ChIP) or gel shift assays will be done on nuclear extracts of human neuroblastoma cells. Carrying out these experiments will shed some light on how these proteins influence the transcription and/or splicing of the human tau gene and how this varies with different genotypes.

Furthermore, TFs that regulate *MAPT* transcription will also be candidates for genetic studies in order to ascertain if polymorphisms are associated with the *MAPT* H1c risk variant and lead to disease pathogenesis in AD and PSP. It is also important to remember that rs242557 lies within a 182-bp conserved region and looking at this region more closely might shed more light on the effects of rs242557 on *MAPT*.

4 Chapter 2 Results

4.1 Validation of DNA binding proteins of interest identified affinity binding assay

To validate the 13 shortlisted proteins as well as one predicted binding protein and a control protein, we analysed these interactions with antibodies raised against these proteins (Table 4.1.1). HDAC1 had been predicted through the bioinformatics tool STRING to be able to interact with hnRNPD0 from the shortlisted proteins as well as vimentin and brain specific tubulin from the rest of the proteins identified. Hence, a subunit of the replicative S-phase helicase MCM complex, MCM2, was used as a nuclear control protein instead. In the first place, since mostly untested, we had to optimise the conditions of these commercially available antibodies for correct dilutions as well as to assess the specificity of the antibodies.

Figure 4.1.1, shows the optimization of nine of the antibodies (PARP1, PTBP1, NONO, U2AF2, DEK, actin, hnRNPD0, ALY, HMGB2) at different antibody dilutions and illustrates the specificity of the antibodies. hnRNPD0 is produced in human cell lines as four main isoforms. The rest of the shortlisted antibodies (hnRNPU, nucleolin, RPA70 and H1) titrated fairly satisfactorily except that they had varying levels of background with anti-H1 antibody having the worst background. The antibody against the predicted protein CP2 not only titrated well but had minimal background as well. Lastly, the antibody against the control nuclear protein MCM2 was not as specific compared to the other antibodies but was sufficient for the purpose of a control protein.

No.	Antibody	Supplier	Cat. No.
1	Monoclonal mouse anti-human PARP1	BD Pharmigen	556494
2	Monoclonal mouse anti-human MCM2	Abcam	Ab6153
3	Monoclonal mouse anti-human hnRNPU	Abcam	Ab10297
4	Monoclonal mouse anti-human Nucleolin	Abcam	Ab13541
5	Polyclonal rabbit anti-human RPA70	Abcam	Ab12320
6	Monoclonal mouse anti-human PTBP1	Abcam	Ab30317
7	Polyclonal rabbit anti-human CP2	Abcam	Ab42973
8	Polyclonal rabbit anti-human NONO	Abcam	Ab70335
9	Polyclonal rabbit anti-human U2AF2	Abcam	Ab37483
10	Monoclonal mouse anti-human DEK	BD Biosciences	610948
11	Monoclonal rabbit anti-human Actin	Sigma	A2066
12	Polyclonal rabbit anti-human hnRNPD0	Millipore	07-260
13	Monoclonal mouse anti-human ALY	ImmuQuest	IQ221
14	Monoclonal mouse anti-human HMGB2	Abcam	Ab11354
15	Monoclonal mouse anti-human H1	Abcam	Ab11079

Table 4.1.1: List of antibodies used.

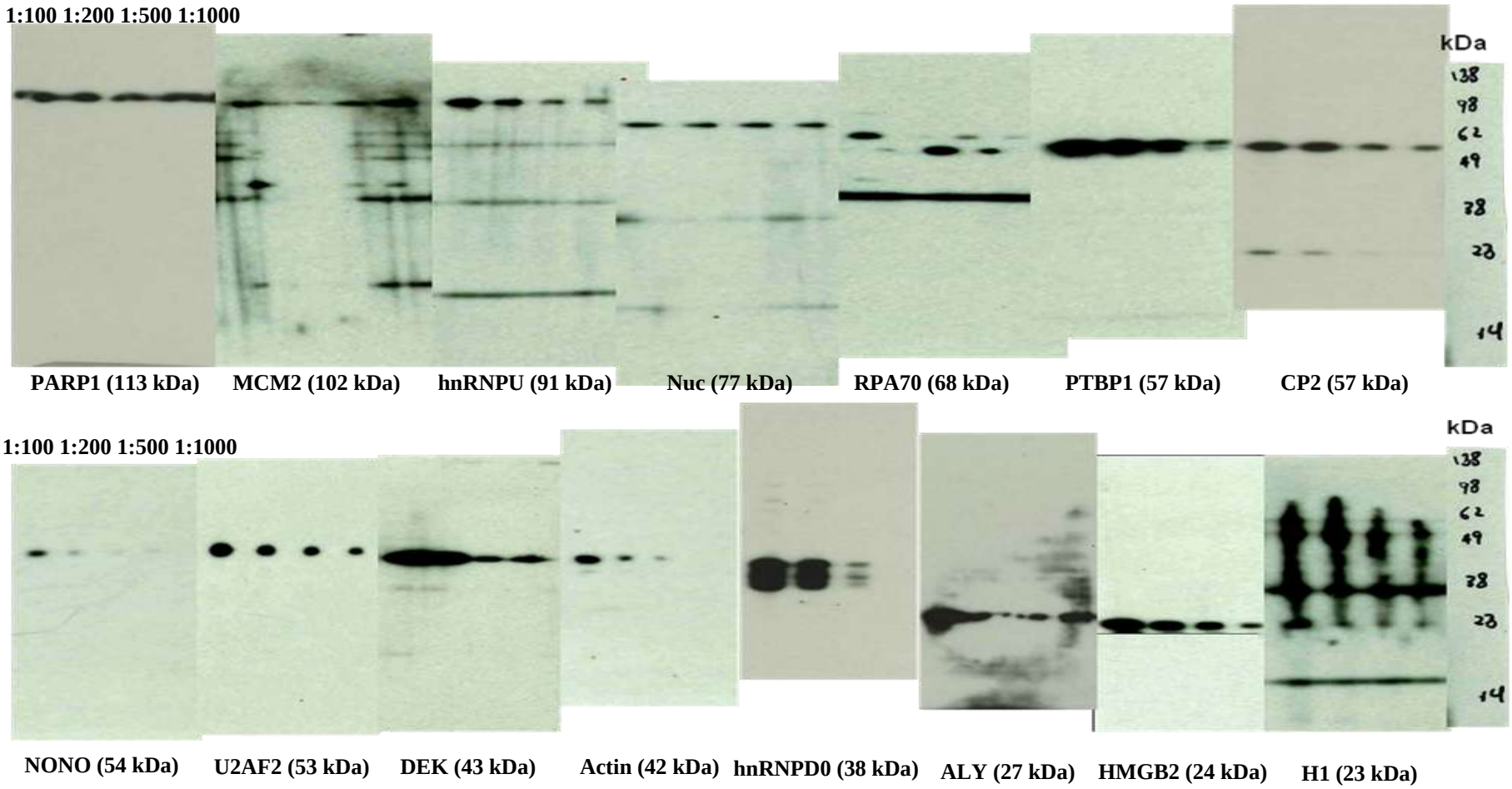


Figure 4.1.1: Antibody titrations for proteins of interest.

Different antibody concentrations such as 1:100, 1:200, 1:500 and 1:1000 were tested on SH-SY5Y nuclear extracts. Extracts were run on a 1-well gel before a 14-slot press was used so sparing amounts of antibodies were used

Once the antibodies had been tested and optimised, a much smaller-scale affinity pull-down experiment similar to that described in Chapter 3.3 was carried out. This was because the Western blot analysis that the fractions were subject to is more sensitive than the Coomassie protein staining. Eight of the 13 shortlisted proteins (U2AF2, hnRNPU, ALY, PTBP1, hnRNPD0, HMGB2, DEK, and PARP1) and the predicted transcription factor CP2 yielded reproducible results for their supernatant fractions (Figure 4.1.2) as well as their elution fractions (Figures 4.1.3 – 4.1.6). The supernatant fraction refers to all the proteins that remained in the cell lysate after incubation with the DNA concatamers of the rs242557 region and the elution fraction refers to all the proteins that are eluted after washing. Although nucleolin, RPA70, NONO, actin and H1 had reproducible staining in their supernatant fractions (Figure 4.1.2 continued), their elution fractions however, were not as reproducible and hence they were not included. None of the wash fractions were included as they were of a larger volume, and hence none of the proteins could be detected because they were present at a very low concentration either due to the dilution effect or tight binding to the DNA concatamers.

MCM2 was chosen as the nuclear control protein as HDAC1 might be involved with some of the proteins mentioned previously. No MCM2 was detected in the elution fractions. In the supernatant fraction, the intensity of the MCM2 bands were fairly even except for the R2 (randomly designed control concatamers 2) pull down which had the least intense band. Taking into account this Western blot profile of the control protein for the supernatant, most of the proteins probed for in the supernatant had fairly even intensities for the various pull downs. The only marked difference was U2AF2. The pull downs with H1c-A and H1b-G for U2AF2 in the supernatant were markedly reduced compared to the randomly designed control concatamer probes R2 and R3 consistently and as low as the beads only pull down for two out of the three replicates.

For those protein pull-downs with beads only i.e. those that did not have any concatamers ligated to them and still resulted in a high background, this is probably not just due to non-specific binding to the beads but also probably due to binding to the linker DNA attached to those beads.

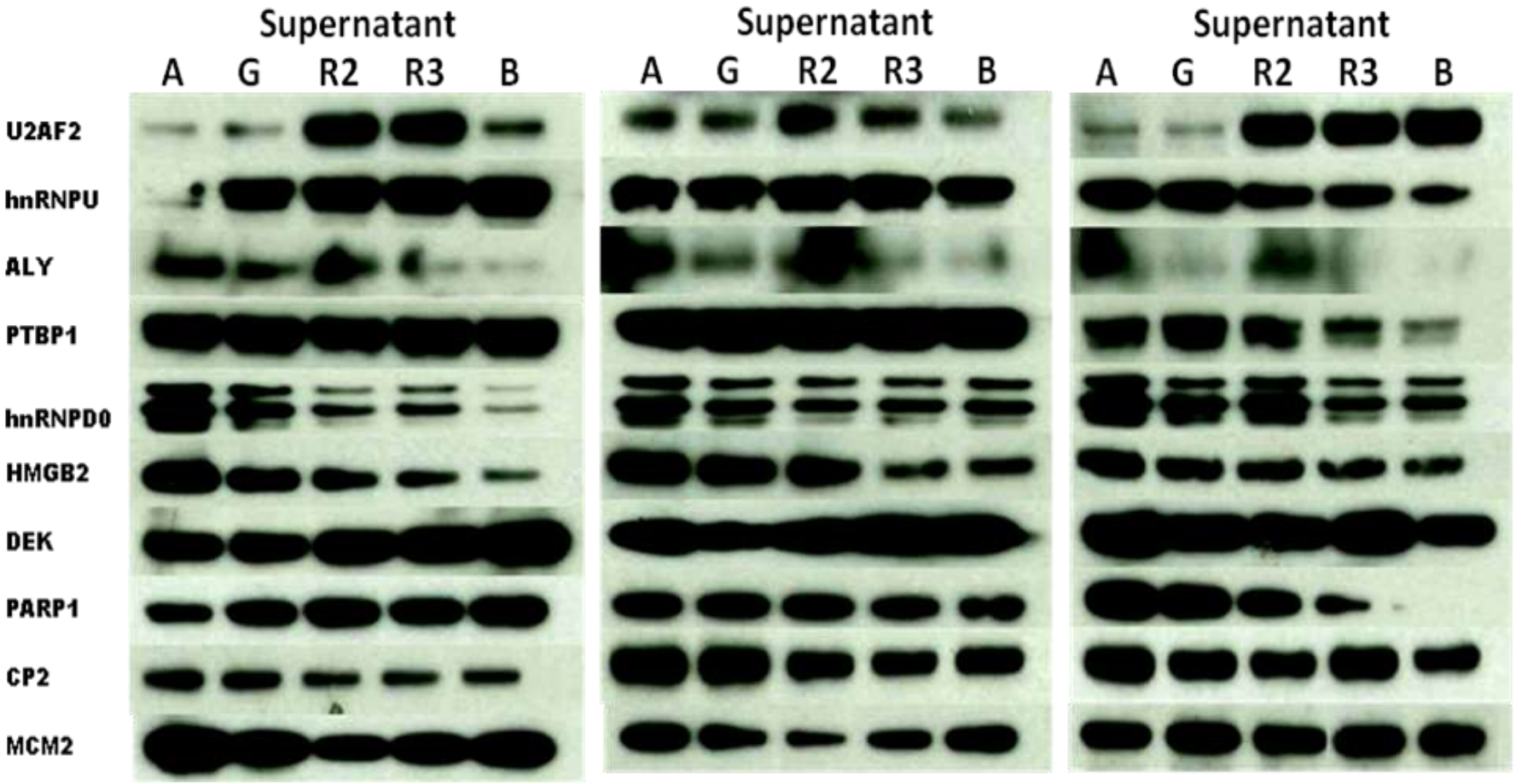


Figure 4.1.2: Supernatant fractions of rs242557 pull-down assay.

SH-SY5Y nuclear extracts were incubated with concatamers or rs242557-containing oligonucleotides attached to magnetic beads (H1c-A, H1b-G, R2 & R3 – random sequences, B – magnetic beads only). Supernatant fractions were probed with antibodies against the proteins of interest (that were detected in the eluates too) as well as control nuclear protein. These were done in triplicate.

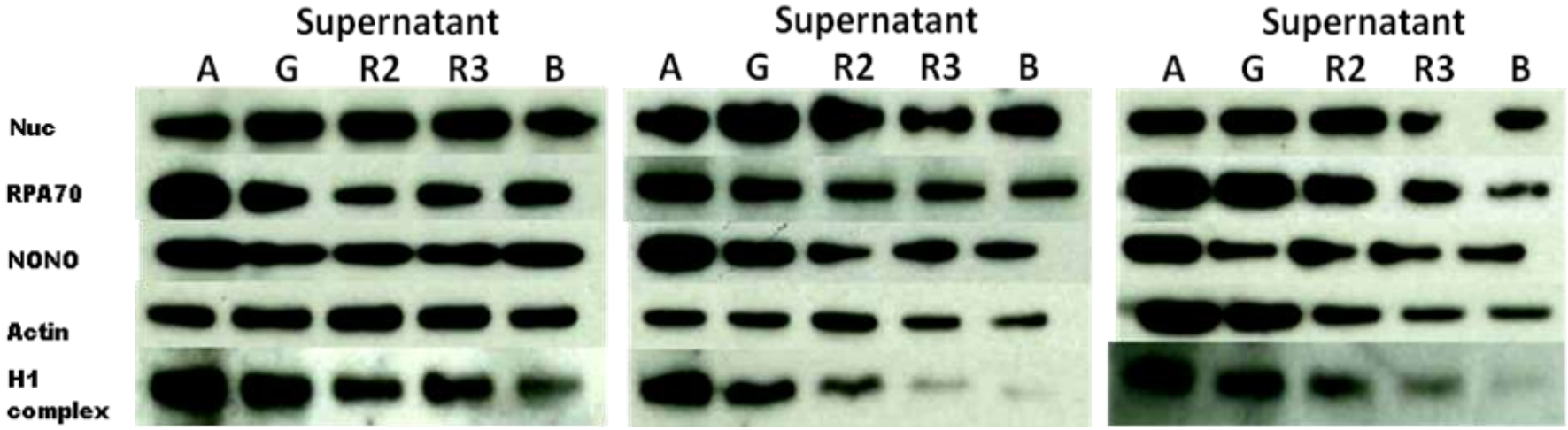


Figure 4.1.2 continued: Supernatant and wash fractions of rs242557 pull-down assay.

These proteins of interest were grouped together as subsequent elution fractions for these proteins did not yield any detectable signal or reproducible bands even though their supernatant profiles were fairly similar.

Looking at the Western blots of the elution fractions (Figures 4.1.3 – 4.1.6) and starting with the shortlisted proteins with the highest confidence (Figure 4.1.3a) first, it can be seen that all the proteins exhibited more binding to the H1c-A concatamer compared to the H1b-G concatamer. They did differ, however, on the binding to the random concatamers R2 and R3 and control beads only pull down as well as the ratio of binding to H1c-A compared to H1b-G.

The protein that showed the most specific binding to the rs242557 region, H1c-A or H1b-G, or in other words, the proteins that had the least non-specific interaction with the random probes and the beads was PTBP1 (Figure 4.1.3c). This was closely followed by hnRNPD0, U2AF2, hnRNPU, HMGB2 and ALY. Taking into account their respective non-specific interactions, the protein that had the largest difference in binding between H1c-A and H1b-G was U2AF2 (Figure 4.1.3b) where the H1c-A band had almost double the intensity (approximately 1.75x) of the H1b-G band. This was closely followed by hnRNPU and ALY which were very similar (approximately 1.6x). PTBP1 had a ratio of approximately 1.45x while hnRNPD0 had the smallest difference of about 1.2x. HMGB2 which was not directly predicted to interact with the rest of the proteins came in at a ratio of about 1.5x but had the largest standard deviation.

DEK protein in Figure 4.1.4a, was one of the shortlisted proteins that had been predicted to interact with the other proteins when the confidence level was relaxed from highest to high (refer to Chapter 3.5). The level of the protein in the eluate, however, was very low. Although DEK, like the rest of the shortlisted proteins, was easily detected in the supernatant fraction, levels of bound protein in the eluate were very low, even with increase in concentration of detection antibody (Figure 4.1.4c). Nevertheless, the interaction appears specific to the rs242557 probe with no detectable bands with the random probes, R1 and R2 and beads only. However, with the weak intensity, it is possible that non-specific interaction, if present, could not be detected. In any case, density analysis revealed the largest difference in allele-specific binding to the rs242557 probes (Figure 4.1.4b).

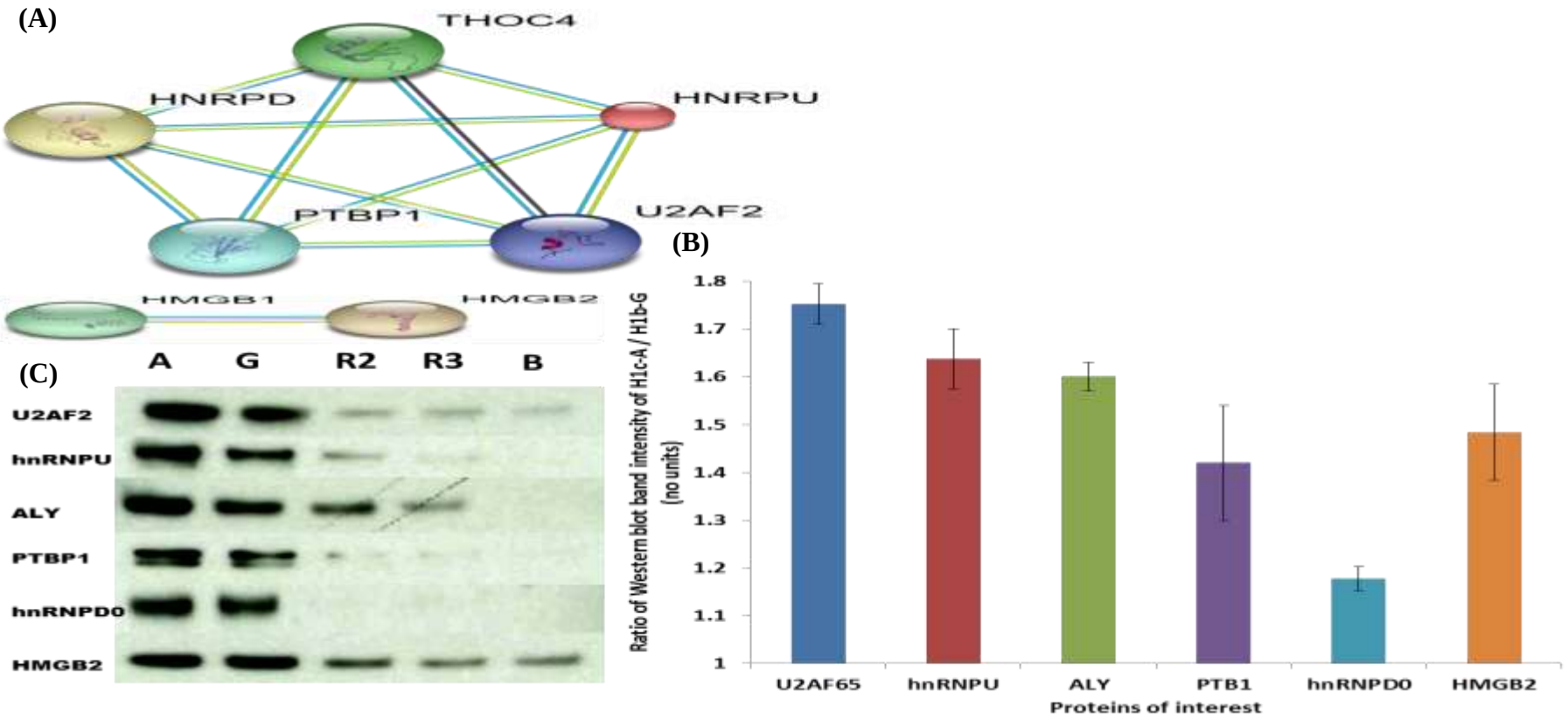


Figure 4.1.3: Elution fractions of rs242557 pull-down assay U2AF2 – HMGB2.

SH-SY5Y nuclear extracts were incubated with concatamers of rs242557-containing oligonucleotides attached to magnetic beads (H1c-A, H1b-G, R2 & R3: random sequences, B: magnetic beads only). Eluates were probed with antibodies against the proteins of interest. (A) This set of proteins was shown by STRING to be of highest confidence in protein-protein interactions. (B) Plot of the ratio of band intensity of H1c-A against H1b-G after taking account of non-specific binding for each protein of interest. (C) Western blot analysis of eluates fractions in triplicate, n = 3.

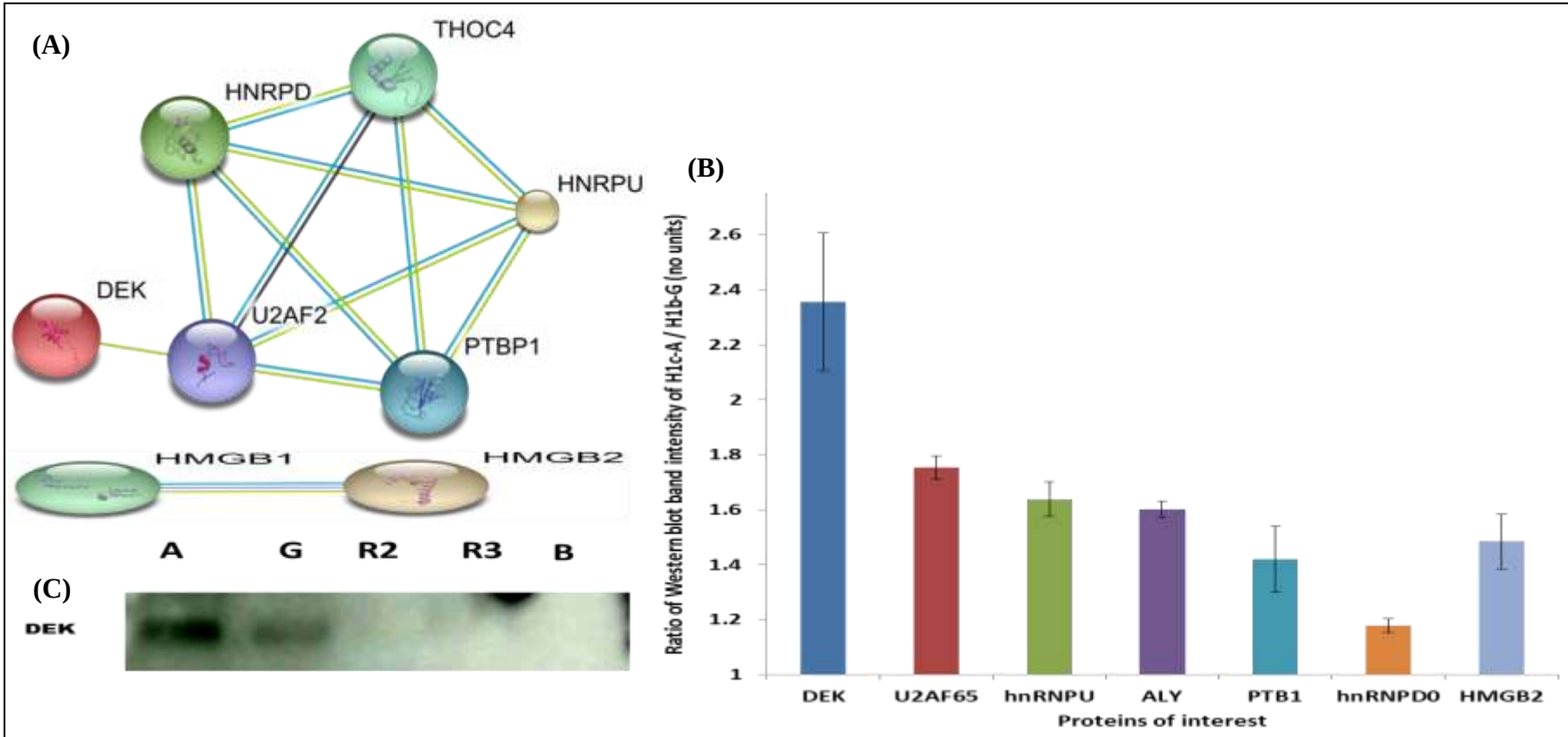


Figure 4.1.4: Elution fractions of rs242557 pull-down assay DEK.

SH-SY5Y nuclear extracts were incubated with concatamers of rs242557-containing oligonucleotides attached to magnetic beads (H1c-A, H1b-G, R2 & R3: random sequences, B: magnetic beads only). Eluates were probed with antibodies against the proteins of interest. (A) This set of proteins was shown by STRING to be of high confidence in protein-protein interactions. (B) Plot of the ratio of band intensity of H1c-A against H1b-G after taking account of non-specific binding for each protein of interest. (C) Western blot analysis of eluates in triplicate, n=3

The last of the shortlisted proteins that could be reproducibly detected was PARP1 (Figure 4.1.5). It was, however, the least reliably reproducible protein blot with the largest standard deviation. It also had substantial non-specific binding to the random probes and beads (Figure 4.1.5c). Even though unlike the rest of the proteins, PARP1 seemed to favour binding to H1b-G, this difference is not reliable when the level of background is considered.

Lastly, CP2, the transcription factor not found in our interactor hunt but predicted *in silico* to bind to the rs242557 region was probed for in the eluate (Figure 4.1.6). It showed specific binding to the rs242557 probes and very little non-specific binding (Figure 4.1.6c). Like most of the shortlisted proteins tested, there appeared to be allelic binding to rs242557 favouring H1c-A by about 1.5x (Figure 4.1.6b).

Looking at the summary of Western blot density analysis in Figure 4.1.6, apart from PARP1, the rest of the proteins in this denaturing pull-down experiment seem to exhibit the same allelic preferential binding and at approximately to the same extent. Whether these results will hold in the native condition will still need to be elucidated. A direct method is to probe the proteins as they natively associate with the DNA probes as they would do in an EMSA experiment.

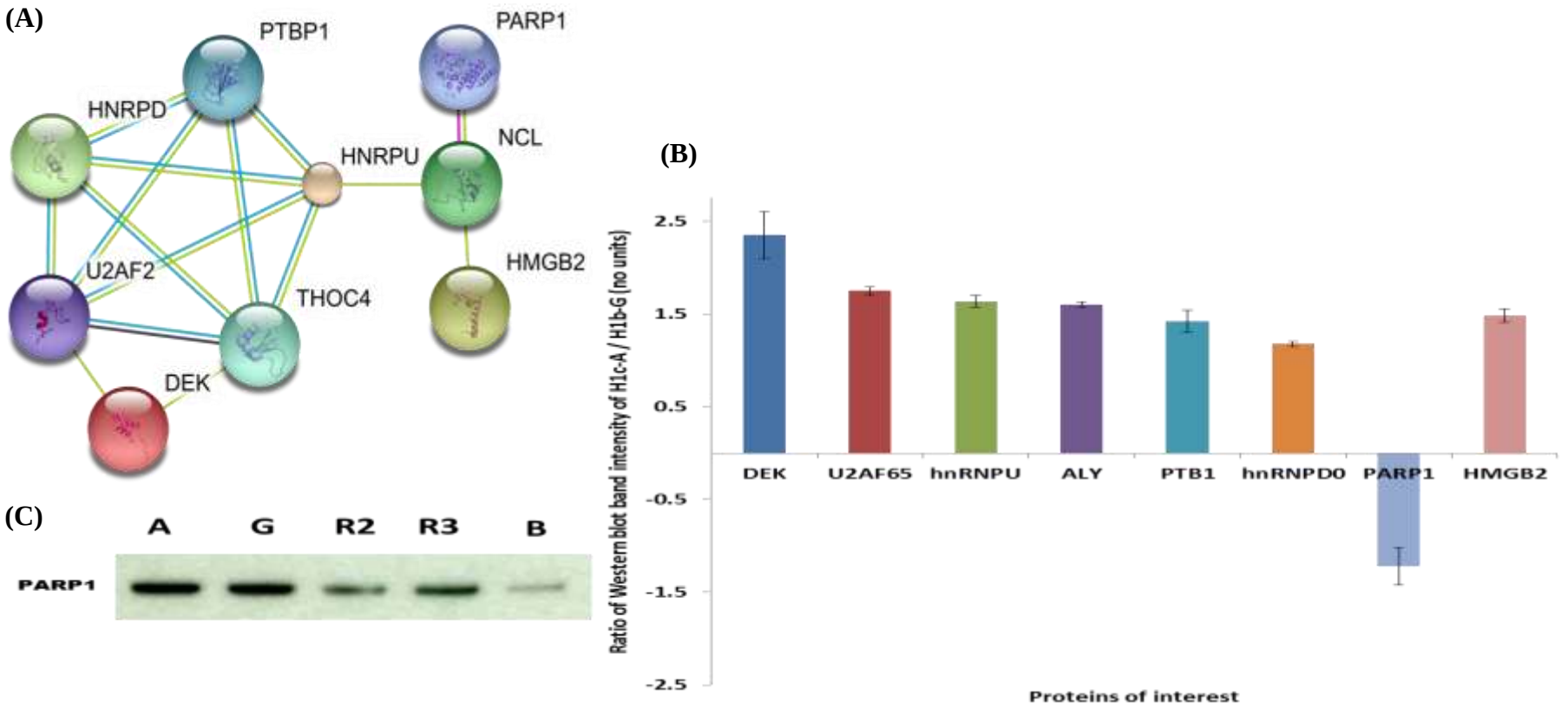


Figure 4.1.5: Elution fractions of rs242557 pull-down assay PARP1.

SH-SY5Y nuclear extracts were incubated with concatamers or rs242557-containing oligonucleotides attached to magnetic beads (H1c-A, H1b-G, R2 & R3 – random sequences, B – magnetic beads only). Elution fractions were probed with antibodies against the proteins of interest. (A) Theses set of proteins were shown by STRING to be of medium confidence in protein-protein interactions. (B) Plot of the ratio of band intensity of H1c-A against H1b-G after taking account for non-specific binding for each protein of interest. (C) Western blot bands of elution fractions in triplicate, n=3.

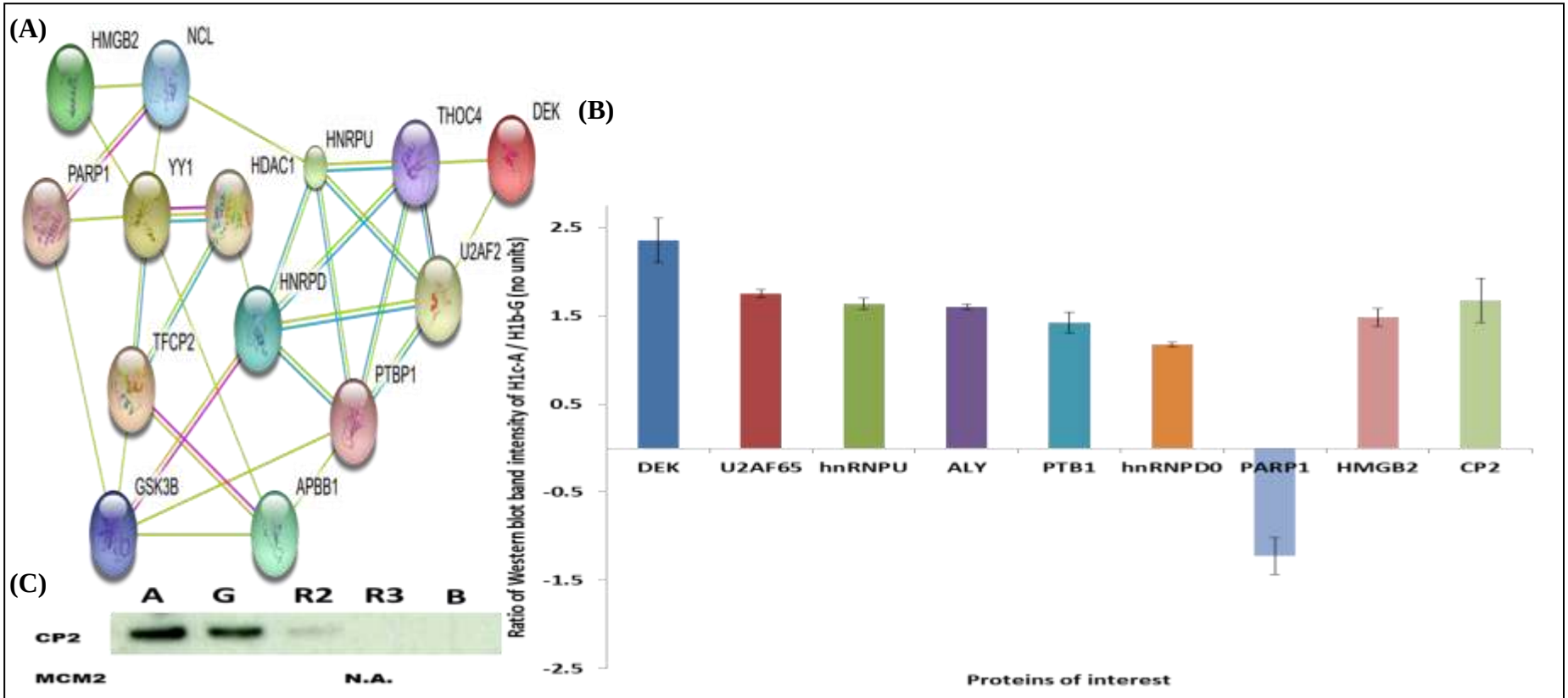


Figure 4.1.6: Elution fractions of rs242557 pull-down assay CP2.

SH-SY5Y nuclear extracts were incubated with concatamers or rs242557-containing oligonucleotides attached to magnetic beads (H1c-A, H1b-G, R2 & R3 – random sequences, B – magnetic beads only). Elution fractions were probed with antibodies against the proteins of interest. (A) Theses set of proteins were shown by STRING to be of medium confidence in protein-protein interactions with CP2. (B) Plot of the ratio of band intensity of H1c-A against H1b-G after taking account for non-specific binding for each protein of interest. (C) Western blot bands of elution fractions in triplicate, n=3.

4.2 Validating proteins of interest binding in native conditions

Before the shortlisted and predicted proteins can be probed for in their native conditions interacting with the DNA region of interest, the Western blotting protocol had to be adapted to allow for the transfer of large protein complexes in their native conditions from native gels to PVDF membranes. A titration experiment using the cytoplasmic extract from SH-SY5Y cells was separated on native gels and visualized with both silver staining and Coomassie staining. An unstained native protein ladder from 1236 kDa to 66 kDa was used too. The range of protein extract used from 270 ng to 3 ng tested the limits of both silver and Coomassie stainings.

Two different approaches for transferring native proteins onto PVDF membranes were tested. A gentle, SDS-free method where the native gel was soaked in 2x transfer buffer was compared to a method involving 0.1% SDS. The blot was then stained with Coomassie to visualize the proteins transferred as silver staining did not work on PVDF membranes. The gels after native Western blotting were then stained with both silver and Coomassie stains to see how much of the proteins were left on the gels. The gentler approach was able to transfer protein complexes around 200 kDa even when using 9 ng of starting cytoplasmic extract while protein complexes around 720 kDa were transferred down to 18 ng of starting material. Even complexes as big as 1048 kDa were detected on the membrane when using 44 ng or more of cytoplasmic extracts. The stained gels after transfer lacked the protein complex bands seen before blotting. Thus, the gentler approach would fit for the purpose of this experiment.

Of all the highest confidence shortlisted protein, U2AF65 was the only protein stained by its antibody on a native blot (Figure 4.2.1) i.e. without having to denature the blot. This meant that the U2AF65 antibody epitope was accessible in the native condition when it was part of a complex approximately greater than 1000 kDa. Since this was a native EMSA-like experiment, there was a certain amount of non-specific binding to random probes R2 and R3. Also, as U2AF65 was one of the main components of the major splicing pathway, it appears that within the nuclear extract, it could be present in various complexes. As the gel run was optimized to visualize protein complexes, it was not possible to see the levels of individual proteins within the same blot as they would have run

out of the gel if they were detectable as individual proteins in native nuclear extracts. As a further control, the DNA probes on their own without the nuclear extracts but with all the EMSA reagents used [(A) & (G)] were not able to create any false positive bands when exposed to the native Western blotting procedure. From the high level of non-specific binding to R2 and R3, it is hard to make any judgment on allele-specific binding.

The only other protein that responded to the experiment was hnRNPD0. However, its antibody could not detect the epitope in the native complex. After denaturing the blot briefly with 6M guanidine hydrochloride, however, the epitope was sufficiently exposed for the blot to work. Like U2AF65, hnRNPD0 displayed a reverse in allelic preference in the native condition from H1c-A to H1b-G. The level of non-specific binding was higher than U2AF65 but was similar to the denaturing version of the experiment (Figure 4.1.3) and very little of it existed in the complex form without any DNA input. Again, from the high level of non-specific binding to R2 and R3, it is hard to make any judgment on allele-specific binding.

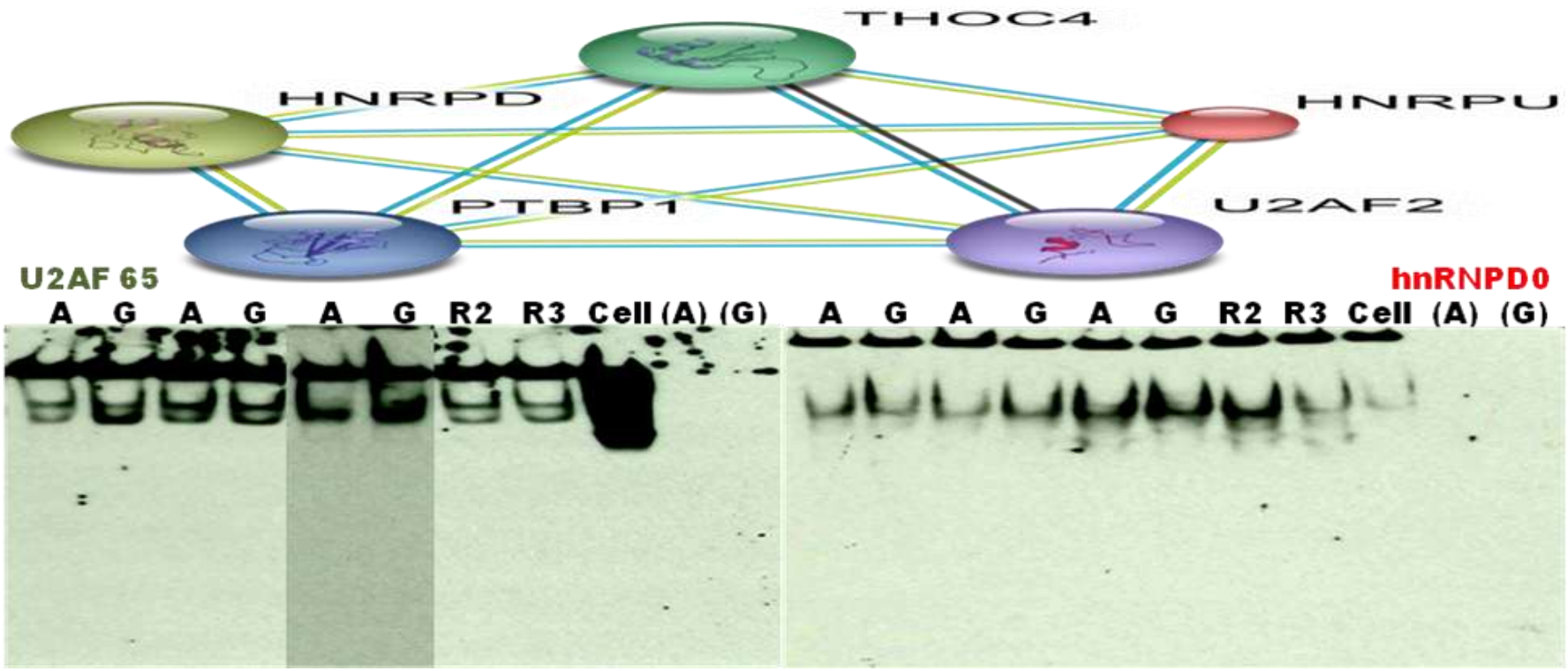


Figure 4.2.1: Native Western blotting of U2AF65 and hnRNPD0.

Biotin-labelled 16-mer oligonucleotides containing the A (H1c) & G (H1b) allelotypes of rs242557 were incubated with nuclear extracts from undifferentiated SH-SY5Y cells. Incubation with randomised probes (R2 & R3) highlighted the specificity of the interaction between probe and nuclear proteins, U2AF65 and hnRNPD0. Other controls included the nuclear extract without any probes (cell), the H1c-A and H1b-G probes without any nuclear extracts [(A) & (G)]. hnRNPD0 is highlighted in red as the blot had to be treated with 6M guanidine hydrochloride before antibody treatment while U2AF65 did not require this pre-treatment. The pairs of allele-specific binding were done in triplicates.

We could not identify the rest of the shortlisted proteins, hnRNPU and PTBP1, H1 and HMBG2 (Figure 4.2.2), ALY, DEK and PARP1 on the native Western blots either without denaturing or denaturing the blots. This did not necessarily mean that these proteins were not present in the native complex. One possibility is that the antibody epitopes are obscured within the complex and that the denaturation was not sufficient to expose these epitopes. The nature of this native method makes the results less clear to interpret and should be cautiously analysed and taken into consideration with the rest of the results in this thesis.

During the course of this work, our group members were also investigating the protein TDP43 which they showed influences *MAPT* transcription by binding to the upstream core promoter region of the gene (Roberto Simone; personal communication). As a control, we assessed the binding of TDP-43 to a probe (T) representing the predicted GC-rich binding domain of the *MAPT* promoter and random probes R2 and R3. Our STRING analysis also showed that TDP-43 interacts with the shortlisted protein hnRNPD0 through ubiquitin although this might not be applicable in this case. Even under such native condition, there was no significant preferential binding above binding to random oligonucleotides.

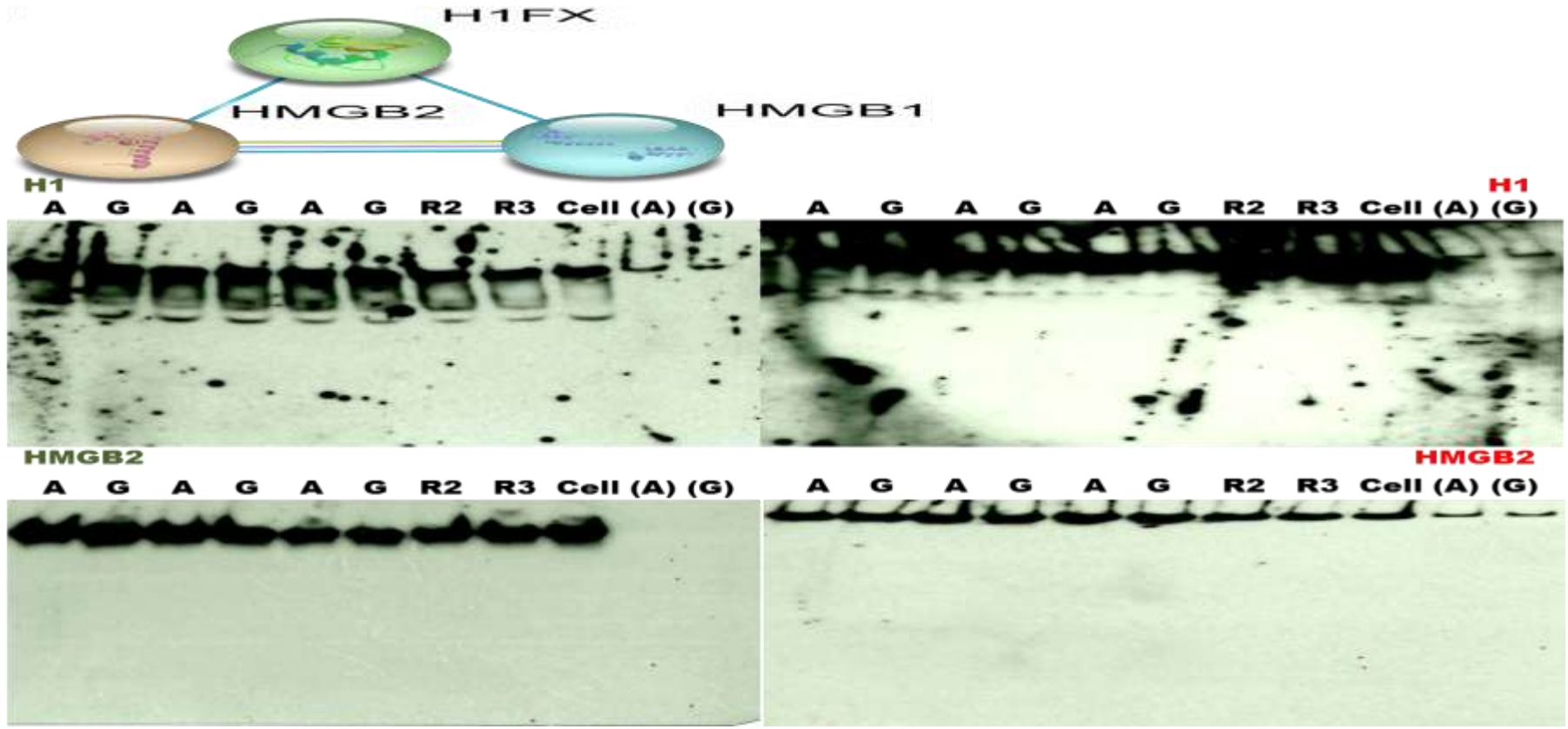


Figure 4.2.2: Native Western blotting of H1 and HMGB2.

Biotin-labelled 16-mer oligonucleotides containing the A (H1c) & G (H1b) allelotypes of rs242557 were incubated with nuclear extracts from undifferentiated SH-SY5Y cells. Incubation with randomised probes (R2 & R3) highlighted the specificity of the interaction between probe and nuclear proteins. Other controls included the nuclear extract without any probes (cell), the H1c-A and H1b-G probes without any nuclear extracts [(A) & (G)]. Treating the blots with 6M guanidine hydrochloride (in red) did not have any effect compared to without (in green). They were done in triplicates.

Three of the rest of the shortlisted proteins, nucleolin, actin and RPA70 and the predicted transcription factor, CP2 (Figure 4.2.3) were much easier to be probed for in the native blots without the blots needing to be denatured.

CP2 and nucleolin like U2AF65 and hnRNPD0 that were previously discussed, had showed greater binding to H1c-A compared to H1b-G. Although nucleolin was present as a complex in the nuclear extract on its own and some binding was seen with the random oligonucleotides, it was interesting to note that in a native blot, no CP2 was seen in the nuclear extract on its own, no binding to the random probe R2 and very little binding to R3. Actin and RPA70 were the only proteins observed in the native blots to show a difference in allelic binding with greater binding to H1b-G rather than H1c-A. RPA70 was like CP2 in that it had specific binding to the DNA probes of interest compared to the random probes and nuclear extract alone whereas actin had a basal level of non-specific binding.

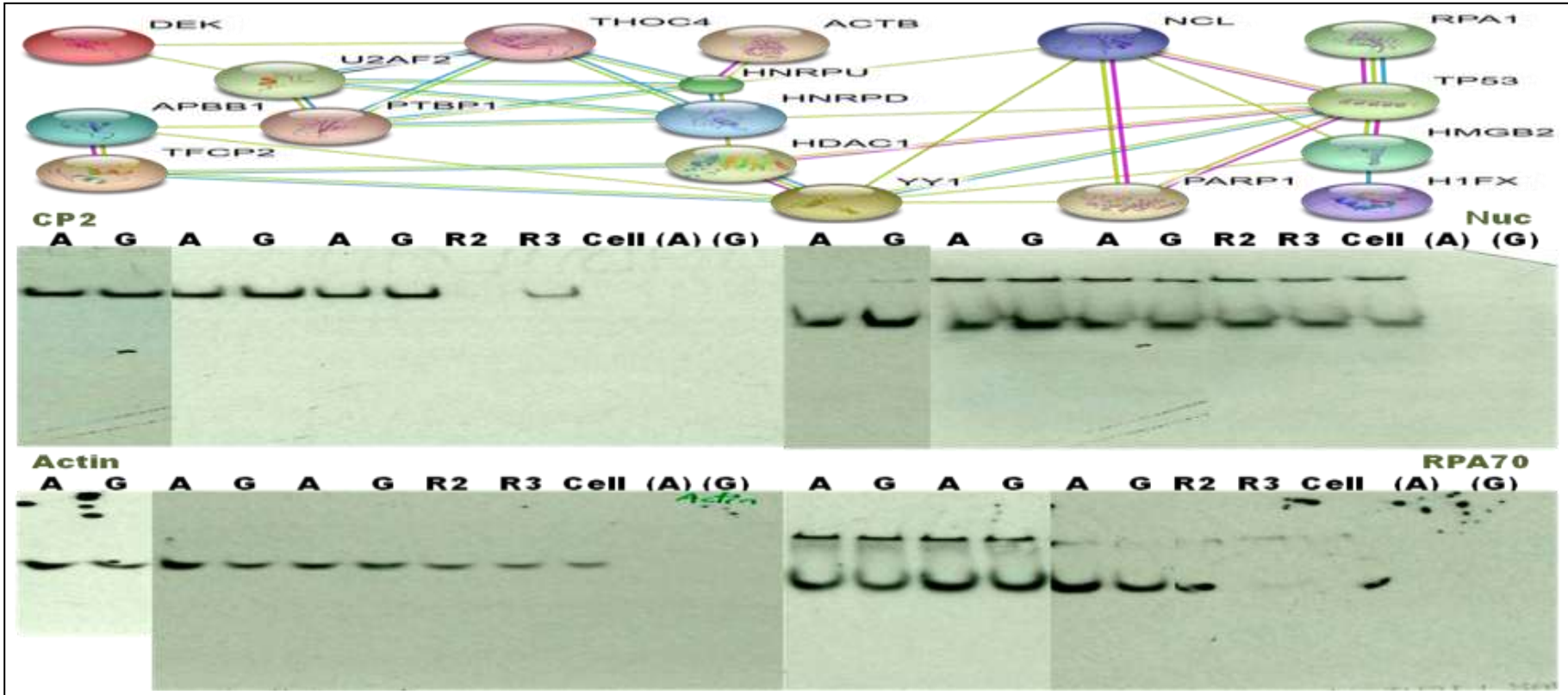


Figure 4.2.3: Native Western blotting of CP2, nucleolin, actin and RPA70.

Biotin-labelled 16-mer oligonucleotides containing the A (H1c) & G (H1b) allelotypes of rs242557 were incubated with nuclear extracts from undifferentiated SH-SY5Y cells. Incubation with randomised probes (R2 & R3) highlighted the specificity of the interaction between probe and nuclear proteins, CP2, nucleolin, actin, RPA70. Other controls included the nuclear extract without any probes (cell), the H1c-A and H1b-G probes without any nuclear extracts [(A) & (G)]. All these blots did not need 6M guanidine hydrochloride treatment. They were done in triplicates.

4.3 Discussion on validating shortlisted DNA-binding proteins

In the work described in this chapter, we aimed to verify the proteins that had been predicted by the mass spectrometry data through the use of antibodies. Besides identifying these proteins through the pull-down assays, they were also probed in native gels that were blotted onto PVDF membranes to assess the association of the proteins with a complex that binds to the rs242557 region. It was also the intention to have used chromatin immunoprecipitation (ChIP) to validate these proteins. However, after many trials of optimisation and the time constraints, it was not possible to do so. The work that had been done on ChIP will be discussed further in the conclusion under future work.

In summary, we confirmed 8 out of 13 of the shortlisted interacting proteins by pull-down experiment followed by denaturing gel electrophoresis as described section 4.1. However, analysis of the interactions on native, non-denaturing gels that would keep the DNA-protein complex intact was largely difficult. While some proteins were observed to bind non-specifically to the random probes, with certain allelotypes, the binding seen were still stronger compared to those random probes. Five of the same shortlist were able to be studied in the native Western blots, understandably due to the nature of conformational changes and protein-protein interactions that come about in native conditions. Only two proteins from the shortlist, U2AF65 and hnRNPD0, were common to both experiments. The predicted transcription factor, CP2 was also common to both denaturing and non-denaturing experimental conditions. Three proteins, actin, nucleolin and RPA70, which had not been able to be reproducibly studied in the denaturing pull-down experiment, were probed with greater ease in the native Western blots. All of these are illustrated in Figure 4.3.1.

It was interesting to note that when the complexes were formed natively before being denatured, more epitopes for the antibodies against U2AF2, hnRNPD0 and CP2 were seen with rs242557 H1c-A compared to H1b-G but the reverse was true without any denaturation. This could be due to the different nature of the native complexes formed around the oligonucleotides where H1c-A recruited more U2AF2, hnRNPD0 and CP2 as seen when the complexes were denatured but were less exposed or had their epitopes more buried within the complex compared to H1b-G. Of these three proteins, hnRNPD0 and CP2

had approximately a two-fold difference between the denaturing and non-denaturing experiments while U2AF2 had a four-fold difference.

The fact that under native conditions, different proteins such as actin and RPA70 behaved differently to U2AF2, hnRNPD0 and CP2 toward H1c-A and H1b-G raised the possibility that some of the proteins within the complex might act as coordinators of varying downstream events between the different allelotypes whiles others served as core recognition binding proteins. A more extensive and detailed study would be needed to elucidate this.

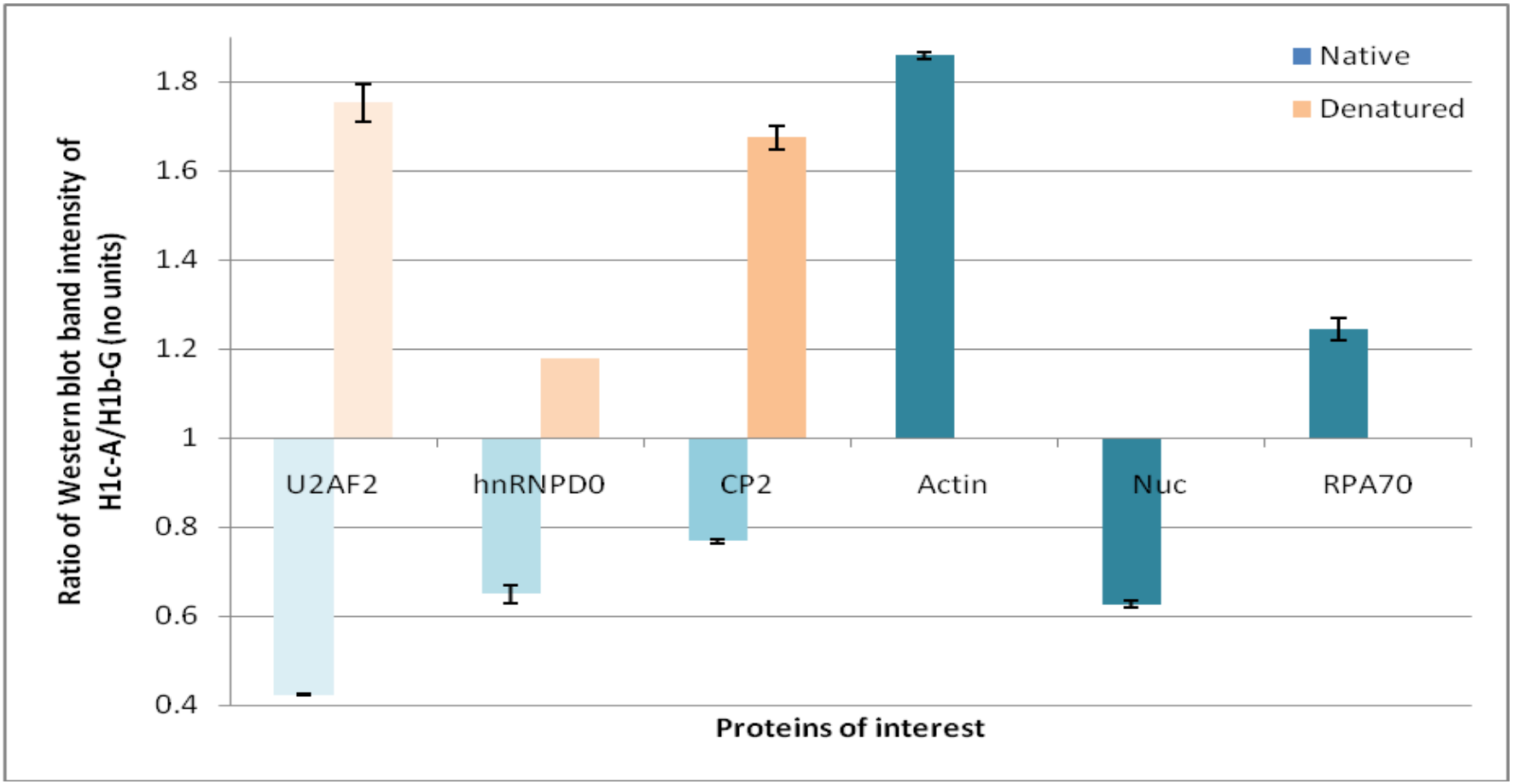


Figure 4.3.1: Density plot of Western blot bands of proteins of interest.

Plot of the ratio of band intensity of H1c-A against H1b-G after taking account for non-specific binding for each protein of interest. Density plots from Western blots arising from native gels are represented in blue while those arising from denaturing gels are represented in orange. Each was done in triplicate. Error bars represent standard deviation. Shades of colour have no significance.

4.4 Transactive Response DNA-binding protein-43 (TDP-43)

In 2006, it was shown that in some frontotemporal lobar degeneration (FTLD) cases, the characteristic ubiquinated inclusions are composed of transactive response DNA-binding protein-43 (TDP-43).[252] Brains of FTLD cases with mutations in progranulin also showed ubiquitin/TDP-43 inclusions as well as tau pathology.[253] It had been shown that mutations in the TDP-43 gene on chromosome 1p36.2 cause familial and sporadic amyotrophic lateral sclerosis (ALS) and are thus sufficient to cause neurodegeneration.[254]

The full extent of the function of the 43-kDa protein, TDP-43, is still being deciphered. TDP-43 is 414 amino acids long and encoded by six exons with two noncoding exons.[255] TDP43 is an hnRNP protein consisting of two RNA recognition motifs, RRM1 and 2, and a C-terminal glycine-rich region.[256] Most of the 38 dominant mutations are localized in the glycine-rich region (Figure 4.4.1 top). TDP-43 was initially identified as a transcriptional repressor that bound to the transactive response (TAR) DNA region in HIV-1.[257] TDP-43 was shown to be highly conserved amongst human, Drosophila and *C. elegans* in terms of their nucleic acid binding specificity while exon recognition and splicing regulation were conserved between human and Drosophila.[258] It was also shown that the protein could bind to the 3' UTR of mRNA such as the low molecular weight neurofilament (NFL) and regulate its mRNA dynamics such shuttling and translation.[259] This gene like numerous other genes have been found to be functionally and directly affected by TDP-43 (Figure 4.4.1 bottom).[260] It is unsurprising then that mutations, mislocalization and protein post-translational modifications would lead to various losses of function and gains in toxic function in cells that would lead to neurodegeneration (Figure 4.4.2 left).

Since it was shown that TDP-43 was found in inclusions in FTLD, there are now a growing number of disorders with similar inclusions (Figure 4.4.2 right). Because of the growing significance of this protein and the lack of understanding of it, TDP-43 was included in the study, especially for the next stage of functional knockdown promoter luciferase assays in order to see if TDP-43 could have a role in *MAPT* expression.

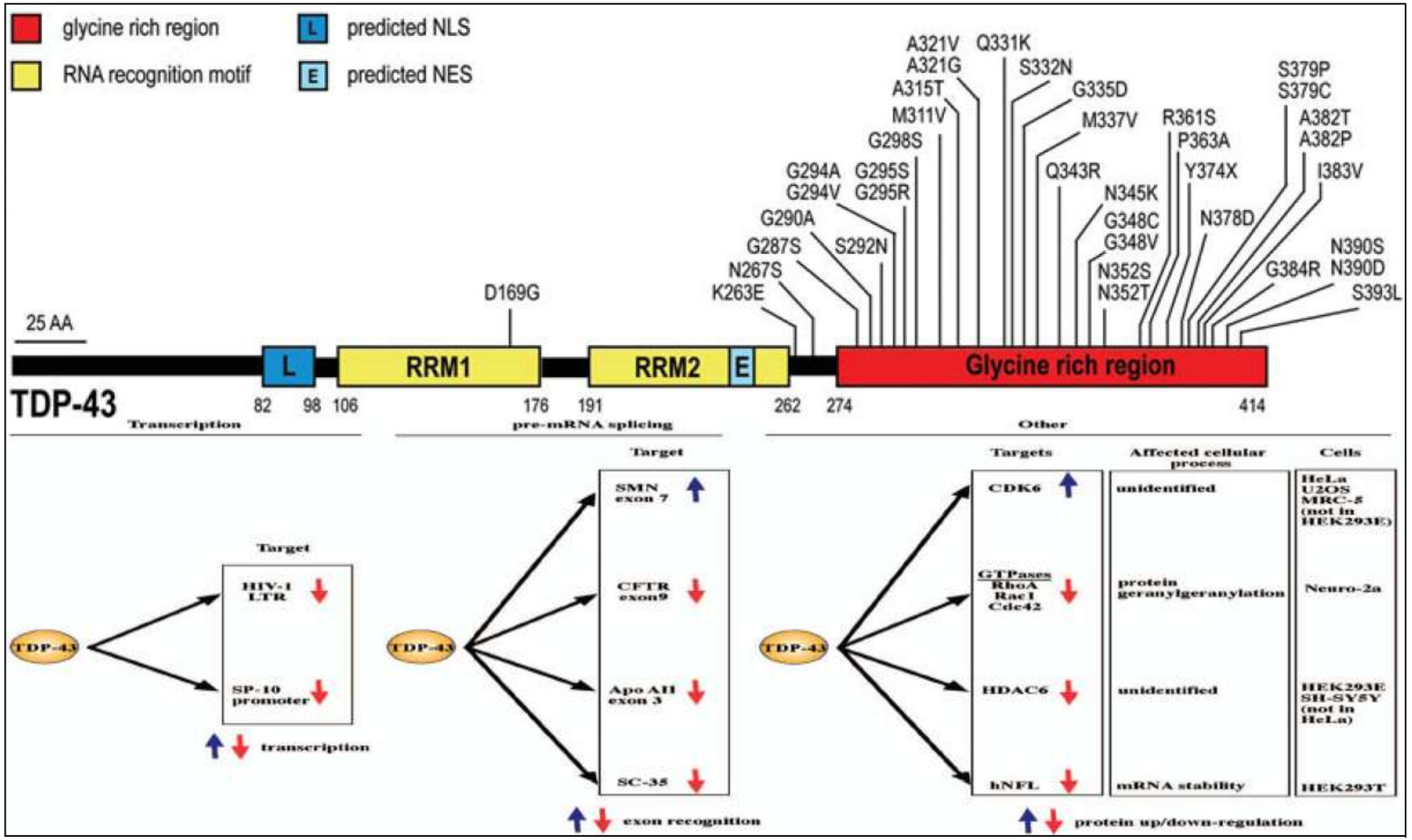
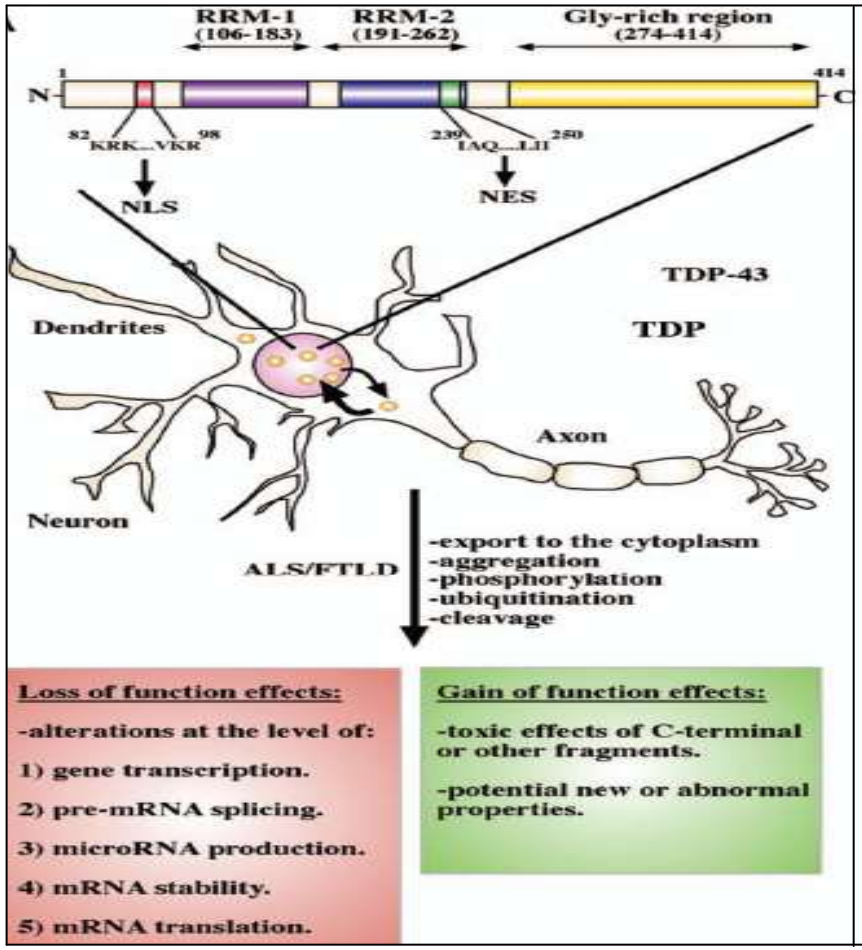


Figure 4.4.1: TDP-43 protein domains, mutations and effects of TDP-43 on genes. Top, schematic representation of the full-length TDP-43 protein. There are two RNA recognition motifs (RRM1, RRM2), the nuclear localization signal (L) and the nuclear export signal (E) and the C-terminal glycine-rich region where most of the mutations lie. Bottom, genes directly controlled by TDP-43 either at the transcription level or exon recognition level other processes.[385,38]



Disease	
Amyotrophic lateral sclerosis	Sporadic ALS
	With *TARDBP* mutation
	With *ANG* mutation
Frontotemporal lobar dementia	
FTLD-TDP-43	Sporadic or familial with or without MND
	With *GRN* mutation
	With *TARDBP* mutation
	With *VCP* mutation*
	Linked to chromosome 9p
FTLD-tau	Argyrophilic grain disease
	Pick disease
	Cortico-basal degeneration
Other Dementias	Alzheimer's disease (sporadic and familial)
	Dementia with Lewy's bodies with or without Alzheimer's disease
	Down syndrome
	Hippocampal sclerosis dementia
	Familial British dementia
Parkinsonism	Parkinson's disease with and without dementia
	Parkinson's disease with *LRKK2* mutation
	Perry syndrome with *DCTN1* mutation
	ALS Parkinsonism–dementia complex of Guam
Myopathies	Sporadic inclusion body myositis
	Inclusion body myopathy with *VCP* mutation*
	Oculo-pharyngeal muscular dystrophy with *PABP2* mutation
	Distal myopathy with rimmed vacuoles
	Myofibrillar myopathies with *MYOT* or *DES* mutation

Figure 4.4.2: Functional effects of TDP-43 in disease and list of pathological conditions showing TDP-43 inclusions. Left, in disease state, TDP-43 is ubiquinated, cleaved and phosphorylated. It accumulates as insoluble aggregates in the cytoplasm. Right, these TDP-43 inclusions can be found in a growing number of pathological conditions including polyglutamine diseases, Huntington's disease and spinocerebellar ataxia (not shown).[385,389]

4.5 Conclusions from validation of proteins identified

Probing the proteins pulled-down with the rs242557 probes and control probes using antibodies validated some of the proteins identified by mass spectrometry. The proteins that showed specificity to the rs242557 probes over the control probes from the denaturing pull-down assays were U2AF2, hnRNPU, PTBP1, hnRNPD0 and CP2. This experiment also showed that these proteins bound more strongly to the H1c risk allele rs242557-A which was similar to the 1-D gel bands that these proteins were identified from except for CP2. The validation of these proteins under native conditions was less straightforward. Although U2AF2, hnRNPD0 and CP2 were identified by the same antibodies under native conditions, the effects observed were different with the protective H1b/H2 allele, rs242557-G. These proteins were seen to bind slightly more to the protective allele. While hnRNPU and PTBP1 could not be identified under native conditions, this did not mean they were not present as more parameters were at play during native experiments such as protein dynamics. Given the extensive research that had been carried out on hnRNPU and PTBP1 as seen from the literature and the results seen from the denaturing pull-down, there was still enough evidence to warrant these proteins to be scrutinized further with functional assays.

Since starting my PhD, the research into TDP-43 has intensified. Having explained its significance, it also made sense to investigate the potential functional role on the *MAPT* promoter. In the following sections, we investigated the six proteins U2AF2, hnRNPU, PTBP1, hnRNPD0, CP2 and TDP43 by specific knock down using siRNA and their effects on *MAPT* promoter activity using promoter-luciferase reporter assays in neuroblastoma cell cultures.

5 Chapter 3 Results

5.1 Optimisation of plasmid and siRNA cotransfection functional assay

To be able to investigate the functional effects of the candidate proteins in relation to the haplotype effects of the *MAPT* promoter, luciferase promoter assays were combined with protein knockdown through transient transfections with siRNA. The luciferase promoter assays were the system to be initially used to show the haplotype effects of rs242557 on the promoter strength of pGL4.10.[124] However, since then, more refined plasmid constructs with more extensive coverage of putative regulatory regions of the *MAPT* promoter region had been created by another PhD student, Victoria Kay. This allowed not only a more extensive and systematic analysis of the promoter region, but also coupling the promoter reporter assays with gene knockdown allowed for a greater understanding of how each downregulated protein specifically exerted their effects.

Before any functional assay could be started, work had to be done to establish the optimal conditions for combined transfections of both plasmid and siRNA. However, both transfection protocols utilized different transfection reagent formulation (Transfast and RNAiMax respectively). Whether one reagent could work for both transfections had to be ascertained. Alternatively, if no one reagent would work for the co-transfection experiments, then a serial assay would have to be set up with either the plasmids or siRNAs transfected first according to their specific protocols followed by the other. The difficulty with the latter approach would be finding the optimized lag time between the two transfections such that there would be sufficient plasmids transfected to yield high enough readings as well as sufficient siRNA and enough time for them to exert a reproducible effect without too much cell death.

Hence, eight different co-transfection conditions were tested. Transfections of both plasmid and siRNA simultaneously using Transfast transfection reagent which was designed for plasmid transfections, were carried out according to the transfection protocol provided by the manufacturer for Transfast (this condition was abbreviated as Transfast in Figure 5.1.1). Transfections of both plasmid and siRNA simultaneously using RNAiMax

transfection reagent, which was designed for siRNA transfections were carried out according to the transfection protocol provided by the manufacturer for RNAiMax (this condition was abbreviated as RNAiMAx in Figure 5.1.1). Serial co-transfections of plasmid and siRNA could be done either with the plasmid or the siRNA transfections first with varying time intervals (6, 12 or 24h) between the two transfections. Plasmid transfections followed by siRNA transfections after 6, 12 or 24 hour intervals are abbreviated as plasmid/si_6h, plasmid/si_12h and plasmid/si_24h. Transfections with siRNA followed by plasmid after 6, 12 or 24 hour intervals are abbreviated as si/plasmid_6h, si/plasmid_12h and si/plasmid_24h.

After the different experimental conditions had been established, a reliable experimental model had to be established. The ideal condition would include a promoter construct that gave a strong luciferase read-out. Also, a siRNA that would target the vector regardless of the construct and elicit a gene knockdown to similar levels with the negative controls had to be done. This would consist of cells being treated to the same experimental conditions but with a scrambled sequence of siRNA as well as without any plasmid constructs or siRNA. At the first instance, the reporter plasmid used for optimization had to be a strong promoter giving robust readings with the luciferase assay. Work by Victoria Kay had showed that the pGL4.10 vector containing the core promoter of *MAPT*, B1, gave the highest luciferase reading of all the constructs she made. Strong luciferase activity from the B1 plasmid enables us to observe the effect of direct siRNA targeting of the plasmid by pGL4si. Hence the model was to observe the luciferase read-out of B1 alone (this was abbreviated as B1 in Figure 5.1.1) and compare it with the co-transfection of B1 plasmid with pGL4si (this was abbreviated as B1&pGL4si in Figure 5.1.1). As a negative, we employed a scrambled siRNA (neg ctrl si) which was co-transfected with the B1 plasmid (this was abbreviated as B1&neg ctrl si in Figure 5.1.1). This should have little effect on the luciferase read-out of the B1 plasmid construct. Lastly, another negative control included untransfected naïve cells (this was abbreviated as “cells only” in Figure 5.1.1).

The right optimized co-transfection condition would have a high luciferase reading for B1 plasmid on its own, a low luciferase reading for cells only transfected without any plasmids or siRNA, a low luciferase reading close to the negative control (cells only) for the co-transfection of the B1 plasmid with the pGL4si siRNA (B1&pGL4si) and a high

luciferase reading close to the B1 plasmid alone for the co-transfection of the B1 plasmid with the negative control siRNA (B1&neg ctrl si). From Figure 5.1.1, the conditions RNAiMax, plasmid/si_6h, si/plasmid_6h, plasmid/si_12h, si/plasmid_12h and plasmid/si_24h did not meet these criteria. Although the negative control of cells only worked as they should, most of the conditions did not yield any significant difference between B1 and B1&pGL4si. Although the experimental condition si/plasmid_6h was an exception, there was no recovery of the luciferase reading with B1&neg ctrl si which should not affect the luciferase reading of the B1 plasmid very much.

Two experimental conditions, however, Transfast and si/plasmid_24h, (Figure 5.1.1 continued) did meet these criteria. Carrying out the co-transfection of plasmid and siRNAs simultaneously with Transfast was slightly better than carrying out the siRNA transfection first for twenty-four hours before the plasmid transfection. Although the luciferase readings varied similarly for both conditions (approximately 80% reduction of B1 when compared to B1&pGL4si), Transfast alone had higher relative luminescence compared to si/plasmid_24h, allowing for a greater window for the actual experiments. Also, two-tailed T-tests revealed that Transfast alone was slightly more significant which again supported the larger experimental window that this condition provided. Using one reagent and adding all the plasmids and siRNAs together at the beginning simplifies the co-transfection experiment which meant less technical errors could be introduced.

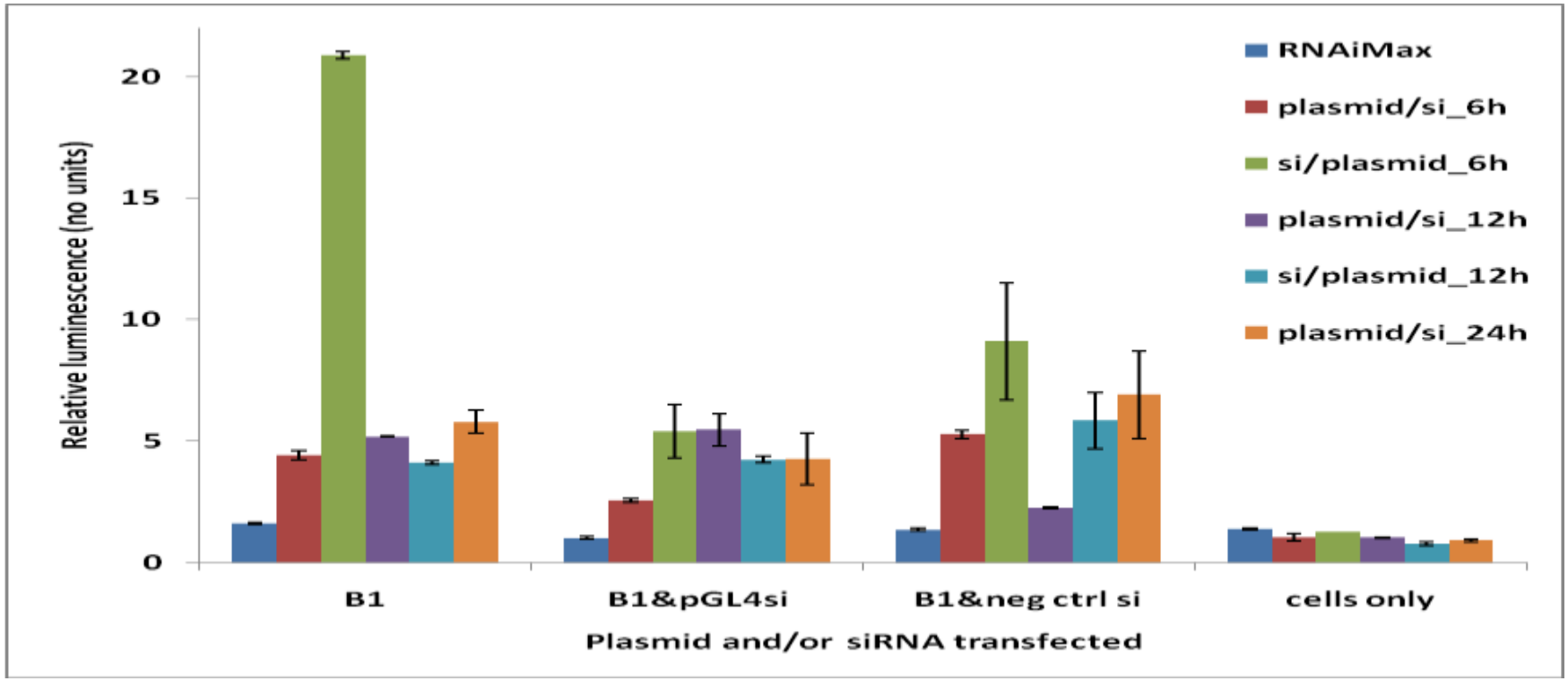


Figure 5.1.1: Time course optimization assay of unsuccessful plasmid and siRNA co-transfections.

Co-transfection was trialed by transfecting plasmid constructs and siRNAs together with siRNA transfection reagent (RNAiMax) only or plasmid constructs were transfected first with Transfast for 6, 12 and 24 hours before siRNAs were transfected using RNAiMax or siRNAs were transfected first with RNAiMax for 6 and 12 hours before plasmid constructs were transfected with Transfast. B1 is a pGL4.10 plasmid *MAPT* promoter-luciferase construct with strong luciferase activity. pGL4si is a siRNA targeting the core promoter of pGL4.10, neg ctrl si is a scrambled siRNA. Each experiment was done in triplicate. Error bars represent standard deviation.

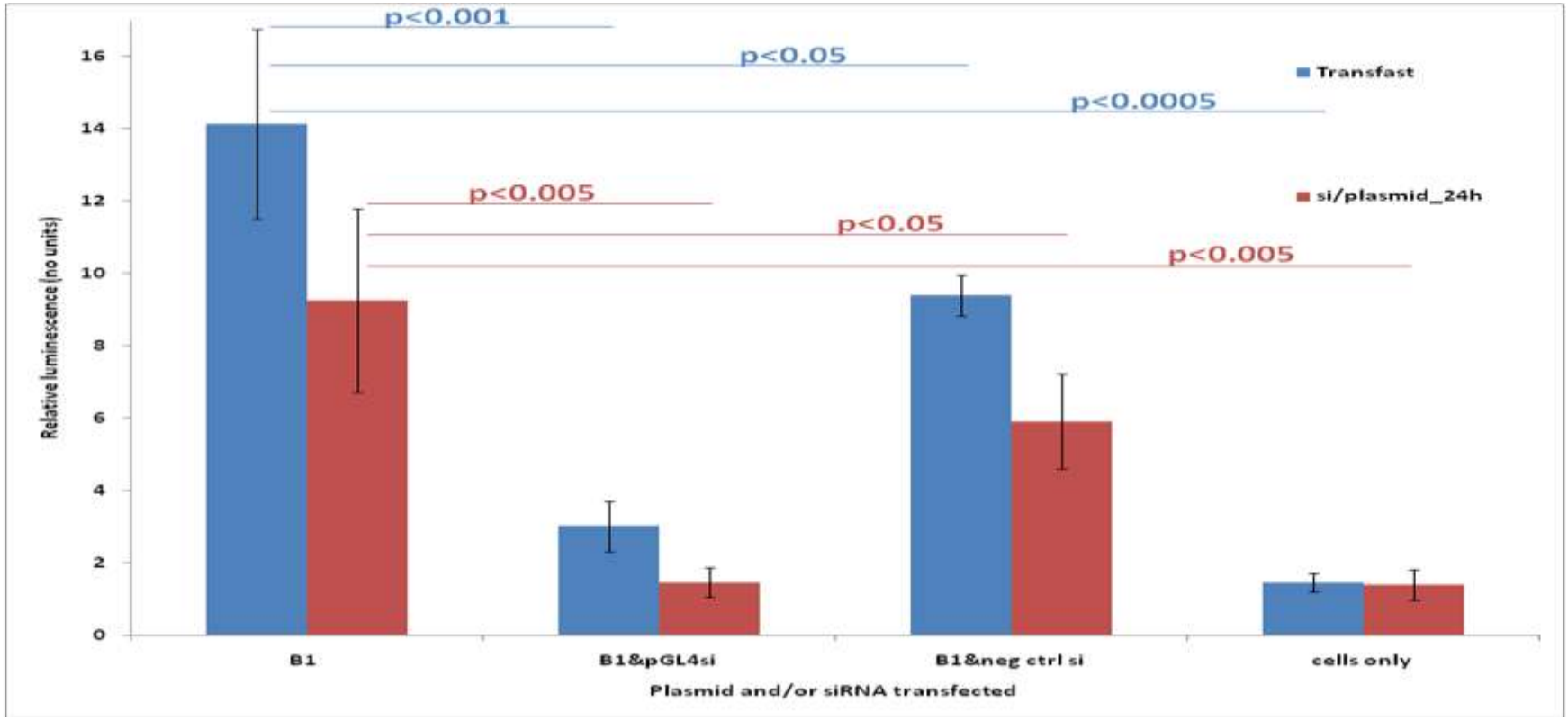


Figure 5.1.1 (continued): Time course optimization assay of successful plasmid and siRNA co-transfection.

Co-transfection was trialed by transfecting plasmids and siRNAs together with plasmid transfection reagent (Transfast) only or siRNAs were transfected first with RNAiMax for 24 hours before plasmid constructs were transfected with Transfast. B1 is a pGL4.10 plasmid *MAPT* promoter-luciferase construct with strong luciferase activity. pGL4si is a siRNA targeting the core promoter of pGL4.10, neg ctrl si is a scrambled siRNA. Each experiment was done in triplicate. Error bars represent standard deviation.

Before the plasmid constructs and siRNA co-transfection experiments could proceed, the siRNAs to be used had to be tested and verified for efficacy of targeted gene knockdown. We carried this out on the human neuroblastoma cell line SH-SY5Y. The siRNAs against U2AF2, PTBP1, hnRNPU, hnRNPD0, CP2 and PAX8 were purchased commercially. For each gene, three different siRNAs had been pre-designed to target different parts of the gene. The siRNA against TDP43, as in the case of the siRNA against pGL4.10, was custom designed to target the upstream promoter region and made-to-order (personal communication with Roberto Simone). PAX8 was chosen as a more stringent transcription factor-specific negative control. Knocking down a known transcription factor not known to be involved and observing no effect on the experiment would be more demonstrative of any specific effects seen with the rest of knockdowns than simply using a scrambled negative control siRNA not known to interact with any human DNA sequence.

SH-SY5Y cells were either transfected without any siRNA (abbreviated as Cells or Cells only in Figures 5.1.2 – 5.1.7), or negative control siRNAs (two were commercially available, Neg ctrl 1 or Neg ctrl 2). Both negative control siRNAs were random sequences not known to interact with any human sequences. The neuroblastoma cells were also transfected with either one of the three commercially available target siRNAs alone (#1, #2 or #3) or in combination (#1&2, #1&3 or #2&3). After transfection, cells were lysed and their protein content normalized before separation by SDS-PAGE and identification by Western blotting. Actin was used as an endogenous housekeeping protein for normalisation.

In Figure 5.1.2, taking into account of the levels of actin, there was no significant difference ($p=0.2$) between transfection without any siRNA and transfection with negative control 1 siRNA. This made negative control 1 a better negative control siRNA than negative control 2 siRNA which had a significance of $p<0.0005$ when compared to transfection without any siRNA (Cells only) using a two-tailed T-test. This result was also reflected in Figures 5.1.3 – 5.1.7 where negative control 1 proved to be a better negative control and hence would be adopted as the negative control siRNA to be used in the co-transfection assays.

When it came to analyzing which target siRNA singly or in combination would provide the highest level of protein knockdown, there was variation amongst the different

target proteins. Although for all the target proteins, a combination of two siRNAs appeared to provide the most protein knockdown, the level of this knockdown varied from one target protein to the next. The exception was TDP43, Figure 5.1.7, which had a custom designed siRNA that led to a 63% reduction in TDP43 protein expression level compared to cells transfected with negative control 1 siRNA ($p<3.5x10^{-4}$, two-tailed T-test). The largest amount of protein knockdown that was also the most significant for U2AF2 was a 70.4% reduction (Figure 5.1.2, $p<5.4x10^{-3}$), for hnRNPU, it was a 84.7% reduction (Figure 5.1.3, $p<9.62x10^{-5}$), for PTBP1, it was a 93.2% reduction (Figure 5.1.4, $p<1.1x10^{-4}$), for hnRNPD0, it was a 89.1% (Figure 5.1.5, $p<8.13x10^{-6}$), for CP2 , it was a 79.6% reduction (Figure 5.1.6, $p<0.0025$).

While for most of the protein knockdowns, a combination of two different siRNAs yielded the largest amount of protein reduction, they were not always significantly different when compared to single siRNA transfections. hnRNPU siRNA knockdowns, however, stood out in that the siRNA#2&3 combination had significantly less hnRNPU when compared to negative control 1 siRNA and any of the siRNAs on their own (Figure 5.1.3). PTBP1 siRNA #1&3 was similar to hnRNPU with respect to having significantly less protein when compared to negative control 1 siRNA and any of the siRNAs on their own, but it also had significantly less protein when compared to another double siRNA knockdown #1&2 (Figure 5.1.4). This observations gave more confidence in using a combination of two siRNAs to knockdown a gene for these proteins of interest.

After all the different aspects of the co-transfection assay had been optimized with confidence, this assay could now be proceeded with to analyse the effects each protein knockdown on different aspects of the *MAPT* promoter.

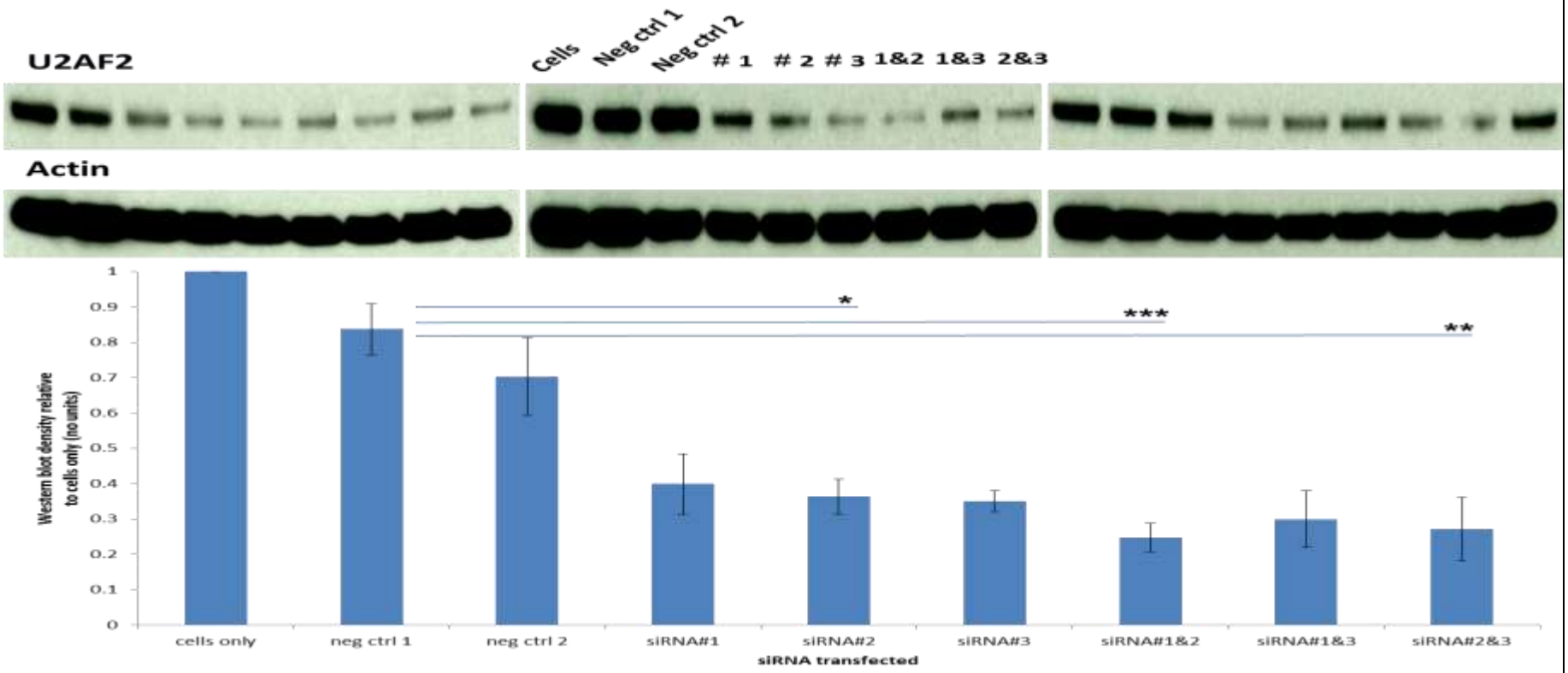


Figure 5.1.2: siRNA optimization assay for target protein U2AF2.

SH-SY5Y cells were transfected with two different scrambled siRNAs (neg ctrl 1 and 2) or three different target siRNAs (#1, #2 and #3) or a combination of two target siRNAs (#1&2, #1&3 and #2&3). Top: Visualisation of target protein siRNA knockdown through Western blots using actin as a housekeeping protein. Bottom: Western blot density plot of target bands normalized against respective housekeeping bands, represented as a fraction of cells only. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.0066$, **: $p<0.0064$, ***: $p<0.0054$ (two-tailed T-test).

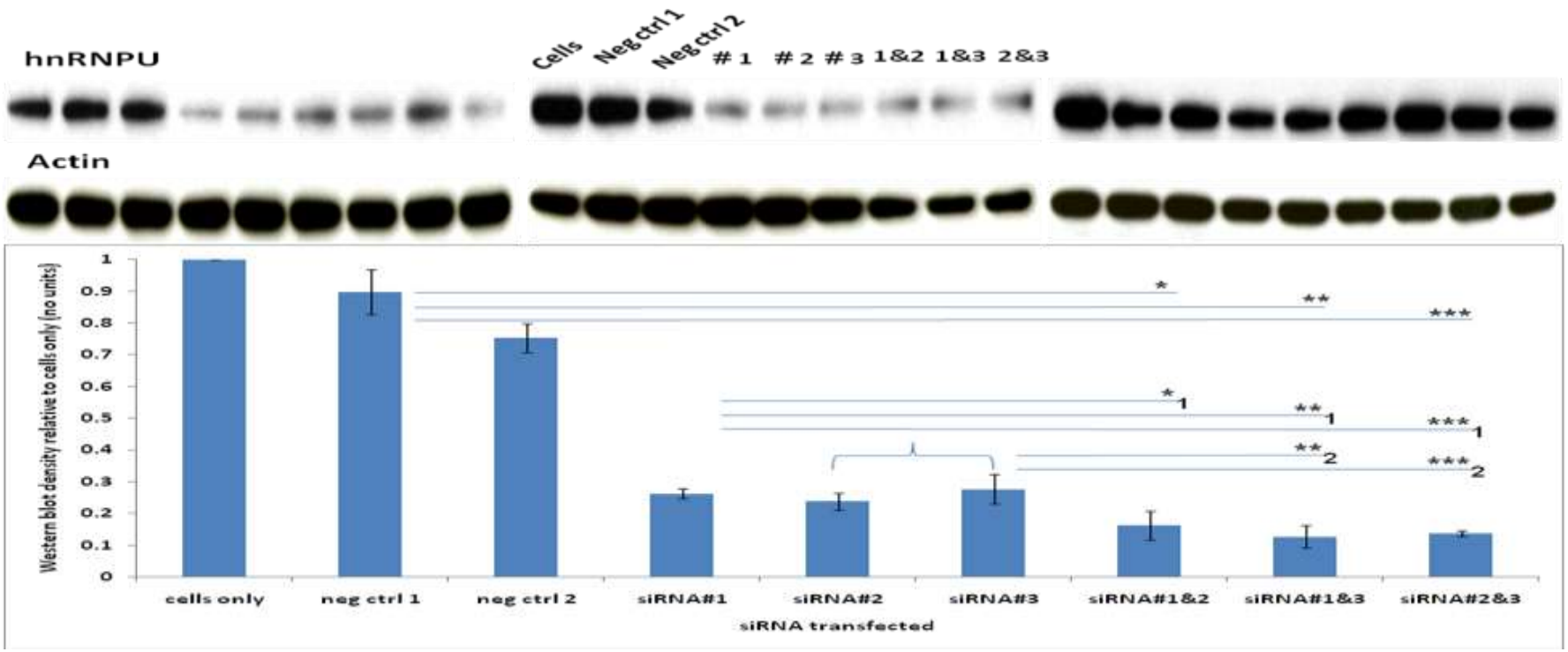


Figure 5.1.3: siRNA optimization assay for target protein hnRNPU.

SH-SY5Y cells were transfected with two different scrambled siRNAs (neg ctrl 1 and 2) or three different target siRNAs (#1, #2 and #3) or a combination of two target siRNAs (#1&2, #1&3 and #2&3). Top: Visualisation of target protein siRNA knockdown through Western blots using actin as a housekeeping protein. Bottom: Western blot density plot of target bands normalized against respective housekeeping bands, represented as a fraction of cells only. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.00018$, **: $p<0.00015$, ***: $p<9.62\text{x}10^{-5}$, $*_1$: $p<0.034$, $**_1$: $p<0.012$, $***_1$: $p<0.0018$, $**_2$: $p<0.032$ / $p<0.033$, $***_2$: $p<0.010$ / $p<0.022$ (two-tailed T-test).

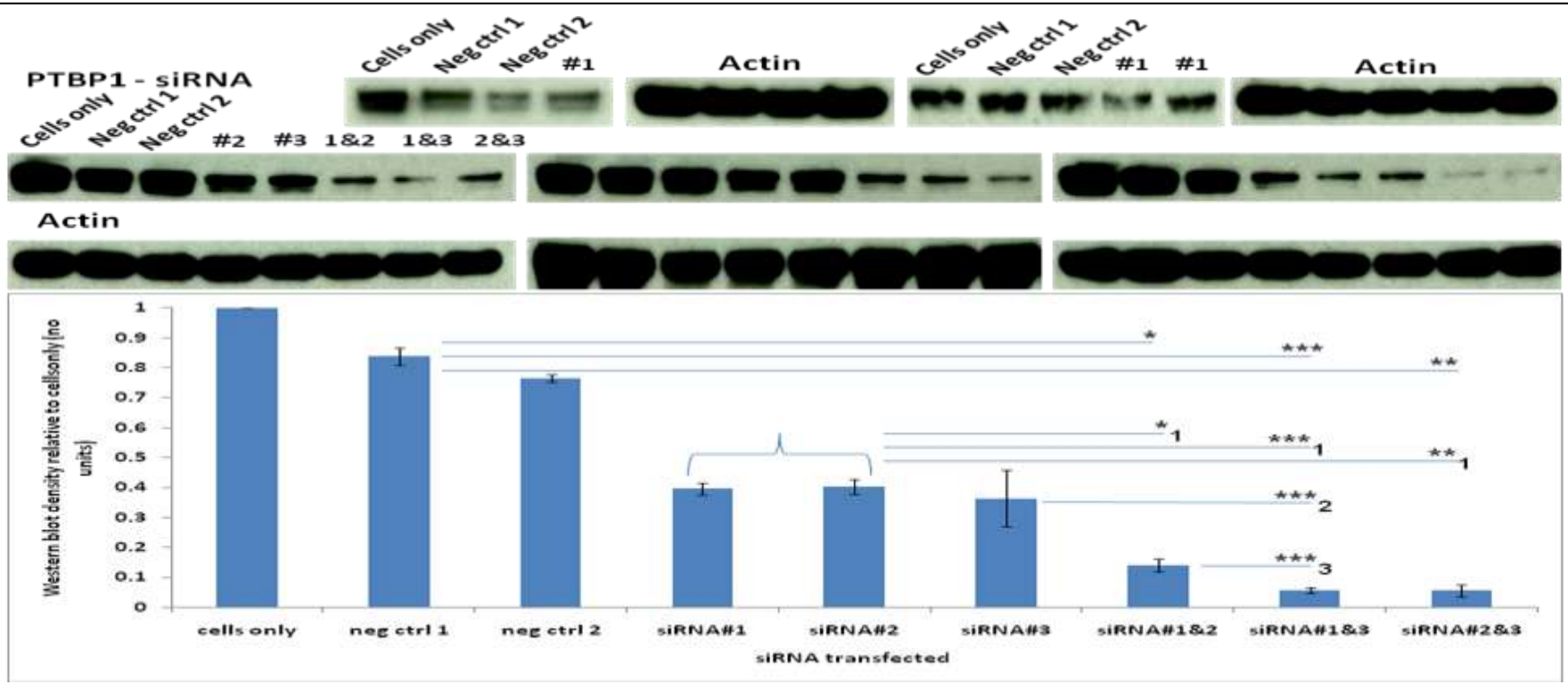


Figure 5.1.4: siRNA optimization assay for target protein PTBP1.

SH-SY5Y cells were transfected with two different scrambled siRNAs (neg ctrl 1 and 2) or three different target siRNAs (#1, #2 and #3) or a combination of two target siRNAs (#1&2, #1&3 and #2&3). Top: Visualisation of target protein siRNA knockdown through Western blots using actin as a housekeeping protein. Bottom: Western blot density plot of target bands normalized against respective housekeeping bands, represented as a fraction of cells only. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.00019$, **: $p<0.00018$, ***: $p<0.00011$, $*_1$: $p<0.0025 / 0.0050$, $**_1$: $p<0.0016 / 0.0027$, $***_1$: $p<0.00059 / 0.0014$, $***_2$: $p<0.048$, $***_3$: $p<0.017$ (two-tailed T-test).

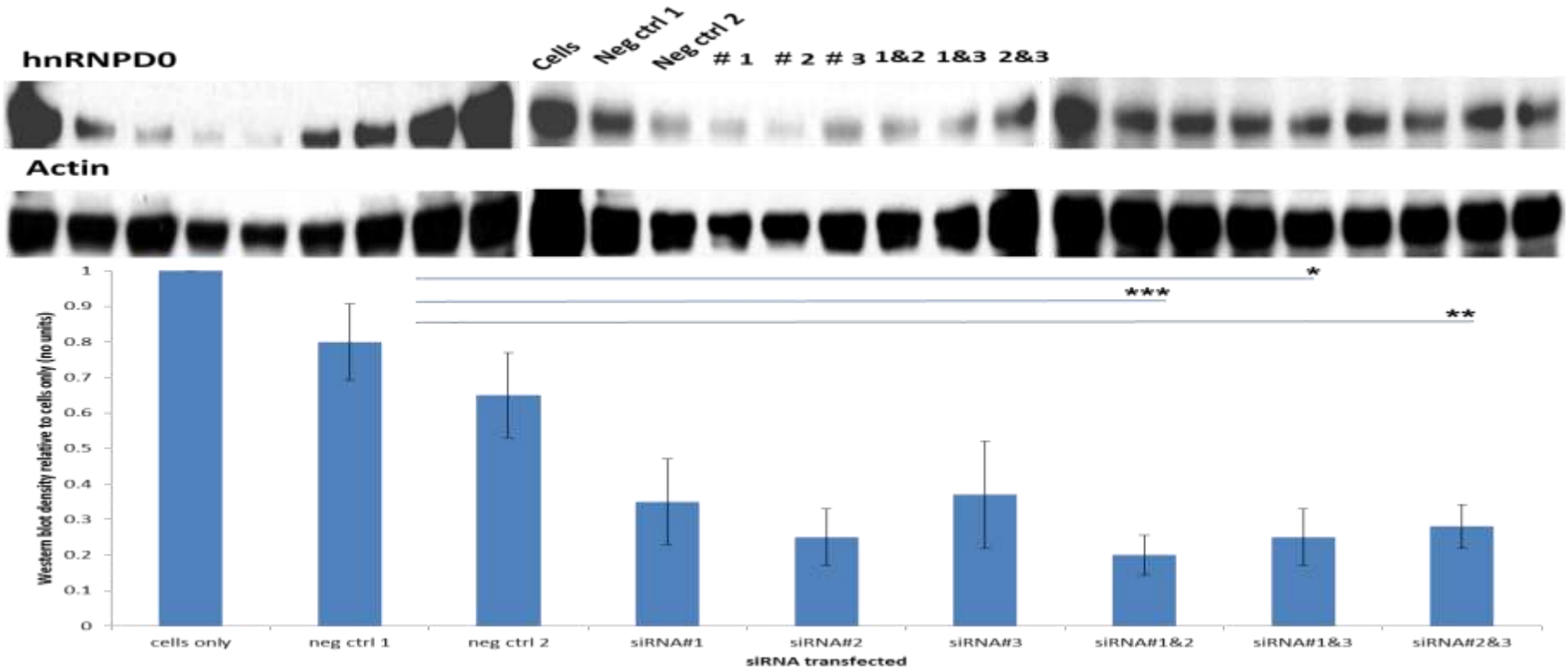


Figure 5.1.5: siRNA optimization assay for target protein hnRNPD0.

SH-SY5Y cells were transfected with two different scrambled siRNAs (neg ctrl 1 and 2) or three different target siRNAs (#1, #2 and #3) or a combination of two target siRNAs (#1&2, #1&3 and #2&3). Top: Visualisation of target protein siRNA knockdown through Western blots using actin as a housekeeping protein. Bottom: Western blot density plot of target bands normalized against respective housekeeping bands, represented as a fraction of cells only. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.0349$, **: $p<0.00421$, ***: $p<0.00213$ (two-tailed T-test).

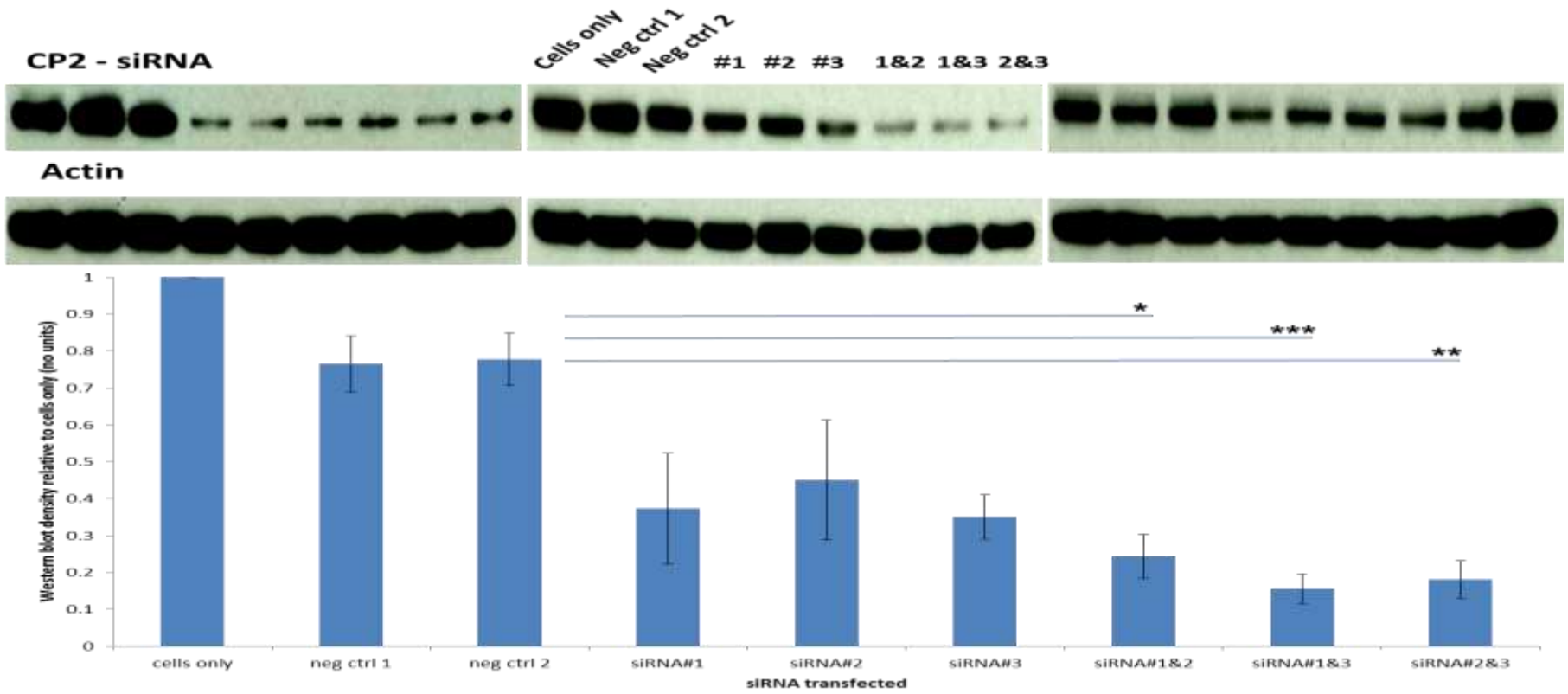


Figure 5.1.6: siRNA optimization assay for target protein CP2.

SH-SY5Y cells were transfected with two different scrambled siRNAs (neg ctrl 1 and 2) or three different target siRNAs (#1, #2 and #3) or a combination of two target siRNAs (#1&2, #1&3 and #2&3). Top: Visualisation of target protein siRNA knockdown through Western blots using actin as a housekeeping protein. Bottom: Western blot density plot of target bands normalized against respective housekeeping bands, represented as a fraction of cells only. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.045$, **: $p<0.018$, ***: $p<0.029$ (two-tailed T-test).

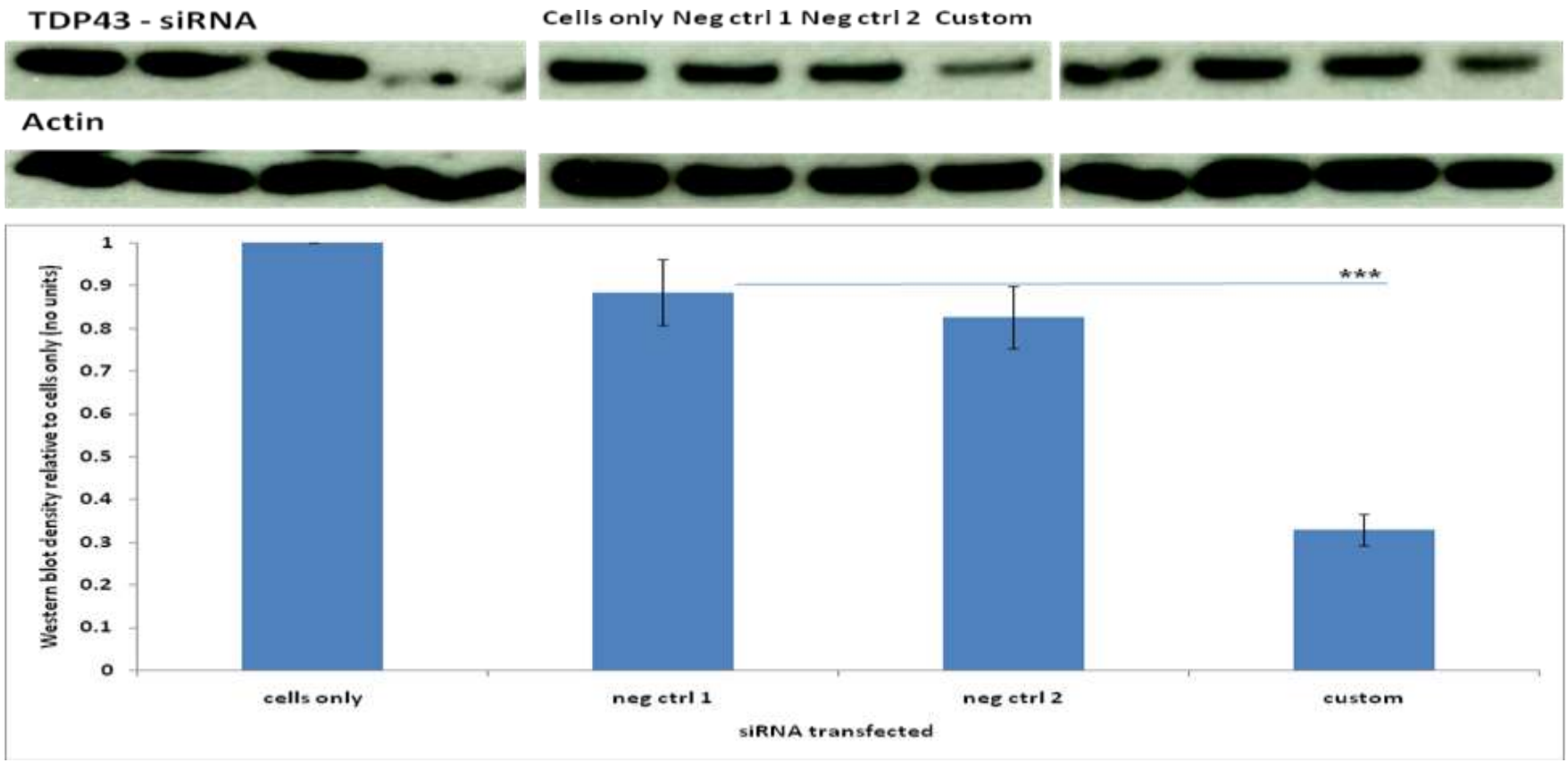


Figure 5.1.7: siRNA optimization assay for target protein TDP43.

SH-SY5Y cells were transfected with two different scrambled siRNAs (neg ctrl 1 and 2) or custom-designed target siRNAs (custom). Top: Visualisation of target protein siRNA knockdown through Western blots using actin as a housekeeping protein. Bottom: Western blot density plot of target bands taking into account respective housekeeping bands, represented as a fraction of cells only. Each was done in triplicate. Error bars represent standard deviation. ***: $p<0.00035$ (two-tailed T-test).

5.2 Effects of shortlisted protein siRNA knockdowns on transcriptional activity of *MAPT* core promoter domains

After establishing the assay protocol, the efficacy of the siRNAs against the proteins of interest, the co-transfection functional assays could proceed. The *MAPT* core promoter plasmid constructs to be used in these assays were created by Victoria Kay and had been tested and sequenced extensively. These promoter plasmid constructs were based on the *MAPT* promoter region (Figure 5.2.1). The 1,398 bp core promoter region is 47,370 bp upstream of the C and D regions and it is in the C region where rs242557 resides. Of multiple haplotypes in the H1 clade, the H1c sub-haplotype drives the association with PSP and CBD, suggesting that it carries the pathogenic variation leading to increased risk of these largely sporadic disorders.

Hence, to fully investigate the effects of gene knockdown of the proteins of interest, various different promoter plasmid constructs were made. Although the underlying impetus for this work was based on the rs242557 SNP located in the C region, it was important to not only study the effect of the proteins of interest on the C region but also on the various different aspects of the *MAPT* promoter region and how they relate and interact to the different gene knockdowns. This would shed as much light on the *in vivo* system *ex vivo*. These plasmid constructs, also shown in Figure 5.2.1, were the core promoter B1, an extended core promoter region termed B1+B2, the downstream promoter region, C, where rs242557 resided (hence the different haplotypes, C-H1b and C-H1c), a combination of the core promoter B1 with the different haplotypes of the C region (B1+C-H1b and B1+C-H1c), the core promoter B1 with an extended sequence of the downstream promoter region (termed B1+D+C-H1b and B1+D+C-H1c) as well as the extended downstream promoter regions on their own (D+C-H1b and D+C-H1c).

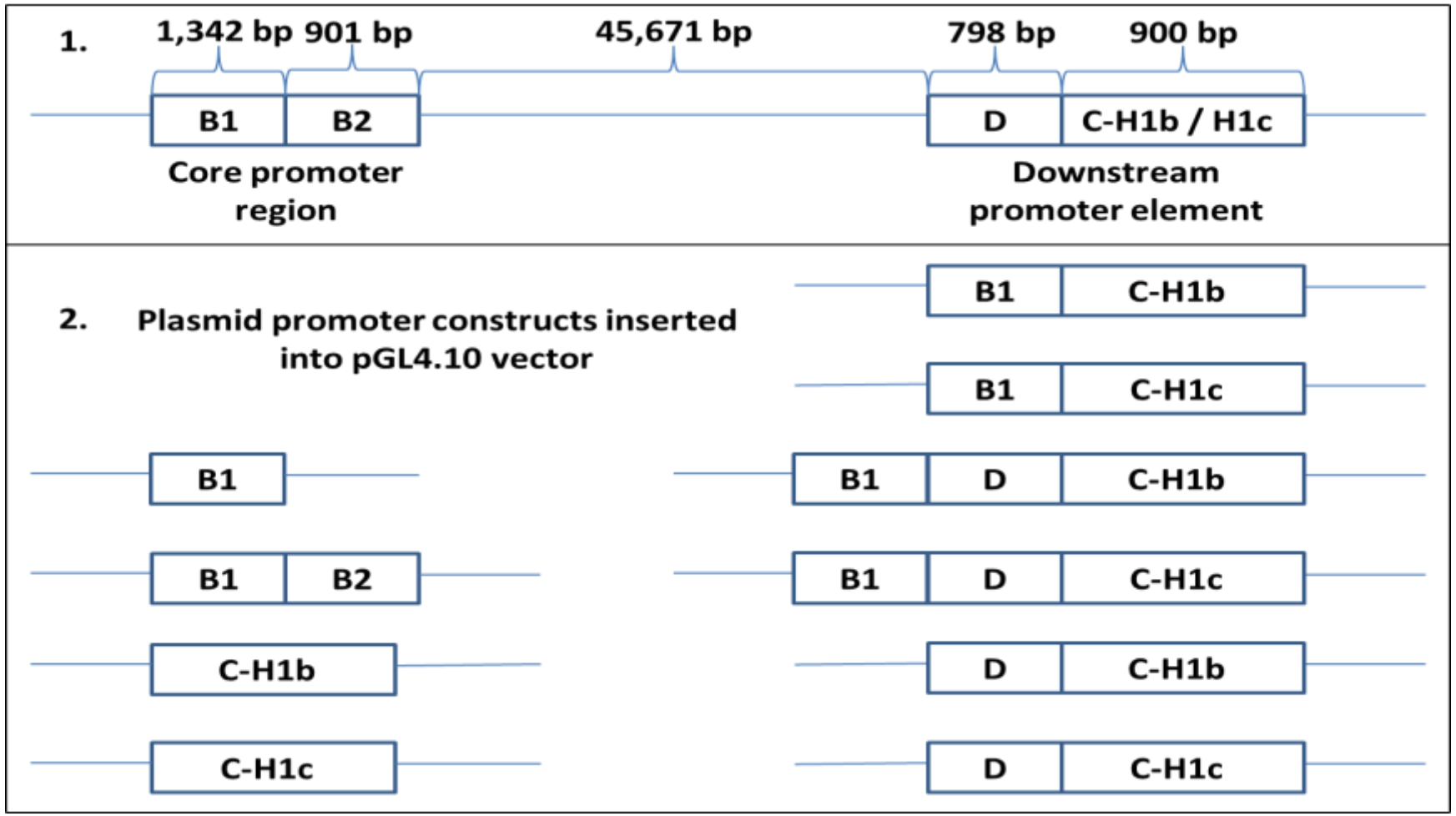


Figure 5.2.1: Organization of the *MAPT* promoter region and promoter plasmid constructs derived from it.

1. Schematic of the *MAPT* promoter region showing the relation of the core promoter with its downstream element. 2. Various promoter constructs inserted to the pGL4.10 vector.

U2AF2

The first protein to be knocked down in the co-transfection assay was U2AF2 (Figure 5.2.2). Before any analysis can begin, checks on whether the co-transfection assay had worked had to be done. The pGL4 siRNA was co-transfected with the plasmid construct containing the B1 core promoter which was the strongest plasmid construct amongst all the other plasmid constructs on their own in a luciferase assay. Compared to the B1 plasmid construct on its own, the co-transfection of the B1 plasmid construct with pGL4 siRNA did manage to knock it down by 84.1%, similar to the 80% seen during the optimization experiment. Next, co-transfection of the B1 plasmid construct with negative control 1 siRNA was compared to transfection with plasmid only to observe whether there was any difference between the two conditions. Out of the ten plasmid constructs used, only two plasmid constructs that had been co-transfected with negative control 1 siRNA (B1+C-H1b and B1+D+C-H1b), did show significant reductions of around 25%.

After ascertaining that the co-transfection assay had largely worked as it should, analysing the effect of knocking down U2AF2 was done by expressing the knock down effect relative to the same plasmid construct co-transfected with negative control 1 siRNA. This would allow any non-specific effects of siRNA transfection on plasmid transfection to be taken into account (bottom panel of Figures 5.2.2 and 5.2.2 continued) and any comparison to co-transfection with negative control 1 siRNA was similar to comparisons to the plasmid constructs on their own. The first observable effect of reducing the level of U2AF2 in the cell was that regardless which *MAPT* plasmid construct was used, all of them performed less effectively. The largest effect of knocking down U2AF2 was on the core promoter, B1, and the extended core promoter, B1+B2, both of which had their functions reduced by about 70%.

The region of most interest was the downstream promoter region, C, where rs242557 resides. C-H1b was significantly more sensitive to the reduction of U2AF2 levels than C-H1c with a 60% reduction in transcription compared to a 50% reduction respectively ($p<0.01$). When the core promoter, B1, was included upstream to the C region however, the allelic effect seen before was abolished. Both plasmid constructs B1+C-H1b and B1+C-H1c were not significantly different from each other with a reduction in

promoter strength of about 50% but were significantly less affected than B1 alone (70%). The extended downstream promoter region D+C, had the opposite effect in that the allelic effect was significantly exacerbated ($p<0.01$). Although D+C-H1b with reduced levels of U2AF2 varied similarly to C-H1b, which was a 60% reduction in the luciferase assay, D+C-H1c was less sensitive to the effects of low levels of U2AF2 (30% reduction) than the 50% reduction with C-H1c.

It was interesting to note that when the different strands of the *MAPT* promoter were put together: the core promoter B1 and the extended downstream promoter region D+C (B1+D+C – the most extensive promoter construct used), the largest and most significant ($p<0.001$) rs242557 allelic effect was seen. Of all the H1b plasmid constructs used, siRNA knockdown of U2AF2 had the least effect on B1+D+C-H1b. This was still a 40% reduction and significantly less than U2AF2 knockdown with B1 alone (70% reduction). No significant effect was seen with U2AF2 knockdown on B1+D+C-H1c . Hence, the reduction of U2AF2 had very little if any effect on the function of B1+D+C-H1c but negatively affected B1+D+C-H1b ten-fold more. Another interesting note was that even though no allelic effect was seen with B1+C, it was only when the D region was included that the full allelic effect was observed. Thus, it was not only important to study the different components of a promoter but also the cumulative effects of these components. All these results seemed to suggest that U2AF2 was required for the normal functioning of *MAPT* promoter except maybe in the case of B1+D+C-H1c.

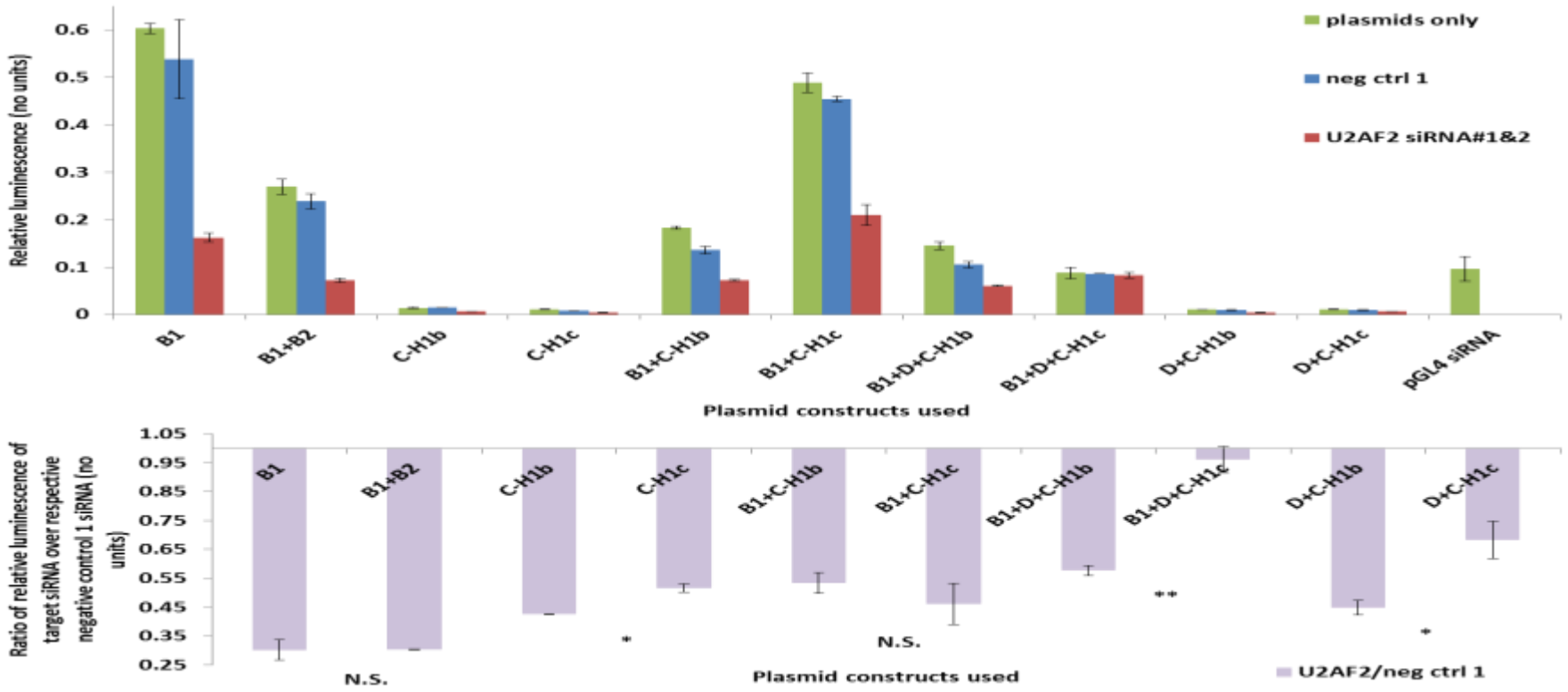


Figure 5.2.2: Effects of U2AF2 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs.

SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with luciferase reporter plasmids for six different combinations of *MAPT* promoter regions (B1, B1+B2, C, B1+C, B1+D+C and D+C), four of which were from a H1b or H1c background (C, B1+C, B1+D+C and D+C). Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$ (two-tailed T-test).

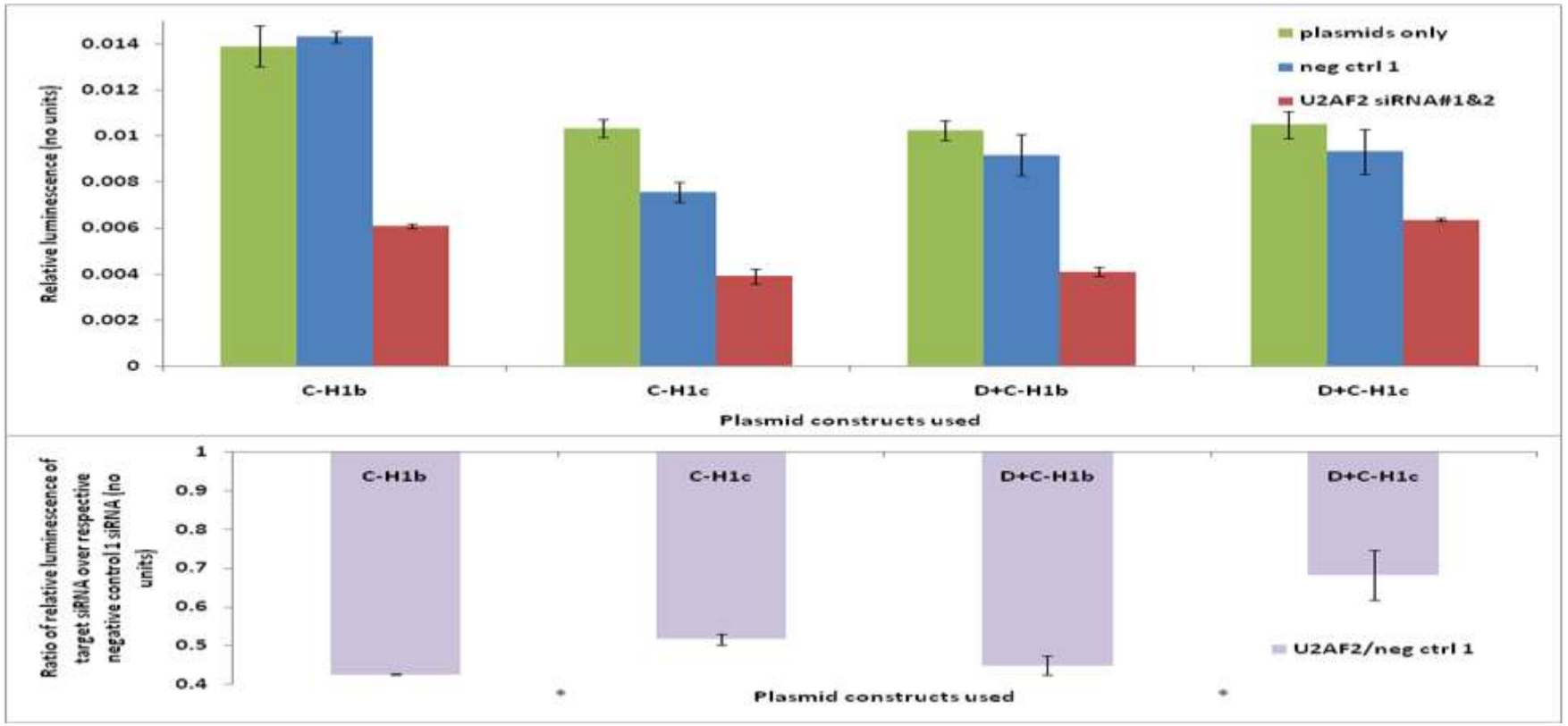


Figure 5.2.2 (continued): Effects of U2AF2 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs. SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with luciferase reporter plasmids for two different combinations of *MAPT* promoter regions (C and D+C), from a H1b or H1c background. Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$ (two-tailed T-test). Different scale to previous graph.

hnRNPU

Before any analysis can begin, checks on whether the co-transfection assay had worked had to be done. The pGL4 siRNA was co-transfected with the plasmid construct containing the B1 core promoter which was the strongest plasmid construct amongst all the other plasmid constructs on their own in a luciferase assay. Compared to the B1 plasmid construct on its own, the co-transfection of pGL4 siRNA with B1 plasmid construct did manage to knock it down by 81.0%, similar to the 80% seen during the optimization experiment (Figure 5.2.3). Next, co-transfection of negative control 1 siRNA with plasmid constructs were compared to transfections with plasmids only to observe whether there was no significance difference between the two conditions. Out of the ten plasmid constructs used, only two plasmid constructs that had been co-transfected with negative control 1 siRNA (B1+B2 and B1+D+C-H1b), did show significant reductions of around 30% and 20% respectively.

The first observable effect of reducing the level of hnRNPU in the cell was that unlike U2AF2, the relationship with protein knockdown and *MAPT* promoter region was less straightforward. The largest effect of knocking down hnRNPU was not only on the core promoter, B1, and the extended core promoter, B1+B2, both of which had their transcriptional activity reduced by about 60% but also on B1+D+C-H1b which had its activity increased by about 60%.

The region of most interest was the downstream promoter region, C, where rs242557 resided. Although the reduction of hnRNPU level did not affect the C region much, it had an opposing effect depending on the allelotype. C-H1b was significantly less sensitive to the reduction of hnRNPU levels than C-H1c with a 10% gain in promoter strength compared to a 20% reduction respectively ($p<0.001$). When the core promoter, B1, was included with the C region however, the allelic effect seen before was substantially exacerbated ($p<0.0001$). While B1+C-H1b had a slight reduction in the gain of promoter strength (4%), B1+C-H1c had a three-fold reduction in promoter strength (60%). Hence, although B1+C-H1b was hardly affected by the reduction of hnRNPU, C- H1b abolished any effect the lack of hnRNPU had on B1 (reduction of 60%). The opposite was true for

B1+C-H1c which showed similar results to the core promoter, B1 affected by hnRNPU knockdown.

The extended downstream promoter region D+C, was differently influenced by the rs242557 alleles ($p<0.01$) on hnRNPU knockdown. Although D+C-H1b had similar gains of about 10% luciferase activity like the C-H1b and B1+C-H1b, D+C-H1c was more sensitive to the effects of hnRNPU knockdown (40% gain in function) than the 20% or 60% reduction with C-H1c and B1+C-H1c respectively. Thus, from these results, it seemed that hnRNPU had very little effect on the H1b haplotype of the C region even with the core promoter or an extended C region. It did, however, have an increased opposing effect on the H1c haplotype depending on whether it was paired with the core promoter or the extended downstream region.

It was interesting to note that when the different domains of the *MAPT* promoter were put together: the core promoter B1 and the extended downstream promoter region D+C (B1+D+C – the most complete promoter construct used), no significant rs242557 allelic effect was observed. While the rest of the H1b plasmid constructs were hardly affected by the reduction of hnRNPU levels, B1+D+C-H1b had the largest gain (60%) in promoter strength with hnRNPU knockdown. Since very little effect was seen either with B1+C-H1b or D+C-H1b, this result hinted at a possible synergistic effect of combining the core promoter and the extended downstream promoter. The 50% gain in promoter strength with B1+D+C-H1c which was not significant when compared to B1+D+C-H1b might suggest that the haplotype effect is abolished by this synergistic effect. Hence, the reduction of hnRNPU had no haplotype effect on the function of B1+D+C. Another interesting observation was that even though no allelic effect was seen with B1+D+C, it was only when the D region was included to B1+C-H1c or C-H1c that similar, non-significant effects were observed. Thus, it was not only important to study the different components of a promoter but also the cumulative effects of these components. All these results seemed to suggest that hnRNPU had a complex repressive effect on the *MAPT* promoter, largely brought about by the D region which otherwise masked any haplotype effect.

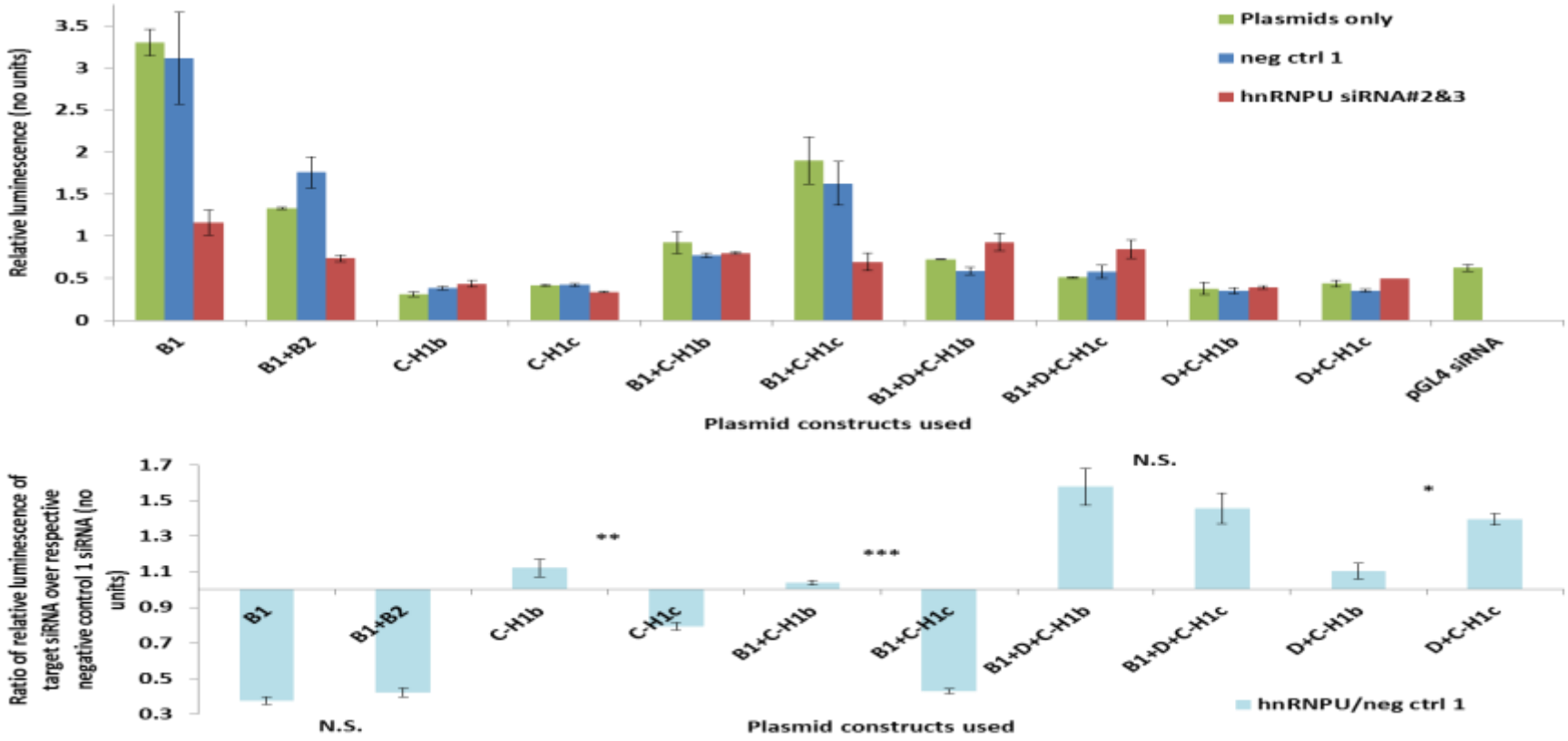


Figure 5.2.3: Effects of hnRNPU siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs.

SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with six different *MAPT* core promoter regions (B1, B1+B2, C, B1+C, B1+D+C and D+C), four of which were from a H1b or H1c background (C, B1+C, B1+D+C and D+C). Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$, ***:$p<0.0001$ (two-tailed T-test).

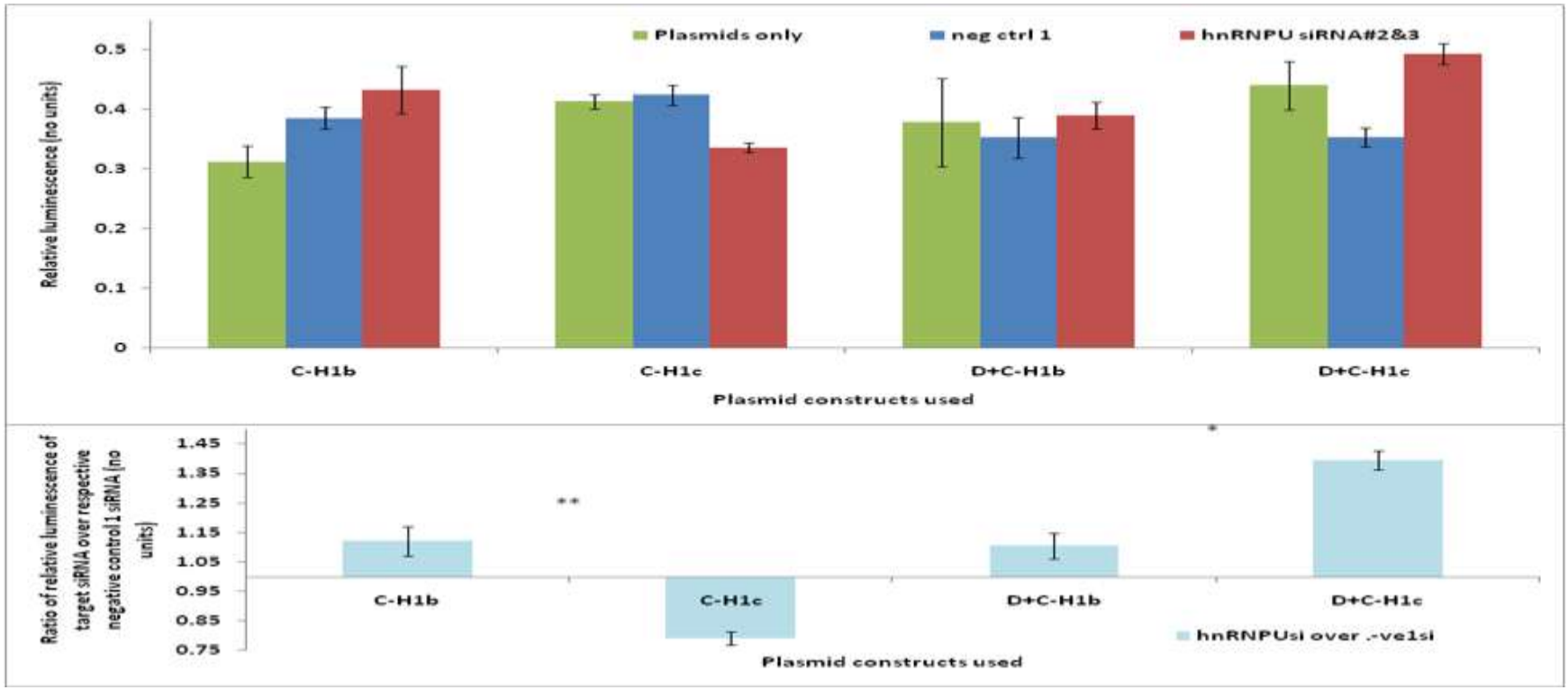


Figure 5.2.3 (continued): Effects of hnRNPU siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs. SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with two different *MAPT* core promoter regions (C and D+C), from a H1b or H1c background. Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$ (two-tailed T-test).

PTBP1

Before any analysis can begin, checks on whether the co-transfection assay had worked had to be done. The pGL4 siRNA was co-transfected with the plasmid construct containing the B1 core promoter which was the strongest plasmid construct amongst all the other plasmid constructs on their own in a luciferase assay. Compared to the B1 plasmid construct on its own, the co-transfection of B1 plasmid construct with pGL4 siRNA did manage to knock it down by 60.1%, not as much as the 80% seen during the optimization experiment (Figure 5.2.4). Next, co-transfection with negative control 1 siRNA was compared to transfection with plasmid only to observe whether there was no significance difference between the two conditions. Out of the ten plasmid constructs used, only two plasmid constructs that had been co-transfected with negative control 1 siRNA (B1 and D+C-H1b), did show significant reductions of around 30%.

The first observable effect of knockdown of PTBP1 in the cells was that like U2AF2 this caused reduction in transcriptional activity for most of the *MAPT* plasmid construct variants. The largest was on the core promoter and the H1c haplotype of the downstream C region, B1+C-H1c, with reductions in transcriptional activity by about 70%. This was a 50% greater reduction that PTBP1 knockdown had on the core promoter, B1, alone. PTBP1 knockdown had the strongest effect on the extended core promoter, B1+B2, with 65% greater reduction in transcriptional activity compared to B1 alone.

Based on the initial premise of this project, the region of greatest interest was the downstream promoter region, C, where rs242557 resided. Although the reduction of PTBP1 level did not affect the C region much, it had an opposing effect depending on the allelotype. C-H1b was significantly more sensitive to the reduction of PTBP1 levels than C-H1c with a 40% reduction in promoter strength compared to a 15% gain respectively ($p<0.01$). When the core promoter, B1, was included with the C region however, the allelic effect seen before was eliminated. Both plasmid constructs B1+C-H1b and B1+C-H1c were not significantly different from each other with a reduction in promoter strength of about 65% like with to B1 alone. The extended downstream promoter region D+C, had a similar effect as B1+C in that the allelic effect was abolished. The effect was less pronounced as there was no significant difference between the activities of D+C-H1b and D+C-H1c, with

a reduction in promoter strength of about 30%. Unlike B1+C where there were significant effects of the core promoter on both haplotypes of the C region, the D region did not have any significant effect on C-H1b but rather only on C-H1c.

It was interesting to note that when the different domains of the *MAPT* promoter were put together – the core promoter B1 and the extended downstream promoter region D+C (B1+D+C – the most extensive promoter construct used), there was still a significant ($p<0.05$) allelic effect of rs242557. However, B1+C and D+C did not show any allelic effect, demonstrating the complex and synergistic effects of the different elements of the *MAPT* promoter.

Compared to the other plasmids representing the H1b sub-haplotype, B1+D+C-H1b had a very similar reduction (40%) in promoter activity with PTBP1 knockdown, similar to the effect on B1 alone. Of all the H1c sub-haplotype variants, the transcription of B1+D+C-H1c was the least affected by PTBP1 knockdown with only a 2% reduction compared to a 50% reduction seen with B1 alone. Hence, PTBP1 knockdown had very little if any effect on the function of B1+D+C-H1c but caused a 20-fold greater reduction of B1+D+C-H1b transcriptional activity. It is also of interest that with PTBP1 knockdown, the effect on H1c haplotype promoter variants varied widely: there was a gain in allelic effect seen with C-H1c alone, a moderate reduction in allelic effect with the D+C-H1c and a substantial reduction in allelic effect with B1+C-H1c whereas B1+D+C-H1c was not affected. Thus, it was not only important to study the different components of a promoter but also the cumulative interactions of these components. All these results seemed to suggest that PTBP1 has an intimate and significant relationship with the H1c risk haplotype of *MAPT* promoter

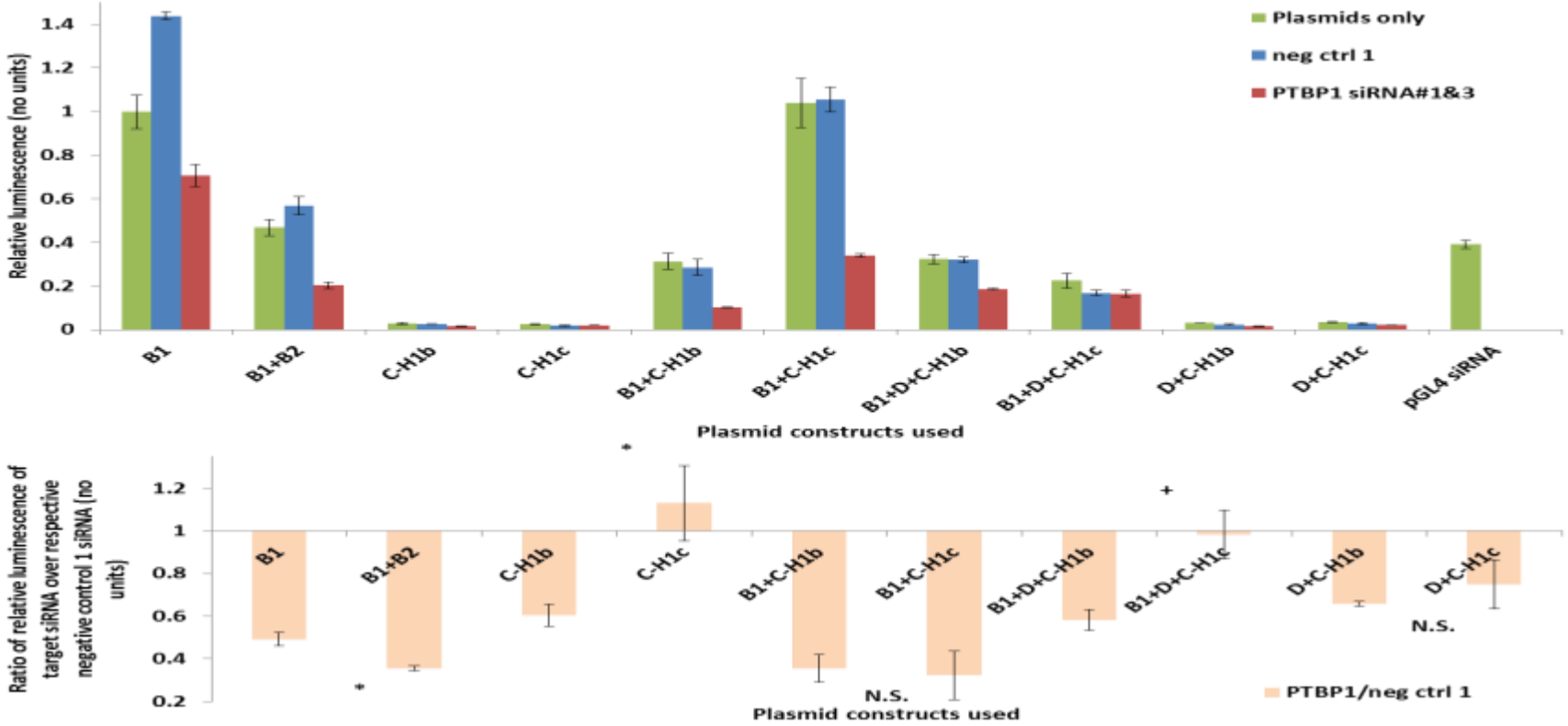


Figure 5.2.4: Effects of PTBP1 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs.

SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with six different *MAPT* core promoter regions (B1, B1+B2, C, B1+C, B1+D+C and D+C), four of which were from a H1b or H1c background (C, B1+C, B1+D+C and D+C). Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation. $^{+}$: $p<0.05$, *: $p<0.01$ (two-tailed T-test).

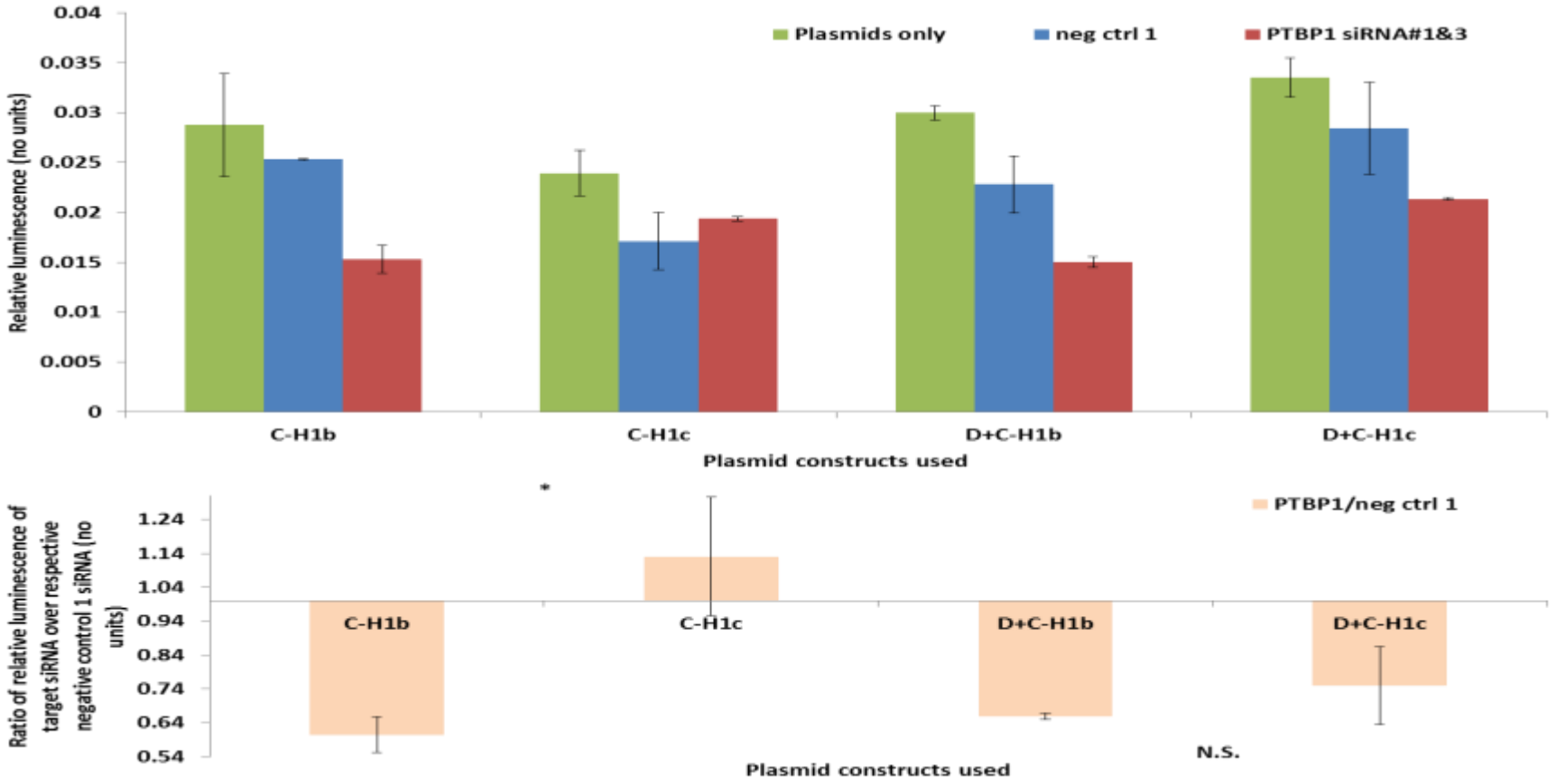


Figure 5.2.4 (continued): Effects of PTBP1 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs. SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with two different *MAPT* core promoter regions (C and D+C), from a H1b or H1c background. Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$ (two-tailed T-test).

hnRNPD0

Before any analysis can begin, checks on whether the co-transfection assay had worked had to be done. The pGL4 siRNA was co-transfected with the plasmid construct containing the B1 core promoter which was the strongest plasmid construct amongst all the other plasmid constructs on their own in a luciferase assay. Compared to the B1 plasmid construct on its own, the co-transfection of B1 plasmid construct with pGL4 siRNA did manage to knock it down by 73.9%, close to the 80% seen during the optimization experiment (Figure 5.2.5). Next, co-transfection of plasmid constructs with negative control 1 siRNA was compared to transfection with plasmid only to observe whether there was no significance difference between the two conditions. Out of the ten plasmid constructs used, only two plasmid constructs that had been co-transfected with negative control 1 siRNA (B1+B2 and B1+C-H1c), did show significant reduction of around 30% and gain of 20% respectively.

The first observable effect of reducing the level of hnRNPD0 in the cell was that it had a close and complex relationship with the various *MAPT* promoter variants. The largest effect of knocking down hnRNPD0 was on the core promoter, B1, and the H1c haplotype of the extended downstream promoter region, D+C-H1c and extended promoter, B1+D+C-H1c, all of which had about 100% increases in transcriptional activity. The extended core promoter, B1+B2, however, was significantly unaffected with only a slight reduction of 6% ($p<0.001$).

Based on the initial premise of this project, the region of greatest interest was the downstream promoter region, C, where rs242557 resided. Although hnRNPD0 knockdown did not affect the C region much, it had an opposing effect depending on the allelotype. C-H1b was significantly more sensitive to hnRNPD0 knockdown than C-H1c with a 30% gain in promoter strength compared to a 60% reduction respectively ($p<0.01$). When the core promoter, B1, was included with the C region however, the allelic effect was reduced and B1 seemed to not affect the C region much. B1+C-H1b was unaffected by the knockdown of hnRNPD0 (reduction of 1%) which was similar to B1+B2 while B1+C-H1c was less affected (reduction of 40%).

Knockdown of hnRNPD0 had the opposite allelic effect on downstream promoter region D+C ($p<0.01$). D+C-H1b and C-H1b activity had 30% reduction and 30% increase respectively. Knockdown of hnRNPD0 had the strongest allele-specific on D+C-H1c promoter activity (100% increase) whereas the C-H1c variant had a 60% reduction in activity.

When the different domains of the *MAPT* promoter were put together: the core promoter B1 and the extended downstream promoter region D+C (B1+D+C – the most extensive promoter construct used), a significant ($p<0.01$) rs242557 allelic effect was still seen. Of all the H1b plasmid constructs used, B1+D+C-H1b had a similar gain in promoter strength as C-H1b (30%) with hnRNPD0 knockdown. B1+D+C-H1c unlike B1+D+C-H1b had the greatest increase in promoter activity of all the H1c plasmid constructs, a 135% gain. Hence, hnRNPD0 knockdown had a moderate effect on the function of B1+D+C-H1b but caused a four-fold increase in B1+D+C-H1c activity. Another interesting note was that even though no knockdown effect was seen with B1+C-H1b, it was only when the D region was included that the full allelic effect was restored. The D region also had an overriding effect on B1+C-H1c. Thus, it was not only important to study the different components of a promoter but also the cumulative effects of these components. All these results seemed to suggest that hnRNPD0 had a haplotypic effect on the various components of the *MAPT* promoter but had an overall repressive haplotypic effect on the *MAPT* promoter.

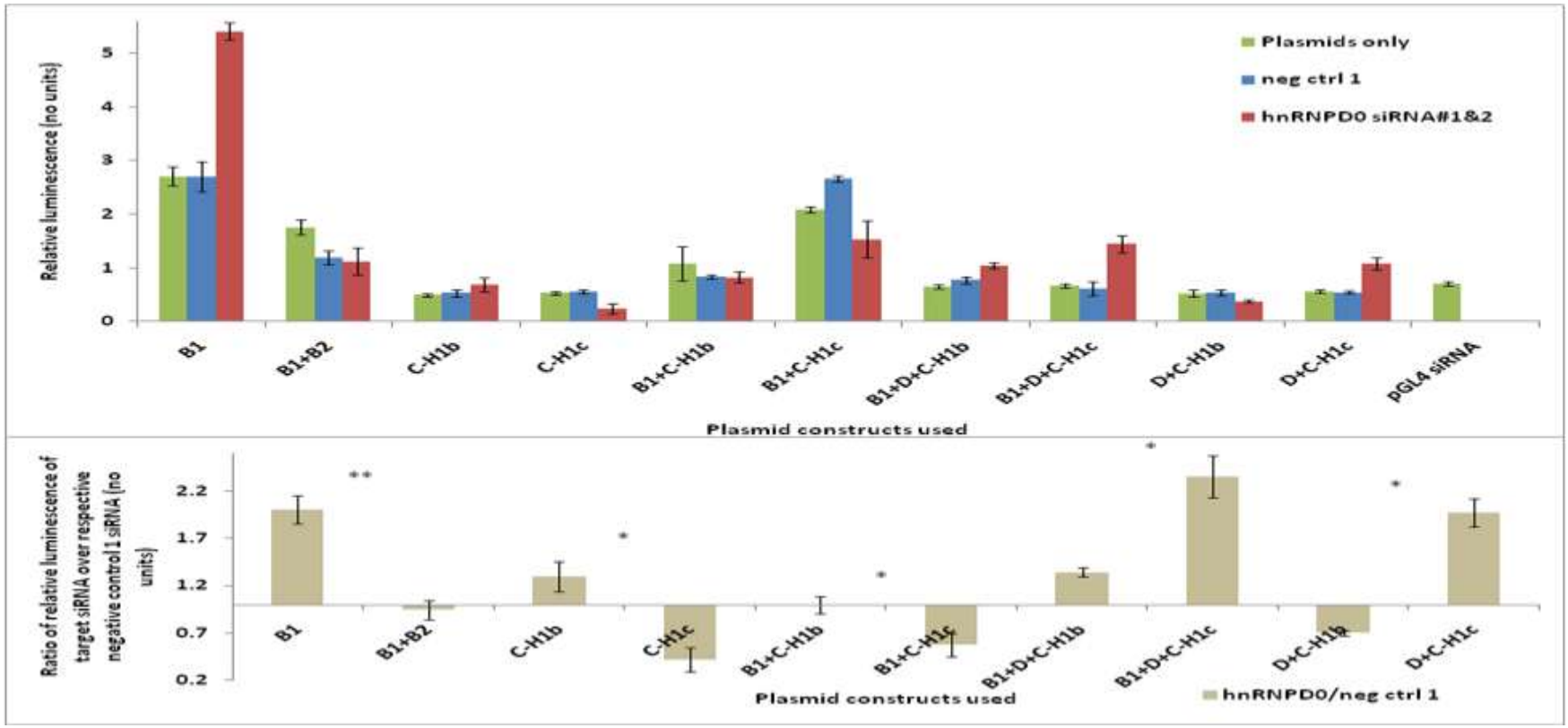


Figure 5.2.5: Effects of hnRNPD0 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs.

SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with six different *MAPT* core promoter regions (B1, B1+B2, C, B1+C, B1+D+C and D+C), four of which were from a H1b or H1c background (C, B1+C, B1+D+C and D+C). Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$ (two-tailed T-test)

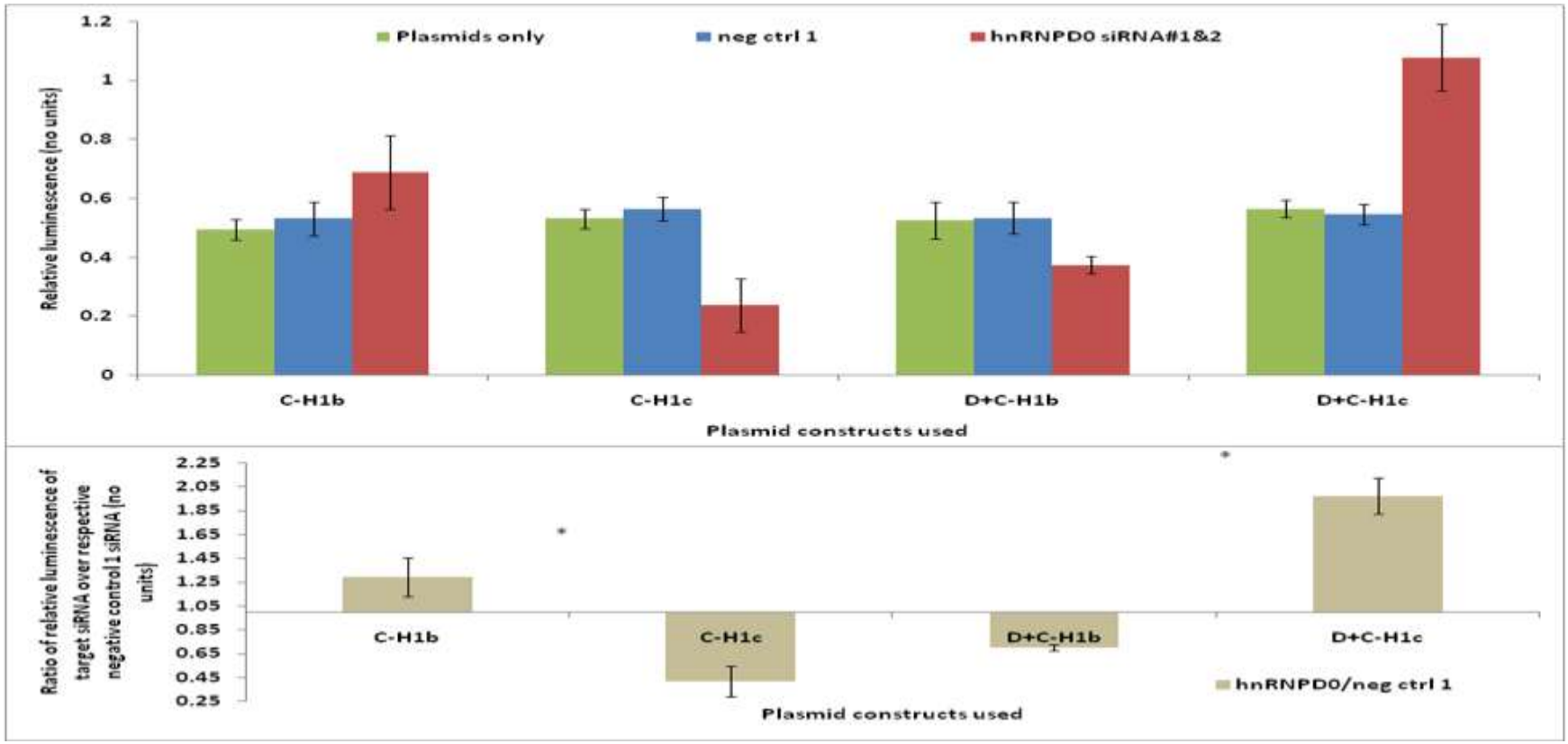


Figure 5.2.5 (continued): Effects of hnRNPD0 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs. SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with two different *MAPT* core promoter regions (C and D+C), from a H1b or H1c background. Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$ (two-tailed T-test).

CP2

Before any analysis can begin, checks on whether the co-transfection assay had worked had to be done. The pGL4 siRNA was co-transfected with the plasmid construct containing the B1 core promoter which was the strongest plasmid construct amongst all the other plasmid constructs on their own in a luciferase assay. Compared to the B1 plasmid construct on its own, the co-transfection of B1 plasmid constructs with pGL4 siRNA did manage to knock it down by 72.3%, similar to the 80% seen during the optimization experiment (Figure 5.2.6). Next, co-transfection of all plasmid constructs with negative control 1 siRNA was compared to transfections with plasmids only to observe whether there was no significance difference between the two conditions. Out of the ten plasmid constructs used, only two plasmid constructs that had been co-transfected with negative control 1 siRNA (C-H1c and D+C-H1c), did show significant reductions of around 30%.

The first observable effect of CP2 knockdown was that it was allele specific and had a moderate effect on most of the *MAPT* promoter variants. The largest effect was on the H1c haplotype of the downstream C region and the extended downstream region, C+D, both had substantial gains in activity of 150% and 200% respectively. This was sharply contrasted with the 57% and 13% reduction in activity of the core promoter, B1 and the extended core promoter, B1+B2.

Based on the initial premise of this project, the region of greatest interest was the downstream promoter region, C, where rs242557 resided. C-H1b was significantly less responsive to CP2 knockdown than C-H1c with an 8% and 150% gains in promoter activity, respectively ($p<0.001$). This was the largest allelic effect seen with almost a twenty-fold difference between the risk and non-risk associated haplotype. Hence, CP2 would exert a much more repressive effect on the risk haplotype. When the core promoter, B1, was included with the C region however, the extreme allelic effect seen before was abolished. Both plasmid constructs B1+C-H1b and B1+C-H1c were not significantly different from each other with a reduction in promoter strength of about 50%, not dissimilar to that seen with B1 alone.

The extended downstream promoter region D+C, had the opposite effect in that the allelic effect was significantly exacerbated ($p<0.001$). D+C-H1b with CP2 knockdown varied moderately compared to C-H1b which had a 36% gain in the luciferase assay. D+C-H1c was more affected by CP2 knockdown (200% gain) than the 150% gain with C-H1c.

When the different domains of the *MAPT* promoter were put together: the core promoter B1 and the extended downstream promoter region D+C (B1+D+C – the most extensive promoter construct used), CP2 knockdown resulted in the smallest significant ($p<0.001$) rs242557 allelic effect and the differences were in the opposite direction compared to the previous results with CP2. Of all the H1b plasmid constructs used, there was little influence on promoter activity and, B1+D+C-H1b like D+C-H1b had a 30% gain in promoter strength with low levels of CP2. Although B1+D+C-H1c like B1+D+C-H1b promoter activity was least affected of all the H1c plasmid constructs, it was much more dramatic as it was a 50% reduction compared to a gain, but comparing to B1+C-H1c, there was little change. This was a big change from the 150 and 200% gain seen with C-H1c and D+C-1c. Hence, the CP2 knockdown had a very modest haplotype effect (two and half-fold difference) on B1+D+C-H1 but of opposing effect. Also of interest was that even though no allelic effect was seen with B1+C, it was only when the D region was included that the full allelic effect was observed and this was completely different from either B1+C or D+C. Thus, it was not only important to study the different components of a promoter but also the cumulative effects of these components. All these results seemed to suggest that CP2 has a highly specific role in the downstream *MAPT* promoter region of C and D+C but this is attenuated by the presence of the core promoter B1.

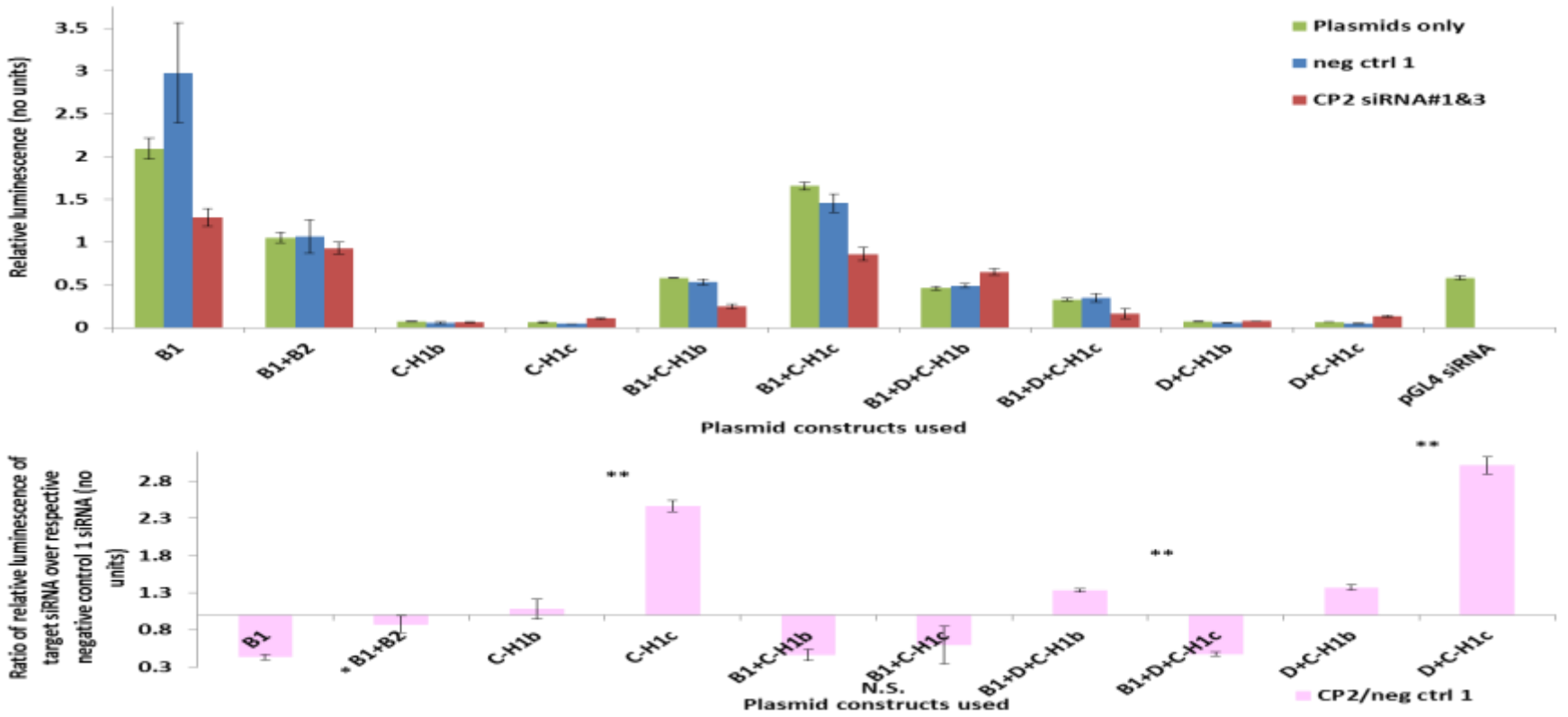


Figure 5.2.6: Effects of CP2 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs.

SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with six different *MAPT* core promoter regions (B1, B1+B2, C, B1+C, B1+D+C and D+C), four of which were from a H1b or H1c background (C, B1+C, B1+D+C and D+C). Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$ (two-tailed T-test).

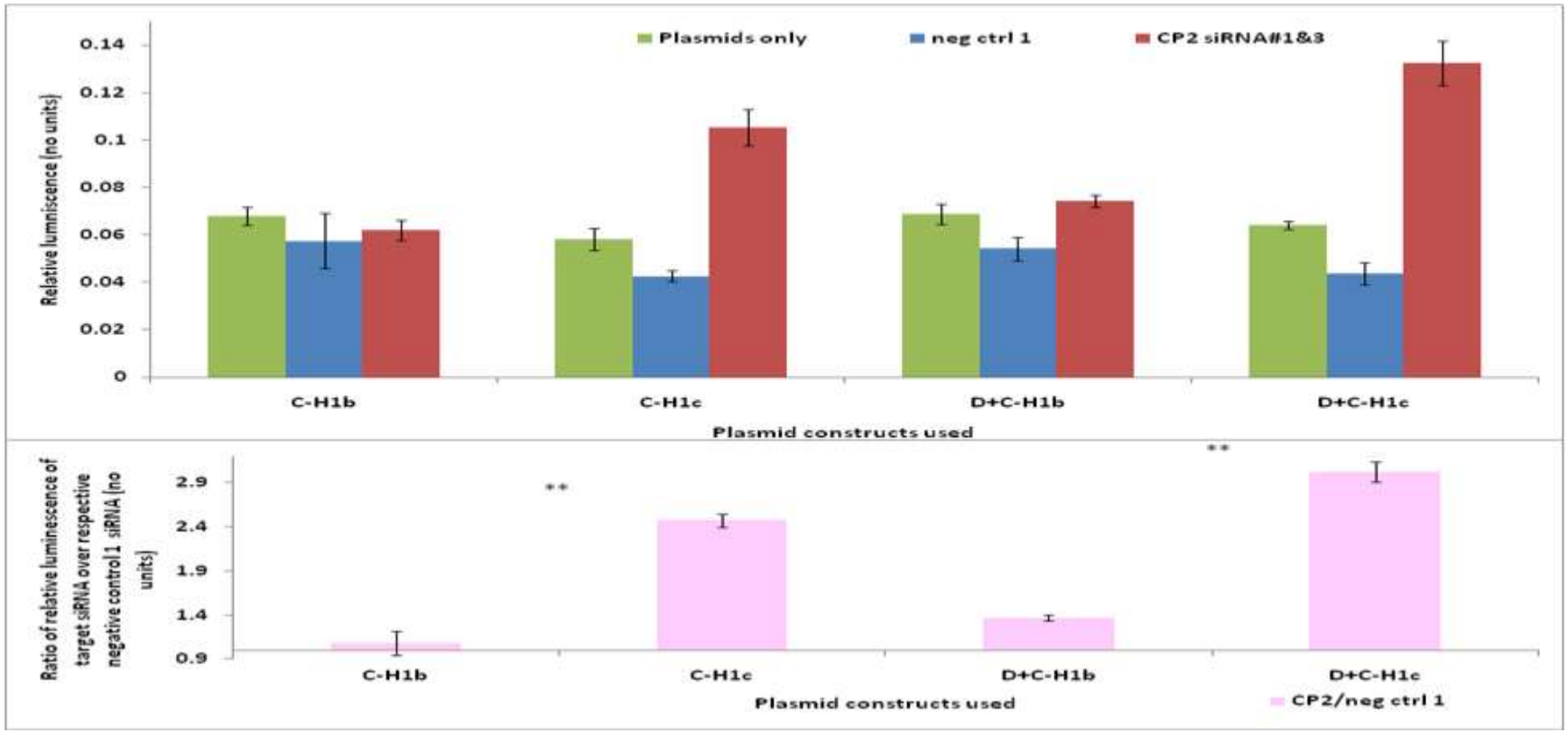


Figure 5.2.6 (continued): Effects of CP2 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs. SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with two different *MAPT* core promoter regions (C and D+C), from a H1b or H1c background. Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation. **: $p<0.001$ (two-tailed T-test).

TDP43

Before any analysis can begin, checks on whether the co-transfection assay had worked had to be done. The pGL4 siRNA was co-transfected with the plasmid construct containing the B1 core promoter which was the strongest plasmid construct amongst all the other plasmid constructs on their own in a luciferase assay. Compared to the B1 plasmid construct on its own, the co-transfection of the B1 plasmid construct with pGL4 siRNA did manage to knock it down by 84.8%, very similar to the 80% seen during the optimization experiment (Figure 5.2.7). Next, co-transfection of plasmid constructs with negative control 1 siRNA was compared to transfections with plasmids only to observe whether there was no significance difference between the two conditions. Out of the ten plasmid constructs used, three plasmid constructs that had been co-transfected with negative control 1 siRNA (B1+C-H1c, B1+D+c-H1b and B1+D+C-H1c), did show significant reductions of around 50%, 20% and 30% respectively.

The first observable effect of TDP43 knockdown was that only those promoter variants containing the core promoter domain, B1, were affected with the strongest effect on the B1, B1+B2, B1+C-H1b, B1+C-H1c, B1+D+C-H1b, B1+D+C-H1c variants. In all cases, TDP43 knockdown resulted in a range of increases in promoter activity (from 170% for B1+D+C-H1b to 240% for B1+D+C-H1c with non-significant differences) between variants under a two-tailed T-test. The rs242557 allelotype did not have any modulatory effect on the changes caused by TDP43 knockdown as seen with B1+C and B1+D+C. Furthermore, plasmid constructs without the core promoter, B1, element (C and D+C) were not affected by the knockdown of TDP43. This demonstrated that TDP43 specifically acted on the core promoter of *MAPT* to repress its activity.

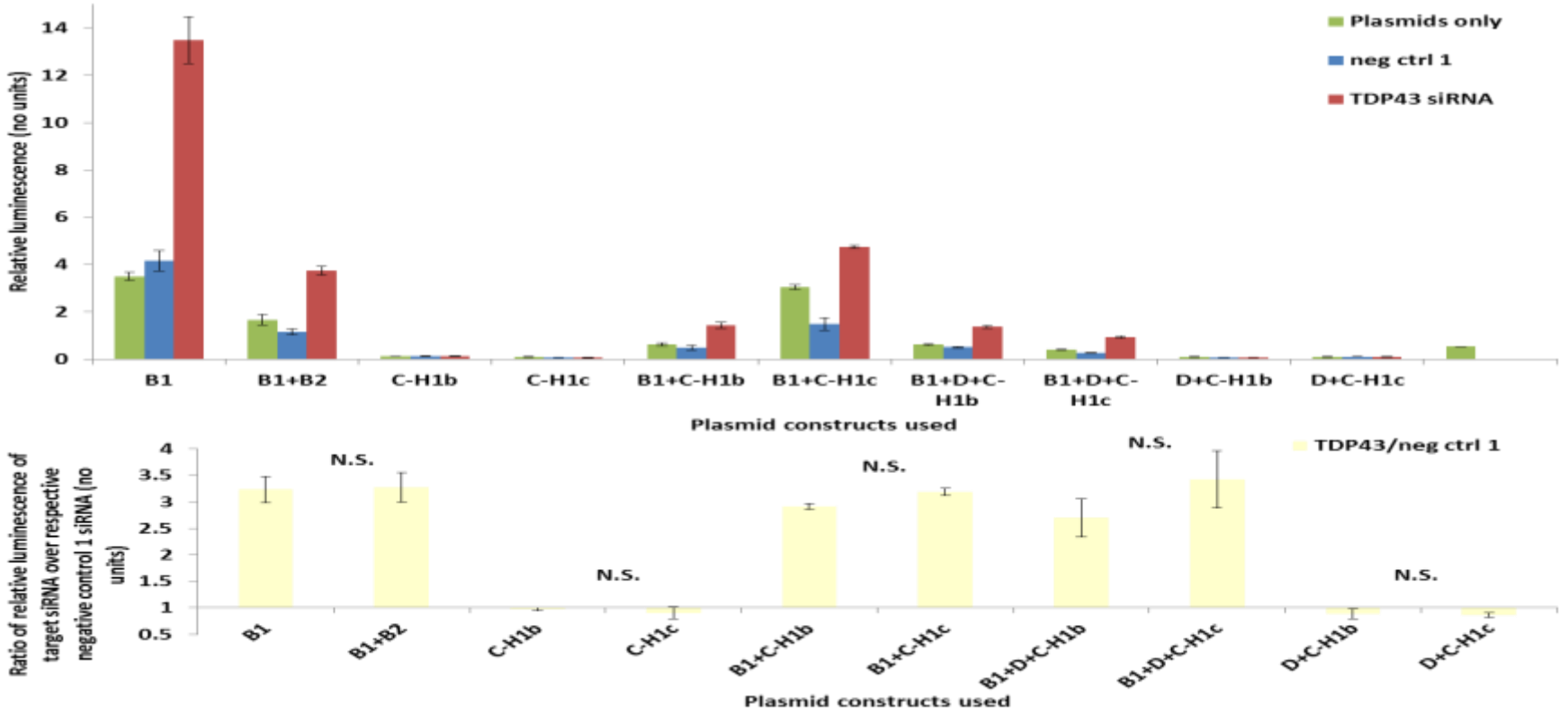


Figure 5.2.7: Effects of TDP43 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs.

SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with six different *MAPT* core promoter regions (B1, B1+B2, C, B1+C, B1+D+C and D+C), four of which were from a H1b or H1c background (C, B1+C, B1+D+C and D+C). Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation.

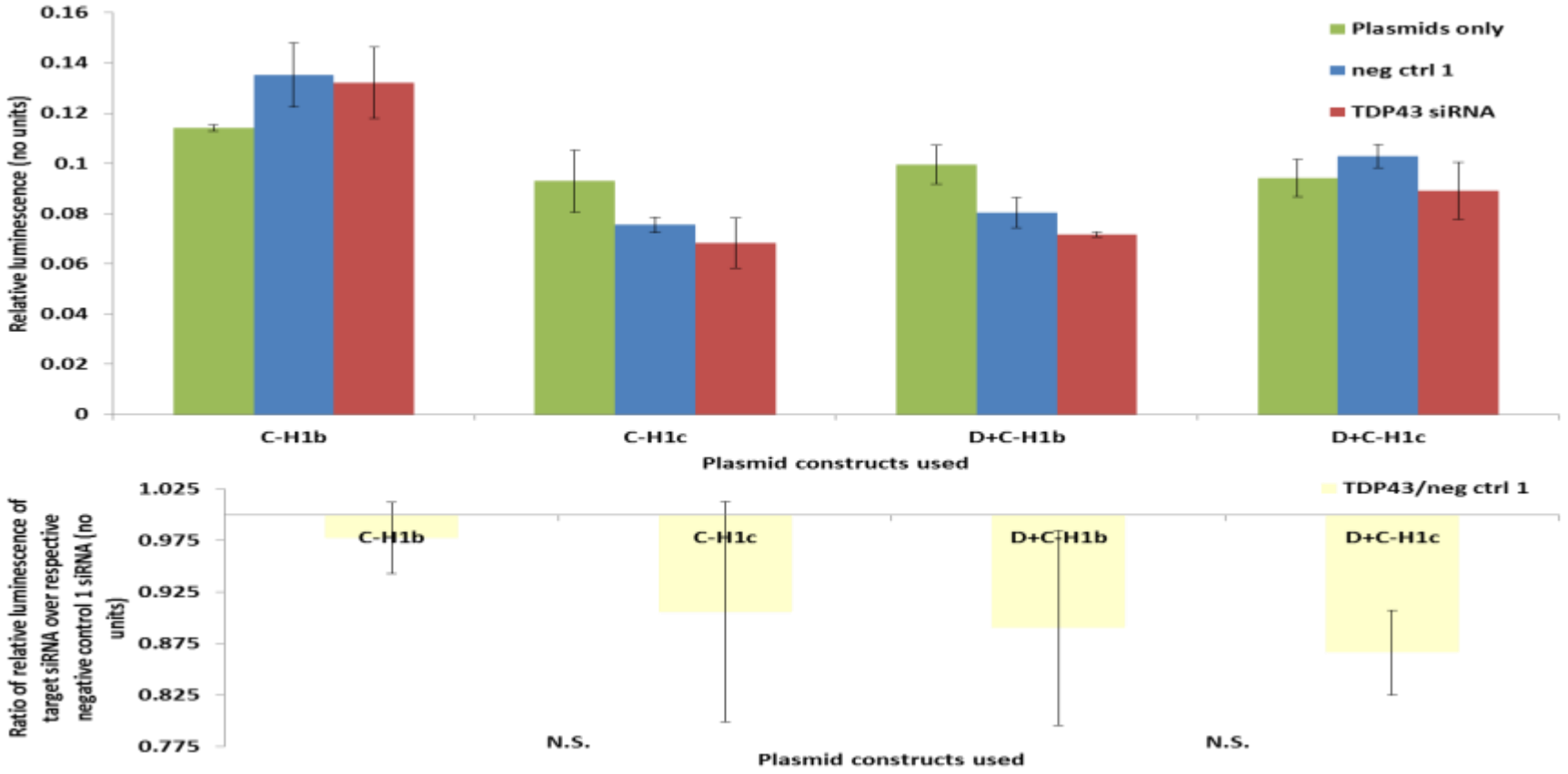


Figure 5.2.7 (continued): Effects of TDP43 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs. SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with two different *MAPT* core promoter regions (C and D+C), from a H1b or H1c background. Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation.

PAX8

Before any analysis can begin, checks on whether the co-transfection assay had worked had to be done. The pGL4 siRNA was co-transfected with the plasmid construct containing the B1 core promoter which was the strongest plasmid construct amongst all the other plasmid constructs on their own in a luciferase assay. Compared to the B1 plasmid construct on its own, the co-transfection of the B1 plasmid constructs with pGL4 siRNA did manage to knock it down by 74.3%, close to the 80% seen during the optimization experiment (Figure 5.2.8). Next, co-transfection of plasmid constructs with negative control 1 siRNA was compared to transfections with plasmids only to observe whether there was no significance difference between the two conditions. Out of the ten plasmid constructs used, none of the plasmid constructs that had been co-transfected with negative control 1 varied significantly from their respective plasmid only co-transfection.

The first observable effect of reducing the level of PAX8 in the cell was that no significant effect was seen for any of the *MAPT* plasmid construct used, regardless of the haplotype effect. This demonstrated that the transcription factor PAX8 did not specifically act on the promoter region of *MAPT*. As a transcription factor specific control, PAX8 had not been previously shown to interact with rs242557 and this was still the case.

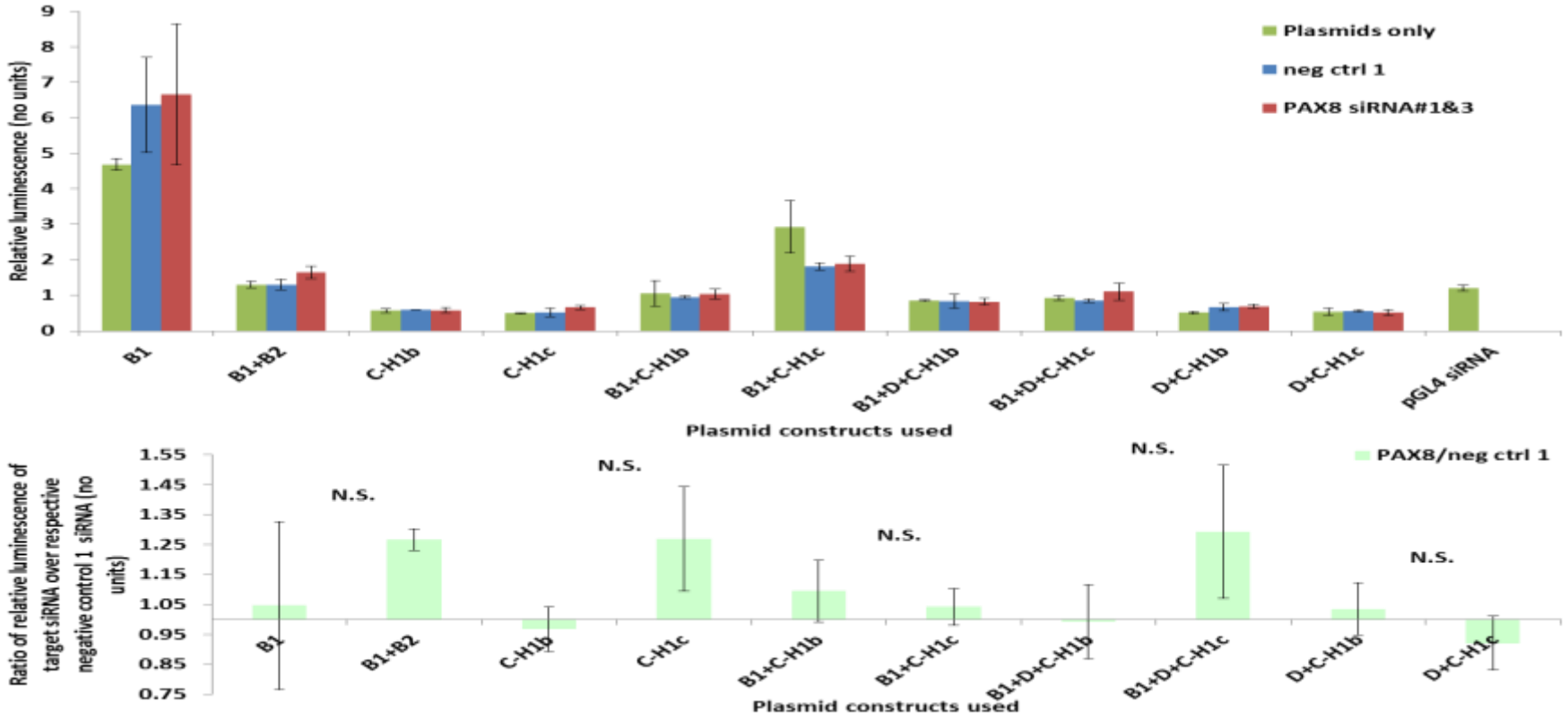


Figure 5.2.8: Effects of PAX8 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs.

SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with six different *MAPT* core promoter regions (B1, B1+B2, C, B1+C, B1+D+C and D+C), four of which were from a H1b or H1c background (C, B1+C, B1+D+C and D+C). Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation.

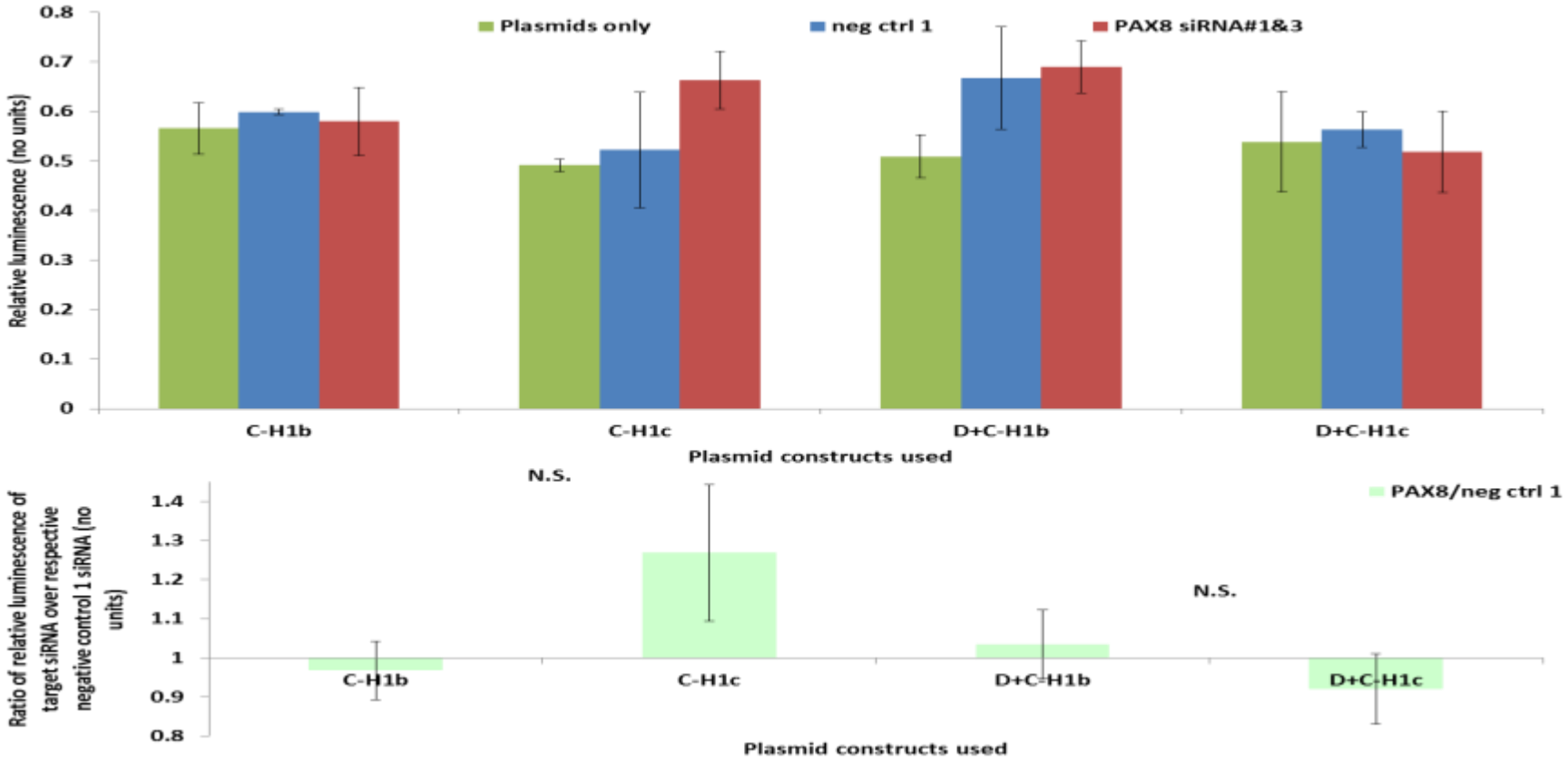


Figure 5.2.8 (continued): Effects of PAX8 siRNA knockdown on various regions of *MAPT* core promoter plasmid constructs. SH-SY5Y cells were transfected with negative control 1 siRNA (neg ctrl 1) or target siRNAs (#1&2) along with two different *MAPT* core promoter regions (C and D+C), from a H1b or H1c background. Top: Relative luminescence for each transfection condition with pGL4 siRNA as a positive control knockdown of B1. Bottom: Representation of target siRNA cotransfection as a fraction of their respective negative control 1 siRNA cotransfection. Each was done in triplicate. Error bars represent standard deviation.

5.3 Discussion on functional assays of shortlisted DNA-binding proteins

The choice of the transcription factor PAX8 as a control transcription factor for the functional assays was completely random. A titration of the PAX8 antibody with nuclear extracts of undifferentiated SH-SY5Y neuroblastoma cells showed that PAX8 was expressed in these cells and could be used as a transcription factor to be knocked down in the functional assays (result not included).

PAX8 belongs to the PAX family of transcription factors primarily involved in the development of the central nervous system.[261] The protein structure consists of two regions of nuclear localization signal as well as an activating and an inhibitory region (Figure 5.3.1A).[262] In humans, there are four isoforms of PAX8 generated through alternative splicing.[263] PAX8 was found to be expressed at the midbrain-hindbrain boundary, kidney and thyroid. In Wilm's tumour, a paediatric renal carcinoma, PAX8 was found to be abnormally upregulated (Figure 5.3.1B).[264] Based on its expression, it was subsequently found that PAX8 regulates the neural cell adhesion molecule (NCAM), Wilms' tumour (WT1) and thyroperoxidase, a primary enzyme in thyroid hormone synthesis, genes (Figure 5.3.1C).[265,266,267,268] More recently, it was shown in glioma using EMSA, quantitative PCR and a telomerase activity assay that PAX8 bound directly to the promoter of both telomerase catalytic subunit (TERT) and the telomerase RNA component (TR).[269]

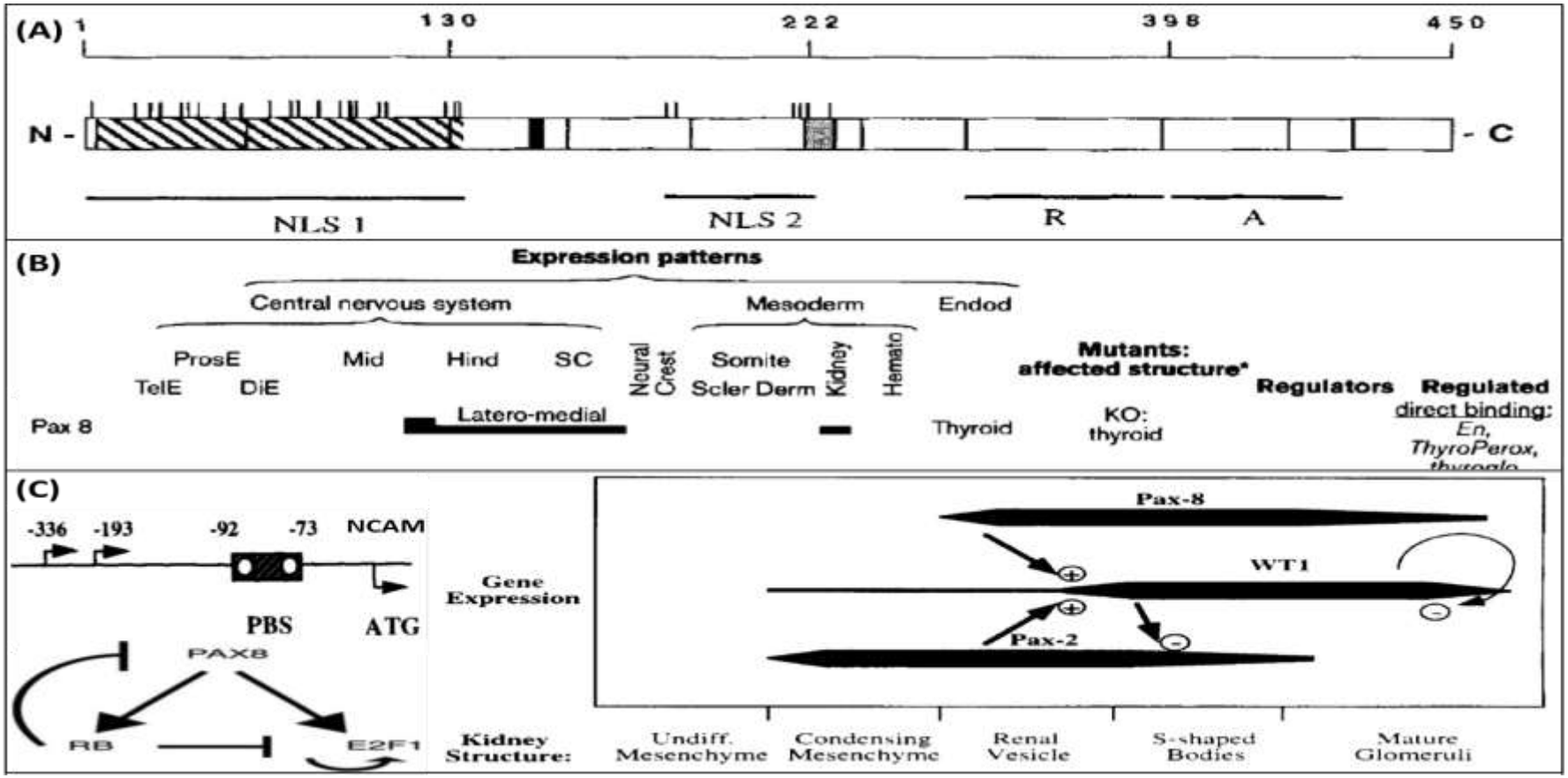


Figure 5.3.1: Schematic of PAX8 protein, its expression pattern and genes it acts on.

(A) PAX8 consists of a paired box domain (striped), octapeptide (filled) and residual paired-type homeodomain (dotted). Nuclear localization signal (NLS), repressor (R) and activation (A) domains indicated. (B) In mice, the PAX8 expression is across the midbrain-hindbrain boundary, kidney and thyroid. (C) Some of the genes known to be regulated by PAX8 include neural cell adhesion molecule (NCAM), E2F1 and Wilms' tumour (WT1) genes.[390,391,394,395,399]

Before we can analyse the effects of the protein knockdowns on the C region (where rs242557 resides), downstream of the *MAPT* promoter region, or in different combinations with different elements of the *MAPT* promoter, we also need to analyse their effects on the core promoter if any. This would allow us to see if these proteins could be part of a transcriptional complex that forms the framework for the interaction of the core promoter and the regulatory (enhancer/repressor) C region. We looked at the effects of the individual protein knockdowns on the core promoter region B1 of *MAPT* and an extended promoter region which included the conserved region (B2) immediately downstream of B1 and of approximately the same size, thus named B1+B2 (Figure 5.3.2). This study did not include the effects of the individual protein knockdowns on B2 on its own. Hence, it would not be possible to attribute any effect seen on B1+B2 to either B1 or B2 or a synergistic effect of both elements. In light of the current results, however, it might be worth to investigate this further.

Since this was a protein knockdown assay, any increase in the promoter luciferase read out would mean that a reduction of that protein led to that increase in activity and that the protein acts as a promoter repressor. An example of this is TDP-43. Reducing the level of TDP-43 in the cell using siRNA knockdown led to approximately a three-fold increase in promoter activity of B1 and this was the same with B1+B2. The control protein PAX8 knockdown did not have any effect on the activity of B1 or B1+B2. All the protein knockdown effects were compared to the effects seen with knocking down with a scrambled control siRNA.

When a reduction in the level of a protein leads to a reduction in the readout of the promoter luciferase assay, the protein could be an enhancer of promoter activity. Proteins knockdowns that reduced the activity of both B1 and B1+B2 were U2AF2, hnRNPU and PTB1. Both promoter elements B1 and B1+B2 were most significantly affected by U2AF2 knockdown. This would suggest that the interplay of the opposing effects U2AF2 and TDP43 might be of importance in regulation of transcription with the possibility of the involvement of hnRNPU and PTB1.

Noting the significant differences between the protein knockdown on B1 compared to B1+B2, keeping in mind no work was done on B2 alone, PTBP1, hnRNPD0 and CP2

stood out. For example, PTBP1 knockdown had a stronger effect on the extended promoter region B1+B2 compared to B1 alone. Knockdown of hnRNPD0 which had an enhancing effect on B1 alone and CP2 which had a repressive effect on B1 alone, both did not have any effect on B1+B2. Further study into B2 would need to be done to explain these effects, especially in light of the work currently being done to ascertain the natural antisense transcript being encoded in that region.

After looking at the effects of the protein knockdowns on the core promoter elements, the next step is to look at the effects on the downstream C region in isolation to see if there were any allele-specific effects based on rs242557-G (H1b / H2) with rs242557-A (H1c) (Figure 5.3.3). From earlier experiments, it was evident that TDP-43 does not interact with the rs242557 region. Hence, knocking this protein down had no significant effect on the transcriptional activity of the C region as was the case with the control protein knockdown of PAX8.

With the neutral allelotype, C-H1b, most of the protein knockdowns had very little effect. U2AF2 and PTBP1 played a role in the transcriptional activity of C-H1b as reducing the levels of these proteins led to a reduction in the luciferase promoter assay. Knockdown of hnRNPU, hnRNPD0 and CP2 either had negligible or mildly enhancing effects on the C-H1b element. Comparing these effects on C-H1b, the proteins of interest had a more pronounced effect on the region C-H1c risk allele. U2AF2 showed least difference between the two alleles. Nevertheless, U2AF2 knockdown did have an effect on C-H1c, albeit less pronounced compared to the other proteins of interest. hnRNPU knockdown enhanced C-H1b, caused a slightly reduced activity of C-H1c. Conversely, knockdown of PTBP1 which repressed C-H1b, had negligible effect on C-H1c. Knockdown of hnRNPD0 like hnRNPU had a slightly enhancing effect on C-H1b but unlike hnRNPU, it repressed the activity of C-H1c more strongly. The biggest effect was seen with CP2 knockdown. Although it had negligible effect on C-H1b, it resulted in 2.5-fold enhancement of C-H1c activity.

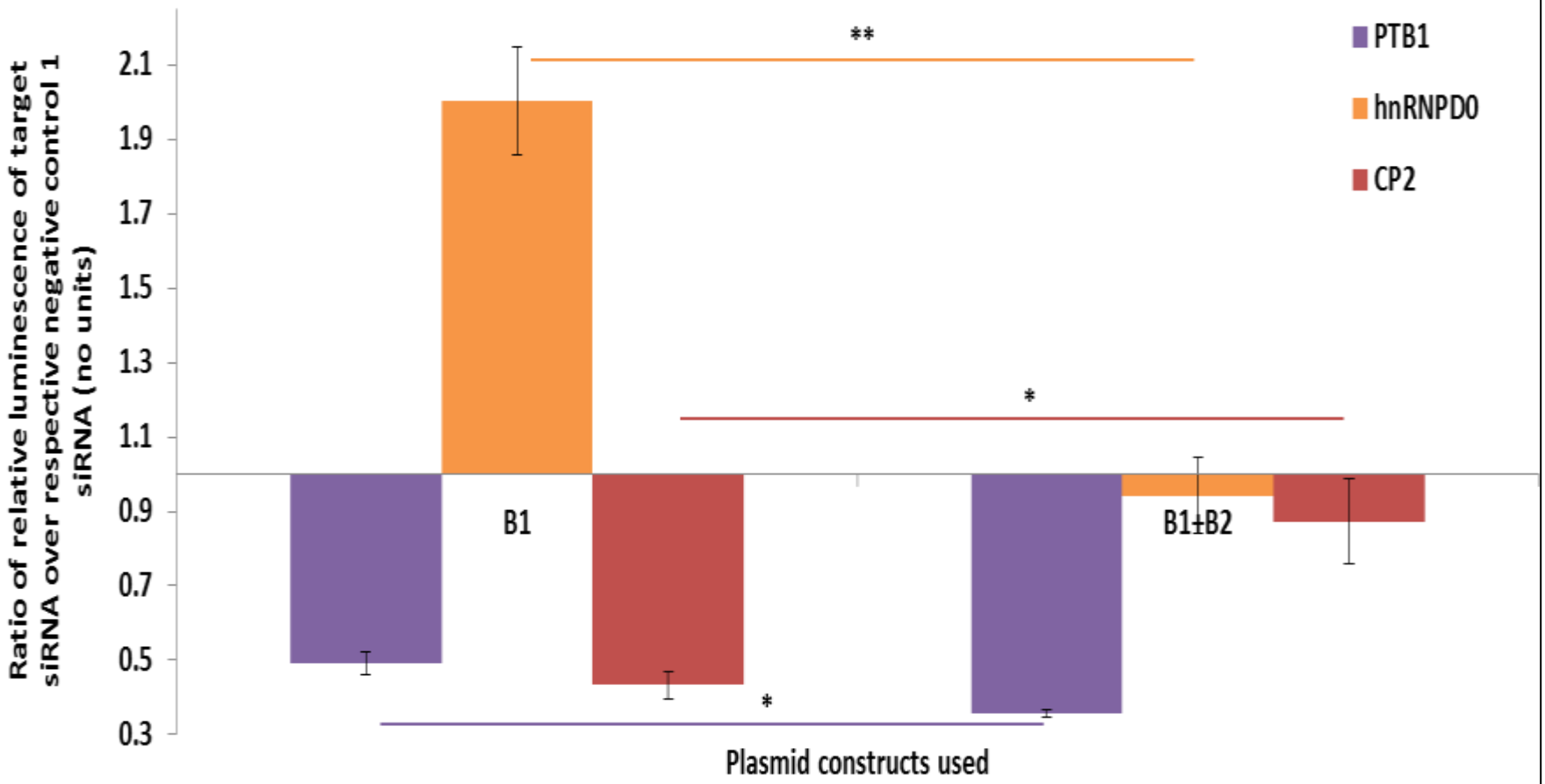


Figure 5.3.2: Effects of shortlisted DNA-binding proteins on the core promoter of *MAPT* plasmid constructs.

Only shortlisted proteins that showed significant difference between the different plasmid constructs were included. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$ (two-tailed T-test).

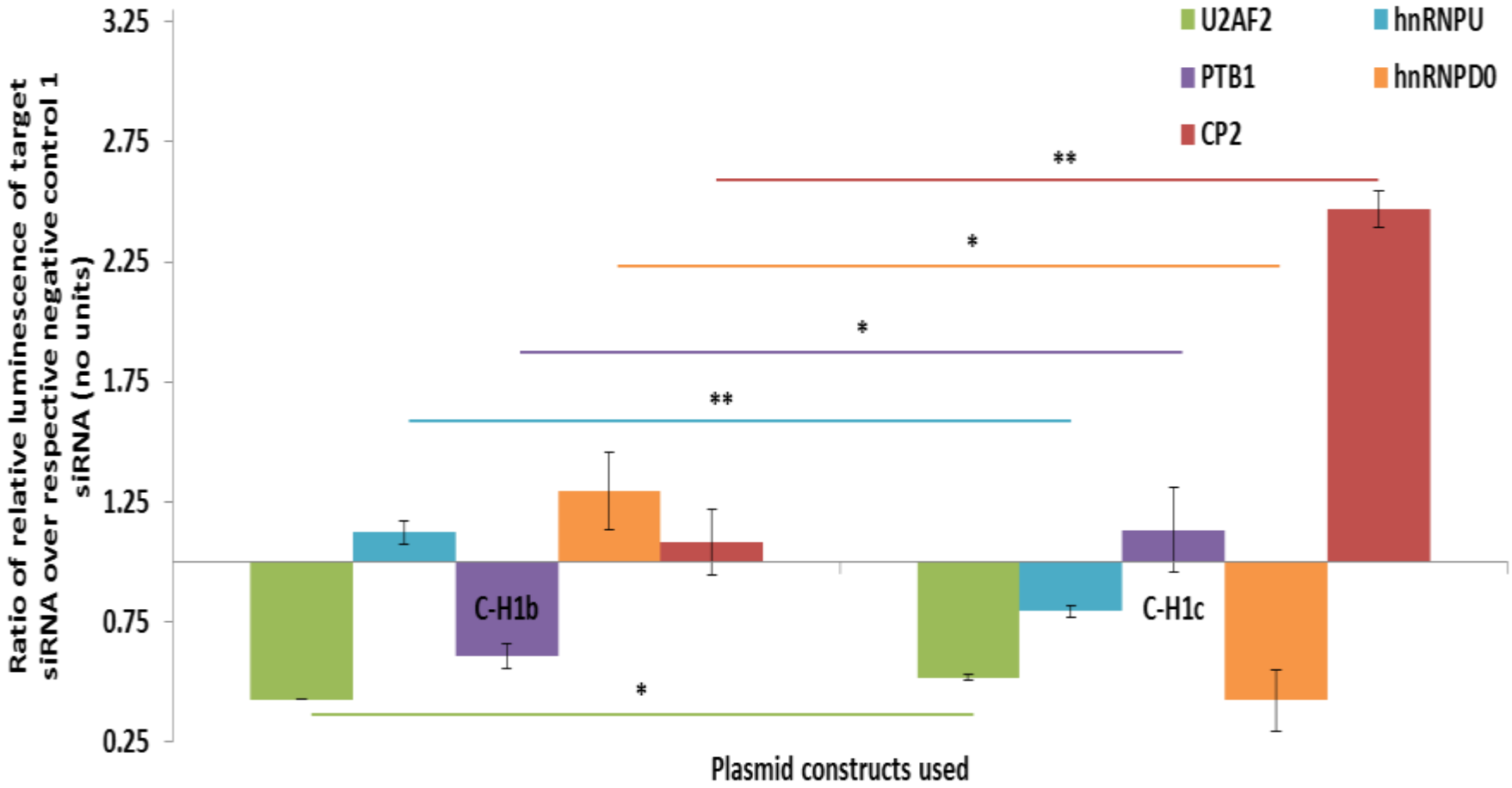


Figure 5.3.3: Effects of shortlisted DNA-binding proteins on the haplotype difference of the downstream promoter region of *MAPT* plasmid constructs. Only shortlisted proteins that showed significant difference between the different plasmid constructs were included. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$ (two-tailed T-test).

After analyzing the effects of protein knockdown on the individual elements of the core promoter B1, B1+B2 and the C region, the effect on these elements combined were assessed (Figure 5.3.4). Although this is not representative of the native gene promoter structure, which is at over 70kb is prohibitively large, it nevertheless provides a simplified model to assess the allele-specific influence of the repressor C region on the core promoter. With the B1 core promoter region combined with the C region, knockdown of U2AF2 caused a reduction of promoter activity like with B1 alone and this effect was not influenced by allelotype of the C region.

With knockdown of other proteins of interest such as hnRNPU, PTBP1 and CP2, there was a clear allele-specific effect of the C region on B1. For example, the repressive effect of hnRNPU knockdown on B1 was not modified by the presence of C-H1c but when combined with C-H1b, this effect was abolished. The reverse was true with knockdown of PTBP1 and CP2 in that their effects on B1 were not modified by C-H1b but in combination with C-H1c, the effect of PTBP1 knockdown was exacerbated whereas the effect of CP2 was diminished.

The only protein whose effect was modified by both allelotypes of the C region was hnRNPD0. In the first place, enhancing effect on B1 with hnRNPD0 knockdown was completely abolished in the presence of C-H1b. However, when combined with C-H1c, hnRNPD0 knockdown had a strongly opposing effect on the enhancing effect of B1 alone in that it represses of B1+C-H1c activity.

Hence, although the luciferase reporter vectors do not represent the full length promoter and could be missing vital regulatory elements in the omitted regions, these data show that the C region could be an allele-specific regulator of the upstream core promoter and this regulatory function is mediated by interaction with various *trans*-acting DNA-binding proteins. These proteins are possibly part of a complex that juxtaposes the C region with the core promoter, thereby manifesting the allele-specific effects of rs242557.

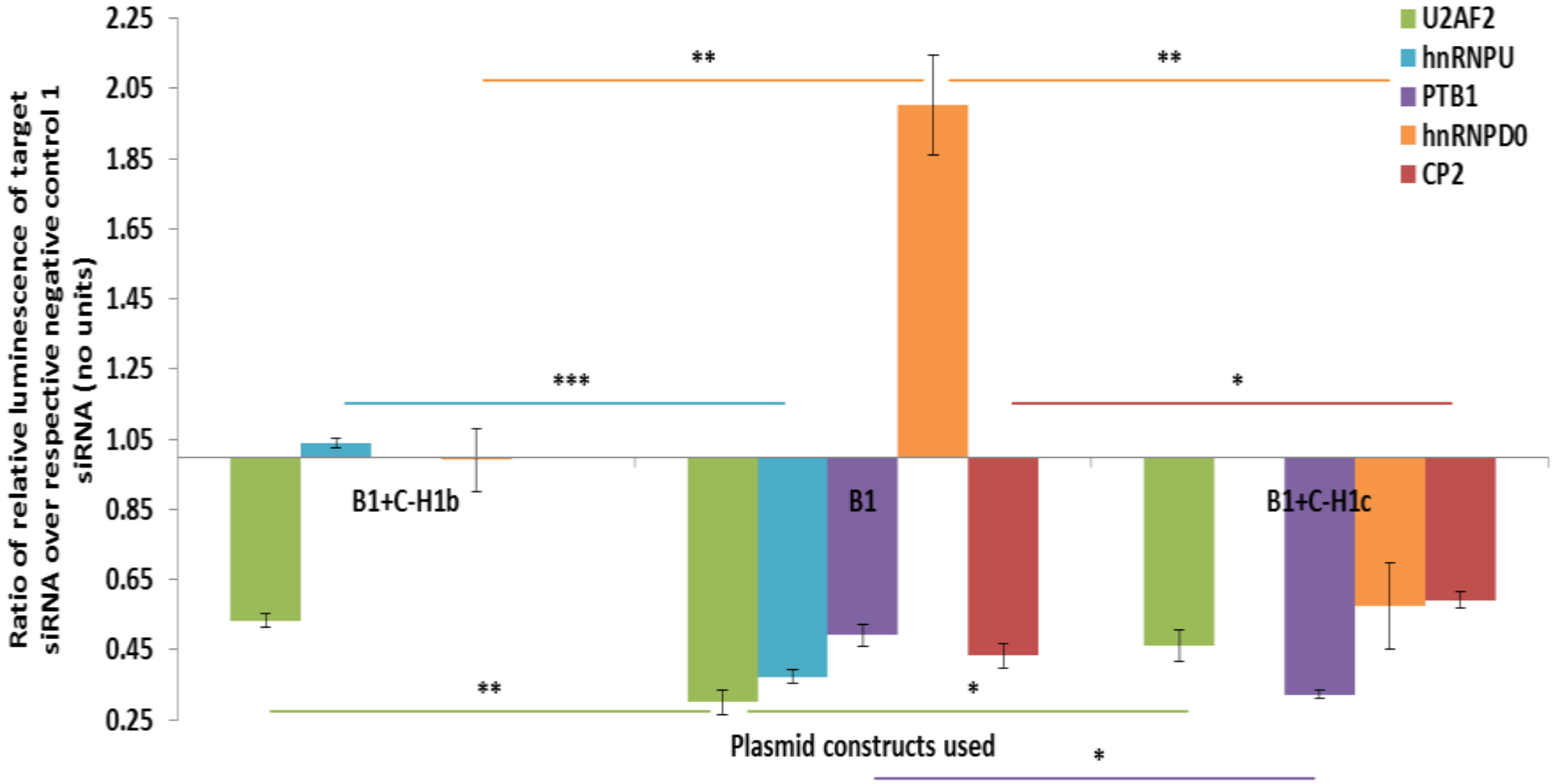


Figure 5.3.4: Effects of shortlisted DNA-binding proteins on the haplotype difference of the downstream promoter region on the core promoter of *MAPT* plasmid constructs.

Only shortlisted proteins that showed significant difference between the different plasmid constructs were included. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$, ***: $p<0.0001$ (two-tailed T-test).

Another region of the promoter highlighted from work in our group as being potentially significant in *MAPT* gene regulation was entitled the D-region (Victoria Kay & Roberto Simone; personal communication). The D-region is located directly upstream to the C-region that contains rs242557. It has a high propensity to form strong secondary structures that might stop the progression of the RNA polymerase activity. The presence of structures was deduced from bioinformatic predictions made by TTS mapping software and from primary sequence analysis. When the D-region was combined with the C-region in the correct context, upstream to the C region, it has a stronger influence on the risk haplotype C-H1c than C-H1b (Figure 5.3.5).

With knockdown of most of the proteins of interest, excepting hnRNPD0 and CP2, the D region had little effect on their previously described effects on C-H1b. It is possible that other trans factors are recruited to the D-region which lead to the reversal of the effect of hnRNPD0 knockdown from mildly enhancing to mildly repressive accounting for D+C-H1b's activity. These factors might also be involved in the increased enhancing effect on D+C-H1b by knockdown of CP2. Interestingly, similar effects that the D-region had on C-H1b were also seen with C-H1c i.e. the role of hnRNPD0 knockdown on D+C-H1c was reversed and the enhancing effect of CP2 knockdown on D+C-H1c was exacerbated.

Unlike the limited effect of the D region on C-H1b variants, addition of the D region in context to C-H1c resulted in responsiveness to knockdown of all the proteins of interest. In the first place, addition of the D-region resulted in reversal of the response seen with knockdown of hnRNPU, hnRNPD0 and PTBP1 on C-H1c were reversed while the effect of U2AF2 knockdown in repressing the activity of D+C-H1c was diminished. It was interesting to note that the allelic effect of the C-region was more pronounced when the D-region was introduced (Figure 5.3.6). This was the first instance where the activity of D+C-H1c was much more muted through the series of increased repressive effects from the proteins of interest hnRNPU, hnRNPD0 and CP2 as well as the smaller role U2AF2 had in stimulating its activity. This was in contrast with D+C-H1b where U2AF2 was the main driver for its function. The role of PTBP1 was the only protein that was not influenced by the haplotype effect as a result of the D-region.

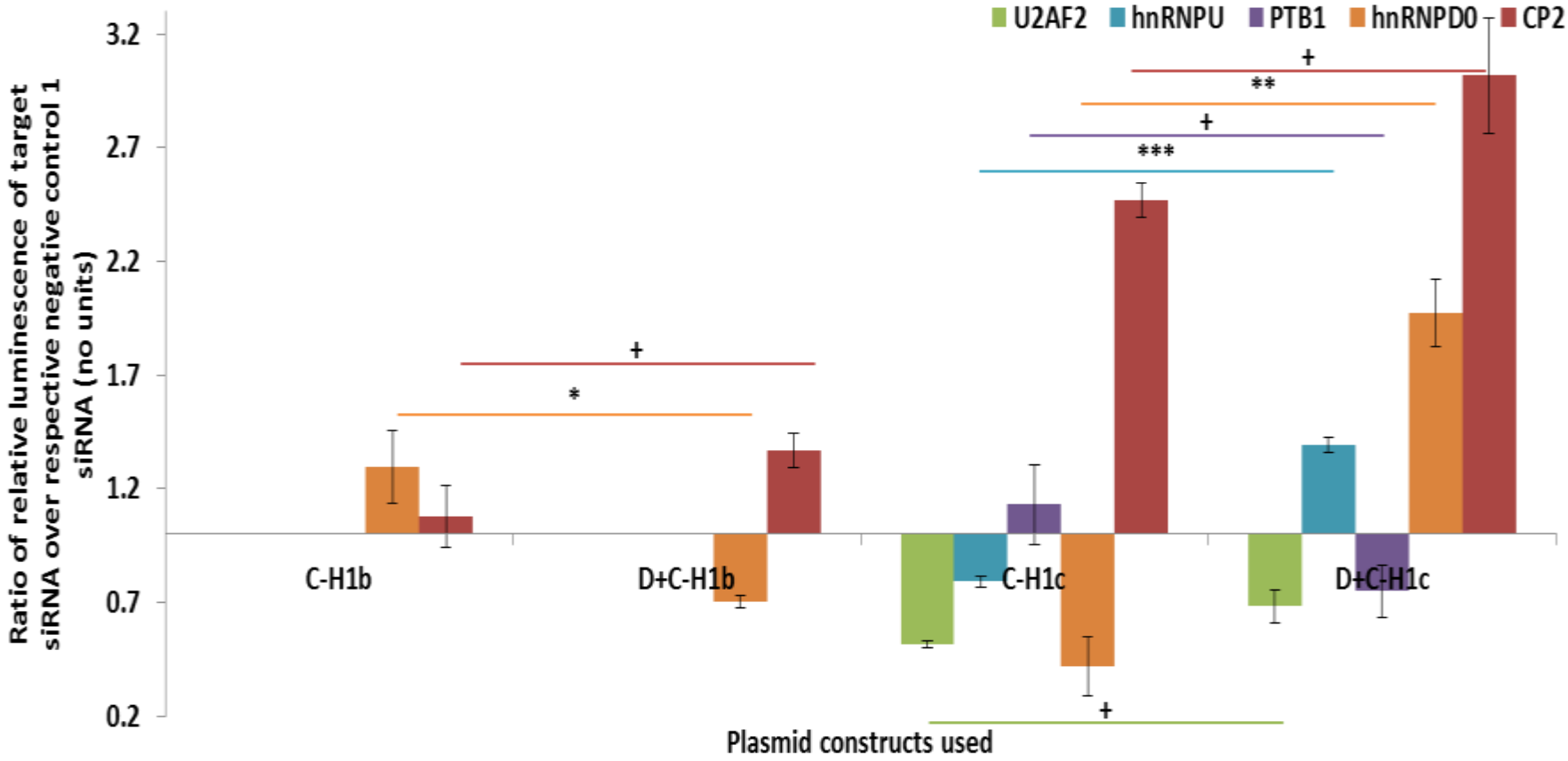


Figure 5.3.5: Effects of shortlisted DNA-binding proteins on the longer downstream promoter region for each downstream haplotype of *MAPT* plasmid constructs.

Only shortlisted proteins that showed significant difference between the different plasmid constructs were included. Each was done in triplicate. Error bars represent standard deviation. +: $p<0.05$, *: $p<0.01$, **: $p<0.001$, ***, $p<0.0001$ (two-tailed T-test).

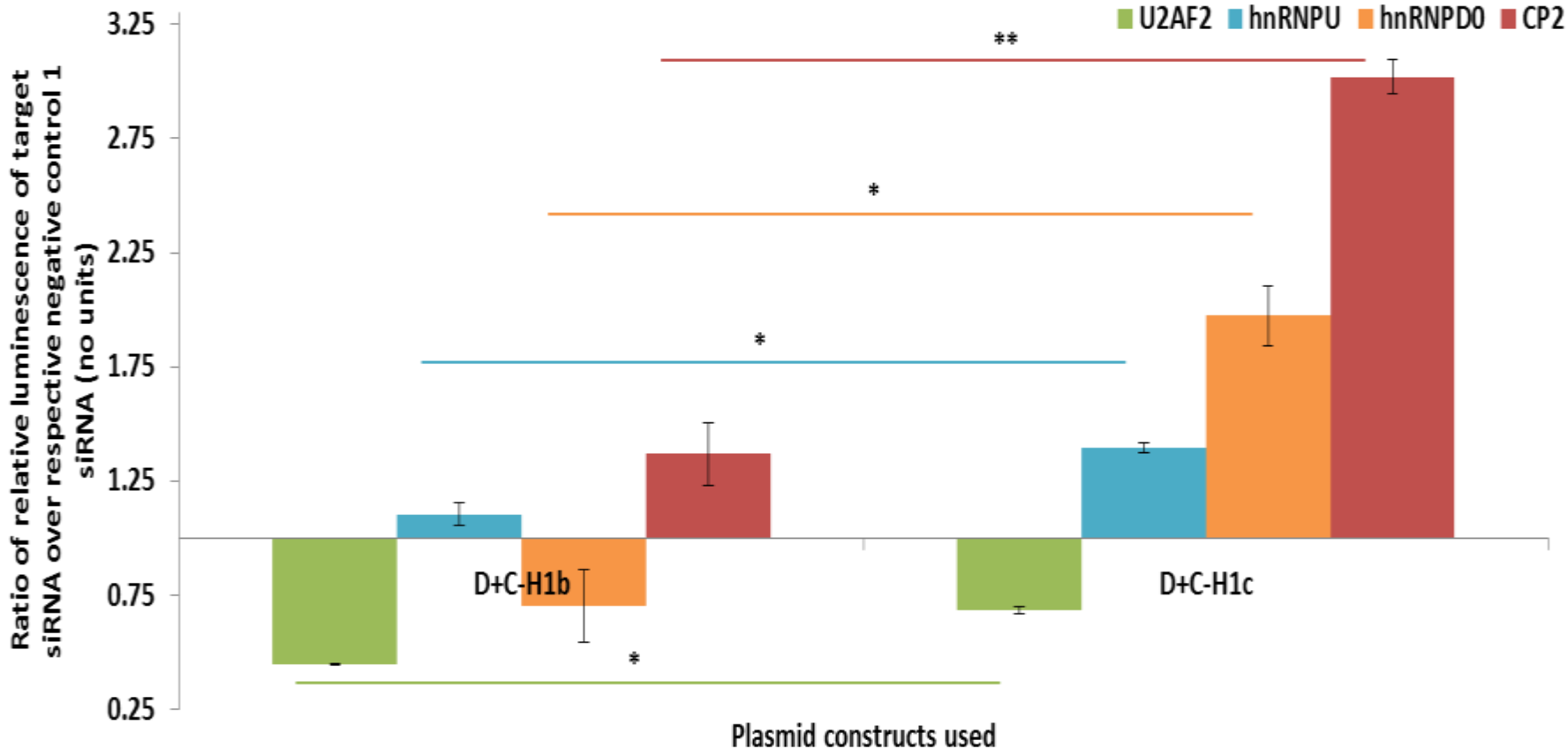


Figure 5.3.6: Effects of shortlisted DNA-binding proteins on the haplotype difference of the longer downstream promoter region of *MAPT* plasmid constructs.

Only shortlisted proteins that showed significant difference between the different plasmid constructs were included. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$ (two-tailed T-test).

The observation that addition of the D-region to the region C-H1c risk allele caused stronger activity enhancement with knockdown of proteins of interest, compared to the D-region in conjunction with C-H1b was also seen when the core promoter B1 region was included (Figure 5.3.7). Although the addition of the D region did not change the repressive effect of U2AF2 on B1+C-H1c, it completely abolished this. The same effect was also seen with PTBP1. While the enhancing effects of hnRNPD0 and TDP43 knockdown on B1 alone were also seen with B1+D+C-H1c as opposed to repression with B1 alone. The effect of D+C-H1b on the B1 core promoter was less straightforward. The neutral H1b haplotype seemed to result in the genetic elements to be more resistant to the effects of the trans factors being investigated. Whether the proteins of interest were demonstrating a positive or negative effect on the construct, the intensity of the effect was reduced. Knockdown of U2AF2 and PTBP1 were necessary for the repression of the construct while knockdown of hnRNPU, hnRNPD0 and CP2 were responsible for the enhancement of the construct.

The inclusion of the D-region has proven to be insightful and promising as without the inclusion of it, a very different discussion and conclusion would have arisen. The D region adds complexity to the genetic regulation of *MAPT*. However, this is not to say that all these elements provide a complete picture as other elements that could be included might change the dynamics again. This stresses the shortcoming of the luciferase reporter constructs used in this study in that they only include small elements of the large *MAPT* promoter and the ultimate importance of including as many functional elements when studying the interaction between genetic elements and trans factors. Due to the large size of the *MAPT* promoter region (>70 kb), this was not possible within the framework of this project.

From Figure 5.3.8, it is clear that the D-region modulates the allelic effects of B1+D+C-H1b and B1+D+C-H1c similar to the effects of the D-region on the C-region alone. However, with the inclusion of the core promoter as well, the modulation that D-region exerted was different, highlighting the possibility of crosstalk amongst the different genetic elements. While CP2 was unaffected, U2AF2 and PTBP1 no longer had any effect on B1+D+C-H1c. Knockdown of hnRNPU and hnRNPD0 led to the enhancement of the construct with the inclusion of the D-region than without. The neutral H1b was less susceptible to the modulation exerted by the D-region. The results from the B1+D+C-H1b

construct showed that U2AF2 was unaffected and the knockdown of PTBP1 was only slightly enhanced. hnRNPU and hnRNPD0 were affected to lesser extents as well. The only notable effect the D-region had on B1+D+C-H1b was that where CP2 used to be important for the activity of the construct, it now had an inhibitory role.

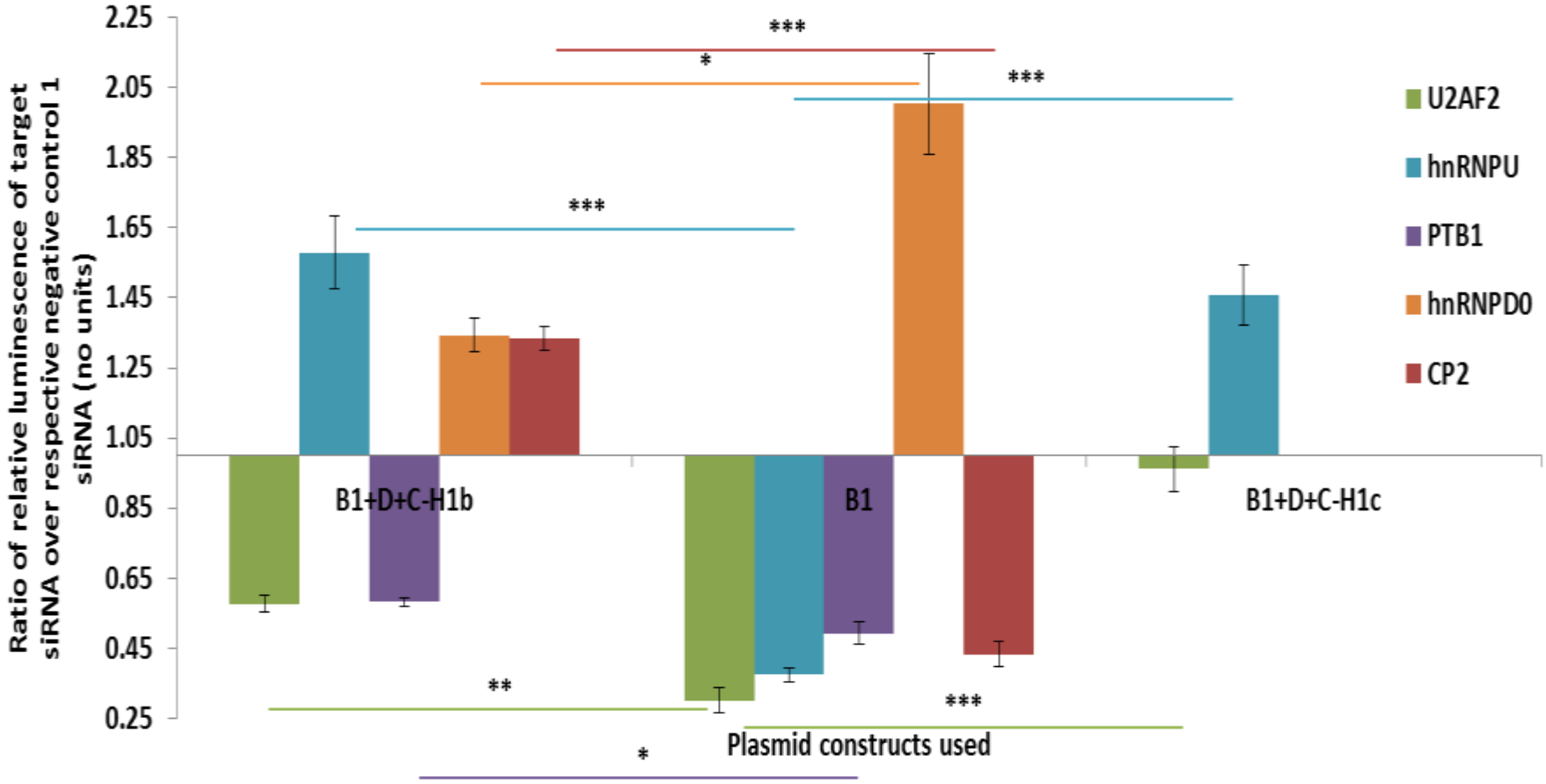


Figure 5.3.7: Effects of shortlisted DNA-binding proteins on the haplotype difference of the longer downstream promoter region on the core promoter of *MAPT* plasmid constructs.

Only shortlisted proteins that showed significant difference between the different plasmid constructs were included. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$, ***, $p<0.0001$ (two-tailed T-test).

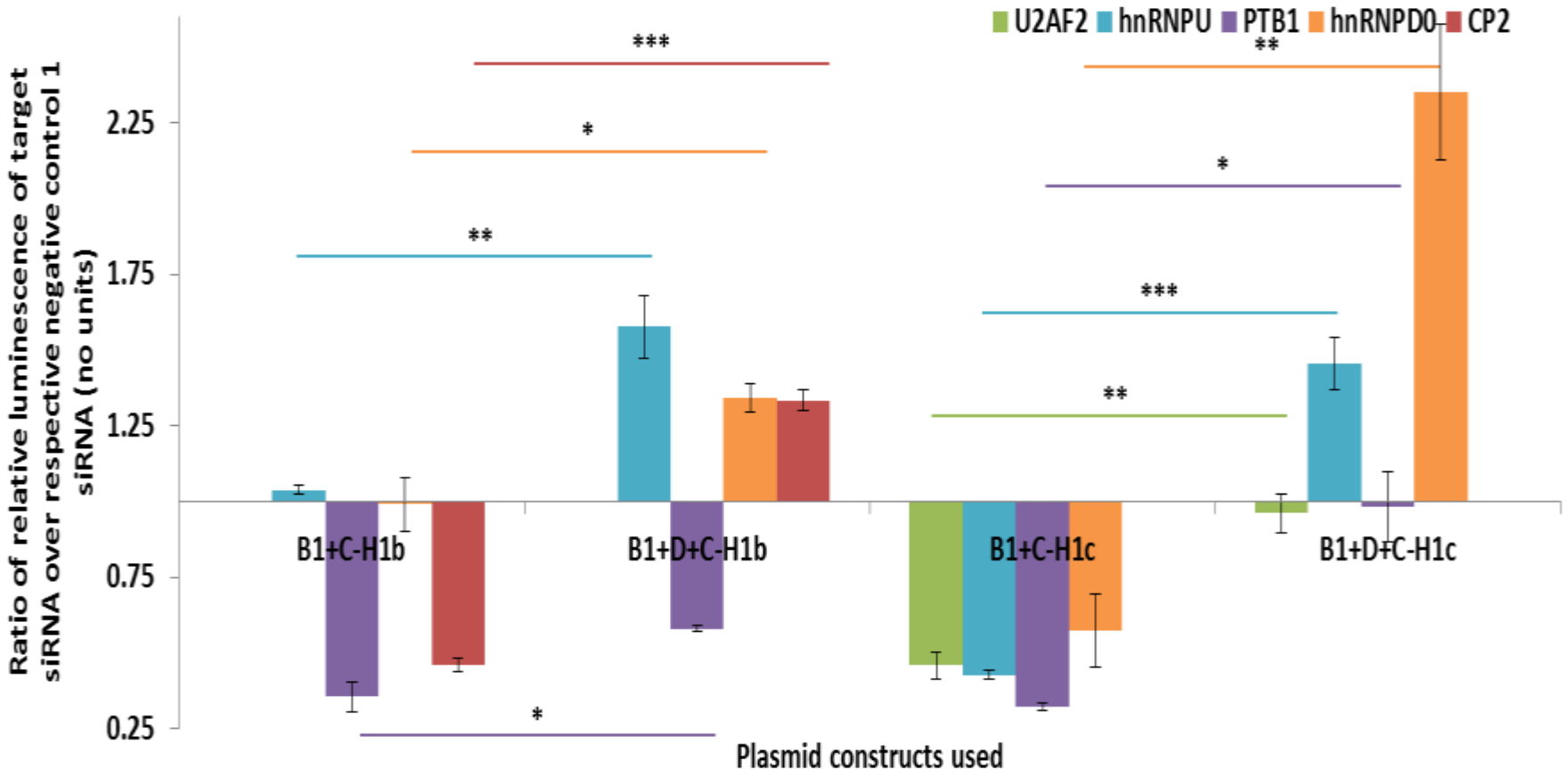


Figure 5.3.8: Effects of shortlisted DNA-binding proteins on the longer downstream promoter region for each haplotype of core promoter and downstream of *MAPT* plasmid constructs.

Only shortlisted proteins that showed significant difference between the different plasmid constructs were included. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$, ***, $p<0.0001$ (two-tailed T-test).

Comparing the allele-specific differences between the non-risk H1b/H2 haplotype with the risk H1c haplotype as in Figure 5.3.9, it would seem that without the D-region (B1+C-H1b and B1+C-H1c); the allele-specific effect was very slight. Knockdown of hnRNPU and hnRNPD0 showed the strongest allele-specific differences with B1+C variants with negligible effect on the H1b allele and repression on the H1c allele. However, addition of the D region caused enhancement for both alleles. In the case of CP2 knockdown, an allele-specific difference was only seen in the presence of the D-region. For the most extensive construct used to study the haplotypic effect i.e. B1+D+C-H1b and B1+D+C-H1c, the haplotype effect was much more substantial. As been previously mentioned, the risk allelotype served to elicit a more repressive state with the knockdown of U2AF2 and PTBP1. This led to an increase in the repressive function of hnRNPD0 and TDP43 which was not seen in any previous analyses. The role of hnRNPU was unchanged between the two allelotypes while the repressive effect CP2 had on the neutral H1b haplotype was removed and played an active role on the risk H1c haplotype.

In summary, the knockdown of U2AF2 and PTBP1 showed the most consistent effects on all ten promoter reporter variants that were studied (Table 5.3.10), supporting their possible importance in regulating the tau promoter. This was especially apparent with their allele-specific effects on B1+D+C. The strong repressive effect of TDP-43 was limited to the core promoter B1 and was not influenced by the other promoter elements other than an allele-specific effect on the B1+D+C. The effects of hnRNPU, hnRNPD0 and CP2 knockdown are less consistent and clear cut. Apart from the promoter only constructs (B1, and B1+B2), these proteins had more variable effects in the presence of the C and D regions. They seem to be much more sensitive to various genetic elements as well as haplotype effects. Besides their dynamic nature, for the majority of the constructs excluding promoter only constructs, they behave in a repressive manner towards them.

This work shows that the regulation of *MAPT* gene regulation is a highly complex process involving the different promoter regulatory motifs. More work still needs to be done to elucidate the extended core promoter B1+B2 effects on that haplotype differences as well as what makes the D-region such an influential genetic element. All the knockdown work was done on individual proteins, a combinatorial approach might highlight the intricacies of the relationship between *MAPT* and its trans factors.

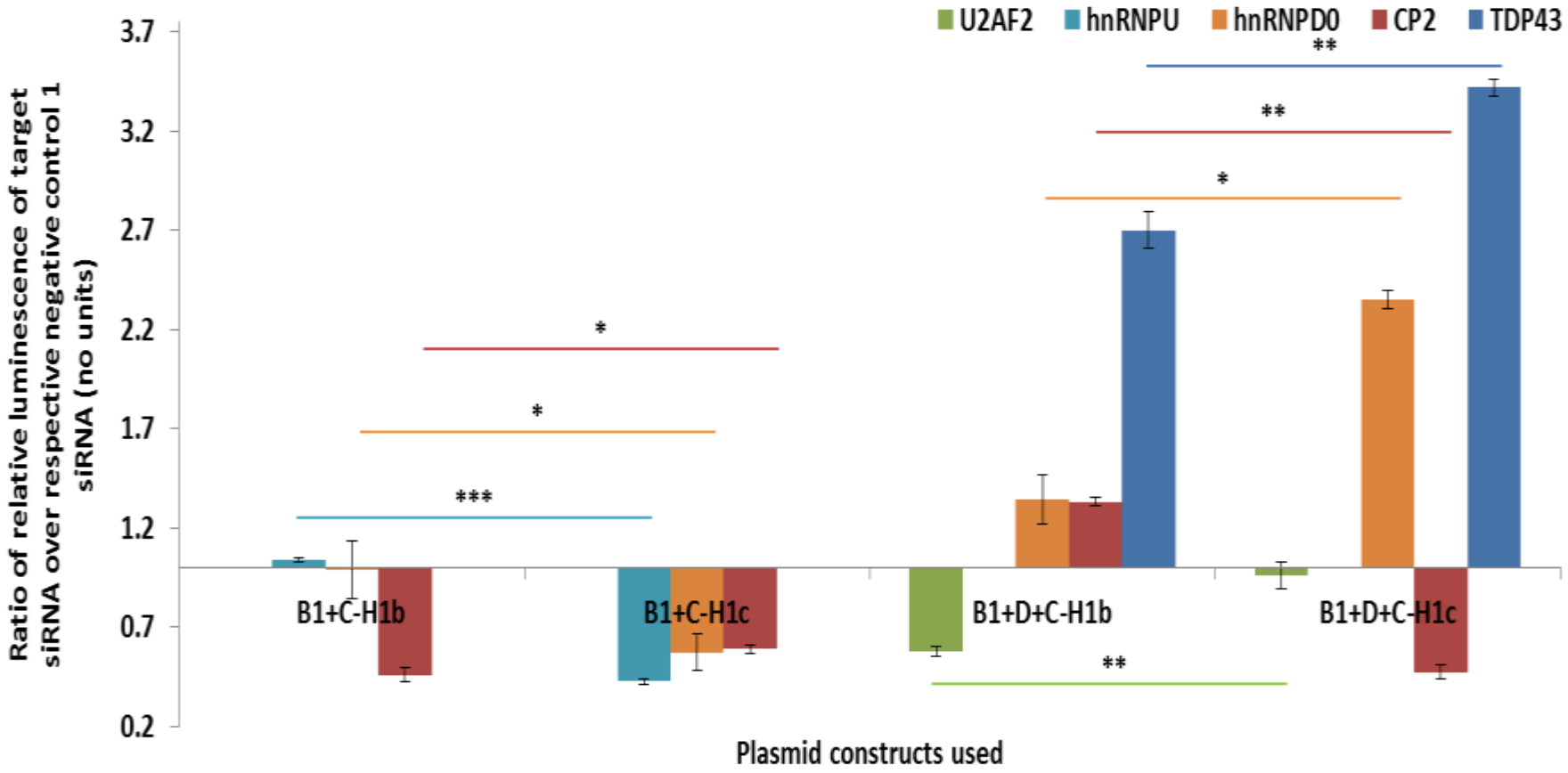


Figure 5.3.9: Effects of shortlisted DNA-binding proteins on the haplotype difference of the downstream promoter regions with the core promoter of *MAPT* plasmid constructs.

Only shortlisted proteins that showed significant difference between the different plasmid constructs were included. Each was done in triplicate. Error bars represent standard deviation. *: $p<0.01$, **: $p<0.001$, ***, $p<0.0001$ (two-tailed T-test).

% Δ in siRNA / neg ctrl 1	B1	B1+B2	C-H1b	C-H1c	B1+C-H1b	B1+C-H1c	B1+D+C-H1b	B1+D+C-H1c	D+C-H1b	D+C-H1c
U2AF2	-70	-70	-57	-48	-47	**-54**	-42	-4	-55	-32
hnRNPU	-63	-58	12	-21	**4**	**-57**	58	46	**10**	39
PTBP1	-51	-64	-40	**13**	-64	-68	-42	-2	-34	-25
hnRNPD0	100	-6	30	-58	**-1**	**-43**	34	135	-30	97
CP2	**-57**	-13	**8**	147	-54	-41	33	-53	37	202
TDP43	224	227	**-2**	**-9**	191	219	170	242	**-11**	-13
PAX8	5	27	**-3**	**27**	**9**	4	-1	29	**3**	-7

Table 5.3.10: Summary of effects of shortlisted DNA-binding proteins on different *MAPT* promoter constructs. Top: Table of all percentage change (% Δ) in the promoter strength of all the *MAPT* plasmid constructs as a result of the siRNA knockdown of the protein of interest relative to their respective negative control 1 siRNA co-transfection. Bottom: Table of all percentage change (% Δ) of the function of the protein of interest had on all the *MAPT* plasmid constructs. Red – Most negative % Δ. Yellow – Moderate negative % Δ. Green – Moderate positive % Δ. Blue – Most positive % Δ for each column. Numbers centered showed non-significance with another value in the same column. Numbers centered and bold showed non-significance with more than one value in the same column. Pairs of adjacent values in a row that had been boxed together showed significance difference with each other

U2AF2 and PTBP1 have been shown previously to be involved with *MAPT* exon 6 and its splicing.[270] They were both shown to act on the branch point / polypyrimidine tract region, overlapping but distinct sites. The effect was seen in both COS cells and SKN cells. PTBP1 is potent in preventing the inclusion of exon 6 while U2AF2 favours the inclusion of exon 6, seemingly by favouring the strongest branch point when the 3' splice site was present. Another isoform of PTBP1, PTBP2 or neuronal PTBP (nPTBP), was also shown to exclude exon 6.[271] However, since nPTBP was not seen in our protein identification, it was not pursued. In our study, we also did see a close relationship between U2AF2 and PTBP1, probably due to overlapping and distinct sites as well. A tau expression and / or splicing assay would be needed to investigate whether they have opposing effects and what their downstream effects are. Work on the splicing of exon 10 showed that U2AF2 and PTBP1 both strongly inhibit exon 10 splicing.[272] Hence, depending on their location and sites of interaction, their effects and outcomes could be very different. Since they are splicing proteins, they might be recruited by RNA polymerase II at the promoter region and follow the course of transcription.

Another group of proteins that have been shown to interact with *MAPT* is the hnRNP family of proteins. Although hnRNPU and hnRNPD0 have not been specifically shown to associate with the tau gene, other members have been shown to interact. hnRNPG was shown to inhibit *MAPT* exon 10 splicing, most likely in concert with other trans factors as a complex with SR proteins.[273] Using tau exon 10 as bait in a yeast three-hybrid screen, hnRNPE2 was found to interact with it.[274] It is interesting to note that the screen also picked up 60S ribosomal protein L23 which was also found in our protein pull-down interacting with rs242557. Co-transfection assays later showed that hnRNPE2 moderately activates exon 10 splicing. Through similar experiments, hnRNPE3 was also found to modestly activate exon 10 splicing.[275] Lastly, hnRNPK could post-transcriptionally regulate the tau mRNP complex by affecting its nuclear export and translation but not transcription.[276] It is unsurprising then comparing these results with our own that hnRNPU and hnRNPD0 were able to give such a varied response depending on the constructs that were used. Since members of the hnRNP family can interact with each other and exert their influence in the nucleus as well as in the cytoplasm, more detailed experiments would be needed to ascertain their full nature with regards to the tau gene and the co-regulation of transcription and splicing.

Although no previous work in the literature showed that the transcription factor CP2 interacted with the tau gene, it had been predicted to bind the rs242557 region by other groups and studied intensively within the field of Alzheimer's disease. For example, it was shown that the intracellular domain (ICD) of amyloid precursor like protein 2 (APLP2) interacted with CP2 in the nucleus and induced the expression of GSK-3β.[277] The identification of CP2 with the tau gene might open up a new range of upstream signalling and interactions that might subsequently impinge on the genetic regulation of *MAPT* and coordination of this disease-associated gene with other disease-associated genes and pathways. Furthermore, TDP-43, the protein of a disease-associated gene which was also found in the inclusions of FTLD-U, was found to be highly involved in the state of the core promoter of *MAPT*. Again, more work will be needed to study the various trans factors regulating tau gene expression.

6 Conclusions and Future Work

6.1 Conclusions

The tau protein is a dynamic architectural protein that is involved in various cellular processes. In neurology, the tau protein and the tau gene have been implicated in causing or playing a role in various neurological disorders. The H1c sub-haplotype of the tau gene, *MAPT*, drives the association of increased risk in PSP and CBD and is also weakly associated with AD. rs242557 is a H1c-specific SNP in a highly conserved C-region downstream of the *MAPT* core promoter and is associated with significantly higher transcription and elevated levels of mRNAs containing the alternatively spliced exon 10. It is one of nine H1c-specific SNPs. This PhD set out to elucidate how rs242557 might affect transcription by identifying DNA-binding proteins that bind to it and their effects on the C region and the core promoter.

Nuclear extracts from SH-SY5Y neuroblastoma cell line were used to carry out EMSA. This experiment showed that the H1b/H2-G allelotype of rs242557 was able to recruit more DNA-binding proteins as compared to the A allelotype. Concatamers of rs242557-containing oligonucleotides were used to purify the DNA-binding proteins and separated on a 1-D SDS PAGE gel. The bands were then identified using mass spectrometry. 36 unique proteins were identified and these were stratified using bioinformatics software such as STRING and Reactome. An additional protein predicted by MatInspector and Transfac to bind to the rs242557 region was also included. Hence, 14 promising DNA-binding protein candidates were pursued further.

These proteins were validated using antibodies raised specific to those proteins. The antibodies were used to identify the DNA-binding proteins from the pull-down experiments as well as in native blots. The pull-down experiments worked much better than the native blots. As a result, five DNA-binding proteins of interest (U2AF2, hnRNPU, PTBP1, hnRNPD0 and CP2) were progressed to the co-transfection luciferase assays to begin to understand the trans effects of these proteins on genetic regulatory regions of *MAPT*. TDP43 was also included in the co-transfection assays through its growing importance as a

neurologically relevant protein and gene as well as personal communication with other members of the lab.

Knock down of U2AF2 and PTBP1 on the B1+D+C-H1b demonstrated that they have a role in C-H1b transcription but they were not observed to participate in the transcription of B1+D+C-H1c. This allele specificity was not observed from the knockdown of hnRNPU and hnRNPD0, both of which enhanced the transcription of B1+D+C-H1b&c. Knock down of TDP43, too, did not show any allele specificity but it did show core promoter specificity, indicating that TDP43's role is intimately linked with the core promoter. Lastly, the only other protein that was knocked down but was not identified through mass spectrometry was CP2. It was a transcription factor predicted to interact with rs242557. Knock down of CP2 enhanced the transcription of B1+D+C-H1b but repressed the transcription of B1+D+C-H1c. Hence, the three proteins that demonstrated rs242557 allele specificity were the splicing factor U2AF2, the splicing repressor PTBP1 and the transcription factor CP2.

Understanding how genes function requires an appreciation of the dynamic nuclear architecture (Figure 6.1.1).[278] Studies on various animal models have using next generation sequencing after micrococcal nuclease digestion consistently revealed that nucleosomes occupy exons 1.5-fold more than introns.[279] In humans and mice, exons were also found to be enriched in post-translational histone modifications.[280] Hence, these findings suggest the plausibility of an RNAPII mediated cross-talk between chromatin and the exon-intron architecture (Figure 6.1.2). The results discussed so far are promising in highlighting the relationship between the risk associations of a SNP with neurological diseases and how a SNP influences the biology and subsequently the pathology of a neurological disease. Furthermore, the nucleus is a three-dimensional structure and is filled with chromosomes and proteins, interacting in three dimensions.[281] From the co-transfection luciferase assays, the inclusion of the core promoter had modulated the knockdown results of the C-region alone or D+C regions. This might suggest that in the *in vivo* context, a complex or various complexes or chromatin modifications could account for this interaction. hnRNPU and hnRNPD0 are versatile proteins that are able to bind to DNA sequences as well as RNA such A+U-rich element in the 3'-UTR of mRNA and other proteins. This makes them highly likely candidate proteins to bridge RNAPII function with chromatin dynamics.

Specifically, hnRNPU had already been shown to be involved with the CTD of RNAPII as well as complexes involved in histone acetylation.[221,223,224]

U2AF2 and PTBP1, besides already been shown to be involved in the splicing of *MAPT* exon 6, have also been studied together with other genes.[185,186,366] From our results, more is known about how the splicing of *MAPT* is coordinated by these splicing proteins. Although direct binding of these proteins to the C region containing rs242557 through ChIP had not been completed, all our results point towards U2AF2 and PTBP1 being functionally involved in an allele-specific manner. How this interaction with rs242557 ties in with their function with the splicing of exon 6 or other splicing events within *MAPT* will require further investigation.

The transcription factor CP2 is a protein that was predicted to bind to the rs242557 region. It might have been pulled down by rs242557 as a band similar to its molecular weight was seen on a SDS-PAGE gel. The mass spectrometer was not able to identify that gel band, possibly due to degraded peptides or insufficient peptides. However, with the blots of the native gels, CP2 came across as binding to the rs242557 probe in a highly specific manner as no band was seen with one of the random probes. Moreover, the binding was allele-specific. While a lot of study involving CP2 protein had been done in Alzheimer's disease, they revolved around the amyloid pathway. Although other groups have postulated that CP2 would bind to the C region, our work has been the first in progressing that further. The allele-specific functionality surrounding CP2 binding warrants further study into this transcription factor, especially in light of the allele-specific binding of U2AF2 and PTBP1 as well.

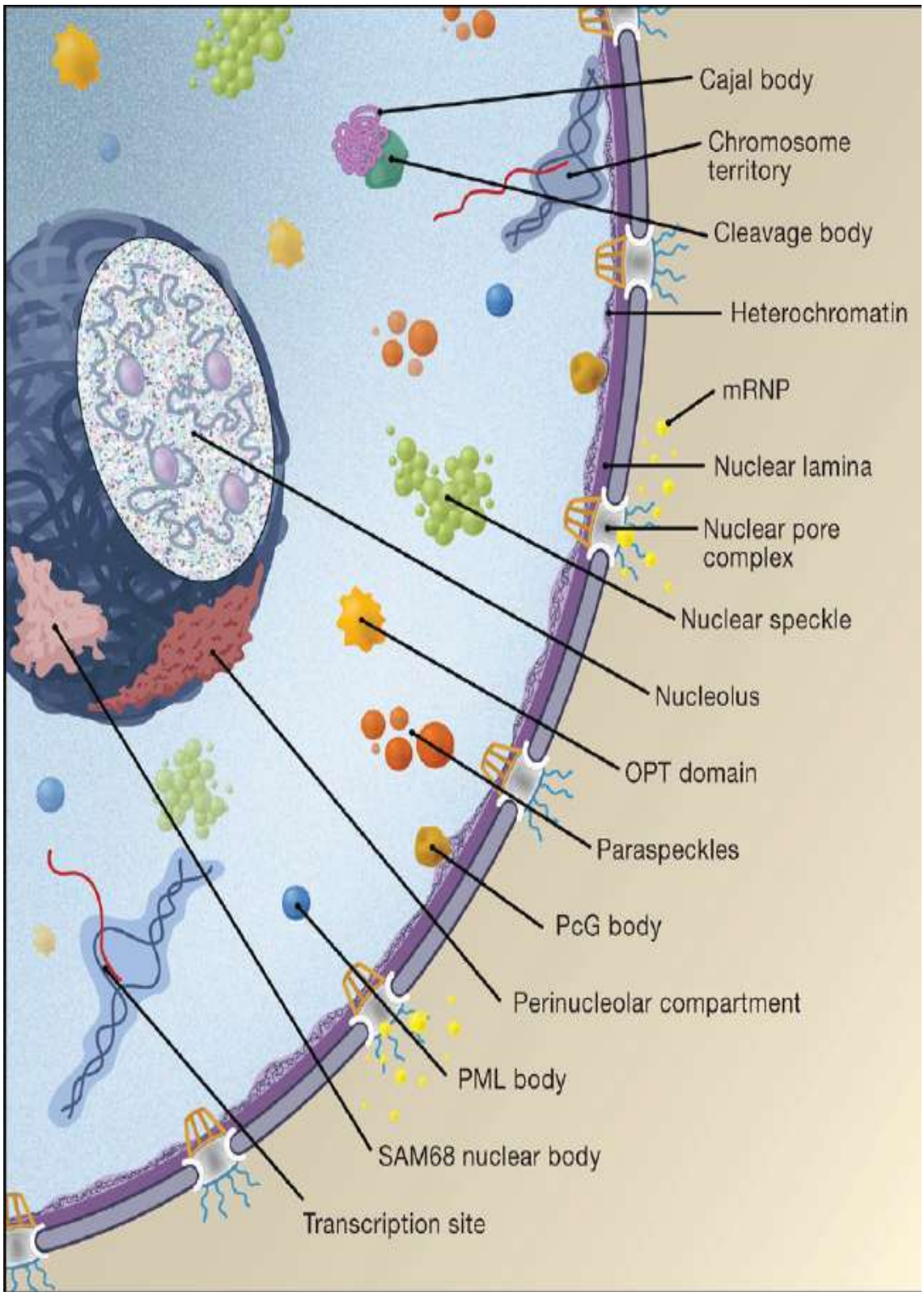


Figure 6.1.1: Chromosomal territories and nuclear bodies present in the nucleus. Modified from [374].

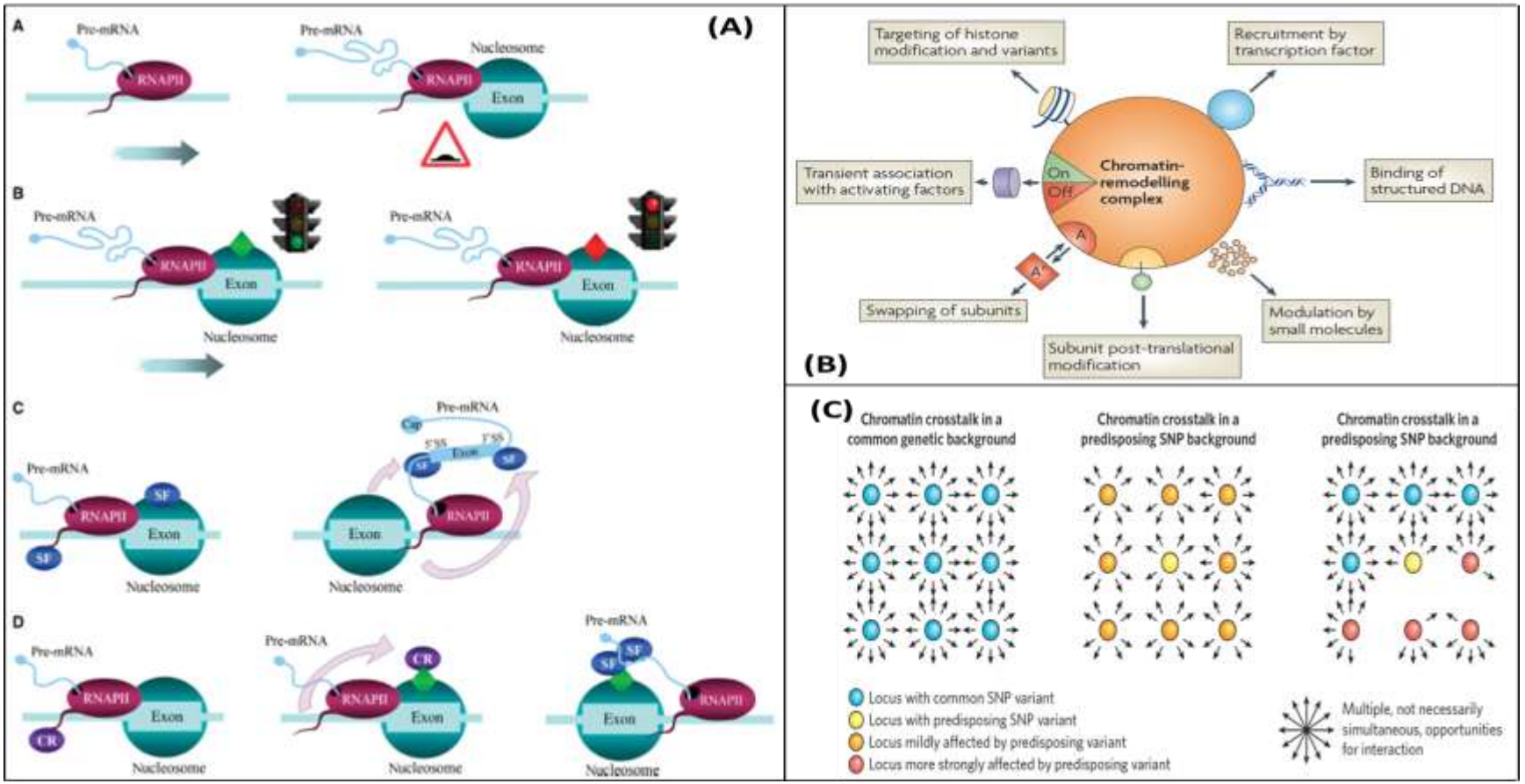


Figure 6.1.2: The interplay of chromatin with SNPs, transcription and splicing. (A) Mechanisms of how nucleosomes may add to the complexity of the exon-intron architecture: slowing transcription sterically, **A**, or through histone modifications, **B**, or the CTD domain of RNAPII and nucleosome interacting with splicing factors, **C**, or chromatin remodelers, **D**. (B) Chromatin remodeling is an intricate process with various entities attenuating the process. (C) SNPs may influence chromatin crosstalk through chromosome loops and or bridges. Modified from[375,377,282]

6.2 Future work

All the work carried out in this PhD had been done on undifferentiated SH-SY5Y neuroblastoma cells. Although these cells in their undifferentiated form do retain certain neuronal characteristics such as dopamine-β-hydroxylase activity and converting glutamate to GABA, their morphology does not resemble neurons. To fully differentiate these cells, they would be treated with retinoic acid before treating them with brain derived neurotrophic factor (BDNF). This process however is expensive due to the BDNF, time consuming and leaves with a very low yield of fully differentiated SH-SY5Y cells. For the purpose of this study, it would have been impractical. However, retinoic acid is relatively inexpensive compared to BDNF and the process of partially differentiating these cells with retinoic acid is much speedier than the full differentiation process. There is also a greater yield of partially differentiated cells.

Retinoic acid, a vitamin A metabolite and a lipohilic signaling molecule, binds to its receptor and retinoid X receptor to regulate transcription.[283] Treatment of SH-SY5Y cells with retinoic acid had been shown to induce the expression of the neurotrophin receptor TrkB and that it was functional.[284] In our own experiment, a 1.5-fold increase in the expression of TrkB was seen in both the nuclear and cytoplasmic extracts after treatment with retinoic acid (Figures 6.2.1 and 6.2.2). Although it was known that undifferentiated SH-SY5Y cells produced significant levels of neurotransmitters such as dopamine, acetylcholine and GABA, how these would be affected upon retinoic treatment was not known. Here, our treatment of cells with retinoic acid led to an approximately four- and two-fold increase in dopamine-β-hydroxylase in the nuclear and cytoplasmic fractions respectively. Vasoactive intestinal polypeptide (VIP), a cholinergic marker had been previously shown to be upregulated in retinoic acid-treated SH-SY5Y cells especially in the neuritic processes through immunohistochemistry.[285] In our experiment, nuclear VIP was significantly decreased with a concomitant increase in cytoplasmic VIP reflecting that immunohistochemistry result. The oncoprotein Bcl-2 was also seen to be moderately increased upon differentiation with retinoic acid in the nuclear extracts. This was confirmed by previous published work from other groups as a key marker of SH-SY5Y differentiation.[286,287] A key architectural protein in neuronal cells is neurofilament-M

(NFM). From figures 6.2.1 and 6.2.2, the expression of this protein was induced further with differentiation especially in the cytoplasmic fractions with a 1.5-fold increase.

Besides analyzing neuron-specific cytoplasmic markers, the methyl CpG transcription factor MeCP2 was also investigated. Besides the increase in expression with retinoic acid treatment, MeCP2 had been shown to indirectly control BDNF expression.[288,289] This protein like Bcl-2 was seen to be moderately upregulated in the nuclear extract of retinoic acid-differentiated cells in our experiment. As MeCP2 is a transcription factor, a slight increase might be sufficient to cause significant changes in its gene targets. Finally, as this PhD revolves around the tau gene *MAPT*, it was also important to look at the expression of the tau protein before and after retinoic acid treatment. It had been previously shown by various groups that upon differentiation of SH-SY5Y cells, levels of tau and its phosphorylation state had increased, in part because of the increased expression of GSK-3β.[290,291,292] In our experiment however, we did not see this effect. In fact, the levels of tau were observed to decrease in both the nuclear and cytoplasmic fractions. It could be that other isoforms of tau was expressed but was undetected.

As controls, the neuron-specific enolase, GAPDH, actin and MCM2 were used (Figure 6.2.1 continued). No change in protein levels was detected either in the nuclear or cytoplasmic fractions after treatment with retinoic acid. However, the other nuclear control protein, HDAC1, was upregulated upon differentiation. Hence, it would be important in going forward to put the results of this PhD to the test with a variety of cell culture models, either retinoic-acid differentiated neuroblastoma cells or primary neuronal cell cultures.

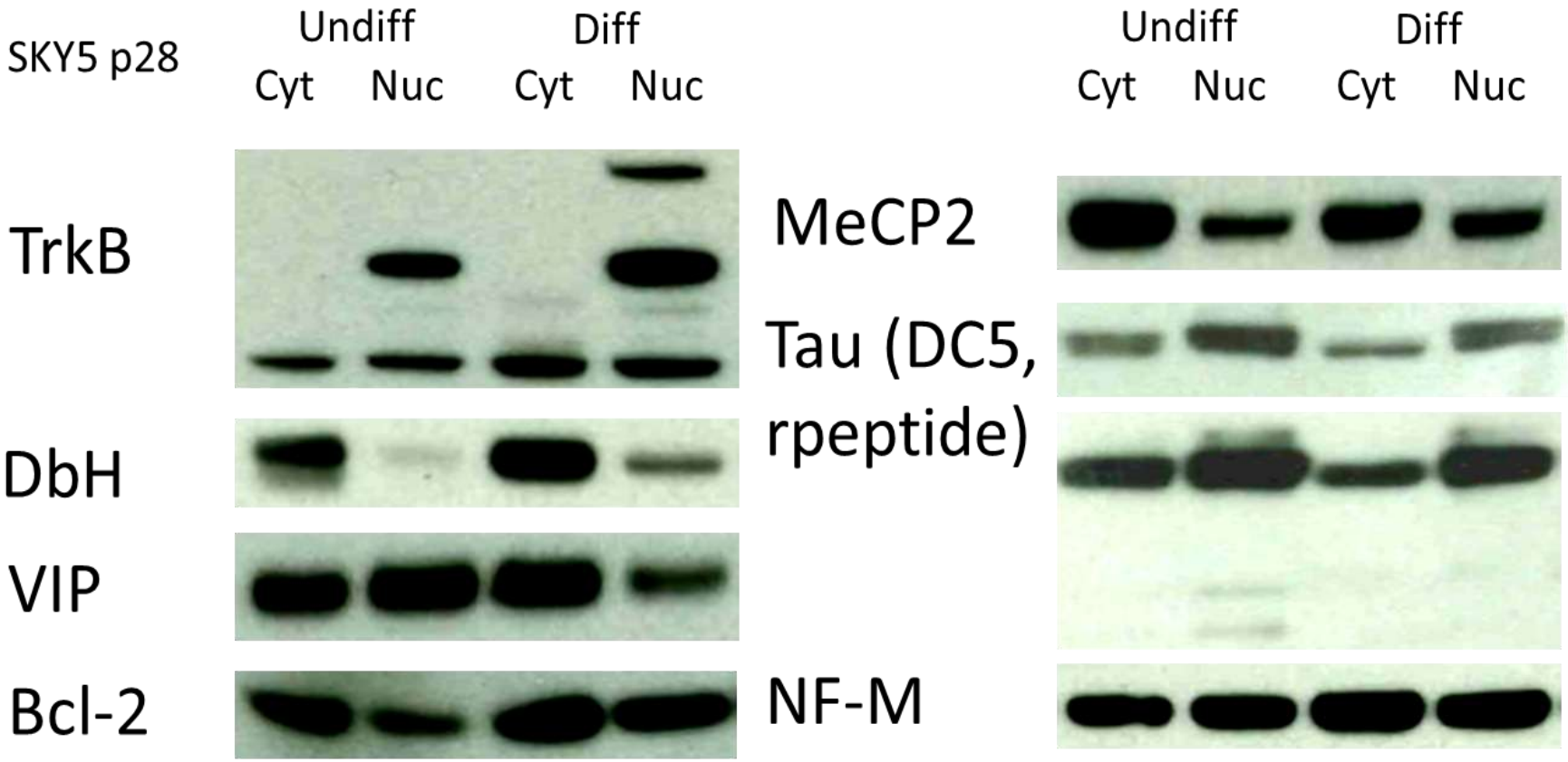


Figure 6.2.1: Western blot analysis of neuronal specific differentiation protein markers of SH-SY5Y cells treated with retinoic acid. SH-SY5Y neuroblastoma cells were treated with retinoic acid for a week before cytoplasmic and nuclear fractions were extracted (Diff Cyt and Diff Nuc) and analysed. The same cells with the same passage number were left untreated before being extracted (Undiff Cyt and Undiff Nuc) and analysed too. N =3

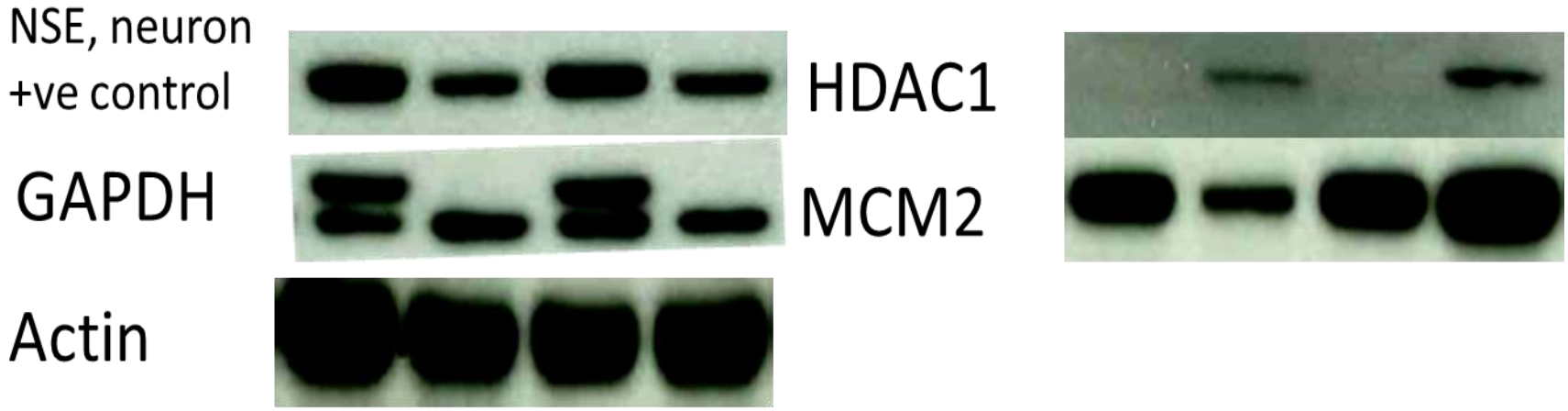


Figure 6.2.1 continued: Western blot analysis of neuronal specific differentiation protein markers of SH-SY5Y cells treated with retinoic acid. SH-SY5Y neuroblastoma cells were treated with retinoic acid for a week before cytoplasmic and nuclear fractions were extracted (Diff Cyt and Diff Nuc) and analysed. The same cells with the same passage number were left untreated before being extracted (Undiff Cyt and Undiff Nuc) and analysed too. N = 3

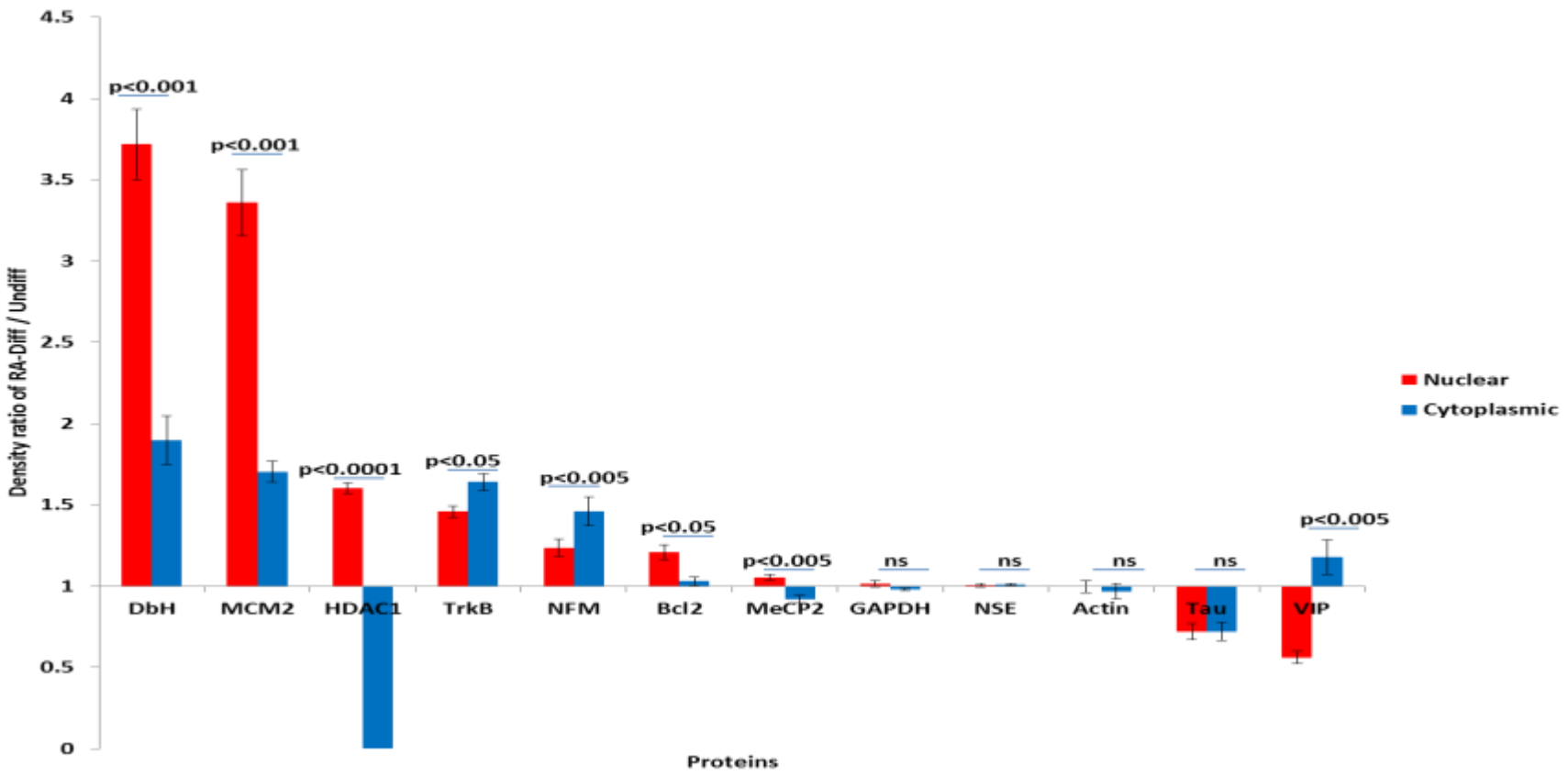


Figure 6.2.2: Plot of the blot density of proteins from retinoic acid differentiated fractions relative to the blot density of proteins from undifferentiated fractions.

Red bars represent the blot density of proteins from the nuclear extracts of retinoic acid differentiated cells relative to the blot density of the same proteins from nuclear extracts of undifferentiated cells. Blue bars represent similar comparisons of the same proteins found in the cytoplasmic fractions.

Another key experiment that was started during my PhD but was not successfully completed during the period was chromatin immunoprecipitation (ChIP). As this was a technique new to the laboratory, a lot of optimisation needed to be done. It was a bold step to start off with ChIP of frozen human brain tissue. After many months, this route did not prove successful, possibly due the amount of cells needed and the condition of the samples. Another attempt was pursued with SH-SY5Y cells and various ChIP protocols were tested. However, because of restraints on time and resources, the ChIP results shown in Figures 6.2.3 and 6.2.4 were not repeated and should be seen as part of the further work that needs to be undertaken.

The ChIP experiment was designed to include various controls. Besides the input control where 10% of the sheared genomic DNA was used to show selective enrichment, there was also an antibody negative control consisting of IgG and a positive antibody control against RNAPII and looking at its binding to the GAPDH promoter. These were the controls suggested by the ChIP protocol. On top of this, we designed primers for regions directly upstream and downstream of the rs242557 region, primers for the B region as well as a predicted Sp1-binding consensus site. These additional primers designed were tested on genomic DNA so that they only amplified unique regions of comparable sizes (Figure 6.2.3 genomic DNA).

Although quantitative PCR had not been done and no replicates were done either, early indications show that there is potential for the proteins of interest U2AF2, hnRNPU, PTBP1, hnRNPD0 and CP2 to bind to the rs242557 region. These binding seem to be stronger than the positive control of RNAPII binding to the GAPDH promoter. Furthermore, there seem to be modest binding of U2AF2 and PTBP1 to the core promoter B region while CP2 seem to bind to the B region as much as it does to the rs242557 region. Repeat experiments, however, need to be done to see if these indications of binding hold up.

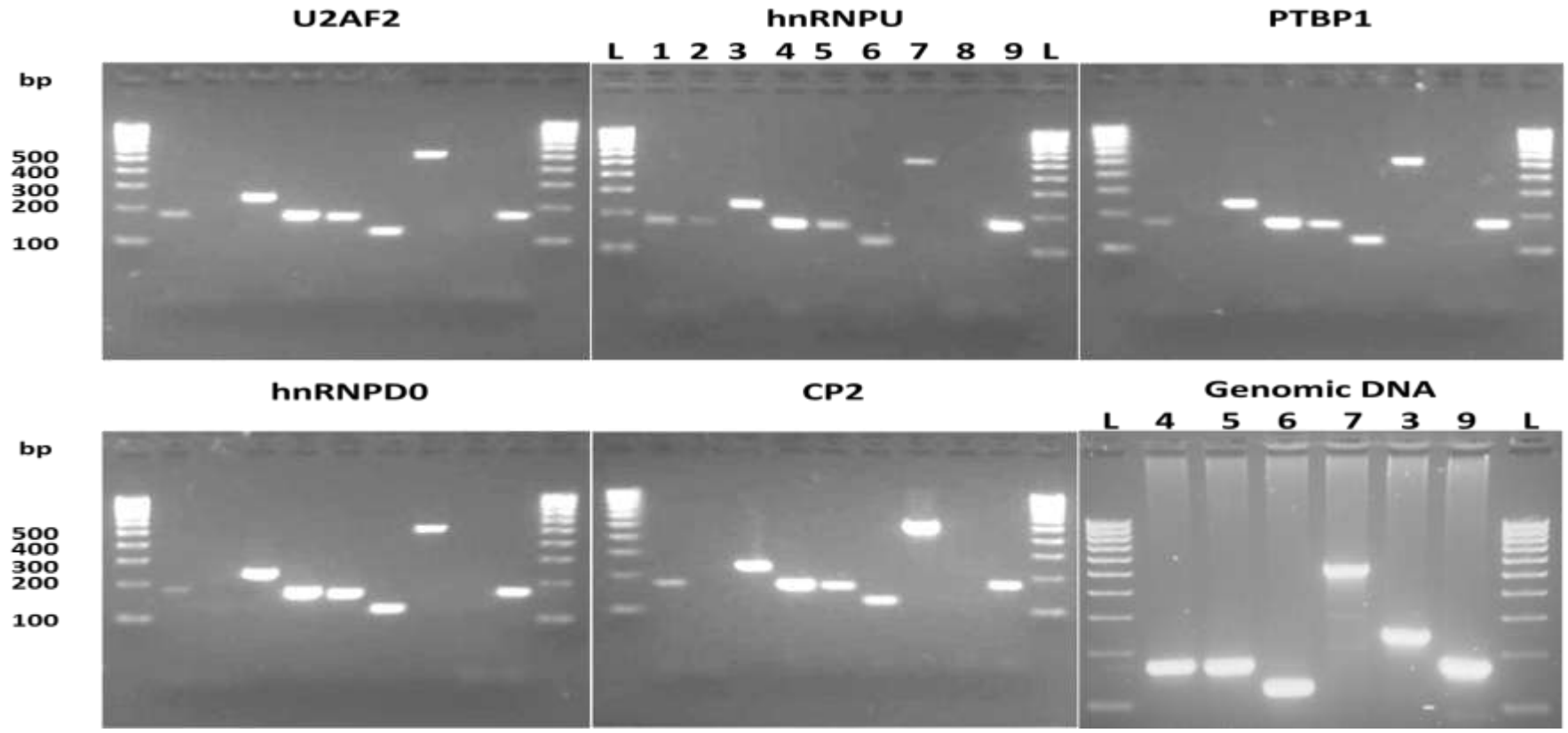


Figure 6.2.3: Agarose separation of PCR products of ChIP from proteins of interest on different regions of *MAPT*.

Genomic DNA and proteins were crosslinked *in vivo* with formaldehyde, extracted and sheared enzymatically. Antibodies targeting proteins of interest were incubated with the sheared DNA-protein complex. The enriched DNA fragments were purified and identified using PCR. L: 1kb DNA ladder, 1: Input DNA, 2: IgG negative control antibody, 3: Sp1 predicted site, 4: rs242557, 5: region directly upstream of rs242557, 6: region directly downstream of rs242557, 7: B region, 8: GAPDH negative control, 9: RNAPII-GAPDH positive control. Genomic DNA panel shows that designed primers work accordingly.

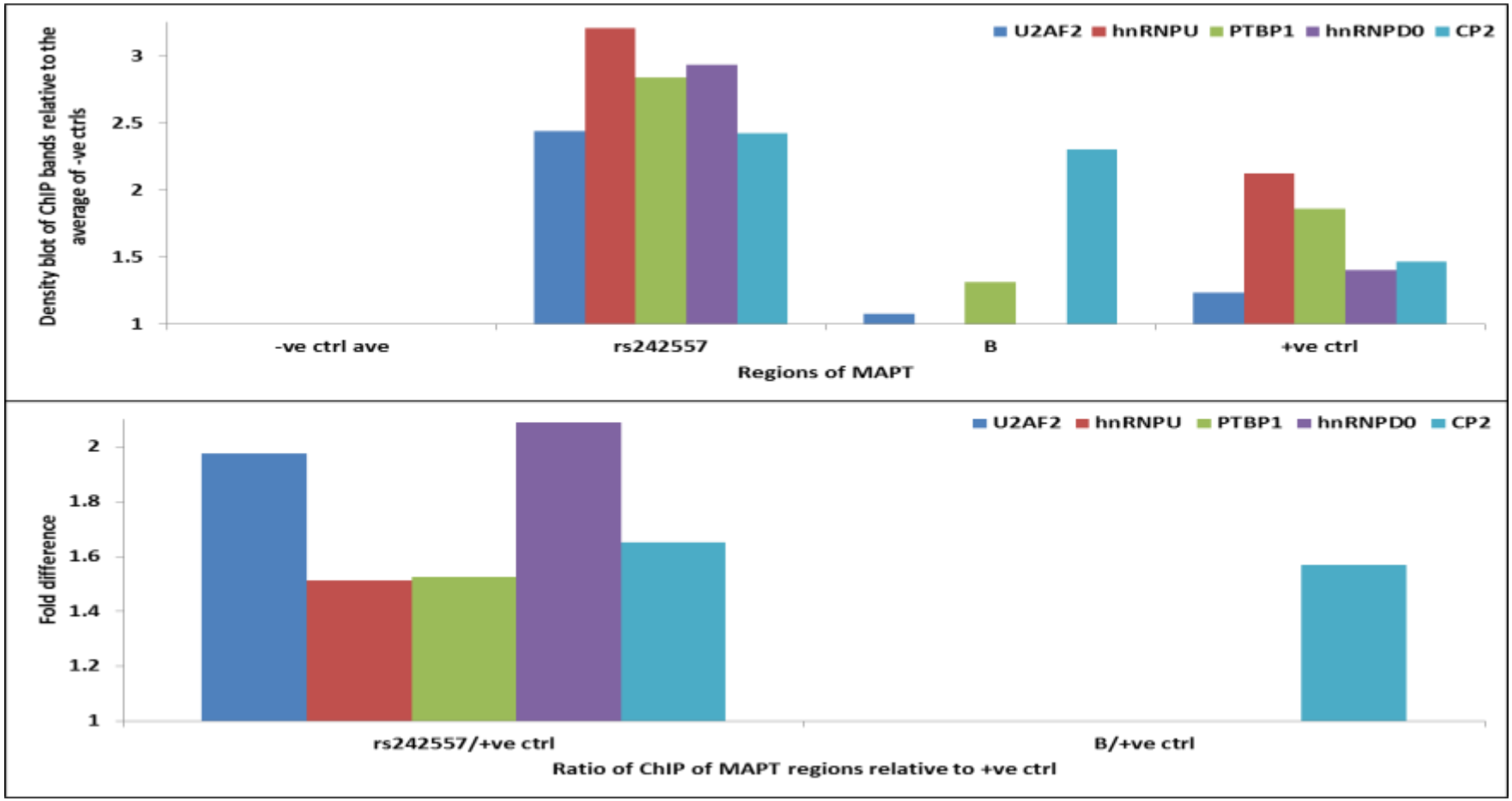


Figure 6.2.4: Analysis of ChIP from proteins of interest on different regions of *MAPT*.

Top panel: The density blot for each ChIP band was measured and was plotted relative to the average density of the stringent negative controls of directly upstream and downstream of rs242557. No repeats were done hence no error bars. Bottom panel: The ChIP values were expressed relative to the positive control of RNAPII antibody-GAPDH PCR values.

7 Appendix

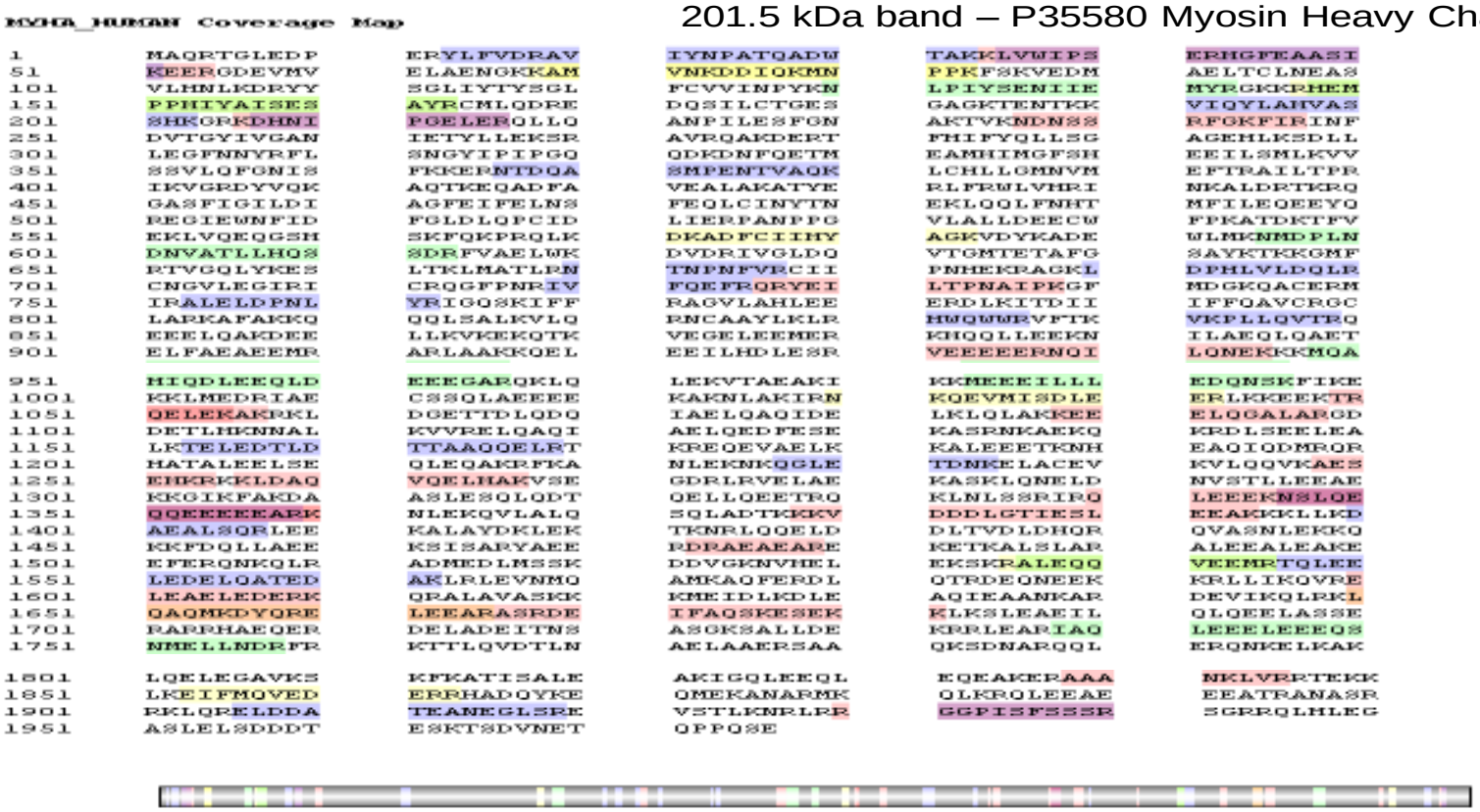


Figure 3.4.2: Coverage of MS peptide data generated for 201.5 kDa band. 27 Blue – matched to a peptide, 14 red – matched to a partial peptide, 7 green – matched to a modified peptide, 7 yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

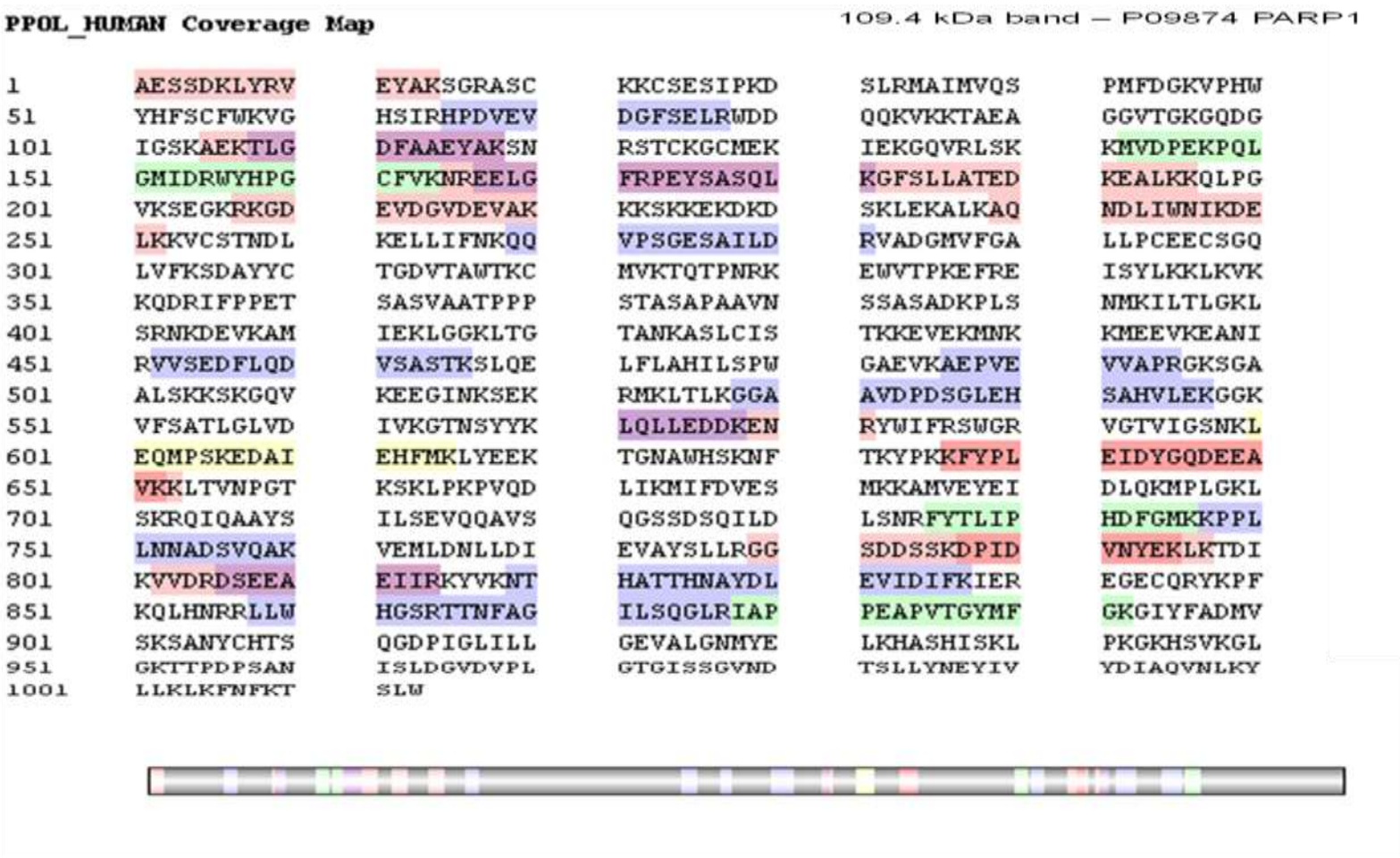


Figure 3.4.3: Coverage of MS peptide data generated for 109.4 kDa. 15 Blue – matched to a peptide, 9 red – matched to a partial peptide, 3 green – matched to a modified peptide, 3 yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

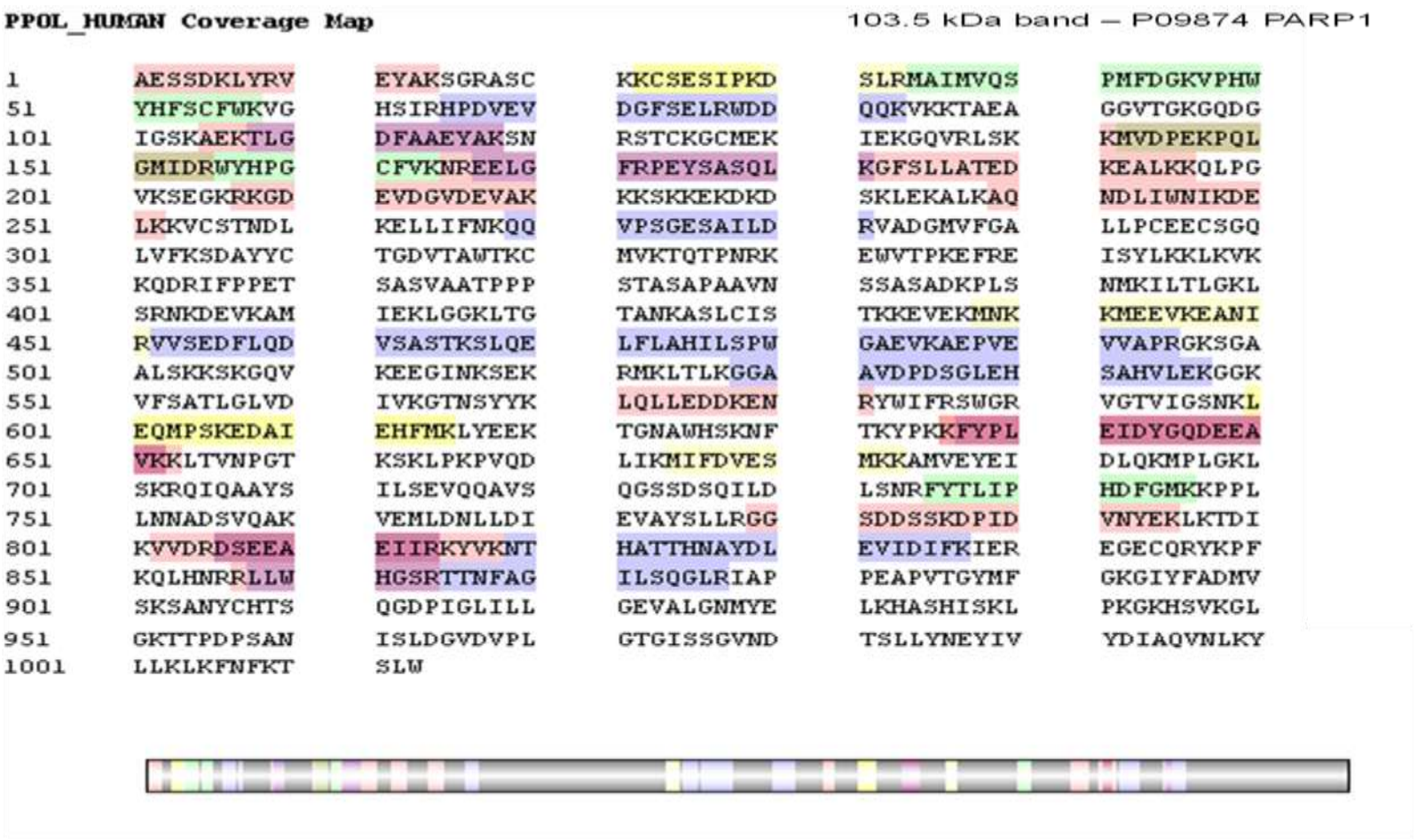


Figure 3.4.4: Coverage of MS peptide data generated for 103.5 kDa. 21 Blue – matched to a peptide, 10 red – matched to a partial peptide, 3 green – matched to a modified peptide, 5 yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

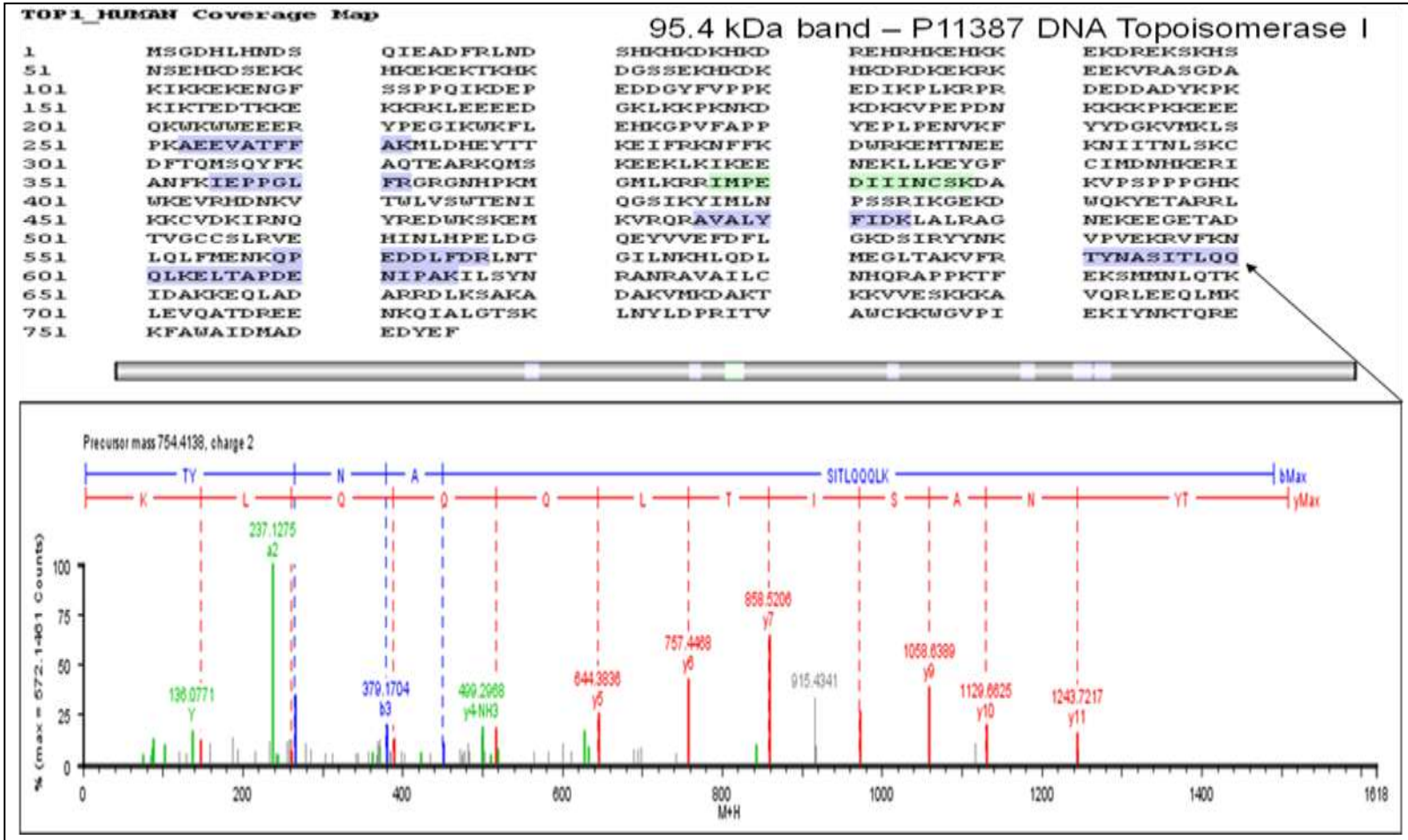


Figure 3.4.5: Coverage of MS peptide data generated for 95.4 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

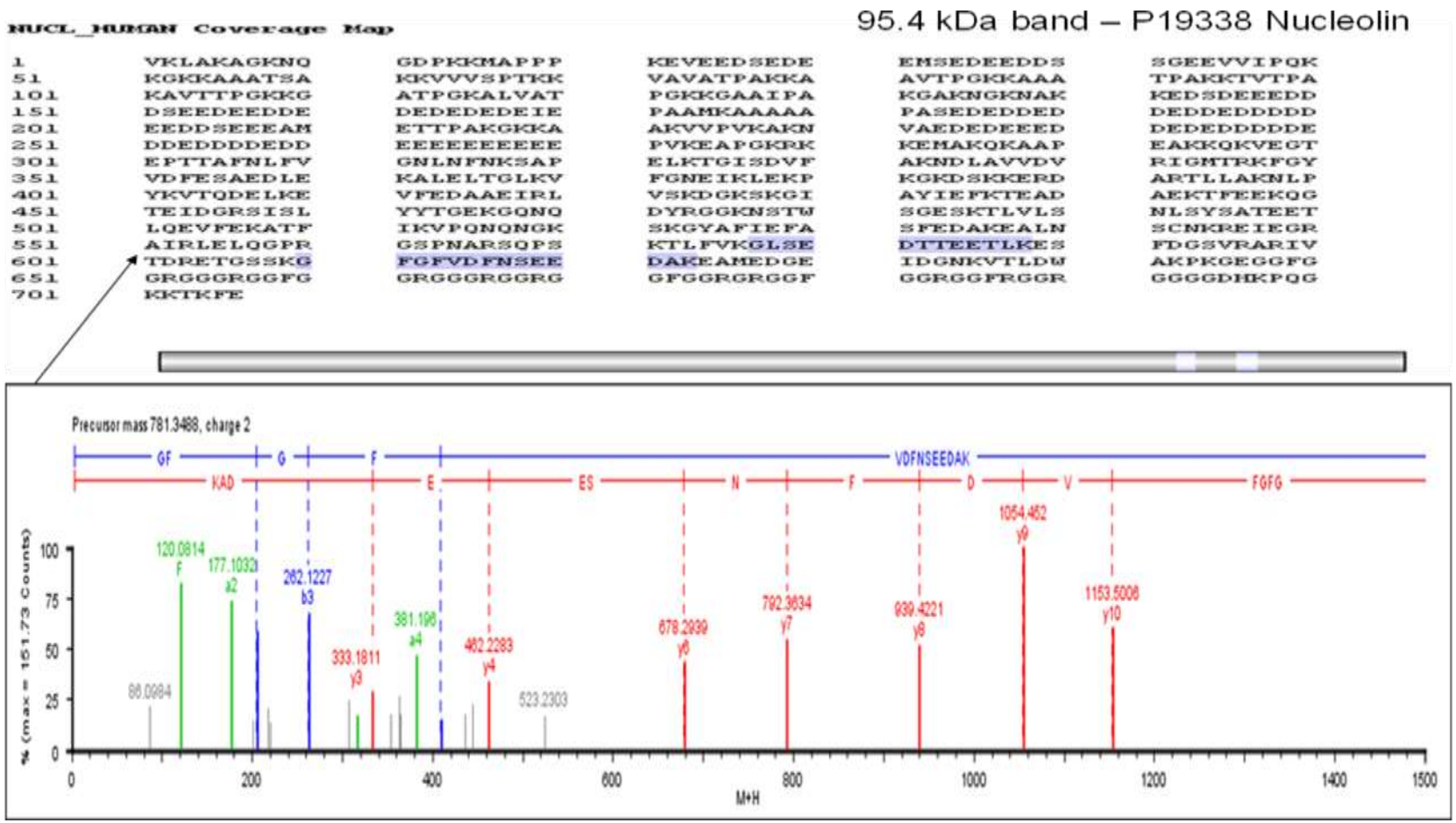


Figure 3.4.5 continued: Coverage of MS peptide data generated for 95.4 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

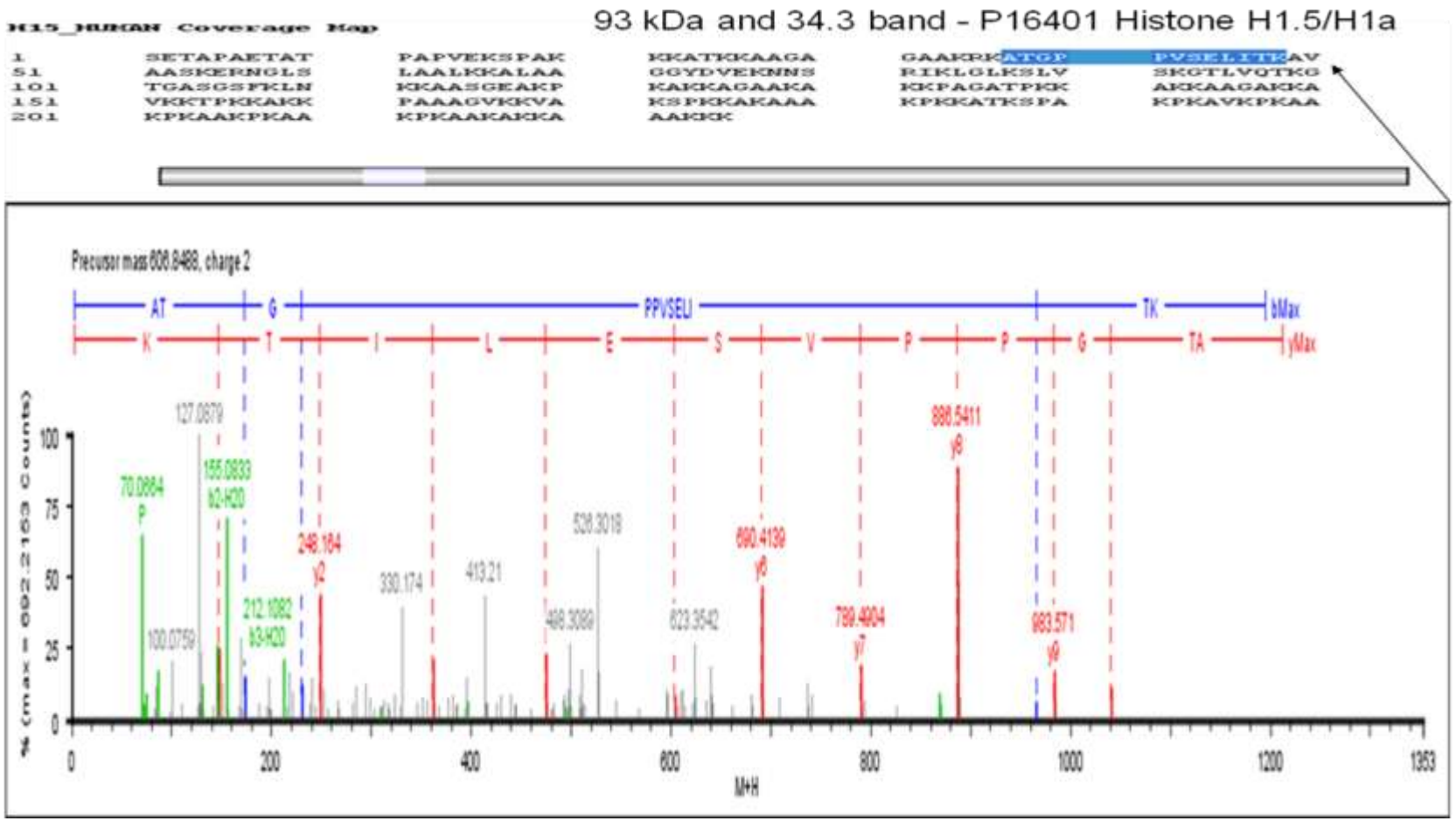


Figure 3.4.6: Coverage of MS peptide data generated for 93 & 34.3 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

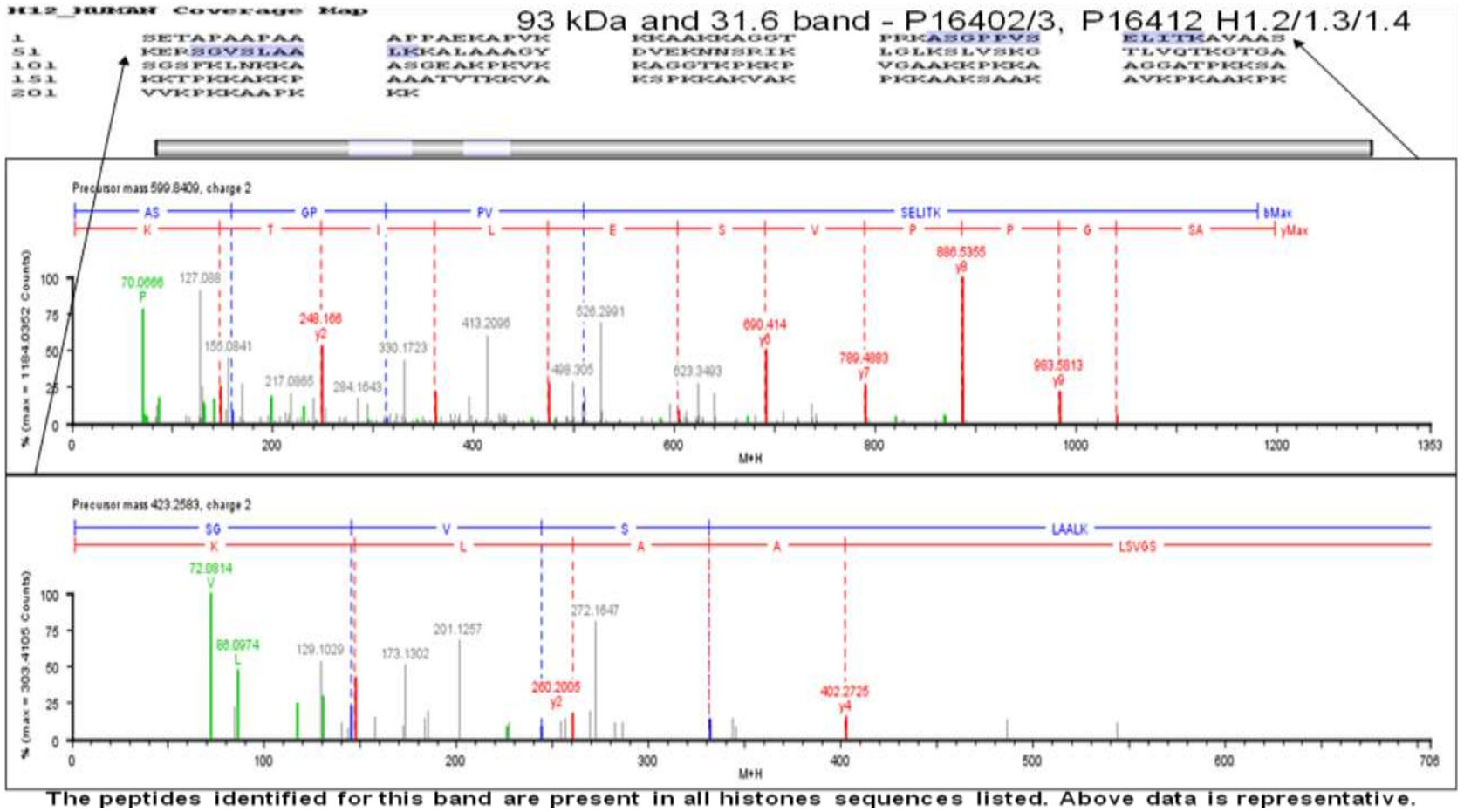


Figure 3.4.7: Coverage of MS peptide data generated for 93 & 31.6 kDa and fragmentation data for two of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

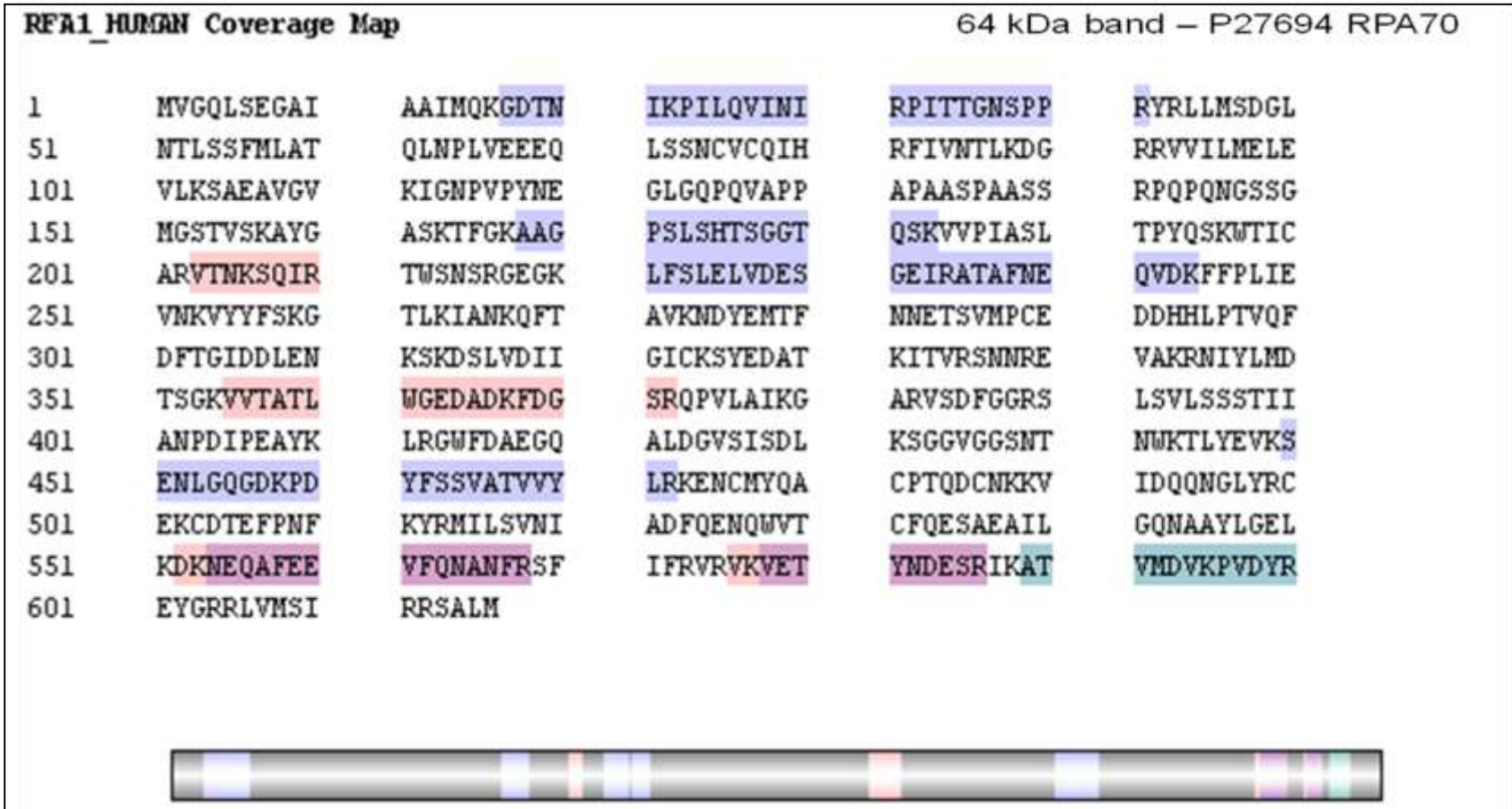


Figure 3.4.8: Coverage of MS peptide data generated for 64 kDa. 7 Blue – matched to a peptide, 4 red – matched to a partial peptide, 1 green – matched to a modified peptide, 1 yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

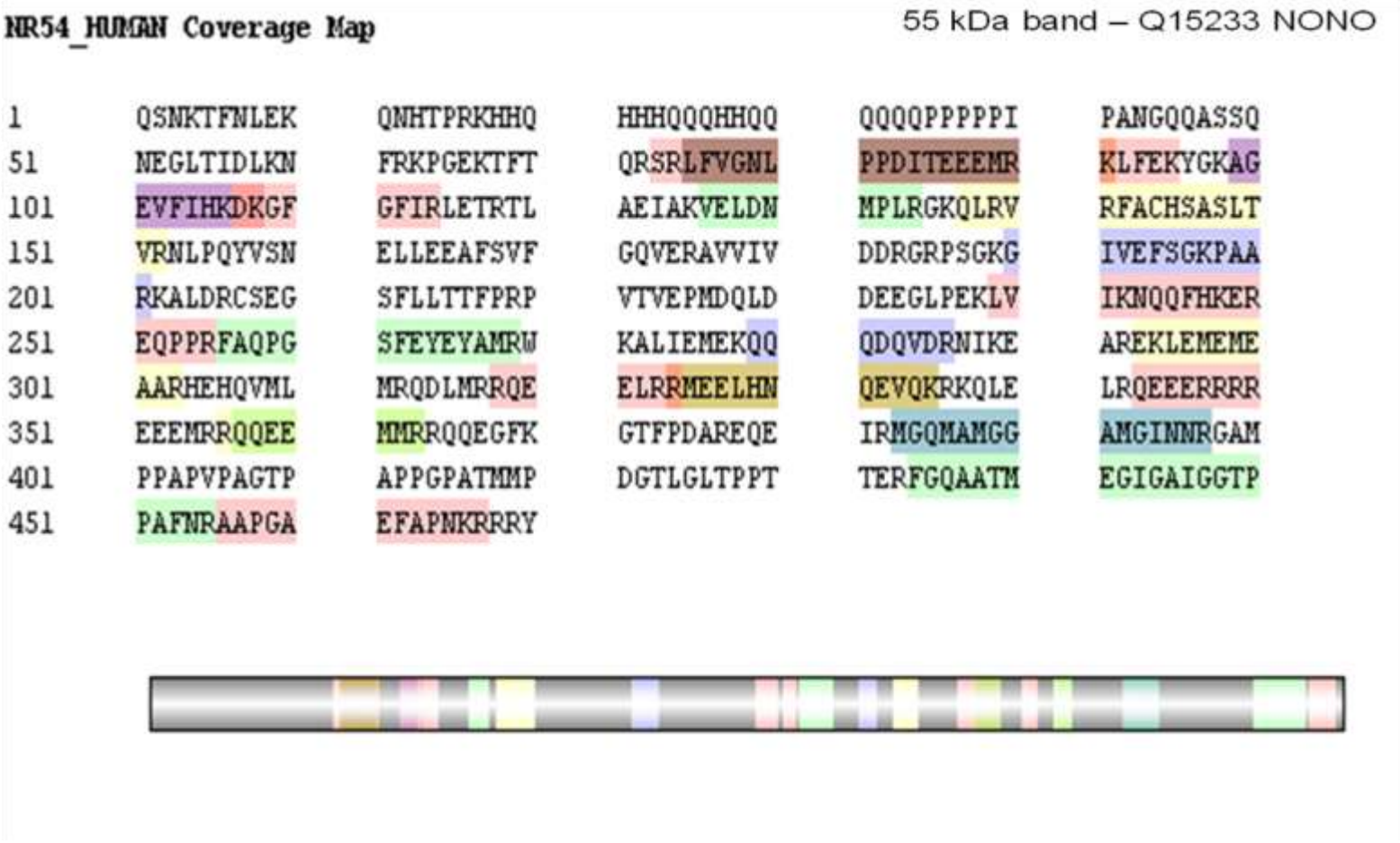


Figure 3.4.9: Coverage of MS peptide data generated for 55 kDa. 6 Blue – matched to a peptide, 6 red – matched to a partial peptide, 6 green – matched to a modified peptide, 9 yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

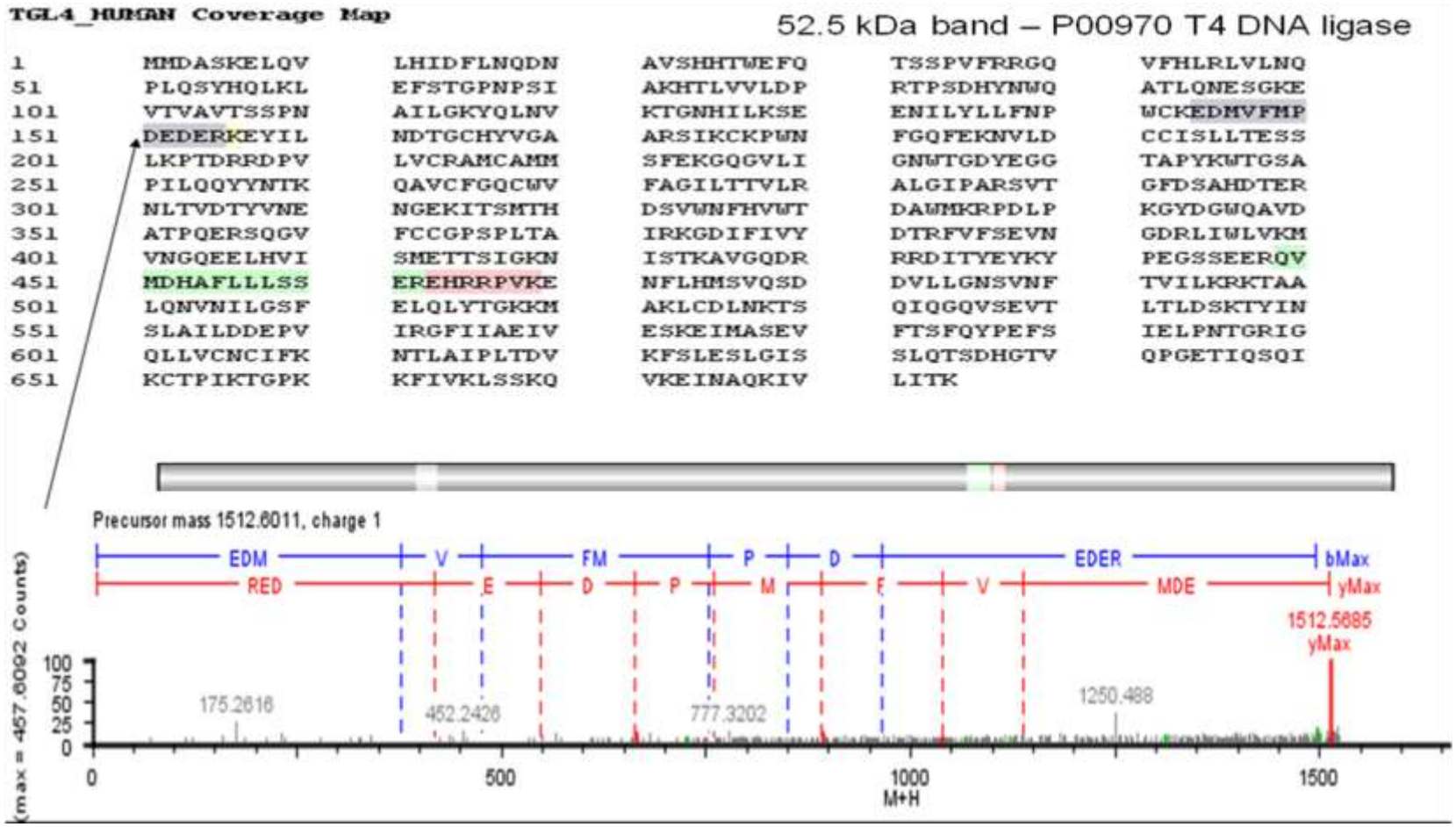


Figure 3.4.10: Coverage of MS peptide data generated for 52.5 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

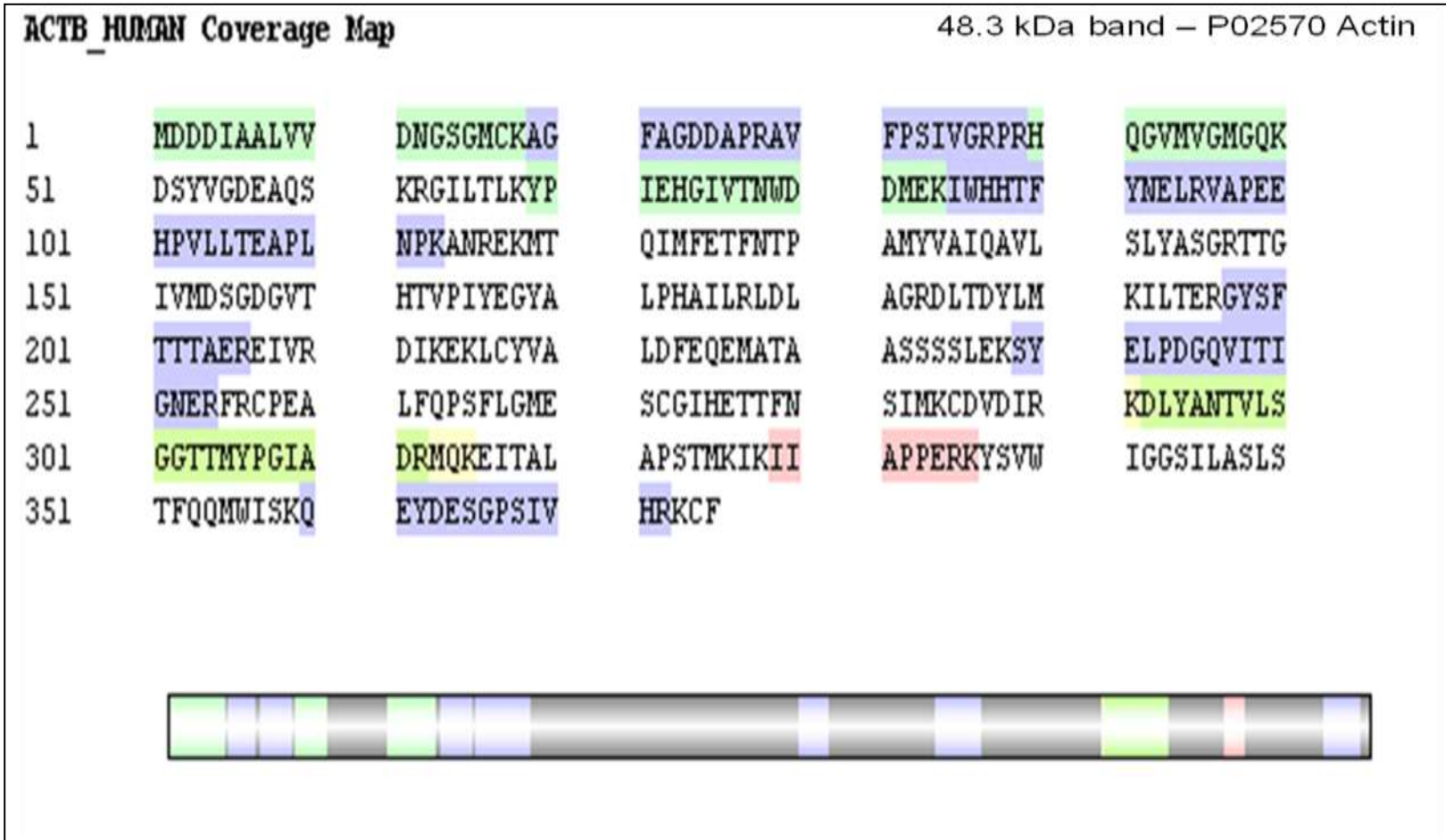


Figure 3.4.11: Coverage of MS peptide data generated for 48.3 kDa. 5 Blue – matched to a peptide, 1 red – matched to a partial peptide, 4 green – matched to a modified peptide, 3 yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

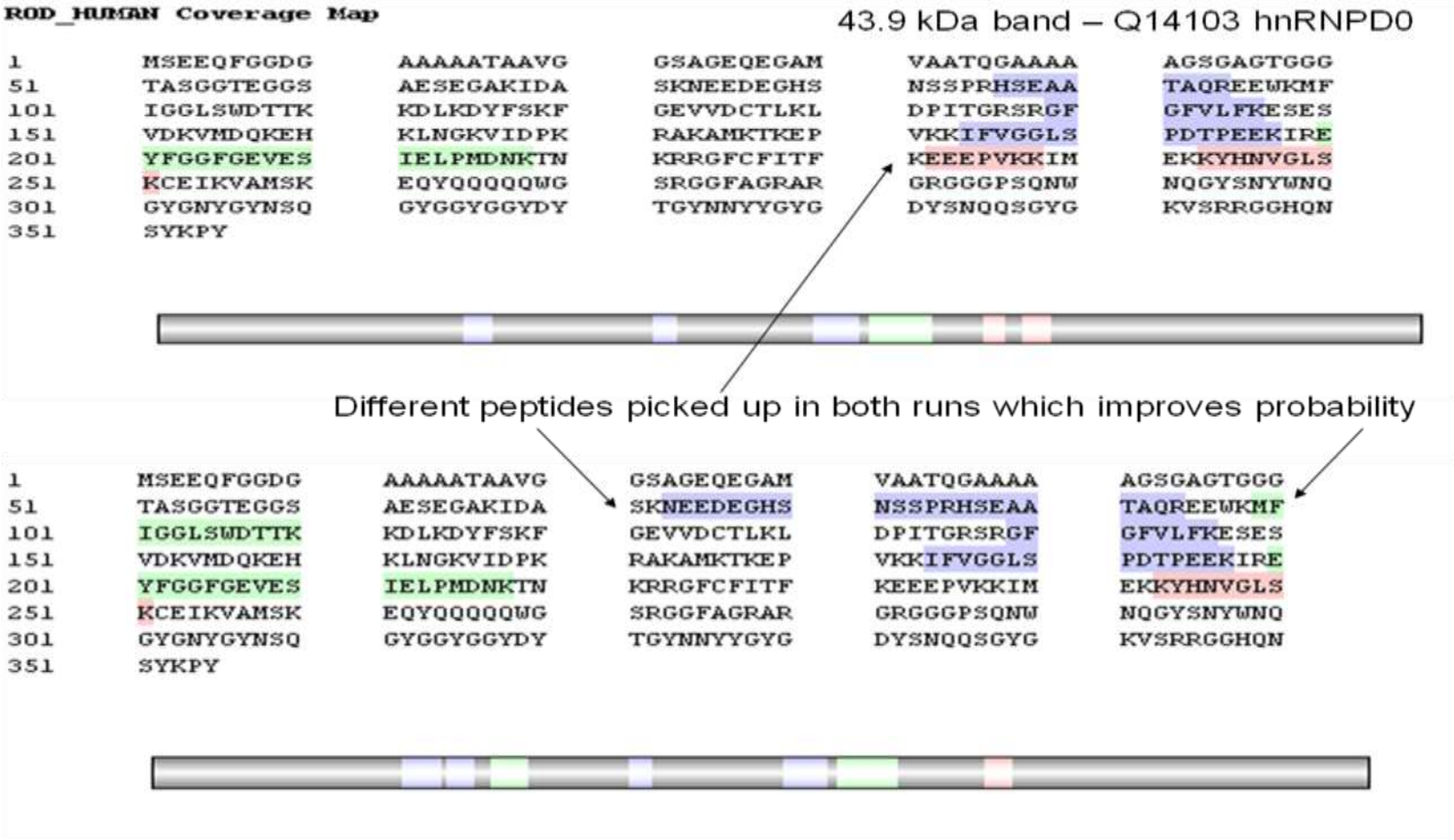


Figure 3.4.12: Coverage of MS peptide data generated for 43.9 kDa. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

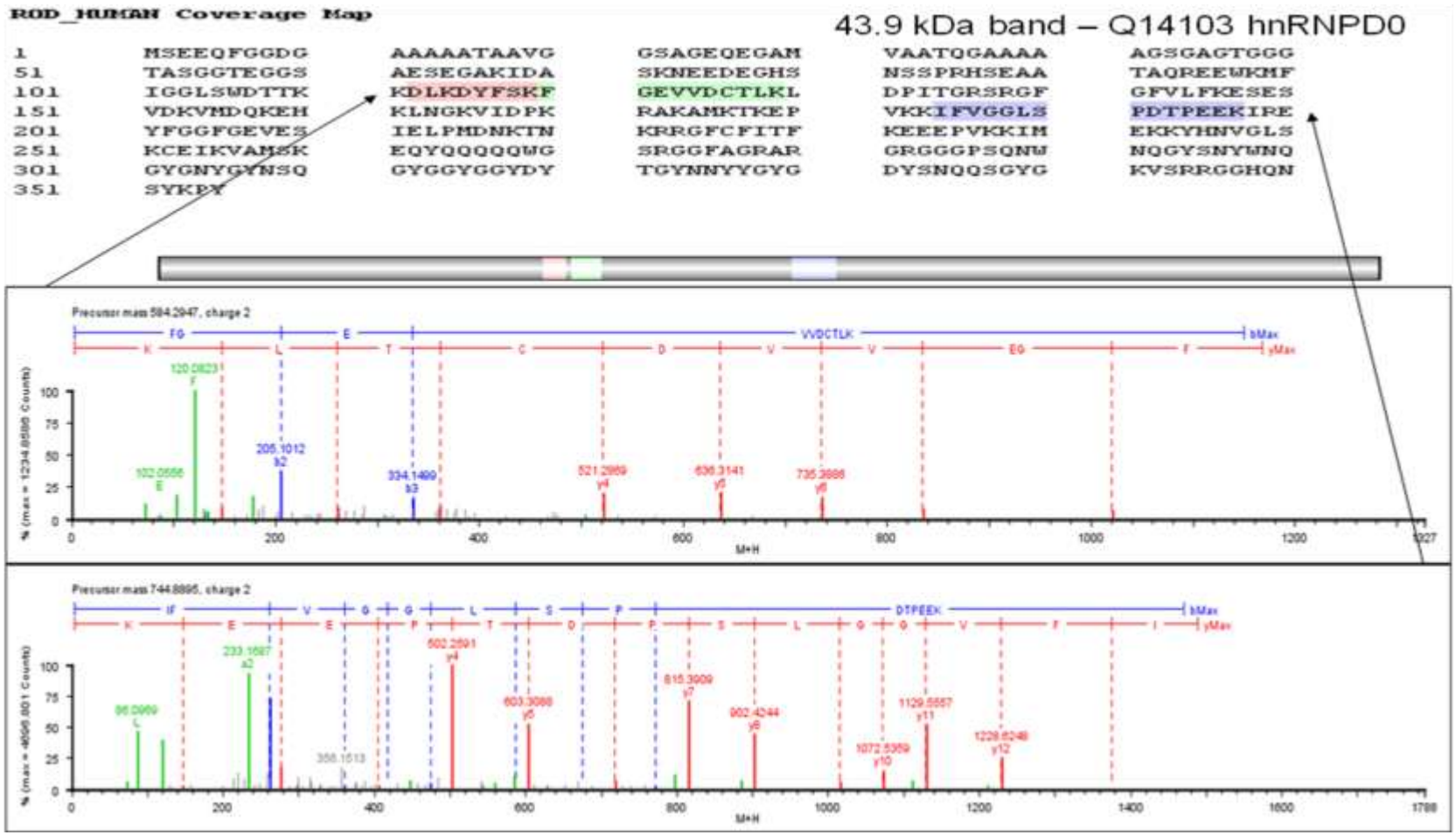


Figure 3.4.12 continued: Coverage of MS peptide data generated for 43.9 kDa and fragmentation data for two of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

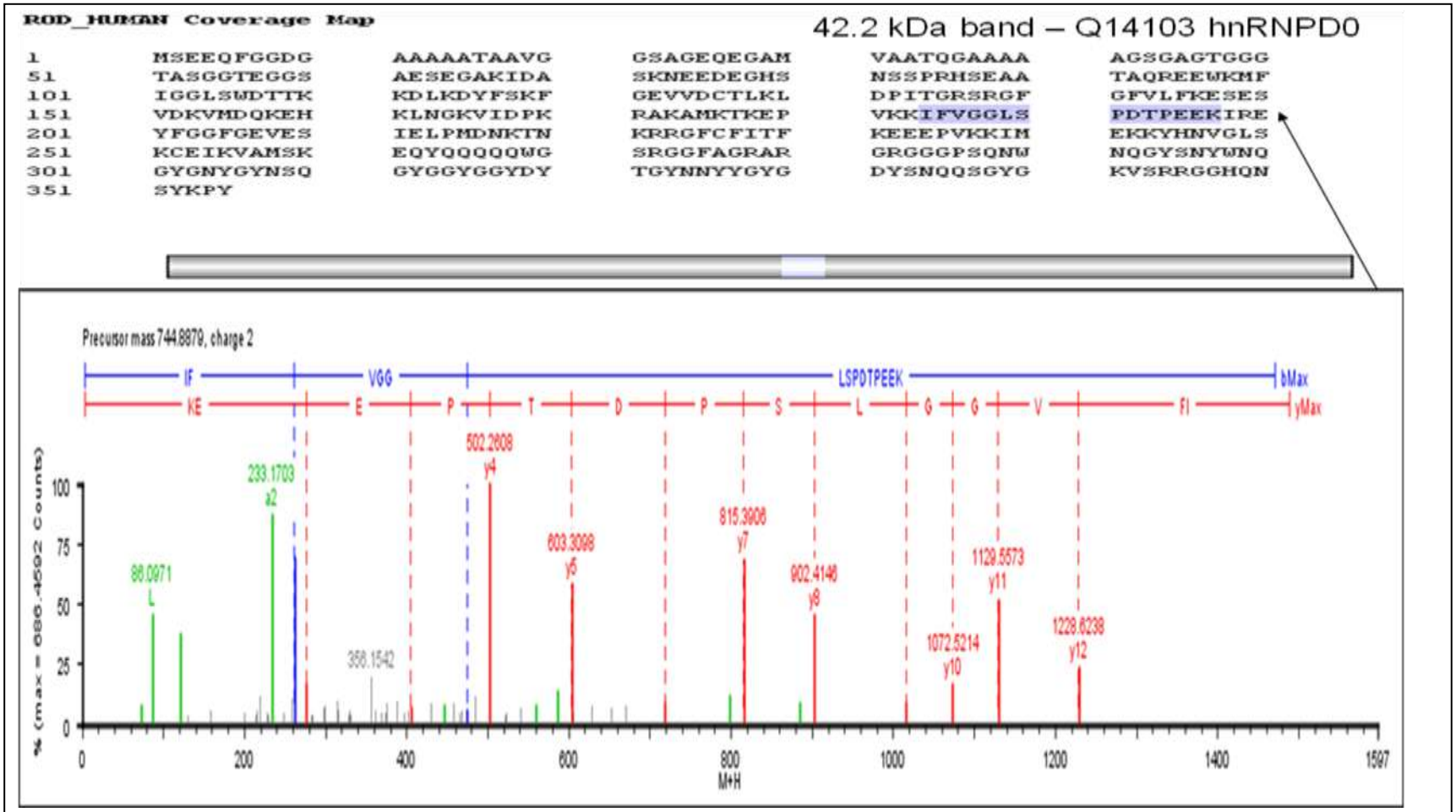


Figure 3.4.13: Coverage of MS peptide data generated for 42.2 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

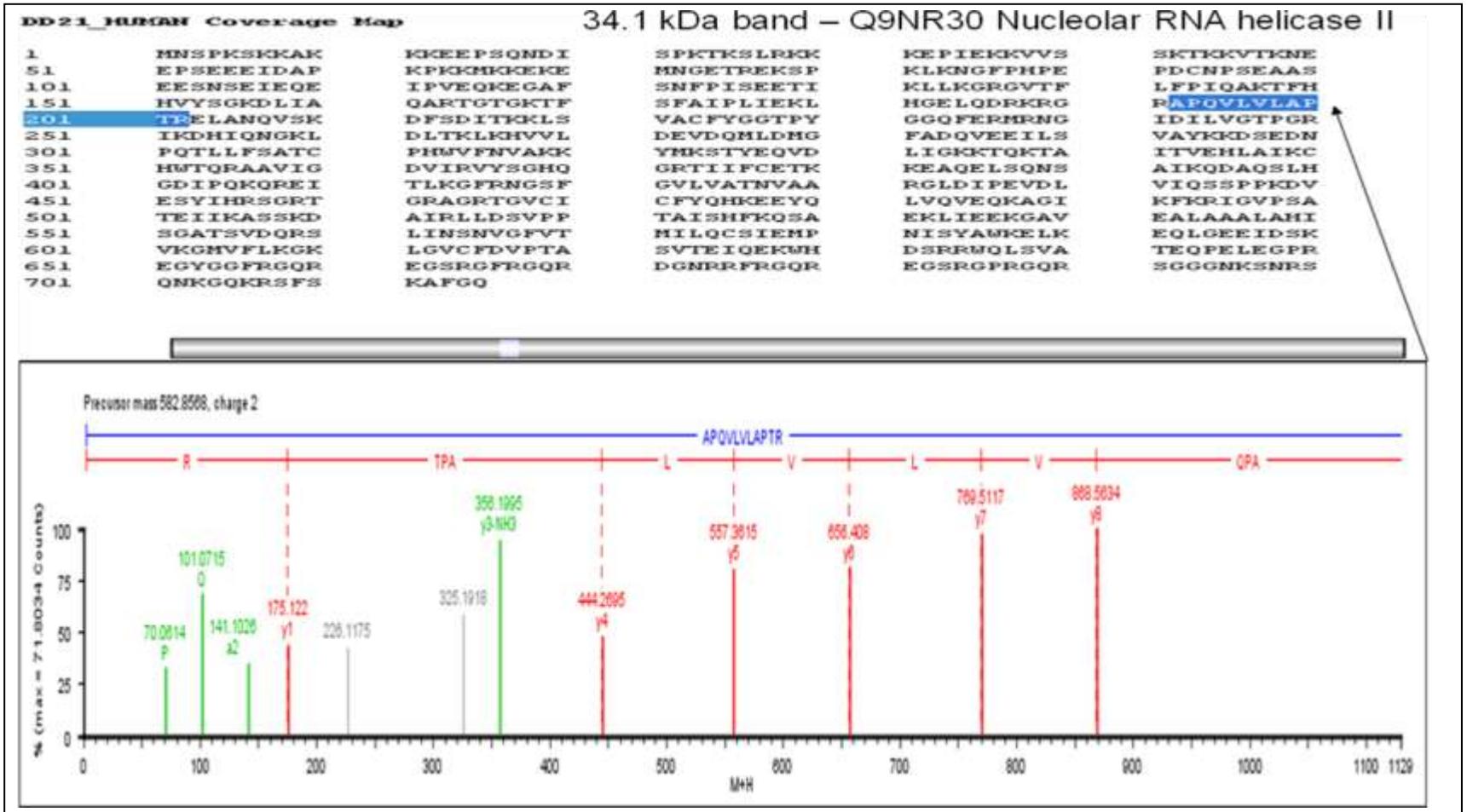


Figure 3.4.14: Coverage of MS peptide data generated for 34.1 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

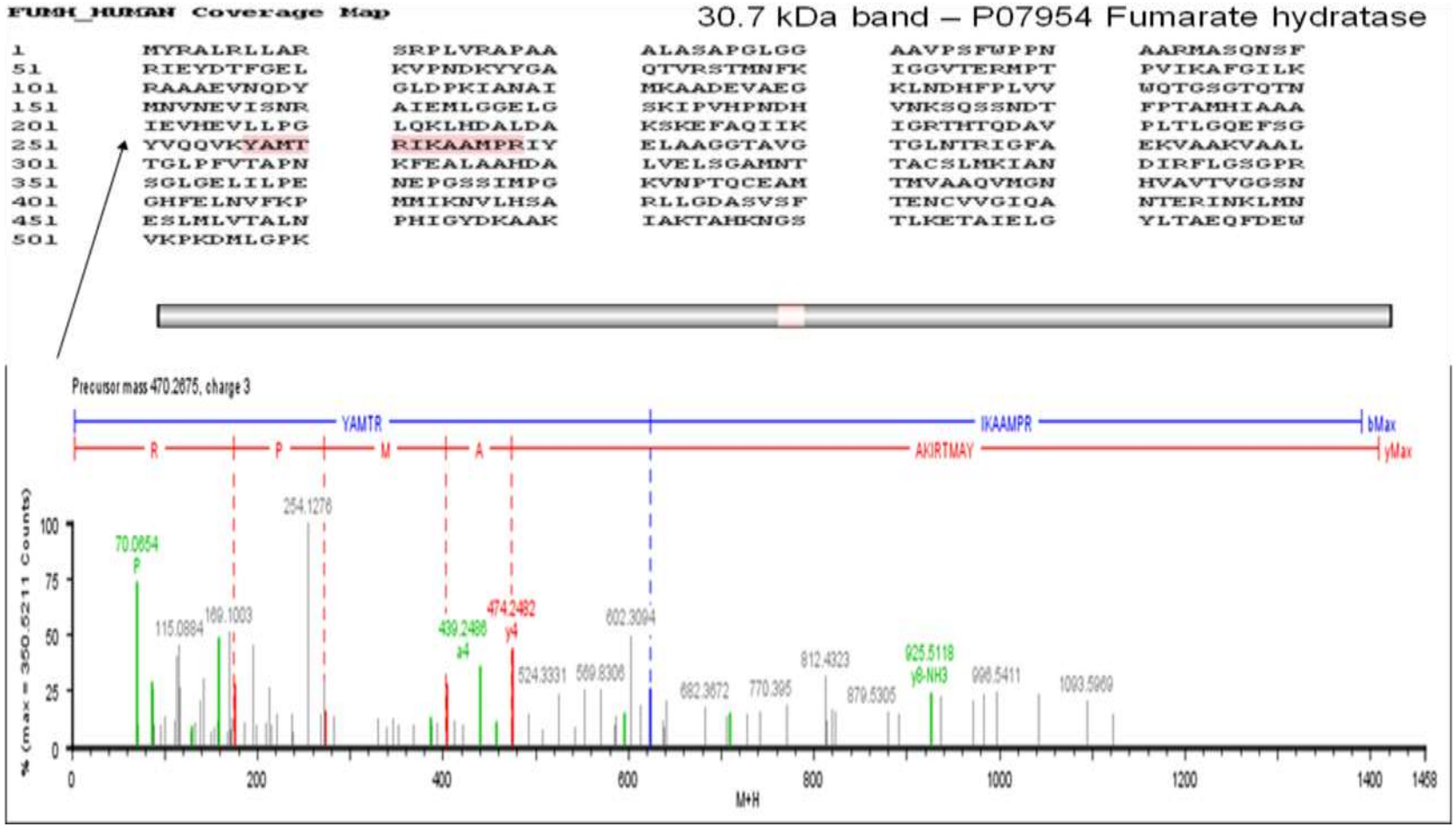


Figure 3.4.15: Coverage of MS peptide data generated for 30.7 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

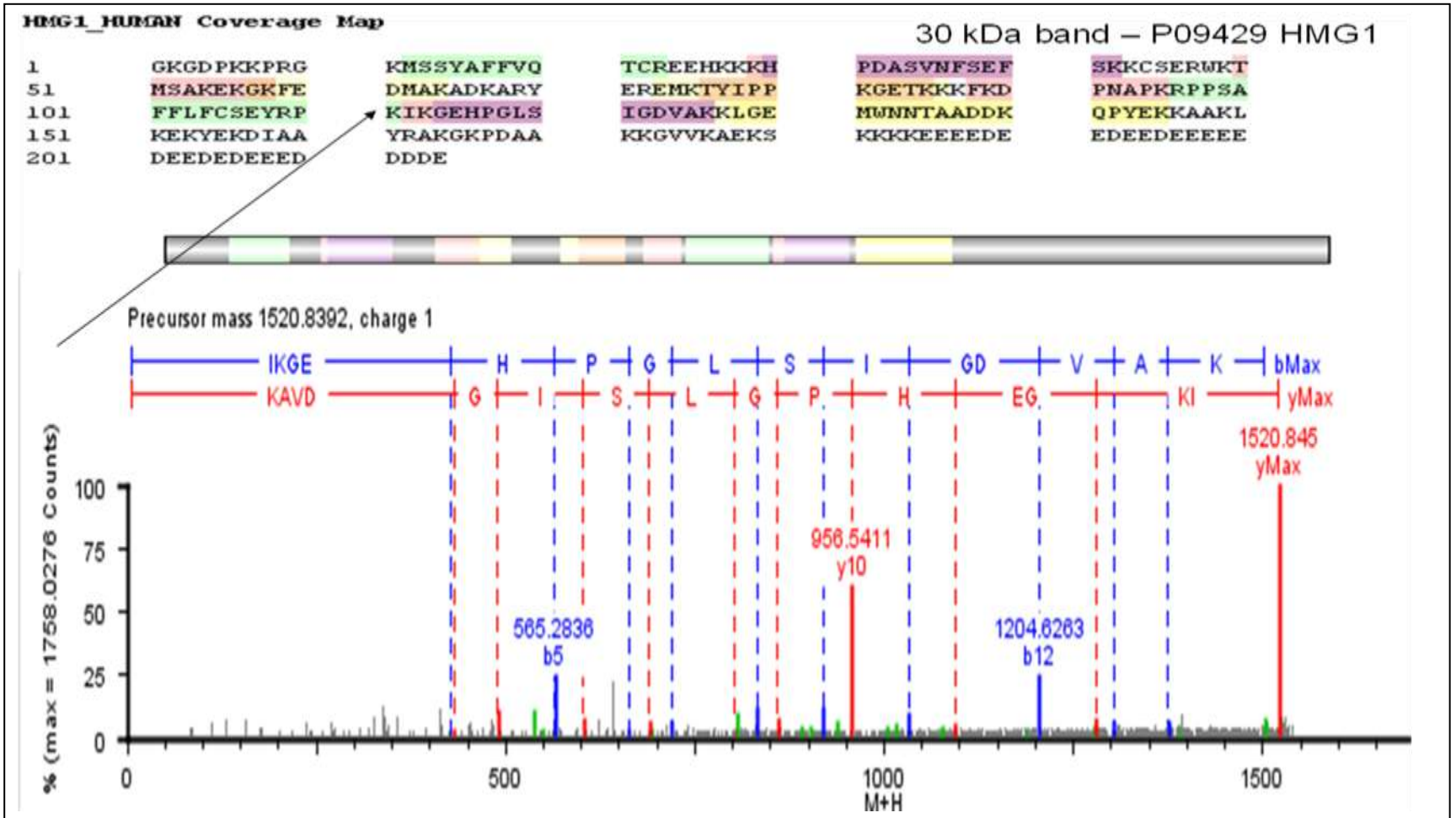


Figure 3.4.16: Coverage of MS peptide data generated for 30 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

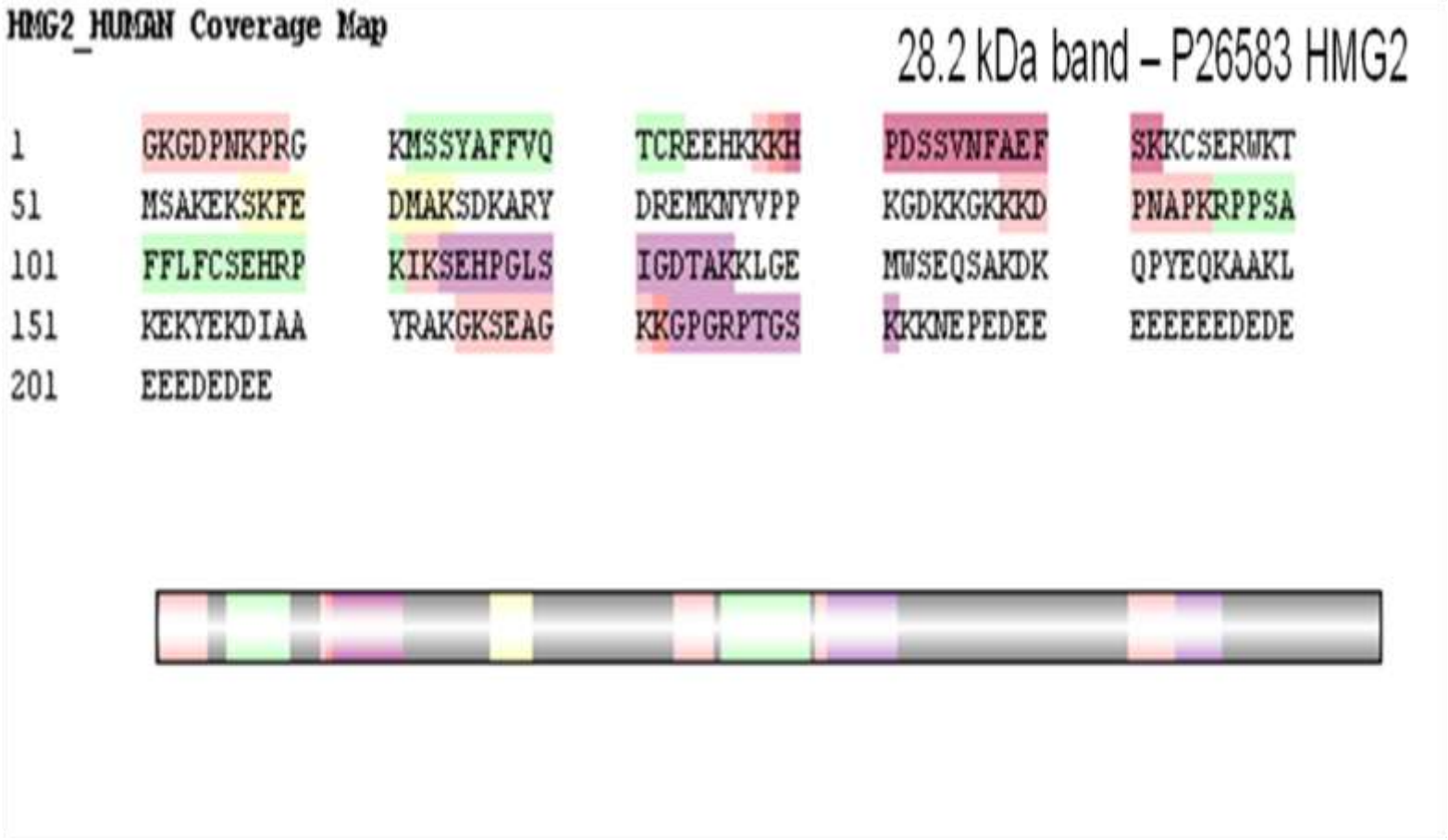


Figure 3.4.17: Coverage of MS peptide data generated for 28.2 kDa. 3 Blue – matched to a peptide, 5 red – matched to a partial peptide, 2 green – matched to a modified peptide, 3 yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

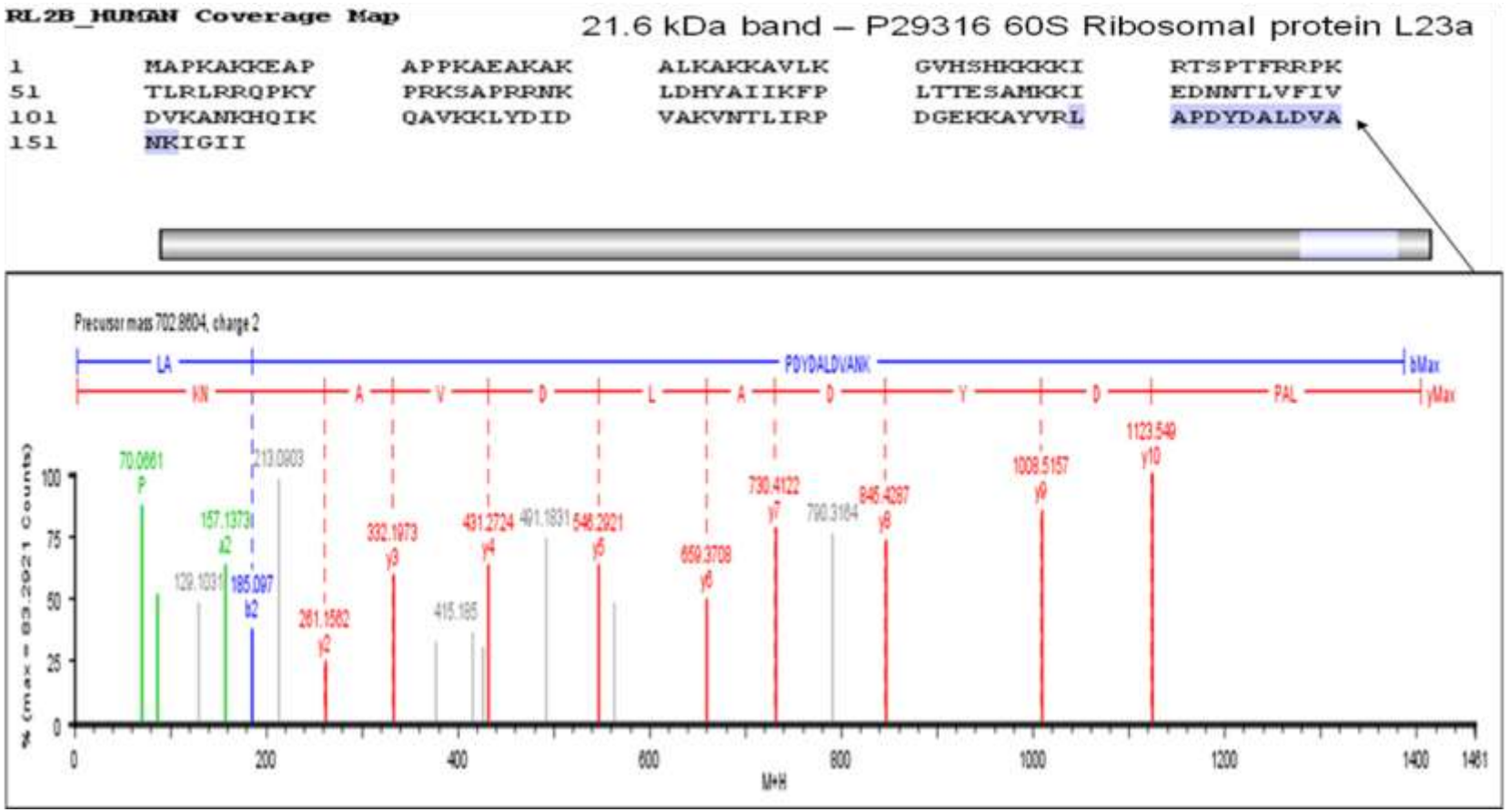


Figure 3.4.18: Coverage of MS peptide data generated for 21.6 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

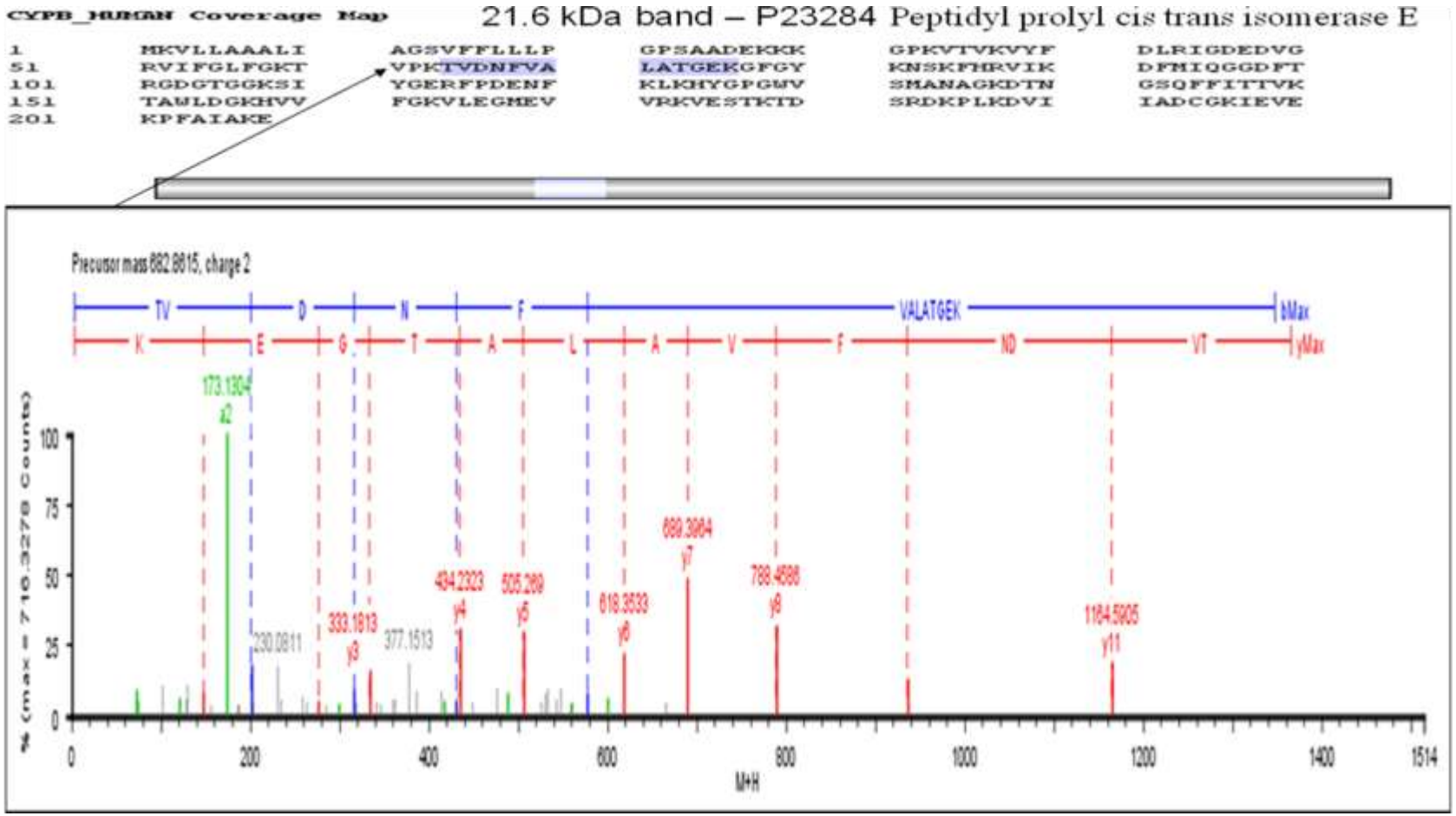


Figure 3.4.18 continued: Coverage of MS peptide data generated for 21.6 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

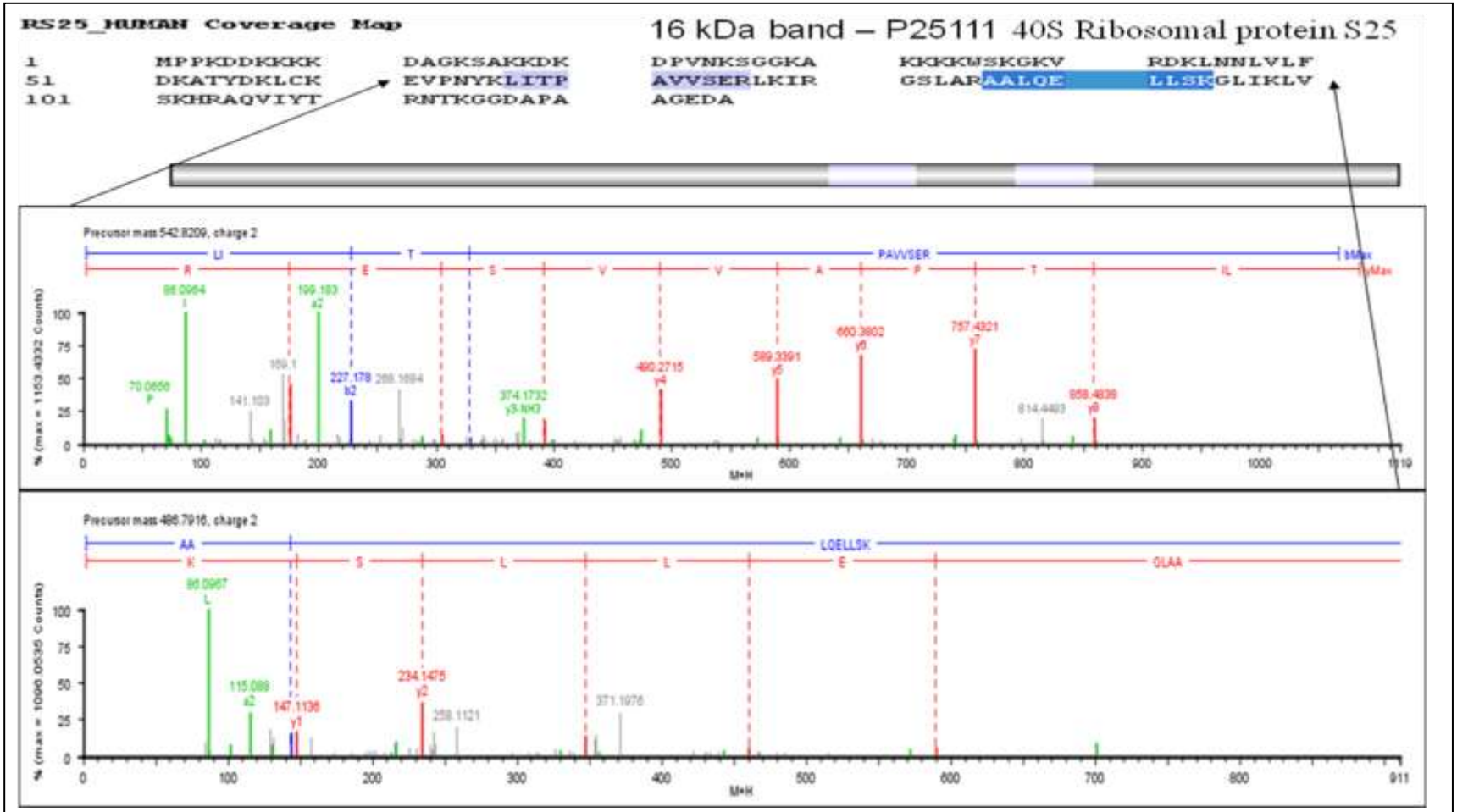


Figure 3.4.19: Coverage of MS peptide data generated for 16 kDa and fragmentation data for two of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

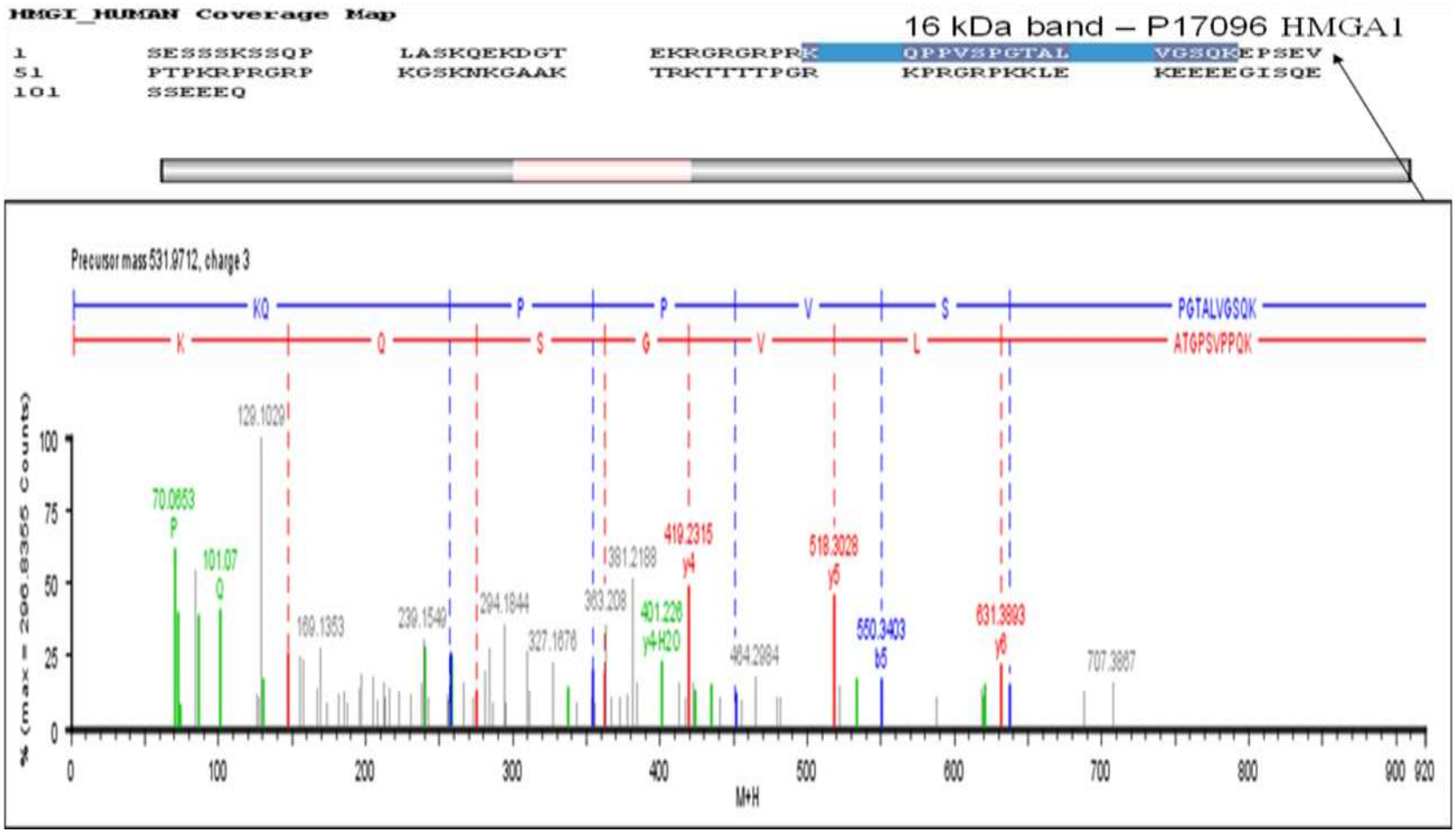


Figure 3.4.19 continued: Coverage of MS peptide data generated for 16 kDa and fragmentation data for one of the peptides. Blue – matched to a peptide, red – matched to a partial peptide, green – matched to a modified peptide, yellow – matched to a partial modified peptide. Other colours arise from different peptide types with overlapping sequences. List of modifications can be found at http://www.abrf.org/index.cfm/dm.home?AvgMass=all

8 References

1 Weingarten M.D., Lockwood A.H., Hwo S.Y. and Kirschner M.W. (1975) A protein factor essential for microtubule assembly. Proceedings in the National Academy of Science U.S.A. **72** 1858-1862

2 Cleveland D.W., Hwo S.Y. and Kirschner M.W. (1977) Purification of tau, a microtubule-associated protein that induces assembly of microtubules from purified tubulin. Journal of Molecular Biology **116** 207-225

3 Maccioni R.B. and Cambiazo V. (1995) Role of microtubule-associated proteins in the control of microtubule assembly. Physiological Reviews **75** 835-862

4 Mandelkow E. and Mandelkow E.M. (1995) Microtubules and microtubule-associated proteins. Current Opinion in Cell Biology **7** 72-81

5 Dehmelt L. and Halpain S. (2004) The MAP2/tau family of microtubule-associated proteins. Genome Biology **6** 204-213

6 Pittman A.M., Fung H.C. and de Silva R. (2006) Untangling the tau gene association with neurodegenerative disorders. Human Molecular Genetics **15** R188-R195

7 Kosik K.S. (1993) The molecular and cellular biology of tau. Brain Pathology **3** 39-43

8 Goedert M., Crowther R.A. and Garner C.C. (1991) Molecular characterization of microtubule-associated proteins tau and MAP2. Trends in Neuroscience **14** 193-199

9 Heidary G. and Fortini M.E. (2001) Identification and characterization of the *Drosophila tau* homolog. Mechanisms of Development **108** 171-178

10 Cleveland D.W., Hwo S.Y. and Kirschner M.W. (1977) Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. Journal of Molecular Biology **116** 227-247

11 Johnson G.V.W. and Stoothoff W.H. (2004) Tau phosphorylation in

neuronal cell function and dysfunction. Journal of Cell Science **117** 5721-5729

12 Avila J., Lucas J.J., Perez M. and Hernandez F. (2004) Role of tau protein in both physiological and pathological conditions. Physiological Reviews **84** 361-384

13 Mandelkow E.M., Biernat J., Drewes G., Gustke N., Trinczek B. and Mandelkow E. (1995) Tau domains, phosphorylation, and interactions with microtubules. Neurobiology of Aging **16** 355-363

14 Mandelkow E.M., Schweers O., Drewes G., Biernat J., Gustke N., Trinczek B. and Mandelkow E. (1996) Structure, Microtubule interactions, and phosphorylation of tau protein. Annals of the New York Academy of Sciences **777** 96-106

15 Goedert M. (2004) Tau protein and neurodegenration. Seminars in Cell & Developmental Biology **15** 45-49

16 Alonso A.d.C. and Iqbal K. (2005) Tau-induced neurodegeneration: a clue to its mechanism. Journal of Alzheimer's Disease **8** 223-226

17 Hardy J., Pittman A., Myers A., Fung H.C., de Silva R. and Duckworth J. (2006) Tangle diseases and the tau haplotypes. Alzheimers Disease Association Disorders **20** 60-62

18 Kalinderi K., Fidani L. and Bostantjopoulou (2009) From 1997 to 2007: a decade journey through the H1 haplotype on 17q21 chromosome. Parkisonism and Related Disorders **15** 2-5

19 Kar A., Kuo D., He R., Zhou J. and Wu J.Y. (2005) Tau alternative splicing and frontotemporal dementia. Alzheimers Disease and Associated Disorders **19** S29-36

20 Fidzianska A. and Glinka Z. (2006) Rimmed vacuoles with beta-amyloid and tau protein deposits in the muscle of children with hereditary myopathy. Acta Neuropathology **112** 185-193

21 Munoz D.G., Ros R., Fatas M., Bermejo F. and de Yebenes J.G. (2007)

Progressive nonfluent aphasia associated with a new mutation V363I in tau gene. American Journal of Alzheimer's Disease and Other Dementias **22** 294-299

22 Matsushita S., Miyakawa T., Maesato, Matsui T., Yokoyama A., Arai H., Higuchi S. and Kashima H. (2008) Elevated cerebrospinal fluid tau protein levels in Wernicke's encephalopathy. Alcoholism: Clinical and Experimental Research **32** 1091-1095

23 Gupta N., Fong J., And L.C. and Yucel Y.H. (2008) Retinal tau pathology in human glaucomas. Canadian Journal of Ophthalmology **43** 53-60

24 Johnson G.V.W. and Bailey C.D.C. (2002) Tau, where are we now? Journal of Alzheimer's Disease **4** 375-398

25 Buee L. and Delacourte A. (2001) Neuronal subpopulations and genetic background in tauopathies: a catch 22 story? Neurobiology of Aging **22** 115-118

26 Pickering-Brown S.M., Baker M., Gass J., Boeve B.F., Loy C.T., Brooks W.S., Mackenzie I.R.A., Martins R.N., Kwok J.B.J., Halliday G.M., Kril J., Schofield P.R., Mann D.M.A. and Mike Hutton (2006) Mutations in progranulin explain atypical phenotypes with variants in *MAPT*. Brain **129** 3124-3126

27 Roks G., Darmut B., Heutink P., Julliams A., Backhovens H., Van de Broeck M., Serneels S., Hofman A., Van Broeckhoven C., van Duijin C.M. and Cruts M. (1999) Mutation screening of the tau gene in patients with early-onset Alzheimer's disease. Neuroscience Letters **277** 137-139

28 Andreadis A., Brown W.M. and Kosik K.S. (1992) Structure and novel exons of the human tau gene. Biochemistry **31** 10626-10633

29 Wilhelmsen K.C., Lynch T., Pavlou E., Higgins M. and Nygaard T.G. (1994) Localization of disinhibition-dementia-parkinsonism-amyotrophy complex to 17q21–22. American Journal of Human Genetics **55** 1159–1165

30 Tsuboi Y. (2006) Neuropathology of familial tauopathy. Neuropathology

26 471-474

31 Goedert M. (2005) Tau gene mutations and their effects. Movement disorders **20** S45-52

32 Poorkaj P., Muma N.A., Zhukareva V., Cochran E.J., Shannon K.M., Hurtig H., Koller W.C., Bird T.D., Trojanowski J.Q., Lee V.M. and Schellenberg G.D. (2002) An R5L tau mutation in a subject with a progressive supranuclear palsy phenotype. Annals of Neurology **52** 511–516

33 Hayashi S., Toyoshima Y., Hasegawa M., Umeda Y., Wakabayashi K., Tokiguchi S., Iwatsubo T. and Takahashi H. (2002) Late-onset frontotemporal dementia with a novel exon 1 (Arg5His) tau gene mutation. Annals of Neurology **51** 525–530

34 Bronner I.F., ter Meulin B.C., Azmani A., Severijnen L.A., Willemsen R., Kamphorst W., Ravid R., Heutink P. and van Swieten J.C. (2005) Hereditary Pick's disease with the G272V tau mutation shows predominant three-repeat tau pathology. Brain **128** 2645-2653

35 Rizzini C., Goedert M., Hodges J.R., Smith M.J., Jakes R., Hills R., Xuereb J.H., Crowther R.A. and Spillantini M.G. (2000) Tau gene mutation K257T causes a tauopathy similar to Pick's disease. Journal of Neuropathology and Experimental Neurology **59** 990–1001

36 Murrell J.R., Spillantini M.G., Zolo P., Guazzelli M., Smith M.J., Hasegawa M., Redi F., Crowther R.A., Pietrini P., Ghetti B. and Goedert M. (1999) Tau gene mutation G389R causes a tauopathy with abundant Pick body-like inclusions and axonal deposits. Journal of Neuropathology and Experimental Neurolgoy. **58** 1207–1226

37 Kobayashi T., Mori H., Okuma Y., Dickson D.W., Cookson N., Tsuboi Y., Motoi Y., Tanaka R., Myashita N., Anno M., Narabayashi H. and Mizuno Y. (2002) Contrasting genotypes of the tau gene in two phenotypically distinct patients with P301L mutation of frontotemporal dementia and parkinsonism linked to chromosome 17. Journal of

Neurology **249** 669–675

38 Bugiani O., Murrell J.R., Giaccone G., Hasegawa M., Ghigo G., Tabaton M., Morbin M., Primavera A., Carella F., Solaro C., Grisoli M., Savoiardo M., Spillantini M.G., Tagliavini F., Goedert M. and Ghetti B. (1999) Frontotemporal dementia and corticobasal degeneration in a family with a P301S mutation in tau. Journal of Neuropathology and Experimental Neurology **58** 667–677

39 Stanford P.M., Halliday G.M., Brooks W.S., Kwok J.B.J., Storey C.E., Creasy H., Morris J.G.L., Fulham M.J. and Schofield P.R. (2000) Progressive supranuclear palsy pathology caused by a novel silent mutation in exon 10 of the tau gene: expansion of the disease phenotype caused by tau gene mutations. Brain **123** 880–893

40 Momeni P., Pittman A., Lashley T., Vandrovcova J., Malzer, E., Luk C., Hulette C., Lees A., Revesz T., Hardy J. and de Silva R. (2009) Clinical and pathological features of an Alzheimer's disease patient with the *MAPT* ΔK280 mutation. Neurobiology of Aging **30** 388-393

41 Pastor P., Pastor E., Carnero C., Vela R., Garcia T., Amer G., Tolosa E. and Oliva R. (2001) Familial atypical progressive supranuclear palsy associated with homozygosity for the delN296 mutation in the tau gene. Annals of Neurology **49** 263–267

42 Zarranz J.J., Ferrer I., Lezcano E., Forcadas M.I., Eizaguirre B., Atares B., Puig B., Gomez-Esteban J.C., Fernandez-Maiztegui C., Rouco I., Perez-Concha T., Fernandez M., Rodriguez O., Rodriguez-Martinez A.B., Martinez de Pancorbo M., Pastor P. and Perez-Tur J. (2005) A novel mutation (K317M) in the *MAPT* gene causes FTDP and motor neuron disease. Neurology **64** 1578-1585

43 van Herpen E., Rosso S.M., Serverijnen L.A., Yoshida H., Breedveld G., van de Graaf R., Kamphorst W., Ravid R., Willemsen R., Dooijes D., Majoor-Krakauer D., Kros J.M., Crowther R.A., Goedert M., Heutink P. and van Swieten J.C. (2003) Variable phenotypic expression and extensive

tau pathology in two families with the novel tau mutation L315R. Annals of Neurology **54** 573–581

44 Rosso S.M., van Herpen E., Deelen W., Kamphorst W., Severijnen L.-A., Willemsen R., Ravid R., Niermeijer M.F., Dooijes D., Smith M.J., Goedert M., Heutink P. and van Swieten J.C. (2002) A novel tau mutation, S320F, causes a tauopathy with inclusions similar to those in Pick's disease. Annals of Neurology **51** 373–376

45 Pickering-Brown S.M., Baker M., Nonaka T., Ikeda K., Sharma S., Mackenzie J., Simpson S.A., Moore J.W., Snowden J.S., de Silva R., Revesz T., Hasegawa M., Hutton M. and Mann D.M.A. (2004) Frontotemporal dementia with Pick-type histology associated with Q336R mutation in the tau gene. Brain **127** 1415-1426

46 Lippa C.F., Zhukareva V., Kawarai T., Uryu K., Shafiq M., Nee L.E., Grafman J., Liang Y., St. George-Hyslop P.H., Trojanowski J.Q., Lee V.M. (2000) Frontotemporal dementia with novel tau pathology and a Glu342Val tau mutation. Annals of Neurology **48** 850-858

47 Neumann M., Diekmann S., Bertsch U., Vanmassenhove B., Bogerts B. and Kretzschmar H.A. (2005) Novel G335V mutation in the tau gene associated with early onset familial frontotemporal dementia. Neurogenetics **6** 91-95

48 Spillantini M.G. and Goedert M. (1998) Tau protein pathology in neurodegenerative diseases. Trends in Neuroscience **21** 428-433

49 Munoz D.G., Ros R., Fatas M., Bermejo F. and de Yebenes J.G. (2007) Progressive nonfluent aphasia associated with a new mutation V363I in tau gene. American Journal of Alzheimer's Disease and Other Dementias **22** 294-299

50 Rossi G., Marelli C., Farina L., Laura M., Basile A.M., Ciano C., Tagliavini F. and Pareyson D. (2008) The G389R mutation in the *MAPT* gene presenting as sporadic corticobasal syndrome. Movement Disorders **23** 892-895

51 Rizzu P., Joosse M., Ravid R., Hoogeveen A., Kamphorst W., van Swieten J.C., Willemsen R. and Heutink P. (2000) Mutation-dependent aggregation of tau protein and its selective depletion from the soluble fraction in brain of P301L FTDP-17 patients. Human Molecular Genetics **9** 3075-3082

52 Myasaki T., Morishima-Kawashima M., Ravid R., Heutink P., van Swieten J.C., Nagashima K. and Ihara Y. (2001) Molecular analysis of mutant and wild-type tau deposited in the brain affected by the FTDP-17 R406W mutation. American Journal of Pathology **158** 373-379

53 Connell J.W., Gibb G.M., Betts J.C., Blackstock W.P., Gallo J.-M., Lovestone S., Hutton M. and Anderton B.H. (2001) Effects of FTDP-17 mutations on the in vitro phosphorylation of tau by glycogen synthase kinase 3β identified by mass spectrometry demonstrate certain mutations exert long-range conformational changes. FEBS Letters **493** 40-44

54 Mack T.G.A., Dayanandan R., van Slegtenhorst M., Whone A., Hutton M., Lovestone S. and Anderton B.H. (2001) Tau proteins with frontotemporal dementia-17 mutations have both altered expression levels and phosphorylation profiles in differentiated neuroblastoma cells. Neuroscience **108** 701-712

55 Sakaue F., Saito T., Sato Y., Asada A., Ishiguro K., Hasegawa M. and Hisanaga S.-I. (280) Phosphorylation of FTDP-17 mutant tau by cyclin-dependent kinase 5 complexed with p35, p25, or p39. The Journal of Biological Chemistry **280** 31522-31529

56 Tatebayashi Y., Planel E., Chui D.-H., Sato S., Miyasaka T., Sahara N., Murayama M., Kikuchi N., Yoshioka K., Rivka R. and Takashima A. (2006) c-jun N-terminal kinase hyperphosphorylates R406W tau at the PHF-1 site during mitosis. The FASEB Journal **20** 762-764

57 Goedert M., Satumtira S., Jakes R., Smith M.J., Kamibayashi C., White III C.L. and Sontag E. (2000) Reduced binding of protein phosphatase 2A to tau protein with frontotemporal dementia and parkinsonism linked to chromosome 17 mutations. Journal of Neurochemistry **75** 2155-2162

58 Goedert M., Jakes R. and Crowther R.A. (1999) Effects of frontotemporal dementia FTDP-17 mutations on heparin-induced assembly of tau filaments. FEBS Letters **450** 306-311

59 Chang E., Kim S., Yin H., Nagaraja H.N. and Kuret J. (2008) Pathogenic missense *MAPT* mutations differentially modulate tau aggregation propensity at nucleation and extension steps. Journal of Neurochemistry **107** 1113-1123

60 Li L., von Bergen M., Mandelkow E.-M. and Mandelkow E. (2002) Structure, stability, and aggregation of paired helical filaments from tau protein and FTDP-17 mutants probed by tryptophan scanning mutagenesis. The Journal of Biological Chemistry **277** 41390-41400

61 Gomez-Ramos A., Abad X., Fanarraga L., Bhat R., Zabala J.C. and Avila J. (2004) Expression of an altered form of tau in Sf9 insect cells results in the assembly of polymers resembling Alzheimer's paired helical filament. Brain Research **1007** 57-64

62 Bunker J.M., Kamath K., Wilson L., Jordan M.A. and Feinstein S.C. (2006) FTDP-17 mutations compromise the ability of tau to regulate microtubule dynamics in cells. The Journal of Biological Chemistry **281** 11856-11863

63 Fasulo L., Ugolini G. and Cattaneo A. (2005) Apoptotic effect of caspase-3 cleaved tau in hippocampal neurons and its potentiation by tau FTDP-mutation N279K. Journal of Alzheimer's Disease **7** 3-13

64 LeBoeuf A.C., Levy S.F., Gaylord M., Bhattachaya A., Singh A.K., Jordan M.A., Wilson L. and Feinstein S.C. (2008) FTDP-17 mutations in tau alter the regulation of microtubule dynamics. The Journal of Biological Chemistry **283** 36406-36415

65 D'Souza I. and Schellenberg G.D. (2000) Determinants of 4-repeat tau expression. Coordination between enhancing and inhibitory splicing sequences for exon 10 inclusion. Journal of Biological Chemistry **275** 17700–17709

66 Black D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing, Annual Review of Biochemistry **72** 291–336

67 Varani L., Hasegawa M., Spillantini M.G., Smith M.J., Murrell J.R., Ghetti B., Klug A., Goedert M. and Varani G. (1999) Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome17. Proceedings of the National Acadamy of Sciences USA **96** 8229–8234

68 Hutton M., Lendon C.L., Rizzu P., Baker M., Froelich S., Houlden H., Pickering-Brown S., Chakraverty S., Isaacs A., Grover A., Hackett J., Adamson J., Lincoln S., Dickson D., Davies P., Petersen R.C., Stevens M., de Graaff E., Wauters E., van Baren J., Hillebrand M., Joosse M., Kwon J.M., Nowotny P., Che L.K., Norton J., Morris J.C., Reed L.A., Trojanowski J., Basun H., Lannfelt L., Neystat M., Fahn S., Dark F., Tannenberg T., Dodd P.R., Hayward N., Kowk J.B.J., Schofield P.R., Andreadis A., Snowden J., Craufurd D., Neary D., Owen F., Oostra B.A., Hardy J., Goate A., van Swieten J., Mann D., Lynch T. and Heutink P. (1998) Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. Nature **393** 702–705

69 Spillantini M.G., Murrell J.R., Goedert M., Farlow M.R., Klug A. and Ghetti B. (1998) Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. Proceedings in the National Academy of Science U. S. A. **95** 7737–7741

70 Yamashita T., Tomiyama T., Li Q., Numata H. and Mori H. (2005) Regulation of tau exon 10 splicing by a double stem-loop structure in mouse intron 10. FEBS Letters **579** 241-244

71 Reed R. (1996) Initial splice-site recognition and pairing during pre-mRNA Splicing. Current Opinion in Genetics and Development **6** 215–220

72 Berget S.M. (1995) Exon recognition in vertebrate splicing. Journal of Biological Chemistry **270** 2411 –2414

73 Hartmann A.M., Rujescu D., Giannakouros T., Nikolakaki E., Goedert M., Mandelkow E.-M., Gao Q.S., Andreadis A. and Stamm S. (2001) Regulation of alternative splicing of human tau exon 10 by phosphorylation of splicing factors. Molecular and Cellular Neuroscience **18** 80-90

74 Kar A.,Havlioglu N., Tarn W.Y. and Wu, J.Y. (2006) RBM4 interacts with an intronic element and stimulates tau exon 10 inclusion. Journal of Biological Chemistry **281** 24479 – 24488

75 Wang J., Gao Q. S., Wang Y., Lafyatis R., Stamm S. and Andreadis A. (2004) Tau exon 10, whose missplicing causes frontotemporal dementia, is regulated by an intricate interplay of *cis* elements and trans factors. Journal of Neurochemistry **88** 1078 – 1090.

76 D'Souza I. and Schellenberg G.D. (2005) Regulation of tau isoform expression and dementia. Biochimica et Biophysica Acta **1739** 104-115

77 Malkani R., D'Souza I., Gwinn-Hardy K., Schellenberg G.D., Hardy J. and Momeni P. (2006) A *MAPT* mutation in a regulatory element upstream of exon 10 causes frontotemporal dementia. Neurobiology of Disease **22** 401-403

78 Hasegawa M., Smith M.J., Iijima M., Tabira T. and Goedert M. (1999) FTDP-17 mutations N279K and S305N in tau produce increased splicing of exon 10. FEBS Letters **443** 93-96

79 D'Souza I., Poorkaj P., Hong M., Nochlin D., Lee V.M., Bird T.D. and Schellenberg G.D. (1999) Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. Proceedings of the National Acadamy of Sciences U. S. A. **96** 5598–5603

80 Kondo S., Yamamoto N., Murakami T., Okumura M., Mayeda A. and Imaizumi K. (2004) Tra2 beta, SF2/ASF and SRp30c modulate the function of an exonic splicing enhancer in exon 10 of tau pre-mRNA.Genes. Cells **9** 121–130

81 Grover A., DeTure M., Yen S.H. and Hutton M. (2002) Effects on splicing and protein function of three mutations in codon N296 of tau in vitro. Neuroscience Letters **323** 33–36

82 Spillantini M.G. and Goedert M. (2000) Tau mutations in frontotemporal Dementia FTDP-17 and their relevance for Alzheimer's disease. Biochimica et Biophysica Acta **1502** 110-121

83 Kovacs G.G., Pittman A., Revesz T., Luk C., Lees A., Kiss E., Tariska P., Laszlo L., Molnar K., Molnar M.J., Tolnay M. and de Silva R. (2008) *MAPT* S305I mutation: implication for argyrophilic grain disease. Acta Neuropathology **116** 103-118

84 Stanford P.M., Shepherd C.E., Halliday G.M., Brooks W.S., Schofield P.W., Brodaty H., Martins R.N., Kwok J.B. and Schofield P.R. (2003) Mutations in the tau gene that cause an increase in three repeat tau and frontotemporal dementia. Brain **126** 814–826.

85 Jiang H., Mankodi A., Swanson M. S., Moxley R. T. and Thornton C. A. (2004) Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. Human Molecular Genetics **13** 3079 – 3088

86 Leroy O., Dhaenens C. M., Schraen-Maschke S., Belarbi K., Delacourte A., Andreadis A., Sablonniere B., Buee L., Sergeant N. and Caillet-Boudin M. L. (2006) ETR-3 represses tau exons 2/3 inclusion, a splicing event abnormally enhanced in myotonic dystrophy type I. Journal of Neuroscience Research **84** 852 – 859

87 Leroy O., Wang J., Maurage C.-A., Parent M., Cooper T., Buee L., Sergent N., Andreadis A., Caillet-Boudin M.-L. (2006) Brain-specific change in alternative splicing of tau exon 6 in myotonic dystrophy type 1. Biochimica et Biophysica Acta **1762** 460-467

88 Gallo J.M., Noble W. and Rodriguez Martin T. (2007) RNA and protein – dependent mechanisms in tauopathies: consequences for therapeutic

strategies. Cellular and Molecular Life Sciences **64** 1701-1714

89 Rovelet-Lecrux A., Lecourtois M., Thomas-Anterion C., Le Ber I., Brice A., Frebourg T., Hannequin D. and Campion D. (2009) Partial deletion of the *MAPT* gene: A novel mechanism of FTDP-17. Human Mutation **30** E591-602

90 Foelich S., Houlden H., Rizzu P., Chakraverty S., Baker M., Kwon J., Nowotny P., Isaacs A., Nowotny V., Wauters E., van Baren M.J., Oostra B.A., Hardy J., Lannfelt L., Goate A., Hutton M., Lendon C.L. and Heutink P. (1999) Construction of a detailed physical map and transcript map of the FTDP-17 candidate region on chromosome 17q21. Genomics **60** 129-136

91 Baker M., Litvan I., Houlden H., Adamson J., Dickson D., Perez-Tur J., Hardy J., Lynch T., Bigio E. and Hutton M. (1999) Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Human Molecular Genetics **8** 711–715

92 Cruts M., Rademakers R., Gijselinck I., van der Zee J., Dermaut B., de Pooter T., de Rijk P., Del-Favero J. And van Broeckhoven C. (2005) Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region. Human Molecular Genetics **14** 1753-1762

93 Hayesmoore J.B.G., Bray N.J., Cross W.C., Owen M.J., O'Donovan M.C. and Morris H.R. (2008) The effect of age and the H1c *MAPT* haplotype on *MAPT* expression in human brain. Neurobiology of Aging Article in Press

94 Hardy J., Pittman A., Myers A., Gwinn-Hardy K., Fung H.C., de Silva R., Hutton M. and Duckworth J. (2005) Evidence suggesting that *Homo neanderthalensis* contributed to the H2 *MAPT* haplotype to *Homo sapiens*. Biochemical Society Trannsactions **33** 582-585

95 Zody M.C., Jiang Z., Fung H.C., Antonacci F., Hillier L.W., Cardone M.F., Graves T.A., Kidd J.M., Cheng Z., Abouelleil A., Chen L., Wallis J., Glasscock J., Wilson R.K., Reily A.D., Duckworth J., Ventura M., Hardy

J., Warren W.C. and Eichler E.E. (2008) Evolutionary toggling of the *MAPT* 17q21.31 inversion region. Nature Genetics **40** 1076-1083

96 Pittman A.M., Fung H.C. and de Silva R. (2006) Untangling the tau gene association with neurodegenerative disorders. Human Molecular Genetics **15** R188-195

97 Conrad C., Andreadis A., Trojanowski J.Q., Dickson D.W., Kang D., Chen X., Widerholt W., Hansen L., Masliah E., Thal L.J., Katzman R., Xia Y. and Saitoh T. (1997) Genetic evidence for the involvement of tau in progressive supranuclear palsy. Annals of Neurology 41 277–81

98 Pittman A.M., Myers A.J., Duckworth J., Bryden L., Hanson M., Abou-Sleiman P., Wood N.W., Hardy J., Lees A. and de Silva R. (2004) The structure of the tau haplotype in controls and in progressive supranuclear palsy. Human Molecular Genetics **13** 1267-1274

99 Pittman A.M., Myers A.J., Abou-Sleiman P., Fung H.C., Kaleem M., Marlowe L., Duckworth J., Leung D., Williams D., Kilford L., Thomas N., Morris C.M., Dickson D., Wood N.W., Hardy J., Lees A.J. and de Silva R. (2005) Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranunclear palsy and corticobasal degeneration. Journal of Medical Genetics **42** 837-846

100 Melquist S., Craig D.W., Huentelman M.J., Crook R., Pearson J.V., Baker M., Zismann V.L., Gass J., Adamson J., Szelinger S., Corneveaux J., Cannon A., Coon K.D., Lincoln S., Adler C., Tuite P., Calne D.B., Bigio E.H., Uitti R.J., Wszolek Z.K., Golbe L.I., Caselli R.J., Graff-Radford N., Litvan I., Farrer M.J., Dickson D.W., Hutton M. and Stephan D.A. (2007) Identification of a novel risk locus for progressive supranuclear palsy by a pooled genomewide scan of 500,288 single-nucleotide polymorphisms. The American Journal of Human Genetics **80** 769-778

101 Baba Y., Tsuboi Y., Baker M.C., Uitti R.J., Hutton M.L., Dickson D.W., Farrer M., Putzke J.D., Woodruff B.K., Ghetti B., Murrell J.R., Boeve B.F., Petersen R.C., Verpillat P., Brice A., Delisle M.-B., Rascol O.,

Arima K., Dysken M.W., Yasuda M., Kobayashi T., Sunohara N., Komure O., Kuno S., Sperfeld A.D., Stoppe G., Kohlhase J., Pickering-Brown S., Neary D., Bugiani O. and Wszolek Z.K. (2005) The effect of tau genotype on clinical features in FTDP-17. Parkinsonism and Related Disorders **11** 205-208

102 Ghidoni R., Signorini S., Barniero L., Sina E., Cominelli P., Villa A., Benussi L. and Binneti G. (2006) The H2 *MAPT* haplotype is associated with familial frontotemporal dementia. Neurobiology of Disease **22** 357-362

103 Baba Y., Putzke J.D., Tsuboi Y., Josephs K.A., Thomas N., Wszolek Z.K., Dickson D.W. (2006) Effect of *MAPT* and *APOE* on prognosis of progressive supranuclear palsy. Neuroscience Letters **405** 116-199

104 Laws S.M., Perneczky R., Drzezga A., Diehl-Schmid J., Ibach B., Bauml J., Eisele T., Forstl H., Kurz A. and Riemenschneider M. (2007) Association of the tau haplotype H2 with age at onset and functional alterations of glucose utilization in frontotemporal dementia. American Journal of Psychiatry **164** 1577-1584

105 Myers A.J., Kaleem M., Marlowe L., Pittman A.M., Lees A.J., Fung H.C., Duckworth J., Leung D., Gibson A., Morris C.M., de Silva R. and Hardy J. (2005) The H1c haplotype at the *MAPT* locus is associated with Alzheimer's disease. Human Molecular Genetics **14** 2399-2404

106 Mukherjee O., Kauwe J.S.K., Mayo K., Morris J.C. and Goate A.M. (2007) Haplotype-based association analysis of the *MAPT* locus in Late Onset Alzheimer's disease. BMC Genetics **8** 3-9

107 Abraham R., Sims R., Carrol L., Hollingworth P., O'Donovan M.C., Williams J. and Owen M.J. (2009) An association study of common variation at the *MAPT* locus with late-onset Alzheimer's disease. American Journal of Medical Genetics Part B Neuropsychiatric Disorders DOI: 10.1002/ajmg.b.30951

108 Fujino Y., Wand D.-S., Thomas N., Espinoza M., Davies P. and Dickson

D.W. (2005) Increased frequency of argyrophilic grain disease in Alzheimer disease with 4R tau-specific immunohistochemistry. Journal of Neuropathology and Experimental Neurology **64** 209-214

109 Jones E.L., Margallo-Lana M., Prasher V.P. and Ballard C.G. (2008) The extended tau haplotype and the age of onset of dementia in Down Syndrome. Dementia and Geriatric Cognitive Disorders **26** 199-202

110 Skipper L., Wilkes K., Toft M., Baker M., Lincoln S., Hulihan M., Ross O.A., Hutton M., Aasly J. and Farrer M. (2004) Linkage disequilibrium and association of *MAPT* H1 in Parkinson disease. American Journal Human Genetics **75** 669-677

111 Fung H.C., Xiromerisiou G., Gibbs J.R., Wu Y.R., Eerola J., Gourbali V., Hellstrom O., Chen C.M., Duckworth J., Papadimitriou A., Tienari P.J., Hadjigeorgiou G.M., Hardy J. and Singleton A.B. (2006) Association of tau haplotype-tagging polymorphisms with Parkinson's disease in diverse ethnic Parkinson's disease cohorts. Neurodegenerative Disease **3** 327-333

112 Winkler S., Konig I.R., Lohmann-Hedrich K., Vieregge P., Kostic V. and Klein C. (2007) Role of ethnicity on the association of *MAPT* H1 haplotypes and subhaplotypes in PD. European Journal of Human Genetics **15** 1163-1168

113 Vandrovcova J., Pittman A.M., Malzer E., Abou-Sleiman P.M., Lees A.J., Wood N.W. and de Silva R. (2007) Association of MAPT haplotype-tagging SNPs with sporadic Parkinson's disease. doi:10.1016/j.neurobiolaging.2007.11.019

114 Zabetian C.P., Hutter C.M., Factor S.A., Nutt J.G., Higgins D.S., Griffith A., Roberts J.W., Leis B.C., Kay D.M., Yearout D., Montimurro J.S., Edwards K.L., Samii A. and Payami H. (2007) Association analysis of *MAPT* H1 haplotype and subhaplotypes in Parkinson's disease. Annals of Neurology **62** 137-144

115 Kwok J.B.J., Teber E.T., Loy C., Hallup M., Nicholson G., Mellick G.D.,

Buchanan D.D., Silburn P.A. and Schofield P.R. (2004) Tau haplotypes regulate transcription and are associated with Parkinson's disease. Annals of Neurology **55** 329-334

116 Canu E., Boccardi M., Ghidoni R., Benussi L., Testa C., Pievani M., Bonetti M., Binetti G. and Frisoni G.B. (2009) H1 haplotype of the *MAPT* gene is associated with lower regional gray matter volume in healthy carriers. European Journal of Human Genetics **17** 287-294

117 Williams D.R., Pittman A.M., Revesz T., Lees A.J. and de Silva R. (2007) Genetic variation at the tau locus and clinical syndromes associated with progressive supranuclear palsy. Movement Disorders **22** 895-897

118 Caffrey T.M., Joachim C., Paracchini S., Esiri M.M. and Wade-Martins R. (2006) Haplotype-specific expression of exon 10 at the human locus. Human Molecular Genetics **15** 3529-3537

119 Caffrey T.M., Joachim C. and Wade-Martins R. (2008) Haplotype-specific expression of the N-terminal exons 2 and 3 at the human MAPT locus. Neurobiology of Aging **29** 1923-1929

120 Sundar P.D., Yu C.-E., Sieh W., Steinbart E., Garruto R.M., Oyanagi K., Craig U.-K., Bird T.D., Wijsman E.M., Galaskko D.R. and Schellenberg G.D. (2007) Two sites in the *MAPT* region confer genetic risk for Guam ALS/PDC and dementia. Human Molecualr Genetics **16** 295-306

121 Laws S.M., Friedrich P., Diehl-Schmid J., Muller J., Eisele T., Bauml J., Forstl H., Kurz A. and Riemenschneider M. (2007) Fine mapping of the *MAPT* locus using quantitative trait analysis identifies possible causal variants in Alzheimer's disease. Molecular Psychiatry **12** 510-517

122 Rademakers R., Melquist S., Cruts M., Theuns J., Del-Favero J., Poorkaj P., Baker M., Sleegers K., Crook R., de Pooter T., bel Kacem S., Adamson J., van den Bossche D., van den Broeck M., Gass J., Corsmit E., de Rijk P., Thomas N., Engelborghs S., Heckman M., Litvan I., Crook J., de Deyn P.P., Dickson D., Schellenberg G.D., van Broeckhoven C. and Hutton M.L. (2005) High-density SNP haplotyping suggests altered regulation of tau

gene expression in progressive supranuclear palsy. Human Molecular Genetics **14** 3281-3292

123 Myers A.J., Pittman A.M., Zhao A.S., Rohrer K., Kaleem M., Marlowe L., Lees A., Leung D., McKeith I.G., Perry R.H., Morris C.M., Trojanowski J.Q., Clark C., Karlawish J., Arnold S., Forman M.S., van Deerlin V., de Silva R. and Hardy J. (2007) The *MAPT* H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. Neurobiology of Disease **25** 561-570

124 Sun W. and Jia J. (2009) The +347C promoter allele up-regulates *MAPT* expression and is associated with Alzheimer's disease among the Chinese Han. Neuroscience Letters **450** 340-343

125 Cruchaga C., Vidal-Taboada J.M., Ezquerra, Lorenzo E., Martinez-Lage P., Blazquez M., Tolosa E., The Iberian Atypical Parkinsonism Study Group Researchers, Pastor P. (2009) 5'-upstream variants of *CRHR1* and *MAPT* genes associated with age at onset in progressive supranuclear palsy and corticobasal degeneration. Neurobiology of Disease **33** 164-170

126 Kobayashi H., Ujike H., Hasegawa J., Yamamoto M., Kanzaki A. and Sora I. (2006) Correlation of tau gene polymorphism with age at onset of Parkinson's disease. Neuroscience Letters **405** 202-206

127 Tobin J.E., Latourelle J.C., Lew M.F., Klein C., Suchowersky O., Shill H.A., Golbe L.I., Mark M.H., Growdon J.H., Wooten G.F., Racette B.A., Perlmutter J.S., Watts R., Guttman M., Baker K.B., Goldwurm S., Pezzoli G., Singer C., Saint Hilaire M.H., Hendricks A.E., Williamson S., Nagle M.W., Wilk J.B., Massood J., Laramie M., DeStefano A.L., Litvan I., Nicholson G., Corbett A., Isaacson S., Burn D.J., Chinnery P.F., Pramstaller P.P., Sherman S., Al-hinti J., Drasby E., Nance M., Moller A.T., Ostergaard K., Roxburgh R., Snow B., Slevin J.T., Cambi F., Gusella J.F. and Myers R.H. (2008) Haplotypes and gene expression implicate the *MAPT* region for Parkinson disease: The GenePD Study. Neurology **71** 28-34

128 http://www.nanodrop.com/nd-1000-overview.html

129 http://www.nanodrop.com/techsupport/nd-1000-users-manual.pdf

130 K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, H. Erlich (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor Symposia on Quantitative Biology **51** Pt 1 263-273

131 Hemat F. and McEntee K. (1994) A rapid and efficient PCR-based method for synthesizing high-molecular weight multimers of oligonucleotides. Biochemical and Biophysical Research Communications **205** (1) 475-481

132 Dignam J.D., Lebovitz R.M. and Roeder R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Research **11** 1475-1489

133 Fried M. and Crothers D.M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gele electrophoresis. Nucleic Acids Research **9** 6505-6525

134 Kerr L.D. (1995) Electrophoretic mobility shift assay. Methods in Enzymology **254** 619-632

135 Ludwig L.B., Hughes B.J. and Schwartz S.A. (1995) Biotinylated probes in the electrophorectic mobility shift assay to examine specific dsDNA, ssDNA or RNA-protein interactions. Nucleic Acids Research **23** 3792-3793

136 Bannister A.J. and Kouzarides T. (1992) Basic peptides enhance protein/DNA interaction *in vitro*. Nucelic Acids Research **20** 3523

137 Biedler J.L., Helson L. and Spengler B.A. (1973) Morphology and growth, tumorigenicity and cytogenetics of human neuroblastoma cells in continuous culture. Cancer Research **33** 2643-2652

138 Biedler J.L., Roffler-Tarlov S., Schachner M. and Freedman L.S. (1978) Multiple neurotransmitter synthesis by human neuroblastoma cell lines and clones. Cancer Research **38** 3751-3575

139 Ciccarone V., Spengler B.A., Meyers M.B., Biedler J.L. and Ross R.A.

(1989) Phenotypic diversification in human neuroblastoma cells: expression of distinct neural crest lineages. Cancer Research **49** 219-225

140 Jämsa A., Hasslund K., Cowburn R.F., Bäcström A. and Vasänge M. (2004) The retinoic acid and brain-derived neurotrophic factor differentiated SH-SY5Y cell line as a model for Alzheimer's disease-like tau phosphorylation. Biochemical and Biophysical Research Communications **319** 993-1000

141 Roche Technical Manual: DNA-Binding Protein Purification Kit (https://www.roche-applied-science.com/pack-insert/1835513a.pdf)

142 Waters Micromass MALDI Q-TOF Premier Operator's Guide (http://www.waters.com/webassets/cms/support/docs/71500106402ra.pdf)

143 Heinemeyer T., Wingender E., Reuter I., Hermjakob H., Kel A.E., Kel O.V., Ignatieva E.V., Ananko E.A., Podkolodnaya O.A., Kolpakov F.A., Podkolodny N.L. and Kolchanov N.A. (1998) Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Research 26 362-367

144 Quandt K., Frech K., Karas H., Wingender E. And Werner T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Research **23** 4878-4884

145 Cartharius K., Frech K., Grote K., Klocke B., Haltmeier M., Klingenhoff A., Frisch M., Bayerlein M. And Wener T. (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics **21** 2933-2942

146 Perkins D.N., Pappin D.J.C., Creasy D.M. and Cottrell J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20** 3551-3567

147 Jensen L.J., Kuhn M., Stark M., Chaffron S., Creevey C., Muller J., Doerks T., Julien P., Roth A., Simonovic M., Bork P. and von Mering C. (2009) STRING 8 – a global view on proteins and their functional

interactions in 630 organisms. Nucleic Acids Reseach **37** D412-D416

148 von Mering C., Jensen L.J., Snel B., Hooper S.D., Krupp M., Foglierini M., Jouffre N., Huynen M.A. and Bork P. (2005) STRING: known and predicted protein-protein associations, integrated and transfected across organisms. Nucleic Acids Research **33** D433-D437

149 von Mering C., Jensen L.J., Kuhn M., Chaffron S., Doerks T., Kruger B., Snel B. and Bork P. (2007) STRING 7 – recent developments in the integration and prediction of protein interactions. Nucleic Acids Research **35** D358-362

150 Vastrik I., D'Eustachio P., Schmidt E., Joshi-Tope G., Gopinath G., Croft D., de Bono B., Gillespie M., Jassal B., Lewis S., Matthews L., Wu G., Birney E. and Stein L. (2007) Reactome; a knowledge base of biologic pathways and processes. Genome Biology **8** R39.1-R39.13

151 Croft D., O'Kelly G., Wu G., Haw R., Gillespie M., Matthews L., Caudy M., Garapati G., Jassal B., Jupe S., Kalatskaya I., Mahajan S., May B., Ndegwa N., Schmidt E., Shamovsky V., Yung C., Birney E., Hermjakob H., D'Eustachio P. and Stein L. (2011) Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Research **39** D691-D697

152 Matthews L., Gopinath G., Gillespie M., Caudy M., Croft D., de Bono B., Garapati P., Hemish J., Hermjakob H., Jassal B., Kanapin A., Lewis S., Mahajan S., May B., Schmidt E., Vastrik I., Wu G., Birney E., Stein L. and D'Eustachio P. (2009) Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Research **37** D619-D622

153 Promega Technical Bulletin: TransFast™ Transfection Reagent (http://www.promega.com/tbs/tb260/tb260.pdf)

154 Promega Technical Manual: Luciferase Assay System (http://www.promega.com/tbs/tb281/tb281.pdf)

155 Promega Technical Manual: Renilla Luciferase Assay System (http://www.promega.com/tbs/tm055/tm055.pdf)

156 Milne T.A., Zhao K. and Hess J.L. (2009) Chromatin immunoprecipitation (ChIP) for analysis of histone modifications and chromatin-associated proteins. Methods in Molecular Biology **538** 409-423

157 Pillai S., Dasgupta P. and Chellappan S.P. (2009) Chromatin immunoprecipitation assays: analyzing transcription factor binding and histone modifications in vivo. Methods in Molecular Biology **523** 323-339

158 Massie C.E. and Mills I.G. (2008) ChIPping away at gene regulation. EMBO Reports **9** 337-343

159 Höglinger G., Mehlem N.M., Dickson D., Sleiman P.M.A., Wang L.S., Klei L., et. al. (2011) Common variants affect risk for the tauopathy progressive supranuclear palsy. Nature Genetics **43** 699-705

160 McCracken S., Fong N., Yankulov K., Ballantyne S., Pan G., Greenblatt J., Patterson S.D., Wickens M. and Bentley D.L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature **385** 357-361

161 Ahn S.H., Kim M. and Buratowski S. (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. Molecular Cell **13** 67-76

162 West A.B., Maraganore D., Crook J., Lesnick T., Lockhart P.J., Wilkes K.M., Kapatos G., Hardy J.A. and Farrer M.J. (2002) Functional association of the *parkin* gene promoter with idiopathic Parkinson's disease. Human Molecular Genetics **11** 2787-2792

163 Ma S.L., Tang N.L.S., Tam C.W.C., Lui V.W.C., Lam L.C.W., Chiu H.F.K., Driver J.A., Pastorino L. and Lu K.P. (2010) A PIN1 polymorphism that prevents its suppression by AP4 associated with delayed onset of Alzheimer's disease. Neurobiology of Aging doi:10.1016/j.neurobiolaging.2010.05.018

164 Cong L. and Jia J. (2011) Promoter polymorphisms which regulate ADAM9 transcription are protective against sporadic Alzheimer's disease. Neurobiology of Aging **32** 54-62

165 Zhou X., Barrett T.B. and Kelsoe J.R. (2008) Promoter variant in the GRK3 gene associated with bipolar disorder alters gene expression. Biological Psychiatry **64** 104-110

166 Scheuch K., Lautenschlager M., Grohmann M., Stahlberg S., Kirchheiner J., Zill P., Heinz A., Walther D.J. and Priller J. (2007) Characterization of a function promoter polymorphism of the human tryptophan hydroxylase 2 gene in serotogenic raphe neurons. Biological Psychiatry **62** 1288-1294

167 Bassett A.S., Buyske S., Millonig J.H., Vieland V.J., and Brzustowicz L.M. (2009) Identification of a schizophrenia-associated functional noncoding variant in *NOS1AP*. American Journal of Psychiatry **166** 434-441

168 Poitras L., Yu M., Lesage-Pelletier C., MacDonald R.B., Gagné J.P., Hatch G., Kelly I., Hamilton S.P., Rubenstein J.L.R., Poirier G.G. and Ekker M. (2010) An SNP in an ultraconserved regulatory regulatory element affects Dlx5/Dlx6 regulation in the forebrain. Development **137** 3089-3097

169 Tapia-Páez I., Tammimies K., Massinen S., Roy A.L. and Kere J. (2008) The complex of TFII-I, PARP1, and SFPQ proteins regulates the *DYX1C1* gene implicated in neuronal migration and dyslexia. The FASEB Journal **22** 3001-3009

170 Zhang H., Gelernter J., Gruen J.R., Kranzler H.R., Herman A.I. and Simen A.A. (2010) Functional impact of a single-nucleotide polymorphism in the *OPRD1* promoter region. Journal of Human Genetics **55** 278-284

171 Matsushita T., Ashikawa K., Yonemoto K., Hirakawa Y., Hata J., Amitani H., Doi Y., Ninomiya T., Kitazono T., Ibayashi S., Lida M., Nakamura Y., Kiyohara Y. and Kubo M. (2010) Functional SNP of *ARHGEF10* confers risk of atherothrombotic stroke. Human Molecular Genetics **19** 1137-1146

172 Garcia-Blanco M.A., Jamison S.F. and Sharp P.A. (1989) Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. Gene and Development **3** 1874-1886

173 Zhang L., Liu W. and Grabowski P.J. (1999) Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB. RNA **5** 117-130

174 Resnick M., Segall A., Rozic-Kotliroff G., Lupowitz Z., Zisapel N. (2008) Alternative splicing of neuroxins: a role for neuronal polypyrimidine tract binding protein. Neuroscience Letters **439** 235-240

175 Polydorides A.D., Okano H.J., Yang Y.Y.L., Stefani G. and Darnell R.B. (2000) A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specfic alternative splicing. Proceedings of the National Academy of Sciences **97** 6350-6355

176 Kikuchi T., Ichikawa M., Arai J., Tateiwa H., Fu L., Higuchi K. and Yoshimura N. (2000) Molecular cloning and characterization of a new neuron-specific homologue of rat polypyrimidine tract binding protein. Journal of Biochemistry **128** 811-821

177 Bonano V.I., Oltean S., Brazas R.M. and Garcia-Blanco M.A. (2006) Imagine the alternative silencing of FGFR2 exon IIIb in vivo. RNA **12** 2073-2079

178 Sauliere J., Sureau A., Expert-Bezancon A. and Marie J. (2006) The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the β-tropomyosin pre-mRNA by directly interfering with the binding of U2AF65 subunit. Molecular and Cellular Biology **26** 8755-8769

179 Sharma S., Falick A.M. and Black D.L. (2005) Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. Molecular Cell **19** 485-496

180 Lou H., Helfman D.M., Gagel R.F. and Berget S.M. (1999) Polypyrimidine tract-binding protein regulates inclusion of an alternative 3'-terminal exon. Molecular Cellular Biology **19** 78-85

181 Spellman R. and Smith C.W.J. (2006) Novel modes of splicing repression by PTB. Trends in Biochemical Sciences **31** 73-76

182 Ma S., Liu G., Sun Y. and Xie J. (2007) Relocalization of the polypyrimidine tract-binding protein during PKA-induced neurite growth. Biochemica et Biophysica Acta

183 Ruskin B., Zamore P.D. and Green M.R. (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell **52** 207-219

184 Valcarcel J., Gaur R.K., Singh R., and Green M.R. (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. Science **273** 1706-1709

185 Zamore P.D., Patton J.G. and Green M.R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. Nature **355** 609-614

186 Thomas J.O. (2001) HMG1 and HMG2: architectural DNA-binding proteins. Biochemical Society Transactions **29** 395-401

187 Das D., Peterson R.C. and Scovell W.M. (2004) High mobility group B proteins facilitate strong estrogen receptor binding to classical and half-site estrogen response elements and relax binding selectivity. Molecular Endocrinology **18** 2616-2632

188 Travers A.A. (2003) Priming the nucleosome:a role for HMGB proteins? EMBO Reports **4** 131-136

189 Balani P., Boulaire J., Zhao Y., Zeng J., Lin J. And Wang S. (2009) High mobility group box 2 promoter-controlled suicide gene expression enables targeted glioblastoma treatment. Molecular Therapy doi: 10.1038/mt.2009.22

190 Vawter M.P., Atz M.E., Rollins B.L., Cooper-Casey K.M., Shao L. and Byerley W.F. (2006) Genome scans and gene expression microarrays converge to identify gene regulatory loci relevant in schizophrenia. Human Genetics **119** 558-570

191 Albig W. and Doenecke D. (1997) The human histone gene cluster at the D6S105 locus. Human Genetics **101** 284-294

192 Garcia B.A., Busby S.A., Barber C.M., Shabanowitz J., Allis C.D. and

Hunt D.F. (2004) Characterization of phosphorylation sites on Histone H1 isoforms by tandem mass spectrometry. Journal of Proteome Research **3** 1219-1227

193 Vignali M. and Workman J.L. (1998) Location and function of linker histones. Nature Structural Biology **5** 1025-1028

194 Wagner B., DeMaria C.T., Sun Y., Wilson G.M. and Brewer G. (1998) Structure and genomic organization of the human AUF1 gene: alternative pre-mRNA splicing generates four protein isoforms. Genomics **48** 195-202

195 DeMaria C.T. and Brewer G. (1996) AUF1 binding affinity to A+U-rich elements correlates with rapid mRNA degradation. The Journal of Biological Chemistry **271** 12179-12184

196 Kukalev A., Nord Y., Palmberg C., Bergman T. and Percipalle P. (2005) Actin and hnRNP U for productive transcription by RNA polymerase II. Nature **12** 238-244

197 Kim M.K. and Nikodem V.M. (1999) hnRNP U inhibits carboxy-terminal domain phosphorylation by TFIIH and represses RNA polymerase II elongation. Molecular and Cellular Biology **19** 6833-6844

198 Kiledjian M. and Dreyfuss G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. EMBO Journal **11** 2655-2664

199 Martens J.H., Verlaan M., Kalkhoven E., Dorsman J.C. and Zantema A. (2002) Scaffold / matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes. Molecular and Cellular Biology **22** 2598-2606

200 Bruhn L., Munnerlyn A. and Grosschedl R. (1997) ALY, a context dependent coactivator of LEF-1 and AML-1, is required for TCRalpha enhancer function. Genes and Development **11** 640-653

201 Virbasius C.M., Wagner S. and Green M.R. (1999) A human nuclear-localized chaperone that regulates dimerization, DNA binding, and

transcriptional activity of bZIP proteinis. Molecular Cell **4** 219-228

202 Reichert V.L., Le Hir H., Jurica M.S. and Moore M.J. (2002) 5' exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly. Genes and Development **16** 2778-2791

203 Swinburne I.A., Meyer C.A., Liu X.S., Silver P.A. and Brodsky A.S. (2006) Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription. Genome Research **16** 912-921

204 Lane N.J. (1969) Intranuclear fibrillar bodies in actinomycin D-treated oocytes. Journal of Cell Biology **40** 286-291

205 Scheer U., Hinssen H., Franke W.W. and Jockusch B.M. (1984) Microinjection of actin-binding proteins and actin antibodies demonstrates involvement of nuclear actin in transcription of Lampbrush chromosomes. Cell **39** 111-122

206 Egly J.M., Miyamoto N.G., Moncollin V. and Chambon P. (1984) Is actin a transcription initiation factor for RNA polymerase B? EMBO Journal **3** 2363-2371

207 Vartiainen M.K., Guettler S., Larijani B. and Treisman R. (2007) Nuclear actin regulates dynamic subcellular localization and activity of the SRF cofactor MAL. Science **316** 1749-1752

208 Shen X., Ranallo R., Choi E. and Wu C. (2003) Involvement of actin related proteins in ATP-dependent chromatin remodeling. Molecular Cell **12** 147-155

209 Dingova H., Fukalova J., Maninova M., Philimonenko V.V. and Hozak P. (2009) Ultrastructural localization of actin and actin-binding proteins in the nucleus. Histochemistry and Cell Biology **131** 425-434

210 Hu P., Wu S. and Hernandez N. (2004) A role for beta-actin in RNA polymerase III transcription. Genes and Development **18** 3010-3015

211 Philimonenko V.V., Zhao J., Iben S., Dingova H., Kysela K., Kahle M.,

Zentgraf H., Hoffmann W.A., de Lanerolle P., Hozak P. and Grummt I. (2004) Nuclear actin and myosin I are required for RNA polymerase I transcription. Nature Cell Biology **6** 1165-1172

212 Percipalle P., Fomproix N., Kylberg K., Miralles F., Bjorkroth B., Daneholt B. and Visa N. (2003) An actin-ribonucleoprotein interaction is involved in transcription by RNA polymerase II. Proceedings of the National Academy of Sciences U.S.A. **100** 6475-6480

213 Sjolinder M., Bjork P., Soderberg E., Sabri N., Ostlund Farrants A.K. and Visa N. (2005) The growing pre-mRNA recruits actin and chromatin modifying factors to transcriptionally active genes. Genes and Development **19** 1871-1884

214 Obrdlik A., Kukalev A. and Percipalle P. (2007) The function of actin in gene transcription. Histology and Histopathology **22** 1051-1055

215 Von Lindern M., Fornerod M., Soekarman N., van Baal S., Jaegle M., Hagemeijer A., Bootsma D. and Grosveld G. (1992) Translocation t(6;9) in acute non-lymphocytic leukaemia results in the formation of a DEK-CAN fusion gene. Baillière's Clinical Haematology **5** 857-879

216 Von Lindern M., Fornerod M., van Baal S., Jaegle M., de Wit T., Buijs A. and Grosveld G. (1992) The translocation (6;9), associated with a specific subtype of acute myeloid leukemia, results in the fusion of two genes, dek and can, and the expression of a chimeric, leukemia-specific dek-can mRNA. Molecular and Cellular Biology **12** 1687-1697

217 Kroes R.A., Jastrow A., McLone M.G., Yamamoto H., Colley P., Kersey D.S., Yong V.W., Mkrdichian E., Cerullo L., Leestma J. and Moskal J.R. (2000) The identification of novel therapeutic targets for the treatment of malignant brain tumors. Cancer Letters **156** 191-198

218 Kappes F., Burger K., Baack M., Fackelmayer F.O. and Gruss C. (2001) Subcellular localization of the human proto-oncogene protein DEK. The Journal of Biological Chemistry **276** 26317-26323

219 Waldmann T., Eckerich C., Baack M. and Gruss C. (2002) The ubiquitous

chromatin protein DEK alters the structure of DNA by introducing positive supercoils. The Journal of Biological Chemistry **277** 24988-24994

220 Le Hir H., Izaurralde E., Maquat L.E. and Moore M.J. (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. The EMBO Journal **19** 6860-6869

221 McGarvey T., Rosonina E., McCracken S., Li Q., Arnaout R., Mientjes E., Nickerson J.A., Awrey D., Greenblatt J., Grosveld G. and Blencowe B.J. (2000) The acute myeloid leukemia-associated protein, DEK, forms a splicing-dependent interaction with exon-product complexes. The Journal of Cell Biology **150** 309-320

222 Kauppinen T.M., Chan W.Y., Suh S.W., Wiggins A.K., Huang E.J. and Swanson R.A. (2006) Direct phosphorylation and regulation of poly(ADP-ribose) polymerase-1 by extracellular signal-regulated kinases 1/2. Proceedings of the National Academy of Sciences **103** 7136-7141

223 Hakmé A., Wong H.-K., Dantzer F. and Schreiber V. (2008) The expanding field of poly(ADP-ribosyl)ation reactions. EMBO Reports **9** 1094-1100

224 Diaz-Hernandez J., Moncada S., Bolanos J.P. and Almeida A. (2007) Poly(ADP-ribose) polymerase-1 protects neurons against apoptosis induced by oxidative stress. Cell death and Differentiation **14** 1211-1221

225 Cosi C. and Marien M. (1998) Decreases in mouse brain NAD^+ and ATP induced by 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP): prevention by the poly(ADP-ribose) polymerase inhibitor, benzamide. Brain Research **809** 58-67

226 Outeiro T.F., Grammatopoulos T.N., Altmann S., Amore A., Standaert D.G., Hyman B.T. and Kazantsev A.G. (2007) Pharmacological inhibition of PARP-1 reduces α-synuclein and MPP^+-induced cytotoxicity in Parkinson's disease *in vitro* models. Biochemical and Biophysical Research Communications **357** 596-602

227 Mandir A.S., Przedborski S., Jackson-Lewis V., Wang Z.-Q., Simbulan-

Rosenthal C.M., Smulson M.E., Hoffman B.E., Guastella D.B., Dawson V.L. and Dawson T.M. (1999) Poly(ADP-ribose) polymerase activation mediates 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-induced parkinsonism. Proceedings of the National Academy of Sciences USA **96** 5774-5779

228 Fox A.H., Lam Y.W., Leung A.K., Lyon C.E., Andersen J., Mann M. and Lamond A.I. (2002) Paraspeckles: a novel nuclear domain. Current Biology **12** 13-25

229 Fox A.H., Bond C.S. and Lamond A.I. (2005) p54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. Molecular Biology of the Cell **16** 5304-5315

230 Shav-Tal Y. and Zipori D. (2002) PSF and p54(nrb)/NonO-multiprotein functional nuclear proteins. FEBS Letters **531** 109-114

231 Kameoka S., Duque P. and Konarska M.M. (2004) p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. EMBO Journal **23** 1782-1791

232 Xu J., Zhong N., Wang H., Elias J.E., Kim C.Y., Woldman I., Pifl C., Gygi S.P., Geula C. and Yankner B.A. (2005) The Parkinson's disease-associated DJ-1 protein is a transcriptional co-activator that protects against neuronal apoptosis. Human Molecular Genetics **14** 1231-1241

233 Lim L.C., Swendeman S.L. and Sheffery M. (1992) Molecular cloning of the α-globin transcription factor CP2. Molecular and Cellular Biology **12** 828-835

234 Lim L.C., Fang L., Swendeman S.L. and Sheffery M. (1993) Characterization of the molecularly cloned murine α-globin transcription factor CP2. The Journal of Biological Chemistry **268** 18008-18017

235 Ramamurthy L., Barbour V., Tuckfield A., Clouston D.R., Topham D., Cunningham J.M. and Jane S.M. (2001) Targeted disruption of the CP2 gene, a member of the NTF family of transcription factor. The Journal of Biological Chemistry **276** 7836-7842

236 Lambert J.C., Goumidi L., Vrièze F.W.D., Frigard B., Harris J.M., Cummings A., Coates J., Pasquier F., Cottel D., Gaillac M., St. Clair D., Mann D.M.A., Hardy J., Lendon C.L., Amouyel P. and Chartier-Harlin M.C. (2000) The transcriptional factor LBP1c/CP2/LSF gene on chromosome 12 is a genetic determinant of Alzheimer's disease. Human Molecular Genetics **9** 2275-2280

237 Taylor A.E., Yip A., Brayne C., Easton D., Evans J.G., Xuereb J., Cairns N., Esiri M.M. and Rubinsztein D.C. (2001) Genetic association of an LBP-1c/CP2/LSF gene polymorphism with late onset Alzheimer's disease. Journal of Medical Genetics **38** 232-233

238 Luedecking-Zimmer E., DeKosky S.T., Nebes R. and Kamboh M.I. (2003) Association of the 3' UTR transcription factor LBP-1c/CP2/LSF polymorphism with late-onset Alzheimer's disease. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) **117B** 114-117

239 Panza F., D'Introno A., Colacicco A.M., Capurso C., Basile A.M., Torres F., Capurso A. and Solfrizzi V. (2004) LBP-1c/CP2/LSF gene polymorphism and risk of sporadic Alzheimer's disease. Journal of Neurology and Neurosurgery **75** 166-168

240 Bertram L., Parkinson M., McQueen M.B., Mullin K., Hsiao M., Menon R., Moscarillo T.J., Blacker D. and Tanzi R.E. (2005) Further evidence for LBP-1c/CP2/LSF association in Alzheimer's disease families. Journal of Medical Genetics **42** 857-862

241 Chang K.-A., Kim H.S., Ha T.Y., Ha J.-W., Shin K.Y., Jeong Y.H., Lee J.-P., Park C.-H., Kim S., Baik T.-K., and Suh Y.-H. (2006) Phosphorylation of amyloid precursor protein (APP) at Thr668 regulates the nuclear translocation of the APP intracellular domain and induces neurodegeneration. Molecular and Cellular Biology **26** 4327-4338

242 Xu Y., Kim H.-S., Joo Y., Choi Y., Chang K.-A., Park C.H., Shin K.-Y., Kim S., Cheon Y.-H., Baik T.-K., Kim J.-H. and Suh Y.-H. (2007) Intracellular domains of amyloid precursor-like protein 2 interact with

CP2 transcription factor in the nucleus and induce glycogen synthase kinase-3β expression. Cell death and Differentiation **14** 79-91

243 Jang S.-M., Kim J.-W., Kim C.-H., An J.-H., Kang E.-J., Kim C.-G., Kim H.-J. and Choi K.-H. (2010) Control of transferring expression by β-amyloid through the CP2 transcription factor. FEBS Journal **277** 4054-4065

244 Schuster C., Myslinski E., Krol A. and Carbon P. (1995) Staf, a novel zinc finger protein that activates the RNA polymerase III promoter of the selenocysteine tRNA gene. The EMBO Journal **14** 3777-3787

245 Schaub M., Krol A. and Carbon P. (2000) Structural organization of Staf-DNA complexes. Nucleic Acids Research **28** 2114-2121

246 Tommerup N. and Vissing H. (1995) Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. Genomics **27** 259-264

247 Wales M.M., Biel M.A., Deiry W.E., Nelkin B.D., Issa J.-P., Cavanee W.K., Kuerbitz S.J. and Baylin S.B. (1995) p53 activates expression of HIC-1, a new candidate tumour suppressor gene on 17p13.3. Nature Medicine **1** 570-577

248 Uhlmann K., Rohde K., Zeller C., Syzmas J., Vogel S., Marczinek K., Thiel G., Nürnberg and Laird P.W. (2003) Distinct methylation profiles of glioma subtypes. International Journal of Cancer **106** 52-59

249 Pinte S., Stankovic-Valentin N., Deltour S., Rood B.R., Guérardel and Leprince D. (2004) The tumor suppressor gene HIC1 (Hypermethylated in Cancer 1) is a sequence-specific transcriptional repressor. The Journal of Biological Chemistry **279** 38313-38324

250 Fleuriel C., Touka M., Boulay G., Guérardel C., Rood B.R. and Leprince D. (2009) HIC1 (Hypermethylated in Cancer 1) epigenetic silencing in tumors. The International Journal of Biochemistry and Cell Biology **41** 26-33

251 The tumor suppressor HIC1 (hypermethylated in cancer 1) is O-GlcNAc

glycosylated. European Journal of Biochemistry **271** 3843-3854

252 Neumann M., Sampathu D.M., Kwong L.K., Truax A.C., Micsenyi M.C., Chou T.T., Bruce J., Schuck T., Grossman M., Clark C.M., McCluskey L.F., Miller B.L., Masliah E., Mackenzie I.R., Feldman H., Feiden W., Kretzschmar H.A., Trojanowski J.Q. and Lee V.M.Y. (2006) TDP-43 is the major disease protein ubiquinated in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science **314** 130-133

253 Leverenz J.B., Yu C.E., Montine T.J., Steinbart E., Bekris L.M., Zabetian C., Kwong L.K., Lee V.M-Y., Schellenberg G.D. and Bird T.D. (2007) A novel progranulin mutation associated with variable clinical presentation and tau, TDP43 and alpha-synuclein pathology. Brain **130** 1360-1374

254 Liscic R.M., Grinberg L.T., Zidar J., Gitcho M.A. and Cairns N.J. (2008) ALS and FTLD: two faces of TDP-43 proteinopathy. European Journal of Neuorology **15** 772-780

255 Lagier-Tourenne C. and Cleveland D.W. (2009) Rethinking ALS: The FUS about TDP-43. Cell **136** 1001-1004

256 Lagier-Tourenne C., Polymenidou M. and Cleveland D.W. (2010) TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Human Molecular Genetics **19** R46-R64

257 Ou S.H., Wu F., Harrich D., Garcia-Martinez L.F., and Gaynor R.B. (1995) Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. Journal of Virology **69** 3584-3596

258 Ayala Y.M., Pantano S., D'Ambrogio A., Buratti E., Brindisi A., Marchetti C., Romano M. and Baralle F.E. (2005) Human, *Drosophila*, and *C. elegans* TDP43: Nucleic acid binding properties and splicing regulatory function. Journal of Molecular Biology **348** 575-588

259 Strong M.J., Volening K., Hammond R., Yang W., Strong W., Leystra-Lantz C. and Shoesmith C. (2007) TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. Molecular and

Cellular Neuroscience **35** 320-327

260 Buratti E. and Baralle F.E. (2010) The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation. RNA Biology **7** 420-429

261 Mansouri A., Hallonet M. and Gruss P. (1996) *Pax* genes and their roles in cell differentiation and development. Current Opinion in Cell Biology **8** 851-857

262 Poleev A., Okladnova O., Musti A.-M., Schneider S., Royer-Pokora B. and Plachov D. (1997) Determination of functional domains of the human transcription factor PAX8 responsible for its nuclear localization and transactivating potential. European Journal of Biochemistry **247** 860-869

263 Kozmik Z., Kurzbauer R., Dörfler P. and Busslinger M. (1993) Alternative splicing of *Pax-8* gene transcripts is developmentally regulated and generates isoforms with different transactivation properties. Molecular and Cellular Biology **13** 6024-6035

264 Poleev A., Fickenscher H., Mundlos S., Winterpacht A., Zabel B., Fidler A., Gruss P. and Plachov D. (1992) PAX8, a human paired box gene: isolation and expression in developing thyroid, kidney and Wilms' tumors. Development **116** 611-623

265 Holst B.D., Goomer R.S., Wood I.C., Edelman G.M. and Jones F.S. (1994) Binding and activation of the promoter for the neural cell adhesion molecule by Pax-8. The Journal of Biological Chemistry **269** 22245-22252

266 Dehbi M. and Pelletier J. (1996) PAX8-mediated activation of the wt1 tumor suppressor gene. The EMBO Journal **15** 4297-4306

267 Frazier G.C., Shimamura R., Zhang X. and Saunders G.F. (1997) PAX8 regulates human *WT1* transcription through a novel DNA binding site. The Journal of Biological Chemistry **272** 30678-30687

268 Esposito C., Miccadei S., Saiardi A. and Civitareale D. (1998) PAX 8 activates the enhancer of the human thyroperoxidase gene. Biochemistry Journal **331** 37-40

269 Chen Y.-J., Campbell H.G., Wiles A.K., Eccles M.R., Reddel R.R., Braithwaite A.W. and Royds J.A. (2008) PAX8 regulates telomerase reverse transcriptase and telomerase RNA component in glioma. Cancer Research **68** 5724-5732

270 Wei M.-L., Memmott J., Screaton G. and Andreadis A. (2000) The splicing determinants of a regulated exon in the axonal MAP tau reside within the exon and in its upstream intron. Molecular Brain Research **80** 207-218

271 Wang J., Tse S.-W. and Andreadis A. (2007) Tau exon 6 is regulated by an intricate interplay if *trans* factors and *cis* elements, including multiple branch points. Journal of Neurochemistry **100** 437-445

272 Wang J., Gao Q.-S., Wang Y., Lafyatis R., Stamm S. and Andreadis A. (2004) Tau exon 10, whose missplicing causes frontotemporal dementia, is regulated by an intricate interplay of *cis* elements and *trans* factors. Journal of Neurochemistry **88** 1078-1090

273 Wang Y., Wang J., Gao L., Stamm S. and Andreadis A. (2011) An SRp75/hnRNPG complex interacting with hnRNPE2 regulates the 5' splice site of tau exon 10, whose misregulation causes frontotemporal dementia. Gene **485** 130-138

274 Broderick J., Wang J. and Andreadis A. (2004) Heterogeneous nuclear ribonucleoprotein E2 binds to tau exon 10 and moderately activates its splicing. Gene **331** 107-114

275 Wang Y., Gao L., Tse S.-W. and Andreadis A. (2010) Heterogeneous nuclear ribonucleoprotein E3 modestly activates splicing of tau exon 10 via its proximal downstream intron, a hotspot for frontotemporal dementia mutations. Gene **451** 23-31

276 Liu Y. and Szaro B.G. (2011) hnRNP K post-transcriptionally co-regulates multiple cytoskeletal genes needed for axonogenesis. Development **138** 3079-3090

277 Xu Y., Kim H.-S., Joo Y., Choi Y., Chang K.-A., Park C.H., Shin K.-Y.,

Kim S., Cheon Y.-H., Baik T.-K., Kim J.-H. and Suh Y.-H. (2007) Intracellular domains of amyloid precursor-like protein 2 interact with CP2 transcription factor in the nucleus and induce glycogen synthase kinase-3β expression. Cell Death and Differentiation **14** 79-91

278 Kumaran R.L., Thakar R. and Spector D.L. (2008) Chromatin dynamics and gene positioning. Cell **132** 929-934

279 Schwartz S. and Ast G. (2010) Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. The EMBO Journal **29** 1629-1636

280 Schwartz S., Meshorer E. and Ast G. (2009) Chromatin organization marks exon-intron structure. Nature Structural & Molecular Biology **16** 990-996

281 Göndör A. and Ohlsson R. (2009) Chromosome crosstalk in three dimensions. Nature **461** 212-217

282 Morrison A.J. and Shen X. (2009) Chromatin remodeling beyond transcription: the INO80 and SWR1 complexes. Nature Reviews Molecular Cell Biology **10** 373-384

283 Kashyap V. and Gudas L.J. (2010) Epigenetic regulatory mechanisms distinguish retinoic acid-mediated transcriptional responses in stem cells and fibroblasts. The Journal of Biological Chemistry **285** 14534-14548

284 Kaplan D.R., Matsumoto K., Lucarelli E. and Thiele C.J. (1993) Induction of TrkB by retinoic acid mediates biologic responsiveness to BDNF and differentiation of human neuroblastoma cells. Neuron **11** 321-331

285 Hashemi S.H., Li J.-Y., Ahlman H. and Dahlström A. (2003) $SSR_{2(a)}$ receptor expression and adrenergic/cholinergic characteristics in differentiated SH-SY5Y cells. Neurochemical Research **28** 449-460

286 Hanada M., Krajewski S., Tanaka S., Cazals-Hatem D., Spengler B.A., Ross R.A., Biedler J.L. and Reed J.C. (1993) Regulation of Bcl-2 oncoprotein levels with differentiation of human neuroblastoma cells. Cancer Research **53** 4978-4986

287 Feng Z. and Porter A.G. (1999) NF-κB/Rel proteins are required for neuronal differentiation of SH-SY5Y neuroblastoma cells. The Journal of Biological Chemistry **274** 30341-30344

288 Jung B.P., Jugloff D.G.M., Zhang G., Logan R., Brown S. and Eubanks J.H. (2003) The expression of methyl CpG binding factor MeCP2 correlates with cellular differentiation in the developing rat brain and in cultured cells. Journal of Neurobiology **55** 86-96

289 Im H.-I., Hollander J.A., Bali P. and Kenny P.J. (2010) MeCP2 controls BDNF expression and cocaine intake through homeostatic interactions with microRNA-212. Nature Neuroscience **13** 1120-1127

290 Smith C.J., Anderton B.H., Davis D.R. and Gallo J.-M. (1995) Tau isoform expression and phosphorylation state during differentiation of cultured neuronal cells. FEBS Letters **375** 243-248

291 Uberti D., Rizzini C., Spano P., Memo M. (1997) Characterization of tau proteins in human neuroblastoma SH-SY5Y cell line. Neuroscience Letters **235** 149-153

292 Jämsä A., Hasslund K., Cowburn R.F., Bäckström A. and Vasänge M. (2004) The retinoic acid and brain-derived neurotrophic factor differentiated SH-SY5Y cell line as a model for Alzheimer's disease-like tau phosphorylation. Biochemical and Biophysical Research Communications **319** 993-1000