Molecular mechanisms underlying haplotype-specific regulation of gene expression at the microtubule associated protein tau locus



Mang Ching Lai

Linacre College
Department of Physiology, Anatomy and Genetics
University of Oxford
Hilary 2016

A thesis submitted towards the degree of Doctor of Philosophy

ABSTRACT

Genome wide association studies (GWAS) have identified the H1 microtubule associated protein tau (*MAPT*) haplotype single nucleotide polymorphisms as leading common risk variants for Parkinson's disease (PD), progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD). Gene expression studies have demonstrated haplotype-specific increases in expression of *MAPT* exon 3-containing transcripts from the protective H2 allele compared to the H1. The difference in alternative splicing between the haplotypes likely contributes risk or protection in the absence of protein coding variants.

Here, we investigate the regulation of *MAPT* exon 3 alternative splicing by common, risk-associated, non-coding, haplotype-specific single nucleotide polymorphisms (SNPs) through a combination of *in silico* analysis of the *MAPT* locus, *in vitro* gene expression and biochemistry studies. Comparative sequence analysis of whole-locus genomic H1 and H2 *MAPT* (143 kb) vectors showed they capture over 86% of the *MAPT* sequence diversity. We generated and expressed haplotype-hybrid H1 and H2 *MAPT* vectors in a human neuroblastoma cell culture model and demonstrated that a functional SNP rs17651213 near the exon 3 5' splice site regulates exon 3 inclusion in a haplotype-specific manner. Using RNA-electrophoretic mobility shift assays (RNA-EMSA), we showed differential RNA-protein complex formation at the H1 and H2 sequence variants of SNP rs17651213. We further identified candidate *trans*-acting splicing factors interacting with functional SNP rs17651213 sequences by RNA-protein pull-down experiment and mass spectrometry. Finally, gene knockdown of candidate splice factors identified by mass spectrometry demonstrated a role for hnRNP F and hnRNP Q in the haplotype-specific regulation of exon 3 inclusion.

In this study, we have dissected the *MAPT* locus to identify sequences regulating the allele-specific alternative splicing of exon 3 and provided mechanistic insights into how common non-coding

H1/H2 *MAPT* haplotype-specific SNPs may contribute to the risk/protection of neurodegeneration at a complex genetic locus.

DECLARATION

Except where acknowledgement is made, all the work presented in this thesis was performed by the author at the University of Oxford. The work reported in this thesis has not been submitted for any other degree in this or any other university or institute of learning.

This thesis contains 48,256 words.

ACKNOWLEDGEMENTS

The completion of this thesis would not have been possible without the tremendous amount of support from many people along this long but amazing adventure.

First and foremost, I would like thank my supervisors Richard and Tara, for offering me this valuable opportunity to join the lab, for providing me with great support and help, and for guiding me throughout the past three and a half years in pursuing my interests and goals. I would also like to thank Julian and Kat for their advice and support on RNA-EMSA techniques, and Svenja for performing the mass spectrometry runs. To previous lab members Fran, Anne-Laure and Maria, who have made this project possible. This work is funded by the Alzheimer's Society.

I am very lucky to have joined this lab as everyone has been incredibly kind, friendly and supportive, and have enjoyed spending time a lot with everyone. There are so many in the lab but a few I would like to thank in particular. Fede, Sarah and Becky, without whom this journey would have been a lot more difficult. Especially the numerous Friday and Saturday nights at the bar, random food time, not really serious and not serious conversations, are some of my best memories in Oxford. Hugo, for listening to me talking the most random things for countless hours at the most random times. Ok, I meant stimulating conversations about science and general stuff. To the amazing people I have spent a lot time with in the lab, and more importantly at the pub, Amy, Natalie, Kathie, Charmaine, Dana, Helle, Nora, Oscar, Hanna and Constanze, it is some of the best times I have had and I am very grateful to have met you all. Also to office people Ruxy, Brent, Lawrence and Liz, who have made my early days at work most enjoyable. And to everyone, Michele, Ben, Ali, Heather, Milena, Dayne, Serena, and Mel in the lab for being really nice to me and for putting up with my spontaneity and craziness.

To Arvind and Antonio, for two years of house sharing accounting for some of the most hilarious moments. For the most awesome evening chill time and hang out in the living room time (of course,

with the company of alcohol, takeaways, Call of Duty, Skyrim, RainbowSix, and epic TV shows). To Jordan and Jolyon from the college, for all the pub evenings and whiskey tasting times.

To Chris, Jamie, Steivan, Alvaro and David, who have always been there for me even though we have been in different places for so many years.

Lastly and most importantly, I would like to express my deepest gratitude to my parents and my sister, for their unconditional love and support, to allow me to become who I am, to allow me to go on my crazy adventures, for which none of the amazing experiences I have had would have been possible without them.

PUBLICATIONS AND PRESENTATIONS ARISING FROM THIS WORK

Lai, M.C., Bechy, A-L., Denk, F., Gavriliouk, M., Ryan, B.J., Wade-Martins, R., Caffrey, T.M. (2016) A common intronic single nucleotide polymorphism associated with neurodegenerative diseases regulates H1/H2 haplotype-specific *MAPT* exon 3 expression. (Manuscript submitted, under review)

Lai, M.C., Bechy, A-L., Denk, F., Gavriliouk, M., Ryan, B.J., Wade-Martins, R., Caffrey, T.M. (2015) Molecular mechanisms underlying haplotype-specific regulation of gene expression at the microtubule associated protein tau locus. Poster presentation: The 12th international conference on Alzheimer's and Parkinson's diseases and related neurological disorder (AD/PD^{TM} 2015).

Lai, M.C., Bechy, A-L., Denk, F., Gavriliouk, M., Ryan, B.J., Wade-Martins, R., Caffrey, T.M. (2014) Molecular mechanisms underlying haplotype-specific regulation of gene expression at the microtubule associated protein tau locus. Poster presentation: Alzheimer's Society UK Research Conference 2014.

Lai, M.C., Bechy, A-L., Denk, F., Gavriliouk, M., Wade-Martins, R., Caffrey, T.M. (2013) Molecular mechanisms underlying haplotype-specific regulation of gene expression at the microtubule associated protein tau locus. Poster presentation: Alzheimer's Society UK Research Conference 2013.

TABLE OF CONTENTS

ABSTRACT **2**

DECLARATION **4**

ACKNOWLEDGEMENTS **5**

PUBLICATIONS AND PRESENTATIONS ARISING FROM THIS WORK **7**

TABLE OF CONTENTS **8**

LIST OF FIGURES **17**

LIST OF TABLES **20**

CHAPTER 1 ..... **21**

Introduction **21**

1.1 The Biochemistry and Physiological Functions of Tau Proteins **21**

1.1.1 Tau Proteins 21

1.1.2 Developmental Regulation of Tau 22

1.1.3 The C-Terminal Microtubule Binding Domains 23

1.1.4 Microtubule Assembly and Dynamics in Neurons 25

1.1.5 The N-Terminal Projection Domains 26

1.1.6 Tau Phosphorylation and the Microtubule Binding Properties 28

1.1.7 Other Physiological Functions of Tau 30

1.1.8 Section Summary 32

1.2 Tau and Neurodegenerative Diseases **33**

1.2.1 Tauopathies 33

1.2.2 The Prion Theory 33

1.2.3 Tau Aggregation in Neurodegeneration 35

1.2.4 The Unfolded Protein Response 37

1.2.5 Overview of Tau Mediated Neurotoxicity 38

1.2.6 Tau and Synaptotoxicity 39

1.2.7 Tau and Axonal Transport Deficits 40

1.2.8 Tau, Energy Failure and the Aging Brain 41

1.2.9 Tau as a Therapeutic Target 42

1.2.10 Section Summary 42

1.3 MAPT H1/H2 Haplotypes and Neurodegenerative Diseases 43

1.3.1 The Complex Genomic Architecture of the MAPT locus 43

1.3.2 The Evolutionary History of the H1/H2 Haplotypes 44

1.3.3 Linkage Analysis and Mendelian Diseases 46

1.3.4 The Theories on Explaining Common Complex Diseases 46

1.3.5 Genome Wide Association Studies 48

1.3.6 GWAS Identified the H1 Haplotype a Risk Locus for PD, PSP and CBD 49

1.3.7 Section Summary 52

1.4 Parkinson's Disease 53

1.4.1 Parkinson's Disease 53

1.4.2 Clinical Features of Parkinson's Disease 53

1.4.3 Treatments for Parkinson's Disease 54

1.4.4 Neuropathological Features of Parkinson's Disease 55

1.4.5 Genetics of Parkinson's Disease 56

1.4.6 Progressive Supranuclear Palsy and Corticobasal Degeneration 57

1.4.7 Section Summary 58

1.5 H1/H2 Haplotype-Specific Exon 3 Expression and Neurodegenerative Diseases 60

1.5.1 Haplotype-Specific Exon 3 Expression in Humans 60

1.5.2 MAPT Exon 3 and Neurodegenerative Diseases 61

1.6 Aims of This Thesis 63

CHAPTER 2 64

MATERIALS AND METHODS 64

2.1 Chapter 3 Methods 64

2.1.1 Sequence Comparison of pMAPT-H1 and pMAPT-H2 Vectors 64

2.1.2 Splice Factor Binding Site Analysis 64

2.1.3 Analysis of Linkage Disequilibrium Structure 64

2.1.4 Mini Preps 65

2.1.5 Maxi Preps 65

2.1.6 Generating Electrocompetent Cells 66

2.1.7 Generating Electrocompetent Cells with Induction of Homologous Recombination Enzymes 67

2.1.8 Electroporation 67

2.1.9 Exchanging H1/H2 SNPs rs117124984, rs118087478 & rs77875796 on pMAPT Vectors 68

2.1.10 Exchanging H1/H2 SNP rs17651213 on pMAPT Vectors 70

2.1.11 Retrofitting pMAPT Vectors with pH-FRT-HY Plasmid 72

2.1.12 Other Primers used to modify pMAPT Vectors 72

2.2 Chapter 4 Methods 74

2.2.1 Primer Design 74

2.2.2 Cell Culture 74

2.2.3 Transfections of SK-N-F1 Neuroblastoma Cells 74

2.2.4 RNA Extraction and cDNA Synthesis 75

2.2.5 qPCR for Total and Exon 3 Expression by pMAPT vectors 75

2.2.6 qPCR analysis for transgenic MAPT expression 76

2.2.7 HSV-1 Packaging Protocol 76

2.3 Chapter 5 Methods 77

2.3.1 Nuclear Lysate Extraction 77

2.3.2 RNA Biotinylation 77

2.3.3 RNA Electrophoretic Mobility Shift Assay (RNA EMSA) 78

2.3.4 RNA-Affinity Pull-Down 79

2.3.5 Mass Spectrometry 80

2.3.6 Mass Spectrometry Data Analysis 80

2.3.7 Western Blot 81

2.4 Chapter 6 Methods 83

2.4.1 Allele-Specific Assay Probes and Primers Design 83

2.4.2 Allele-Specific qPCR Assay 83

2.4.3 qPCR for Measuring Endogenous Gene Expression 84

2.4.4 qPCR Data Analysis 85

2.4.5 siRNA Knockdown 86

2.4.6 Camptothecin Treatments 86

2.4.7 RNA EMSA Supershift Assays 87

CHAPTER 3 88

MAPT H1/H2 Haplotype Genomic Expression PAC Vectors 88

3.1 Introduction 88

3.1.1 Genomic Expression PAC Vectors 88

3.1.2 Cis-Acting Sequence Elements in Alternative Splicing 89

3.1.3 Aims of Chapter 3 90

3.2 Results 91

3.2.1 Comparative Sequence Analysis of the Genomic MAPT H1/H2 Haplotype PAC Vectors 91

3.2.2 Analysis of Splice Factor Binding Sites at the Exon 3 Splice Sites In Silico 96

3.2.3 Candidate SNPs rs1800547 and rs17651213 are in Strong Linkage Disequibrium with the H1/H2 Haplotype Structure 98

3.2.4 The Homologous Recombination Strategy by Gene Bridges for SNPs Modification 100

3.2.5 Generation of Haplotype-Hybrid Vectors H1-2-2-2, H2-1-1-1, H1-1-2 and H2-2-1 .. 104

3.2.6 Retrofitting pMAPT Vectors with pH-FRT-Hy Plasmid 110

3.3 Discussion 113

3.3.1 Generating Haplotype-Hybrid pMAPT Vectors by Engineering Candidate SNPs 113

3.3.2 BAC/PAC Vectors as a Tool for Alternative Splicing Studies 113

3.3.2 Alternative Splicing Beyond Cis-Acting Regulatory Elements 114

CHAPTER 4 116

Expression of pMAPT Genomic Vectors in SK-N-F1 Neuroblastoma Cells 116

4.1 Introduction 116

4.1.1 Exon 3 Splicing Mechanisms 116

4.1.2 Aims of Chapter 4 117

4.2 Results 119

4.2.1 Designing, Optimising and Validating a qPCR Assay for Studying pMAPT Vectors Expression 119

4.2.2 Delivery of pMAPT Genomic Vectors by HSV-1 Amplicons 127

4.2.3 Delivery of pMAPT Genomic Expression Vectors by Transfection Reagents 133

4.2.4 pMAPT–H1WT and –H2WT Vectors Recapitulate Human Physiological Exon 3 Expression 134

4.2.5 Candidate SNPs Near the 5’ and 3’ Splice Sites are Functional and Modulate Exon 3 Inclusion 136

4.2.6 rs17651213 Regulates the Haplotype-Specific Inclusion of MAPT Exon 3 136

4.3 Discussion 138

4.3.1 rs1800547 and rs17651213 Regulate the Allele-Specific Splicing of Exon 3 138

4.3.2 Gene Delivery by HSV-1 138

4.3.3 rs1800547, rs17651213 and Splicing Regulatory Elements 140

CHAPTER 5 142

RNA-Protein Interaction Studies on H1 and H2 rs17651213 and rs1800547 Sequences 142

5.1 Introduction 142

5.1.1 Splicing Regulation by RNA-Binding Proteins 142

5.1.2 Biochemical Studies on RNA-Protein Interaction 144

5.1.3 Aims of Chapter 5 146

5.2 Results 148

5.2.1 SK-N-F1 Nuclear Lysate Extraction 148

5.2.2 Biotinylation of rs1800547 RNA Oligonucleotides for RNA-EMSA Studies 149

5.2.3 Optimising RNA-EMSA Binding Reaction Conditions Using rs1800547 RNA Probes 152

5.2.4 rs1800547-A/G Sequences Alter RNA-Protein Interactions 154

5.2.5 Optimise RNA-EMSA Binding Reaction Condition Using rs17651213 RNA Probes ... 156

5.2.6 rs17651213-G/A Sequences Alter RNA-Protein Interactions 158

5.2.7 Identifying Splice Factors Interacting With rs17651213 H1 and H2 SNP Sequences 160

5.3 Discussion 166

5.3.1 rs1800547 and rs17651213 SNP Sequences Alter RNA-Protein Interactions 166

5.3.2 RNA-EMSA in Studying RNA-Protein Interactions 166

5.3.3 Quantitative Proteomics in RNA-Binding Protein Identification 168

CHAPTER 6 172

Investigating the Allele-Specific Regulation of MAPT Exon 3 Expression 172

6.1 Introduction 172

6.1.1 Understanding the Functional Biological Basis of Non-Coding Polymorphisms 172

6.1.2 Co-Transcriptional Splicing Mechanisms 174

6.1.3 Aims of Chapter 6 175

6.2 Results 176

6.2.1 Developing and Validating a Quantitative MAPT Allele-Specific Expression qPCR Assay 176

6.2.2 Optimisations of qPCR Primers for Measuring Gene Expression Changes 182

6.2.3 Optimisations of siRNA delivery condition in SK-N-F1 cells 185

6.2.4 siRNA Knockdown of Splice Factors in SK-N-F1 Neuroblastoma Cells 188

6.2.5 hnRNP Q and hnRNP F Regulate the Allele-Specific Expression of MAPT Exon 3 190

6.2.6 hnRNP F Regulates Both Exon 3 Inclusion and Total MAPT Expression 192

6.2.7 hnRNP Q and hnRNP F Interact with rs1800547 and rs17651213 Respectively 194

6.2.8 Investigating the Allele-Specific Effects of RNA Polymerase II Elongation on MAPT Exon 3 Inclusion 199

6.3 Discussion 201

6.3.1 hnRNP Q and hnRNP F Regulate the Haplotype-Specific Inclusion of Exon 3 201

6.3.2 Detecting Allele-Specific Alternative Splicing Events 201

6.3.3 Investigating RNA Polymerase II Elongation Rate on the Allelic Expression of MAPT Exon 3 202

6.3.4 Regulation of Exon 3 Inclusion by hnRNP Q and hnRNP F 203

CHAPTER 7 205

Concluding Remarks 205

7.1 A Model for the Differential Regulation of MAPT Exon 3 Splicing 205

7.1.1 A Model for the Haplotype-Specific Regulation of H1 and H2 MAPT Exon 3 Inclusion 205

7.1.2 Allele-Specific Splicing and Disease Susceptibility 206

7.2 The Post GWAS-Era 207

7.2.1 Challenges in Understanding Common Traits and Diseases 207

7.2.2 Genome-Wide Approaches to Understand Common Variation and Genetic Predisposition 208

7.3 Machine Learning in Understanding the Human Genome 210

7.3.1 Human Genetics in the Age of Artificial Intelligence 210

7.3.2 Application of Machine Learning Algorithms in Genetics and Genomics 210

7.3.3 Cross-Disciplinary Research in Understanding Common Diseases 211

REFERENCES 214

LIST OF FIGURES

Figure 1. Tau isoforms in the adult human central nervous system. 22

Figure 2. The N- and C-terminal domains of tau. 24

Figure 3. The MAPT H1/H2 haplotypes. 44

Figure 4. An illustration of a Parkinson's patient. 54

Figure 5. pMAPT-H1 and -H2 genomic expression PAC vectors. 93

Figure 6. Comparative sequence analysis of SNPs present on the H1 and H2 MAPT gene. 94

Figure 7. Splice factor binding sites prediction on five candidate SNP sequences. 97

Figure 8. rs1800547 and rs17651213 are in linkage disequilibrium with the H1/H2 haplotype structure. 99

Figure 9. The selection counter-selection method for PAC modifications. 101

Figure 10. MAPT haplotype-hybrid genomic expression vectors. 103

Figure 11. Generating MAPT haplotype-hybrid vectors H1-2-2-2 and H2-1-1-1. 105

Figure 12. Generating MAPT haplotype-hybrid vectors H1-1-2 and H2-2-1. 106

Figure 13. Confirmation of H1-2-2-2 and H2-1-1-1 MAPT haplotype-hybrid vectors. 108

Figure 14. Confirmation of H1-1-2 and H2-2-1 MAPT haplotype-hybrid vectors. 109

Figure 15. Retrofitting pMAPT vectors with pH-FRT-Hy plasmid. 111

Figure 16. Restriction digests of retrofitted and unretrofitted vectors by NotI and SalI. 112

Figure 17. qPCR assay design for measuring transgenic MAPT expression. 120

Figure 18. Primer pair specificities assessed by melting curve. 121

Figure 19. Amplification plots of qPCRs using cDNA from untransfected cells. 122

Figure 20. Low transfection efficiency in BE(2)-M17 and HEK293 cells by a large PAC vector. 123

Figure 21. Determining the effects of varying forward to reverse primer ratios. 124

Figure 22. qPCR assay validations for selected candidate primer pairs. 126

Figure 23. HSV-1 amplicon production for infectious delivery of pMAPT vectors. 128

Figure 24. Infectious delivery of pMAPT vectors in cultured neuroblastoma cells. 130

Figure 25. Expression of pMAPT-H1WT and -H2WT vectors in HSV-1 transduced cells..... 132

Figure 26. pMAPT vectors delivery using transfection reagents. 133

Figure 27. rs17651213 regulates the haplotype-specific inclusion of MAPT exon 3. 135

Figure 28. Confirmation of nuclear enriched cell lysate fractions by Western blot. 149

Figure 29. Estimation of biotinylation efficiency for rs1800547 H1 and H2 probes. 151

Figure 30. RNA-EMSA optimisations for rs1800547 H1 and H2 probes..... 153

Figure 31. RNA-EMSA analysis demonstrates differential RNA-protein complex formation at the H1 and H2 variants of SNP rs1800547. 155

Figure 32. RNA-EMSA optimisations for rs17651213 H1 and H2 probes..... 157

Figure 33. RNA-EMSA analysis demonstrates differential RNA-protein complex formation at the H1 and H2 variants of SNP rs17651213. 159

Figure 34. RNA-affinity pull-down of SK-N-F1 nuclear proteins interacting with rs17651213 H1 and H2 alleles. 161

Figure 35. Mass spectrometry data analysis to identify rs17651213 interacting splice factors.... 164

Figure 36. Allele-specific Taqman assay probes and primers design. 176

Figure 37. Selecting primer pairs for allele-specific Taqman assay. 177

Figure 38. Optimisations of the MAPT allele-specific Taqman qPCR assay. 179

Figure 39. MAPT allele-specific expression assay standard curve 181

Figure 40. qPCR primer efficiencies determined by serial dilution standard curves 183

Figure 41. qPCR primer specificities assessed by melting curve and gel electrophoresis 184

Figure 42. Lipofectamine RNAiMAX delivers siRNA more efficiently than Lipofectamine 2000 ... 186

Figure 43. Gene expression remains knocked down 4 days post transfection. 187

Figure 44. High gene knockdown efficiencies of the candidate splice factors in SK-N-F1 cells. 189

Figure 45. hnRNP Q and hnRNP F regulate the haplotype-specific inclusion of MAPT exon 3 191

Figure 46. Splice factor knockdown on total MAPT and exon 3 expression. 193

Figure 47. Optimisations of RNA-EMSA supershift binding reaction condition with rs17651213 H1 and H2 probes. 196

Figure 48. hnRNP Q and hnRNP F interact with SNPs rs17651213 and rs1800547 respectively... 198

Figure 49. Camptothecin reduced exon 3 inclusion rate. 200

Figure 50. Model for the haplotype-specific regulation of H1 and H2 MAPT exon 3 inclusion..... 205

Figure 51. The 5 dimensions of genetic variation. 211

LIST OF TABLES

Table 1. Primers for generating H1-2-2-2 & H2-1-1-1 70

Table 2. Primers for generating H1-1-2 & H2-2-1 72

Table 3. Primers previously used to modify pMAPT vectors. 73

Table 4. Primers for measuring total and exon 3 expression from pMAPT vectors 76

Table 5. Allele-specific qPCR assay primers and probes. 84

Table 6. Primers for expression of endogenous genes. 85

Table 7. Ambion silencer select pre-designed siRNA IDs. 86

Table 8. Genotyping five BAC/PAC vectors spanning the whole MAPT locus identified four H1 and one H2 alleles. 92

Table 9. Identifying candidate splice factors interacting with rs17651213-G/A alleles. 165

CHAPTER 1

Introduction

1.1 The Biochemistry and Physiological Functions of Tau Proteins

1.1.1 Tau Proteins

Tau is encoded by the microtubule associated protein tau (*MAPT*) gene located on chromosome 17q21.31. It is predominantly expressed in the central nervous system (CNS) and the peripheral nervous system. In humans, six major tau isoforms are generated in the adult CNS through the alternative splicing of exons 2, 3 and 10 (Figure 1) (Andreadis et al. 1992). Tau was originally identified as a low molecular weight protein that organises the tubular structure of microtubules (Hirokawa et al. 1988; Weingarten et al. 1975). Exons 9 to 12 are four imperfectly repeated sequences forming the microtubule binding domain at the carboxyl-terminal half of the protein that is responsible for the tubulin binding property of tau (G. Lee et al. 1989). The amino-terminal (N-terminal) domain of tau, which can carry two N-terminal inserts encoded by exons 2 and 3, mediates various non-classical functions including the interaction with the plasma membrane, dynactin, actin cytoskeleton, phospholipase C-γ and tyrosine kinase fyn signalling pathways (Cunningham et al. 1997; M. DiTella et al. 1994; Hwang et al. 1996; Jenkins and Johnson 1998; Kempf et al. 1996; G. Lee et al. 1998; Magnani et al. 2007; Shahani and Brandt 2002).

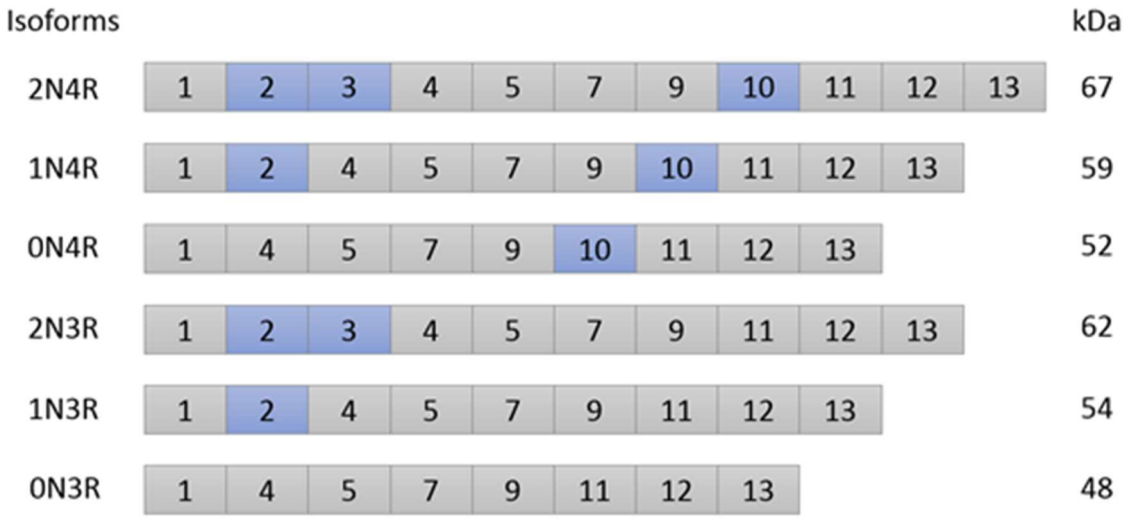


Figure 1. Tau isoforms in the adult human central nervous system.

Six major tau isoforms are expressed in the adult brain and are generated via the alternative splicing of exons 2, 3 and 10 of the *MAPT* gene. Alternative exons in blue, constitutive exons in grey.

1.1.2 Developmental Regulation of Tau

The expression of tau isoforms is developmentally regulated through the alternative splicing of exons 2, 3 and 10. 0N3R (Figure 1) tau is the only isoform expressed in the foetal human brain, while all six tau isoforms (Figure 1) are expressed in the adult human brain where the 4R:3R isoform ratio is approximately 1:1 (Goedert et al. 1989b; Hong et al. 1998; Kosik et al. 1989a). 4R isoforms are thus often referred to as the "mature" tau isoforms. Unlike the C-terminal alternative exon 10, the N-terminal exons 2 and 3 are expressed at different levels in the adult brain, with the 1N isoforms being the most abundant, followed by 0N and 2N isoforms (Hong et al. 1998). The 1N, 0N and 2N isoforms comprise of approximately 53.7%, 36.8% and 9.5%, respectively, of the total tau isoforms expressed (Hong et al. 1998). At the protein level, tau phosphorylation and subcellular localisation are also regulated during development. Tau is phosphorylated at a higher level in the foetal than in the adult brain due to the activation of phosphatases with increasing age (Dudek and Johnson 1995; Kanemaru et al. 1992; Pope et al. 1993; Rosner et al. 1995). The reduced tau phosphorylation, hence increased microtubule binding by tau, and the increased expression of the 4R isoforms have been suggested as an important event during development where reduced

neuronal plasticity, increased neuronal differentiation and axonal growth is required (Bullmann et al. 2010; S. Chen et al. 2010; Sennvik et al. 2007; Takuma et al. 2003).

In young neurons, tau is found in the cell soma and in neurites (Brandt et al. 1995; Kempf et al. 1996). As neurons become mature and polarised, tau is then concentrated in axons but is still present in the somatodendritic compartment (Hamre et al. 1989; Kempf et al. 1996; Kosik et al. 1989b; Litman et al. 1993). While some studies observed the localisation of tau only in axons in mature neurons (Binder et al. 1985; Trojanowski et al. 1989), others have demonstrated that such selective staining of tau only in the axonal region is a false negative discovery due to the phosphorylation state of tau (Papasozomenos and Binder 1987; Tashiro et al. 1997), together these studies showed that tau has different phosphorylation states in different cellular compartments. Several mechanisms of the compartmentalisation of tau have been identified, such as the selective transport of tau protein and mRNA into axons, its selective degradation in dendrites and its increased translation in axons (Aronov et al. 2002; Hirokawa et al. 1996; Morita and Sobue 2009; Nakata and Hirokawa 2003). The distribution of tau in axons is an important cellular process as the missorting of tau into the somatodendritic compartment has been described as an early sign of neurodegeneration in Alzheimer's disease (E. Braak et al. 1994).

1.1.3 The C-Terminal Microtubule Binding Domains

The classical microtubule binding activity of tau has been most extensively studied. Tau interacts reversibly with microtubules via its four microtubule binding domains encoded by exons 9 – 12 (Brandt and Lee 1993; G. Lee et al. 1989). The four microtubule binding domains are repetitive regions that are highly conserved, containing very similar DNA and amino acid sequences (Goedert et al. 1989a; Himmler et al. 1989). Each repeat is 18 amino acid in length and are separated by less conserved inter-repeat regions of 13 or 14 amino acid in length (G. Lee et al. 1989). Tau binds microtubules to promote the nucleation of tubulin subunits, modulate microtubule growth and bundling, and stabilise the microtubules formed (Figure 2) (Brandt and Lee 1993; Cleveland et al. 1977; Drechsel et al. 1992; Trinczek et al. 1995).

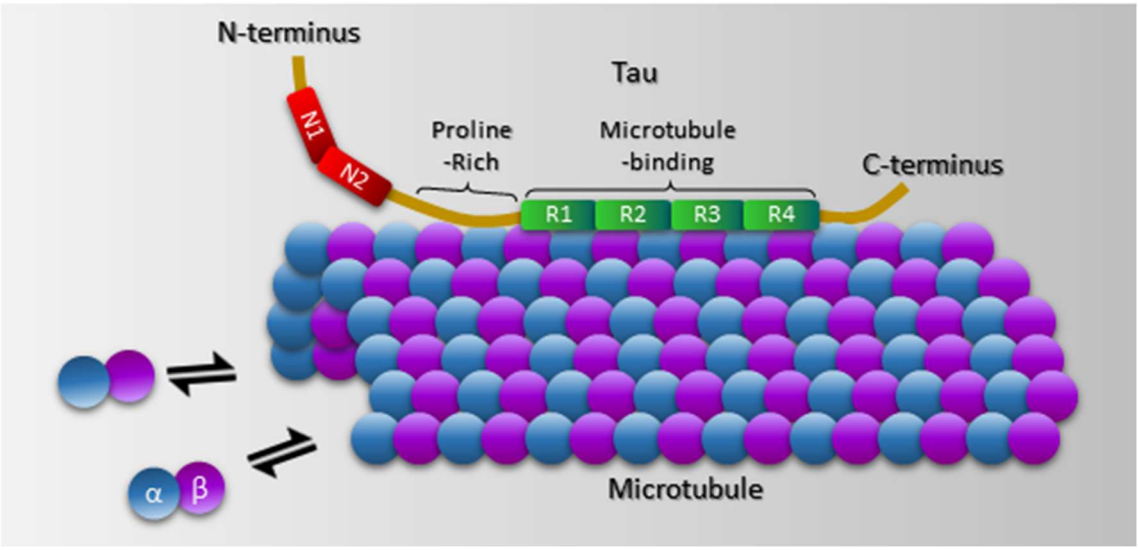


Figure 2. The N- and C-terminal domains of tau.

Tau interacts with microtubules via its C-terminal repeats encoded by exons 9 – 12 (R1 – R4). The N-terminal domains are encoded by exons 2 and 3 (N1 and N2), followed by a proline-rich domain, which projects away from the microtubule when tau is bound. Microtubules are dynamic structures through the constant assembly and disassembly of the α- and β-tubulin subunits.

Tau adopts a relatively disordered structure physiologically even when it is bound to microtubules on the outer surface (Santarella et al. 2004). Interestingly, due to this intrinsically disordered structure of tau, its exact binding site on microtubule has yet to be determined despite attempts using high resolution imaging techniques such as cryo-electron microscopy and atomic force microscopy (Al-Bassam et al. 2002; Santarella et al. 2004; Schaap et al. 2007). The structure of microtubules however, has been identified at high resolution (Nogales et al. 1999). The microtubule biding property of tau is highly regulated by the phosphorylation of serine/threonine residues, which in turn is regulated by specific kinases and phosphatases (Buee et al. 2000). About 30 amino acid residues on tau can be phosphorylated under physiological conditions (Buee et al. 2000; Hanger et al. 2009; Mazanetz and Fischer 2007). In general, tau phosphorylation decreases its binding affinity to microtubules thereby regulating microtubule assembly (Lindwall and Cole 1984). The phosphorylation of some residues locating in or around the microtubule binding domain can lead to the loss of tubulin binding affinity, detaching tau from microtubules (Drewes et al. 1995).

1.1.4 Microtubule Assembly and Dynamics in Neurons

Microtubules are dynamic cytoskeletal structures that can undergo rapid cycles of polymerisation and depolymerisation, or growth and shrinkage, both *in vivo* and *in vitro* (Desai and Mitchison 1997). They are assembled in a head-to-tail fashion by tubulin heterodimer subunits composed of α- and β- tubulin monomers, forming the tubular hollow cylindrical microtubule polymer structure of 25 nm in width (Desai and Mitchison 1997). The tubulin subunits polymerise at the two ends of a microtubule at different rates, conferring with microtubule its polarity where the faster and slower growing ends are referred to as the plus and minus end, respectively (Allen and Borisy 1974; Amos and Klug 1974). Microtubule polymerisation dynamics, or its growth and shrinkage, *in vivo* is an important mechanism in allowing the rapid spatial and temporal reorganisation of microtubules to adapt to the cellular context in order to generate mechanical pulling and pushing forces (S. Inoue and Salmon 1995; Kirschner and Mitchison 1986).

Indeed, neuronal microtubules play a central role in directing the formation of axons and dendrites in neurons (Conde and Caceres 2009). Neuronal polarisation occurs when arrays of short neurites or processes begin to form by a neuronal cell (Dotti et al. 1988). As the neuron becomes mature, one of the neurites is eventually specified as an axon and protrudes away from the cell soma (Dotti et al. 1988). The formation of a growth cone is a major change that accompanies the transformation of a neurite into an axon where major changes to the cytoskeleton occurs, including an increase in the length and number of microtubules that extends and protrudes into the growth cone (Bradke and Dotti 1997, 1999; Kunda et al. 2001). Upon entry into the growth cone, microtubules exhibit complex dynamic remodelling behaviours such as splaying, looping, bending and bundling. Such complex reorganisation of microtubules at the growth cone is needed for the continuous growth cone advancement leading to the elongation of axons (E. M. Tanaka and Kirschner 1991; E. Tanaka et al. 1995). Besides axonal growth, microtubules are also a crucial element in dendritic development. A key difference observed in dendritic and axonal microtubules is their orientation. Dendritic microtubules have their minus end pointing towards the dendritic tip while axonal

microtubules have their positive end pointing towards the axonal terminal (Desai and Mitchison 1997). Axonal and dendritic microtubules also differ in their regulating microtubule associated proteins. For example, microtubule associated protein 2 (MAP2) and tau are mainly found in dendrites and axons, respectively (Desai and Mitchison 1997).

Previous studies have demonstrated that tau is involved in neuronal polarisation, neurite extension and the formation of axons through organising microtubules dynamics (Dawson et al. 2001; M. C. DiTella et al. 1996; Takei et al. 2000). Interestingly, despite the clear role of tau in organising microtubules and regulating axonal growth, experiments in cultured neurons and in mice showed that tau is not absolutely essential for axon formation as tau knockdown is not lethal to neurons and tau knockout does not reduce longevity and critical brain functions (King et al. 2006; Qiang et al. 2006; Takei et al. 2000). This could be explained by the overlapping roles of microtubule associated protein 1B (MAP1B) with tau, where compared to tau knockout mice, MAP1B knockout mice showed much more severe phenotypes of disorganised neuronal layering and premature mortality (Takei et al. 2000). The phenotypes were further worsened in tau and MAP1B double knockout mice, demonstrating tau and MAP1B are required in conjunction for the normal and healthy regulation of axon formation (Takei et al. 2000).

1.1.5 The N-Terminal Projection Domains

Exons 2 and 3 encode two N-terminal inserts that are 29 amino acids in length that contain acidic residues. The acidic inserts are then followed by a proline-rich region that contains basic residues (Buee et al. 2000). The acidic inserts together with the proline-rich region are termed the projection domain as it projects away from microtubule upon tau binding (Figure 2) (Hirokawa et al. 1988). Despite the extensive studies into the C-terminal microtubule binding domain, the functions of the two N-terminal domains are not well understood. Nevertheless, several studies have provided insights into the functions of the N-terminal regions. The N-terminal inserts can associate with the neural plasma membrane, influence microtubule distribution by affecting the spacing between microtubules, and may influence the subcellular distribution of tau in neurons (Al-Bassam et al.

2002; Brandt et al. 1995; J. Chen et al. 1992; Frappier et al. 1994; C. Liu and Götz 2013; Maas et al. 2000). The results from these studies point towards a potential molecular mechanism of how tau regulates the axon formation through stabilising microtubules onto the plasma membrane (G. Lee 2005). In addition, tau has been shown to associate with the Golgi and rough endoplasmic reticulum membranes (Farah et al. 2005; Farah et al. 2006). Together with previous evidence that showed the N-terminal inserts are responsible for localising tau to the plasma membrane which is independent of the microtubule binding domains and is influenced by specific phosphorylation (Arrasate et al. 2000; Brandt et al. 1995; Maas et al. 2000), the N-terminal inserts may therefore play a role in tethering tau onto Golgi and rough endoplasmic reticulum membranes.

The proline-rich domain can interact with the src homology 3 (SH3) domain of src family tyrosine kinases and phospholipase Cγ (PLC-γ), which are both proteins associated with the plasma membrane, indicating the potential role of tau in participating in signalling cascades via src kinases and PLC-γ pathways (Figure 2) (G. Lee et al. 1998; Reynolds et al. 2008). Moreover, the src family kinase fyn has been shown to phosphorylate a tyrosine residue on tau, while tau interacts with the SH3 domain on fyn, further suggesting that tau could interact with signalling pathways via src family kinases (G. Lee et al. 2004). Src and fyn kinases are present in growth cones and can regulate growth cone steering via src signalling pathway that connects cell adhesion molecules and the cytoskeleton, while neurite outgrowth is impaired in neuronal cultures from src or fyn deficient mice (Beggs et al. 1994; Helmke and Pfenninger 1995; Ignelzi et al. 1994; Maness et al. 1988; Suter and Forscher 2001). These experiments support the role of tau in neuronal signal transduction pathways.

Indeed, studies on β-amyloid toxicity in Alzheimer's disease have provided much evidence that tau participates in fyn signalling pathway, where tau is responsible for mediating the synaptotoxic and neurotoxic effects of β-amyloid via fyn kinase (Chin et al. 2004; Lambert et al. 1998; Rapoport et al. 2002; Tomidokoro et al. 2001). Long term potentiation (LTP) of synaptic transmission in the hippocampus is the fundamental molecular model for learning and memory formation (Bliss and

Collingridge 1993). In an experiment that uses hippocampal slices to study the effects of tau on LTP, the authors found that β-amyloid failed to induce LTP deficits in tau knockout mice (Shipton et al. 2011). Furthermore, the authors found blocking tau phosphorylation that is induced by β-amyloid prevented LTP impairment in wild type mice (Shipton et al. 2011), demonstrating that tau is required for the impairment of LTP induced by β-amyloid (Shipton et al. 2011). In another study that examined the postsynaptic excitotoxic effects of β-amyloid in transgenic mice with β-amyloid forming background, the authors found that the excitotoxic effects mediated by fyn is disrupted in the tau knockout background (Ittner et al. 2010). Tau knockout mice with β-amyloid forming background further prevented memory deficits and improved survival (Ittner et al. 2010). Together with previous studies mentioned here on tau mediating β-amyloid toxicity via fyn kinase, the molecular mechanisms between LTP, β-amyloid toxicity and fyn signalling pathway has finally been linked to tau over a decade after the hypothesis where impaired synaptic plasticity and memory deficits in Alzheimer's disease could be caused by β-amyloid and fyn kinase pathways had been proposed (Lambert et al. 1998).

1.1.6 Tau Phosphorylation and the Microtubule Binding Properties

Tau functions are highly regulated by phosphorylation, and there are 85 amino acid residues in the full length 2N4R tau that could be phosphorylated, including 45 serine, 35 threonine and 5 tyrosine. Using either mass spectrometry analysis, Edman peptide degradation or anti-phospho tau antibodies, a total 45 phosphorylation sites have been experimentally identified (Hanger et al. 2009). Under normal conditions, approximately 30 sites can be phosphorylated and have normal physiological functions, while higher phosphorylation states, approximately 10 more sites, of pathological tau in Alzheimer's brains have been identified (Buee et al. 2000; Hanger et al. 2007; Mazanetz and Fischer 2007). Interestingly, most phosphorylation sites are found at the proline-rich domain and at the C-terminus end of the protein, with only 8 sites identified in the microtubule binding repeats and N-terminal inserts (Hanger et al. 2009), suggesting that phosphorylation events

in the functional domains of tau may have significant consequences on affecting the functions of those domains.

Many kinases and phosphatases have been shown to regulate the phosphorylation state of tau (Y. Wang and Mandelkow 2016). Proline-directed protein kinases (PDPK) phosphorylate serine-proline or threonine-proline sites (Ubersax and Ferrell 2007), and most of the kinases identified to phosphorylate tau belong to the family of PDPK, including glycogen synthase kinase 3 (GSK3), mitogen activated protein kinase (MAPK), tau protein kinase I (TPKI) and cyclin-dependent kinases (CDKs) (Baumann et al. 1993; Drewes et al. 1992; Goedert et al. 1997; Hanger et al. 1992; W. K. Liu et al. 1995; Reynolds et al. 1997; Takahashi et al. 1995; Vulliet et al. 1992). Stress-activated protein (SAP) kinases have been reported to act on tau, suggesting phosphorylated tau could have a role in cellular stress response (Goedert et al. 1997; Jenkins et al. 2000; Reynolds et al. 1997). Non serine-proline or serine-threonine sites are phosphorylated by many other kinases, including Ca^{2+}/calmodulin-dependent protein kinase II (CaMPK II), microtubule-affinity regulating kinase (MARK), cyclic-AMP-dependent kinase (PKA) and casein kinase II (Drewes et al. 1997; Greenwood et al. 1994; Jicha et al. 1999; Johnson 1992; Litersky and Johnson 1992).

Dephosphorylation of tau occurs rapidly in post-mortem brain tissue and in neuroblastoma cells (Buee et al. 2000). Serine/threonine phosphatases that are present in the brain include phosphatase proteins 1 (PP1), 2A (PP2A), 2B and 2C (Cohen and Cohen 1989). These phosphatases are developmentally regulated, hence explaining how adult tau is generally less phosphorylated than foetal tau due to the activation of phosphatases during development (Dudek and Johnson 1995). Interestingly, these phosphatases are often associated with microtubules, either by direct interactions or indirectly through other proteins, indicating their functional role in regulating the phosphorylation state of proteins interacting with microtubules (Buee et al. 2000). For instance, tau proteins have been shown to link PP1 and tubulin, while PP2A binds directly to microtubules via ionic interactions (H. Liao et al. 1998; Sontag et al. 1999).

The regulation of tau by a large number of kinases and phosphatases suggests that many physiological functions could be coupled to tau functions, and tau may contain different phosphorylation in different cellular compartments depending on where the kinases and phosphatases are localised in. Indeed, the microtubule binding ability of tau is dependent on its phosphorylation states as phosphorylation reduces microtubule binding affinity (Buee et al. 2000). Previous experiments showed tau phosphorylation reduces microtubule binding affinity, while dephosphorylation promotes tubulin assembly and polymerisation (Bramblett et al. 1993; Cho and Johnson 2003; Cleveland et al. 1977; Drubin and Kirschner 1986; Lindwall and Cole 1984; Yamamoto et al. 1988). In particular, the phosphorylation Ser262 locating within the microtubule binding domain has been shown to be sufficient in causing tau to detach completely from microtubules (Biernat et al. 1993; Seubert et al. 1995). Similar to microtubule assembly and disassembly, the binding and dissociation of tau from microtubules are also in a constant dynamic equilibrium, and this equilibrium is thought to be regulated by phosphorylation state of tau, determined by the actions of various kinases and phosphatases (Ballatore et al. 2007).

1.1.7 Other Physiological Functions of Tau

Besides being a structural cellular component, microtubules are molecular tracks that act like railings for the transport of membrane-bound organelles and macromolecular complexes, it is therefore not surprising that tau also plays a role in regulating axonal transport events (G. Liao and Gundersen 1998; Nakata and Hirokawa 2003; Reed et al. 2006; Stamer et al. 2002). Intracellular trafficking of organelles along microtubules is driven by opposing molecular motors kinesin and dynein where accurate transport of organelles are highly dependent on the regulation of these opposing molecular motors, which often act simultaneously on the same cargo (M. M. Fu and Holzbaur 2014). Kinesin and dynein mediate the anterograde and retrograde transport of cargoes, respectively, along microtubules (M. M. Fu and Holzbaur 2014).

Tau has been shown to influence dynein and kinesin functions through competing with the molecular motors for microtubule binding, thereby slowing down anterograde and retrograde

transport through reducing the binding frequency, motile fraction and run length of the molecular motors (Dixit et al. 2008; Stamer et al. 2002). In addition, tau can interfere with kinesin and dynein molecules that are already bound to cargoes through reducing the reattachment rates of the motors, affecting the travel distance, motive force and disposal of the cargoes (Vershinin et al. 2007). Tau itself can act as a cargo to be transported, competing with other cargoes (Konzack et al. 2007; Utton et al. 2005). As mentioned above, tau couples PP1 to microtubules and is a substrate of GSK3. Tau has been shown to regulate the release of cargoes from kinesin motors by activating PP1 and GSK3 through its N-terminus phosphatase-activating domain in squid axons (Kanaan et al. 2011).

However, it should be noted that the above effects of tau on axonal transport were identified in *in vitro* studies, the *in vivo* mechanisms are still unclear, as tau appears to have little influence on axonal transport in tau knockout or overexpressing mice, indicating additional compensatory mechanisms could be present (Yuan et al. 2008). Understanding the importance of tau on axonal transport is important as axonal transport deficits are implicated in many neurodegenerative diseases, including hereditary spastic paraplegias, motor neuron disease, Charcot-Marie-Tooth disease, Alzheimer's disease, Huntington's disease, amyotrophic lateral sclerosis, Parkinson's, multiple system atrophy and dementia with Lewy bodies (Roy et al. 2005).

Apart from its role in axons and dendrites, tau has also been identified as a nuclear protein (Loomis et al. 1990; Sjoberg et al. 2006). The role of tau in the nucleus however, is not entirely clear. Experiments that places cultured cells and mice under heat stress conditions demonstrate that tau can bind directly to genomic DNA and protect DNA from heat stress induced damage, such as double-strand breaks, in neurons (Sultan et al. 2011; Violet et al. 2014). The authors suggested that the hyperthermia condition induces a transient reactive oxygen species production, and thus the protective role of tau against heat induced nuclear damages could have implications in oxidative damages observed in Alzheimer's disease (Violet et al. 2014). However, the role of nuclear tau in

relation to reactive oxygen species has not yet been directly investigated and therefore the proposed protective effects are purely speculative.

1.1.8 Section Summary

In this section, several physiological aspects of tau proteins were reviewed, including the biochemical structure of the protein isoforms, the role of tau as a microtubule binding protein, the role of tau as a signalling protein, the regulation of tau function in development and normal physiological conditions, and other functions of tau. Tau is intricately involved in neurodegeneration. In the next section, the pathological involvement of tau will be introduced.

1.2 Tau and Neurodegenerative Diseases

1.2.1 Tauopathies

Abnormally aggregated hyperphosphorylated tau in neurons is a pathological hallmark in many neurodegenerative diseases collectively termed tauopathies where tau deposits are one of the most prominent pathological features. To date, tau pathology has been described in at least 25 different neurodegenerative diseases. Examples of tauopathies include Alzheimer's disease (AD), Pick's disease (PiD), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17), Down's syndrome, myotonic dystrophy and amyotrophic lateral sclerosis. Although they share the common feature of abnormal tau aggregation, the compositions and the ultrastructures of tau deposits vary between diseases. For example, in AD, neurofibrillary tangles (NFTs) containing paired helical filament (PHF) tau are composed of both 3R and 4R tau isoforms (de Silva et al. 2003; Kidd 1963). In PSP, NFTs are predominantly composed of 4R tau with a straight filament ultrastructure (Mailliot et al. 1998; Sergeant et al. 1999; Tellez-Nagel and Wiśniewski 1973; Tomonaga 1977). In CBD, NFTs are also predominantly composed of 4R tau with a twisted filament structure that differs from PHF found in AD (Feany et al. 1995; Ksiezak-Reding et al. 1994). In PiD on the contrary, tau aggregations are mainly composed of the 3R isoforms (de Silva et al. 2006).

1.2.2 The Prion Theory

The term prion (proteinaceous infectious particle) disease was first described in the 1980s when the physiological variant of the prion protein (PrP) was first discovered to be responsible for causing transmissible spongiform encephalopathy (TSE), or scrapie, in sheep and goats (Bolton et al. 1982; Prusiner 1982). The term was used to differentiate between scrapie and other infectious diseases where the responsible reagents are often bacterial or viral in nature. In humans, prion diseases include Creutzfeldt-Jakob disease (CJD), variant CJD (vCJD), Kuru, Gerstmann-Straussler-Scheinker (GSS) disease and fatal insomnia (Goedert 2015; Sanders et al. 2016). Prion diseases are rare neurodegenerative disorders, with an incidence of approximately 1 in a million per year, which

none was described as due human to human transmission and are largely sporadic in nature (Wadsworth et al. 2003). Over 20 mutations in the PRNP (prion protein) gene have been described to cause familial CJD (Gabizon et al. 1993; Hsiao et al. 1989; Owen et al. 1989; Prusiner 1998), while a nonsynonymous polymorphism in PRNP has been described to increase prion disease susceptibility (Palmer et al. 1991).

The pathogenic form of prions (PrP^{Sc}) are converted from normal cellular prion proteins (PrP^{C}) where part of the PrP^{C} α-helical structures are folded into β-sheets (Pan et al. 1993). The process which PrP^{C} is converted to PrP^{Sc} is termed seeding, and PrP^{Sc} is known to adopt several pathological conformations (Holmes and Diamond 2012). The β-sheet-rich PrP^{Sc} aggregates can then act as a template where additional PrP^{C} monomers are converted. Once enough PrP^{Sc} is formed, PrP^{C} can be converted readily to its pathological form, and propagates intracellularly by converting normal prions (Sanders et al. 2016). Prions can also "escape" from one cell and spread to another via different pathways. "Escaping" can occur when prions are released directly to the extracellular space during cell death when the cell membrane breaks down, or actively being secreted through exosome release. Once prions have escaped, they could be taken up by another cell via vesicle fusion or macropinocytosis triggered by their interaction with heparan sulphate proteoglycans. Once internalised, the prions could act as a "seed" and lead to their propagation in the new host cell. The cycle of propagation to another cell then repeats, giving prions its infectious nature (Sanders et al. 2016). The formation of prion aggregates occurs as a multistage process where misfolded monomeric proteins first form soluble dimers and trimers. The trimers further propagate to oligomeric species and eventually they form protofilaments. Protofilaments can interact with each other to form fibrils and larger aggregate structures such as plaques and tangles (Aguzzi and Lakkaraju 2016).

The aggregations of prions as a pathogenic mechanism was originally confined to CJD where aggregated β-sheet-rich proteins was a prominent feature of brain pathology. However, the concept of intracellular seeding, templating, propagating, and the subsequent spread to other cells,

has now been applied to neurodegenerative diseases characterised by insoluble protein aggregates, such as AD and PD. It should be noted that AD and PD are not considered as prion diseases as there is little evidence that the protein aggregations in AD and PD can be transmitted from one person to another (Goedert 2015).

1.2.3 Tau Aggregation in Neurodegeneration

As mentioned above, the aggregation of tau into NFTs and paired helical filaments (PHFs) characterises tauopathies. Similar to prions, the formation of NFTs is also a multistage process where abnormal dissociation of tau from microtubules into the cytosol forms the first step in the aggregation process. The dissociation could be caused be many factors, such as increased phosphorylation and reduced dephosphorylation, leading to an increase in cytosolic unbound tau concentration (Ballatore et al. 2007). At this stage, misfolding of unbound tau may occur leading to the formation of "seeds" that are prone to further aggregation, becoming pretangles, or small nonfibrillary deposits. The nonfibrillary deposits, or pretangles, do not appear to contain β-sheet structures as they are not detected by β-sheet-specific dyes such as congo red or thioflavine-T (Galvan et al. 2001; Maeda et al. 2006; Maeda et al. 2007). The transition from pretangles to PHFs involves the structural re-organisation of tau aggregations, leading to the formation of β-sheet structures in PHFs. Lastly, PHFs interact with each other to form large NFT ultrastructures (Ballatore et al. 2007).

FTDP-17 is a dominantly inherited form of FTD, and is a clinically and pathologically heterogeneous class of neurodegenerative disorder with an average age of onset at 55 (Seelaar et al. 2011). Patients often show a dementia-dominant phenotype with parkinsonism-dominant extrapyramidal features that resemble CBS and PSP, changes in behaviour such as disinhibition and obsessive-compulsiveness, and may exhibit mild semantic impairment (Seelaar et al. 2011). The composition of tau inclusions in post-mortem FTDP-17 brains are also heterogeneous, with cases that resemble 4R tauopathies such as PSP and CBD, or 3R tauopathies such as PiD, or 3R/4R tauopathies such as AD (Goedert et al. 2012a). Despite the heterogeneity, all cases of FTDP-17 exhibit pathological

aggregation of hyperphosphorylated filamentous tau inclusions (Ballatore et al. 2007). FTDP-17 mutations have been shown to promote tau aggregation in several different ways. For example, missense mutations can lead to an accelerated aggregation of tau into filaments, increased likelihood to phosphorylation and reduced microtubule binding affinity (Alonso Adel et al. 2004; Dayanandan et al. 1999; Nacharaju et al. 1999). Furthermore, intronic and exonic mutations in exon 10, were identified to cause the overproduction 4R tau isoforms (Goedert and Jakes 2005). Importantly, these mutations provide evidence that tau aggregations are pathogenic and can cause neurodegeneration directly.

AD is the main cause of dementia and accounts for 50% to 60% of all dementia cases worldwide (Blennow et al. 2006). General symptoms include the progressive impairment of episodic memory, aphasia, apraxia, agnosia and cognitive impairment (Blennow et al. 2006). Currently there is no disease modifying therapy available, and treatments of AD are solely symptomatic (Frost et al. 2015). While tau and β-amyloid lesions are a prominent characterising neuropathological feature in AD, no mutation in the *MAPT* gene has been identified to cause AD. The molecular events of how tau contributes to neurodegeneration is thus less well understood in tauopathies without *MAPT* mutation. Studies into AD provide evidence that tau is a required component in AD pathogenesis as individuals with abundant β-amyloid pathology but with no or very few tau positive lesions, do not develop AD (H. Braak and Braak 1991; Nelson et al. 2012). Mutant tau overexpression in transgenic mice recapitulates the pathological features of tauopathies, such as tau filament formation and neurodegeneration. Furthermore, mice expressing both mutant β-amyloid and mutant tau enhanced the tau pathology observed (J. Lewis et al. 2001). Another experiment using a human neural stem cell-derived three-dimensional culture system expressing mutant β-amyloid, showed the formation of β-amyloid plaques and tau inclusions in the culture (Choi et al. 2014). Tau also appears to be needed for mediating β-amyloid-induced toxicity as reduced hyperexcitability of neurons subjected to β-amyloid exposure was observed in the absence of tau (Roberson et al. 2007). The above studies demonstrate tau as an important mediator of β-amyloid toxicity.

1.2.4 The Unfolded Protein Response

The mechanisms of tau aggregation and neurotoxicity is not well understood. The general consensus views the oligomeric and fibrillary tau species as the toxic and are responsible for leading to downstream gain and loss of function toxic effects (Brunden et al. 2009). NFT formation could thus prevent the direct toxic effects of oligomeric and fibrillary tau species by sequestering the toxic species into the rather stable aggregate core, protecting the neurons from various insults (Y. Wang and Mandelkow 2016). While many studies have shown tau oligomers as toxic and are able to induce the loss of dendritic spines, it is not well understood what the consequences of NFTs are (Y. Wang and Mandelkow 2016). Studies on prion disease have provided important insights into how aggregated proteins contribute to cell death via the unfolded protein response (UPR).

The UPR is activated during periods of endoplasmic reticulum (ER) stress to restore protein homeostasis (Halliday and Mallucci 2015). The UPR detects improperly folded proteins by binding immunoglobulin protein (BiP) which targets the hydrophobic regions of unfolded proteins (Gething 1999). BiP is responsible for inactivating three UPR signalling cascades by binding to protein kinase RNA-like ER kinase (PERK), inositol-requiring enzyme 1 (IRE1) and activating transcription factor 6 (ATF6), until it dissociates from the three factors and binds unfolded proteins (Bertolotti et al. 2000; Ron and Walter 2007). When BiP dissociates from PERK, PERK dimerises and autophosphorylates (PERK-P) to activate to kinase active PERK-P. Elongation initiation factor 2α (eIF2α) is a target of PERK-P and when phosphorylated (eIF2α-P), it strongly attenuates translation through inhibiting ternary complex formation by ribosomes (Harding et al. 2000; Spriggs et al. 2010).

The UPR is a cause of neurodegeneration in prion disease. In an elegant study that uses prion-infected mice as a model, the authors found increasing levels of misfolded prions led to the sustained overactivation of the UPR-PERK-eIF2α signalling pathway, causing a global repression of translational activities, synaptic failure and even neuronal death (Moreno et al. 2012). The reduction of eIF2α-P by overexpressing eIF2α-specific phosphatase and by the knockdown of prion expression using lentiviral vectors, significantly prolonged the survival of the mice from

approximately 80 days to 90 days (Moreno et al. 2012). Oral administration of a compound named GSK2606414 which inhibits PERK with high specificity, demonstrated the reversal of cognitive deficits and the prevention of clinical manifestation in the prion mouse model (Moreno et al. 2013). The striking effects of the UPR-PERK pathway on disease progression and prognosis demonstrate how protein misfolding and aggregation lead to neurodegeneration via prolonged translational suppression.

UPR activation is well documented in proteinopathies other than prion disease, including AD, PD, ALS and PSP (Halliday and Mallucci 2015). Through analysing BiP and PERK-P levels in the brains of AD and healthy controls, BiP and PERK-P were found to be upregulated in AD brains but not in controls, suggesting that UPR is activated during AD pathogenesis (Hoozemans et al. 2005; Hoozemans et al. 2009). The authors also identified that PERK-P is most abundant in neurons with diffuse localisation of phosphorylated tau protein (Hoozemans et al. 2009). Similar results of PERK-P upregulation in neurons with hyperphosphorylated tau were identified in several other tauopathies, including PSP, PiD and FTDP-17 (Nijholt et al. 2012). The spreading of tau inclusions through different brain regions suggests that tau could behave as prions do. Tau pathology have been shown to spread along neural pathways from one region to another over the course of time, and the pattern of spreading is known as the Braak stages (H. Braak and Braak 1991; H. Braak and Del Tredici 2011). Tau inclusions generally first appear in the locus coeruleus and entorhinal cortex where they spread towards the hippocampus, the amygdala, and eventually to the neocortex (Blennow et al. 2006). When enough tau and β-amyloid inclusions are found in the neocortex, symptoms of AD begin to manifest (Goedert 2015). While tau inclusions is a prominent pathological feature, it is unknown whether the formation of tau tangles is the cause or consequence of neurodegeneration. To date, it is unknown whether tau causes cell death the same way prions does via prolonged suppression of translation.

1.2.5 Overview of Tau Mediated Neurotoxicity

The distribution of NFTs correlates well with cognitive deficits in the AD brain. However, NFTs are rather inert structures and do not appear to be necessary or sufficient to cause neurodegeneration as the neurons can survive for decades (Morsch et al. 1999; Y. Wang and Mandelkow 2016). Importantly, neuronal death does not correlate with NFT formation but rather most neurons that die without showing NFT formation in several mouse models (Andorfer et al. 2005; Spires-Jones et al. 2008). The results agree with a previous study in AD brains that showed the amount of neuronal loss exceeds the amount of NFTs accumulated by several fold, which implies most neurons die without forming NFTs (Gomez-Isla et al. 1997). These results indicate that soluble tau is the toxic species rather than NFTs. Transgenic animal models expressing human tau have provided important mechanistic insights into how soluble normal and hyperphosphorylated tau species contribute to toxic cellular effects in neurons (LaFerla and Green 2012; Mhatre et al. 2013). Synaptic impairment, axonal transport defects, oxidative stress and mitochondrial dysfunction, over-stabilisation of actin filaments, DNA damage, activation of cell cycle and epigenetic changes have all been implicated in tau mediated toxicity (Frost et al. 2015).

1.2.6 Tau and Synaptotoxicity

Synaptic loss occurs when neurons fail to maintain axonal and dendritic connections, or during neuronal death (Bloom 2014). Studies have indicated that in AD, synaptic loss occurs decades before the death of neurons, as demonstrated by the disproportionate decrease synapse number and density compared to the number of neuronal cell bodies (Davies et al. 1987). Studies from other tauopathies including mild cognitive impairment and early onset AD, have also indicated synaptic loss as an early event in disease pathogenesis (Masliah et al. 2001; Scheff et al. 2006; Scheff et al. 2007). While β-amyloid toxicity has been placed upstream of tau and can trigger synaptic loss, experiments using different β-amyloid forming mouse models showed that tau is the molecular link required for β-amyloid toxicity, and a reduction in tau expression ameliorates harmful effects of β-amyloid such as neuronal and synaptic loss, reduced lifespan, memory impairment and excitotoxic seizures (Ittner et al. 2010; Ittner and Gotz 2011; K. Leroy et al. 2012;

Palop and Mucke 2010; Roberson et al. 2007). Tau mediates β-amyloid toxicity via the actions of src kinase fyn at the dendritic spines. It is a necessary component to localise fyn kinase in the dendritic spine to the N-methyl-D-aspartate (NMDA) –type glutamate receptors to cause the pathological overactiviation of NMDA receptors, leading to downstream excitotoxicity where fyn is eventually inactivated, resulting in the loss of synapses (Boehm 2013; Ittner et al. 2010).

In post-mortem AD brains, Hirano bodies are often observed where cytoskeletal proteins are abnormally organised, composing of bundled filamentous actin (Galloway et al. 1987). Actin bundles are more stable than actin filaments as they are resistant to swinholide-A, a drug that promotes the destabilisation of actin (Fulga et al. 2007). In vitro evidence suggests that tau interacts with actin filaments and promotes actin bundling through the proline-rich domain and the N-terminal domain (Henriquez et al. 1995; Moraga et al. 1993). Tau is implicated in regulating synaptic plasticity as upon synaptic activation in normal conditions, tau was shown to translocate from the dendrite to the postsynaptic compartment where they interact predominantly with filamentous actin (Frandemiche et al. 2014). Toxic β-amyloid species were able to induce the mislocalisation of tau into dendritic spines, demolishing its activity-dependent translocation in the postsynaptic compartment, suggesting a possible role of β-amyloid synaptotoxicity via tau mistranslocation in AD (Frandemiche et al. 2014).

1.2.7 Tau and Axonal Transport Deficits

Under normal conditions, *MAPT* mRNAs are targeted towards the axonal shaft via a localisation signal present at the 3' untranslated region of the *MAPT* transcripts whereas mutations in the localisation signal result in the trapping of *MAPT* transcripts and mRNA in the cell body (Aronov et al. 2001). The missorting of tau from the axons into the somatodendritic compartment is a feature in AD. Tau hyperphosphorylation breaks down the retrograde barrier, termed axon initial segment, which allows for the unilateral entering of tau into the axon, contributing to the missorting of tau (Li et al. 2011). Hyperphosphorylated pathological tau has also been implicated in impairing axonal transport processes. Under pathological conditions, c-Jun N-terminal kinase interacting protein 1

(JIP1), part of the kinesin motor complex, is found to be sequestered and trapped by hyperphosphorylated tau in the cell soma, resulting in the failure of axonal transport of cargoes that are dependent on kinesin motors (Ittner et al. 2009).

1.2.8 Tau, Energy Failure and the Aging Brain

The abnormal morphology of mitochondria is a feature found in various regions in AD brains, and that the distorted morphology is an indication of mitochondrial dysfunction (Baloyannis 2006; David et al. 2005; Kidd 1964; Schagger and Ohm 1995). Various tauopathy models have observed distorted mitochondrial morphology, impaired mitochondrial complex I and V activity, and disturbed mitochondrial fission and fusion (David et al. 2005; DuBoff et al. 2012; Schulz et al. 2012). The disrupted microtubule dynamics in tauopathies appears to be a result from the pathological over-stabilisation of actin filaments as experimental evidence suggests the presence of excess filamentous actin prevents the translocation of the mitochondrial fission protein DRP1 to mitochondria (DuBoff et al. 2012). Furthermore, DRP1 is depleted in post-mortem AD brains, further supporting the role of mitochondrial dysfunction in disease pathogenesis (X. Wang et al. 2009). Experiments in a neuronal model expressing truncated tau cleaved at asparagine 421, which is a pathological form of tau observed in AD, showed significant mitochondrial fragmentations and an enhanced loss of mitochondrial membrane potential by β-amyloid toxicity (Quintanilla et al. 2012). The synergistic role of tau pathology and β-amyloid in disrupting mitochondrial respiratory capacity is further demonstrated in triple transgenic mice expressing mutant tau/β-amyloid/presenilin, where drastic deregulation and loss of complexes I and IV function, impaired mitochondrial membrane potential, defective ATP synthesis and enhanced reactive oxygen species production were observed, establishing the molecular link between β-amyloid and tau in AD pathology (Rhein et al. 2009).

The aging brain is associated with multiple insults both cellular and environmental in nature, such as oxidative stress, DNA and RNA damage and dysfunction of antioxidant enzymes (Schmitt et al. 2012). Likewise, various tauopathy models indicate oxidative stress and free-radical damage as a

prominent feature (Cente et al. 2009; David et al. 2005; DuBoff et al. 2012). Repressor element 1-silencing transcription factor (REST) is a repressor of neuronal genes during embryonic development and its expression is reduced when neuronal differentiation has reached completion (Ballas et al. 2005). It has been shown that in the aging brain, REST expression is upregulated and controls the expression of genes involved in cell death, stress resistance and AD pathology (Lu et al. 2014). Increased REST levels in the aging brain is associated with preservation of cognitive function and increased longevity (Lu et al. 2014). Nuclear localisation of REST is lost in AD, and the loss of nuclear REST is associated with impaired cognition in AD patients compared to other AD patients without mislocalisation of REST (Lu et al. 2014). Interestingly, REST is found in autophagosomes together with phosphorylated tau, suggesting a pathogenic mechanism may be present to cause REST to be sequestered into autophagosomes by pathological proteins (Lu et al. 2014).

1.2.9 Tau as a Therapeutic Target

Several therapeutic strategies have been employed to target various aspects of tau and related mechanisms in preventing neurodegeneration. They include stabilising microtubules, attenuating tau hyperphosphorylation, preventing the formation of soluble toxic tau species and promoting the degradation of misfolding tau (V. M. Lee et al. 2011). While several drugs and immunotherapy have shown promising results in mouse models, a few drug treatments have reached phase III clinical trials and none has reached the clinic (V. M. Lee et al. 2011; Mandelkow and Mandelkow 2012; M. Morris et al. 2011).

1.2.10 Section Summary

Tau is intricately involved in various disease mechanisms, mediating β-amyloid toxicity and contributing to energy deficits, in aging neurons leading to neurodegeneration. The *MAPT* gene locates in a complex haplotypic region of linkage disequilibrium that associates with several neurodegenerative diseases. In the next section, the genetic architecture of the *MAPT* locus and its association with neurodegeneration will be introduced.

1.3 *MAPT* H1/H2 Haplotypes and Neurodegenerative Diseases

1.3.1 The Complex Genomic Architecture of the *MAPT* locus

The *MAPT* gene locates on chromosome 17q21.31, a location with an extended linkage disequilibrium (LD) structure that spans approximately 1.8 Mb (Pittman et al. 2004; Pittman et al. 2005). The 1.8 Mb chromosomal region of extended LD is structurally highly complex due to the presence of multiple low-copy repeats (LCRs), regions of duplications near the *NSF* gene, and microdeletion by non-allelic homologous recombination, and LCR copy number variation (CNV) such as duplications and triplications (Shaw-Smith et al. 2006; Stefansson et al. 2005; Varela et al. 2006). The centromeric edge of the extended LD block was identified to be in a region near rs732589, encompassing several genes including *MAP3K14* and *PLCD3*, while the telomeric edge locates in a region near the *NSF* gene before *WNT3* (Pittman et al. 2004). Within the extended 1.8 Mb LD block, a region of 900 kb chromosomal inversion occurs at palindromic LCR sequences, defining the H1 and H2 haplotypes (Figure 3) (Stefansson et al. 2005).

The 900 kb inversion encompasses 5 genes, *CRHR1*, *IMP5*, *MAPT*, *NSF* 1 – 13 and *STH* (Figure 3). *NSF* 1 – 13 is a tandem duplication of the *NSF* gene containing exons 1 – 13 of the full length gene. The duplication is larger on the H2 haplotype (280 kb) than on the H1 (127 kb), with the duplication spanning that in the H1, extending into the 5' end of the *KIAA1267* gene (Stefansson et al. 2005). The H1 and H2 haplotypes are further defined by a 238 bp deletion in the intron 9 region of the *MAPT* gene in the H2 haplotype that completely segregates with H1 and H2, including several haplotype-tagging SNPs that can capture the genomic LD diversity at this locus (Figure 3) (Baker et al. 1999; Pittman et al. 2005). A recent study has examined the structural features of the LD region with high resolution by using whole-genome sequencing data (Boettger et al. 2012). The authors first identified three common overlapping but distinctive segments, α, β and γ, of CNVs that exhibit population level patterns. The copy numbers of each CNV segment, designated α, β and γ, were then measured by both drop-based digital PCR and read depth analysis of the whole-genome sequence data in the 1000 Genomes Project. The quantified copy numbers were then deconvolved

to match the α, β and γ segments (Boettger et al. 2012). By inferring the number of copies of each CNV (α, β and γ) segments in family trios in relation to the H1 and H2 haplotypes, the authors found that the four features (α, β, γ and H1/H2 inversion) segregated as nine common CNV haplotypes. The authors further inferred the genomic locations of the structural features and identified β and γ segments as tandem duplications, while the α segment duplicated at separate sites that are 600 kb apart, agreeing with previous observations (Boettger et al. 2012; Zody et al. 2008). Using this complex genetic locus as an exemplar, the authors imputed SNP haplotypes with the structural CNV haplotypes and found that SNPs can capture structural chromosomal variations which was previously not thought to be possible (Boettger et al. 2012; Redon et al. 2006; Sharp et al. 2005).

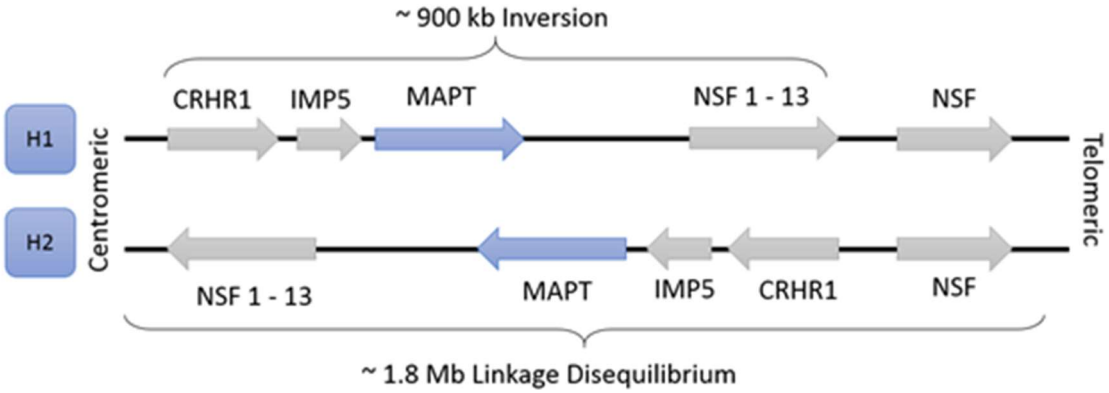


Figure 3. The *MAPT* H1/H2 haplotypes.

MAPT locates in a region of linkage disequilibrium of approximately 1.8 Mb on chromosome 17q21.31. The H1 and H2 haplotypes are characterised by a 900 kb inversion of a chromosomal region with the 1.8 Mb linkage disequilibrium block. A 238 bp insertion/deletion in intron 9 of *MAPT* also completely segregates with the haplotype. Figure not drawn to scale.

1.3.2 The Evolutionary History of the H1/H2 Haplotypes

Phylogenetic network approaches identified a much greater genetic diversity within the H1 haplotype than that compared to the H2, with many subhaplotypes existing in the H1 background and only a single H2 haplotype is present (Pittman et al. 2005). With the generation of large scale genetic variation datasets across different populations from the HapMap Project and the 1000 Genomes Project (1000 Genomes Project Consortium et al. 2012; The International HapMap

Consortium 2003), a more recent study found that a very rare H2 haplotype representing an ancestral form of the H2 allele is also present in the human population (Boettger et al. 2012). The H2 haplotype is extremely homogeneous relative to the H1 haplotype (Stefansson et al. 2005). Evidence also indicates that no recombination has occurred between the H1 and H2 chromosomes as identified by unique markers, including the intron 9 deletion, between the haplotypes (Pittman et al. 2005; Stefansson et al. 2005). Analysis of SNPs in different populations indicates that the lack of recombination between H1 and H2 is a result of the 900 kb chromosomal inversion that prevented the haplotypes from recombining (Stefansson et al. 2005).

Studies analysing the evolutionary history of the haplotypes came to different conclusions on when H1 and H2 diverged. Three studies estimated the divergence occurred approximately 3 million years ago (Cruts et al. 2005; Stefansson et al. 2005; Zody et al. 2008), while two other studies suggested a more recent time approximating between 10,000 to 1 million years for which common ancestry diverged (Donnelly et al. 2010; Hardy et al. 2005). Authors from the most recent study pointed out that both estimations could be correct as different methods of estimation were used (Donnelly et al. 2010; Stefansson et al. 2005). The authors also suggested that their data, together with the previous studies, indicate that the chromosomal inversion occurred recently, but the founding haplotype where that the H2 inversion occurred from was rather ancient, explaining the differences identified in different studies (Donnelly et al. 2010; Stefansson et al. 2005; Zody et al. 2008). H1 and H2 haplotypes are common polymorphisms with distinctive geographical distribution in different populations. The H2 haplotype is found at the highest frequency in the Mediterranean regions of southwest Asia and Europe, moderately frequent in northern Africa, western, central and southeast Europe (Donnelly et al. 2010; Stefansson et al. 2005). The H2 haplotype is present at low frequency in eastern and northern Europe, Russia, the Arabian Peninsula and in Africa, while it is almost absent in east Asia and in the Pacific Islands (Donnelly et al. 2010; Stefansson et al. 2005). Overall, the allele frequencies for the H1 and H2 haplotypes are approximately 66.6% and 33.3% respectively (Donnelly et al. 2010). Studies of the lineage frequencies also suggest that H2 arose in

Africa as indicated by the greater H2 diversity and the presence of the founder haplotypes only in the African population (Donnelly et al. 2010; Stefansson et al. 2005).

1.3.3 Linkage Analysis and Mendelian Diseases

Linkage analyses have had huge success in identifying low allele frequency genetic factors for Mendelian phenotypes where the genotype at a single locus has a one-to-one correspondence with a particular phenotype in affected families (N. J. Risch 2000). The success of linkage analyses is due to the discovery that genetic variations can be assayed and can be used as markers in linkage studies (Botstein et al. 1980). The genetic marker closest to the causal disease gene will exhibit the strongest correlation with familial pattern of inheritance in a disease, allowing for the causal gene to be identified (N. J. Risch 2000). An example of such is the identification of mutations in the *CFTR* gene on chromosome 7 as the cause of cystic fibrosis using restriction fragment length polymorphisms that are in LD with the *CFTR* mutations as genetic markers (Kerem et al. 1989). However, this approach is only feasible if the causal gene has a strong effect and a strong genotype-phenotype correlation, which is rarely the case in most common complex disorders without clear inheritance patterns or non-Mendelian in nature (N. J. Risch 2000). The high false-positive discovery rate of linkage analysis has rendered this approach non-applicable to most common human traits and diseases. This lack of power of linkage analysis in studying non-Mendelian complex traits and diseases, where the relative effects or risks conferred by a genetic locus is often low to modest, has led to an alternative approach of tracking disease loci using SNPs through their LD structure by performing genome-wide case-control association studies, or known today as genome wide association studies (GWAS) (Collins et al. 1997; N. Risch and Merikangas 1996; N. J. Risch 2000).

1.3.4 The Theories on Explaining Common Complex Diseases

Two main hypotheses, the common disease common variant (CDCV) hypothesis and the multiple rare variant (MRV) hypothesis, have been proposed to underlie the genetic risks for common complex non-Mendelian diseases (Lander 1996; Pritchard 2001). The CDCV suggests common genetic variation in the population is likely influence the risks in developing common disease and

that multiple variants contribute together to influence disease phenotypes. The very basis of this idea that common variations contribute to genetic predisposition would by definition mean that each risk associated variant must have a relatively small effect size or penetrance compared to the risk alleles found in Mendelian type disorders (Bush and Moore 2012). The MRV hypothesis argues that population level events, such as genetic drift, mutation and selection, would favour the existence of multiple rare variations that lead to common diseases (Pritchard 2001). Each rare risk associated variation should have a mild effect on its own and hence it cannot be overtly acted on by natural selection. The variation should be present at a high frequency, have an overall high mutation rate and exhibit extensive allelic heterogeneity (Pritchard 2001).

Historically the two hypotheses appear to be contradicting theories (Schork et al. 2009), however, a third hypothesis termed pleomorphic risk loci (PRL) hypothesis has emerged which suggested that the CDCV and MRV hypotheses are not mutually exclusive and in fact they can co-exist to form the PRL theory (Gibson 2011; Singleton and Hardy 2011). The PRL hypothesis proposed that genetic variation of different risk classes can exist in the same locus, for example, common non-coding variants, rare coding high effect size mutations, common coding low risk variants, rare non-coding low risk variants and rare CNVs with a moderate risk. Each risk locus identified is likely to contribute towards diseases pathogenesis where they may act by many different mechanisms including affecting gene expression, alternative splicing and coding changes (Singleton and Hardy 2011). Before the Human Genome Project and the International HapMap Project, it was almost impossible to test for the existence of low risk common alleles as there was a major lack of understanding in the LD structures of the human genomes (N. Risch and Merikangas 1996). In addition, a major technical challenge for performing association tests is the large number of SNPs, which was proposed to be well over 100,000, required to be genotyped (N. J. Risch 2000). However, as genotypic technologies have improved, chip-based microarray platforms such as Affymetrix and Illumina, have allowed the high throughput measurements of SNPs (Distefano and Taverna 2011).

At the same time, our knowledge of the human LD structure is also greatly advanced by the HapMap project, making study designs of GWAS possible (Visscher et al. 2012).

1.3.5 Genome Wide Association Studies

The knowledge and understanding of genetic variation on common complex diseases have hugely increased by the International HapMap project (The International HapMap Consortium 2003). The International HapMap project is an international effort to characterise SNPs in humans across different populations. SNPs are the most common and the smallest unit of genetic variation, and it is estimated that there are approximately 10 million SNPs in the human genome and over 3 million SNPs have been characterised by the project (The International HapMap Consortium 2003; The International HapMap Consortium et al. 2007). The enormous amount of SNP genotype data allowed the genome-wide LD structure to be assessed in great detail and have made GWAS possible. It is now understood that the pattern of LD is different in different populations as the rate of LD decay is dependent on many factors, including the population size, generation number and founding chromosomal number (Bush and Moore 2012). Tag SNPs are selected to capture the genomic LD structure such that the SNPs do not provide redundant information. Most genotyping platforms used in GWAS use approximately 500,000 to 1 million tag SNPs to capture the genomic diversity and the LD structure (Grant and Hakonarson 2008). As the SNPs sampled in GWAS account for less than 10% of all SNPs present in the genome, the causative SNPs or genetic elements are unlikely to be sampled themselves and are thus more likely to be found in linkage disequilibrium (LD) with the GWAS risk SNPs (The International HapMap Consortium 2003).

GWAS is designed to detect commonly occurring genetic variants that are of low penetrance, or effect size, in contributing to common diseases or traits (Bush and Moore 2012). There are two main classes of phenotypes GWAS can measure, namely quantitative and categorical. An example of quantitative phenotype is the levels of high-density and low-density lipoprotein cholesterol as traits for cardiovascular diseases (Teslovich et al. 2010). However, very often complex diseases, such as multiple sclerosis (MS), are clinically heterogeneous and do not have well-established

quantitative phenotypes. In this case individuals can be classified into either affected or unaffected, or essentially a case-control situation, hence a categorical measurement can be used (Habek et al. 2010). An important control in GWAS design is the standardisation of phenotyping criteria. The McDonald diagnostic criteria for MS is a good example where clinical assessments can be unified and applied across different clinics to ensure consistent diagnosis for MS is made by different clinicians (Polman et al. 2005). The standardisation of phenotyping is crucial to minimising diagnostic variability and hence GWAS outcome.

Different types of statistical tests can be performed to analyse the association of genotype and phenotype depending on the measured traits. For example, quantitative traits are usually analysed using the analysis of variance (ANOVA) approach, while categorical traits are generally analysed using the chi-square test or logistic regression methods (C. M. Lewis 2002). Statistical analyses are usually adjusted for clinical covariates such as age and gender, while ancestral information is usually used to prevent population stratification where the dataset is selected for according to the ethnic groups of the samples (Falush et al. 2003; Price et al. 2006). P-values obtained are corrected for multiple testing by using Bonferroni correction, false discovery rate, permutation testing or simply apply the threshold of genome wide significance (Dudbridge and Gusnanto 2008; Hochberg and Benjamini 1990; Purcell et al. 2007; van den Oord 2008). GWAS is usually repeated in subsequent replication studies for validating the findings and meta-analyses are performed to examine and refine the relative risks of the loci and also their significance from multiple studies (NCI-NHGRI Working Group on Replication in Association Studies et al. 2007; Zeggini and Ioannidis 2009).

1.3.6 GWAS Identified the H1 Haplotype a Risk Locus for PD, PSP and CBD

Since the first GWAS reported in 2005 (R. J. Klein et al. 2005), over 2400 GWAS have been conducted on various human traits and diseases, leading to the identification of over 18,000 tag SNPs associating with over 1500 diseases and traits (Burdett et al.). The first large scale, well designed and controlled GWAS was published in 2007 conducted by the Wellcome Trust Case

Control Consortium, which employed approximately 14000 case individuals of 7 major diseases and a shared set of approximately 3000 control individuals (Wellcome Trust Case Control Consortium 2007). The study was the first study to employ the Affymetrix GeneChip which samples 500,000 SNPs, providing a good coverage of the genomic LD structure. The study identified 24 susceptibility loci for bipolar disorder, coronary artery disease, Crohn's disease, rheumatoid arthritis, type 1 and 2 diabetes, providing new important insights into the pathophysiology of the diseases (Wellcome Trust Case Control Consortium 2007). GWAS has since proven to be a powerful tool to identify susceptibility loci for common traits and diseases.

Prior to the GWAS era, genetic association studies on the *MAPT* haplotypes have linked the H1 and H2 haplotypes to several neurodegenerative disease, including Parkinson's disease (PD), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), frontotemporal dementia (FTD) and Alzheimer's disease (AD) (Baker et al. 1999; Conrad et al. 1997; Di Maria et al. 2000; Fung et al. 2006; Houlden et al. 2001; Laws et al. 2007; H. R. Morris et al. 1999; Myers et al. 2007; Oliva et al. 1998; Pastor et al. 2004; Pittman et al. 2005; Refenes et al. 2009; Skipper et al. 2004; Tobin et al. 2008; Vandrovcova et al. 2009; Webb et al. 2008; Wider et al. 2010; Williams et al. 2007; Zabetian et al. 2007). While genetic studies on the association of the *MAPT* haplotypes for FTD and AD has been inconclusive, the H1 polymorphisms have been repeatedly and consistently shown to associate with PD, PSP and CBD. PSP and CBD are tauopathies, where tau inclusions are a prominent pathological feature, therefore the association of the H1 *MAPT* haplotype with the two diseases implies that haplotype-specific features may underlie disease pathogenesis. However, PD is not a tauopathy and is characterised by α-synuclein pathology, thus it is unclear which aspects of the associated haplotypic feature of the *MAPT* gene may underlie disease pathogenesis, or indeed if the *MAPT* gene is in fact at all the main feature implicated in PD given the H1/H2 LD block spans 1.8 Mb. Indeed, it has been shown that different H1 subhaplotypes underlie PSP and PD risks, indicating that different features in the haplotypes may drive the risk factors in diseases (Wider et al. 2010).

The first GWAS in PD was conducted in 2009 employing approximately 1700 PD individuals and approximately 4000 control individuals of European ancestry (Simon-Sanchez et al. 2009). The *MAPT* H1 haplotype was identified as one of the leading loci that underlie increased risks for PD (Simon-Sanchez et al. 2009). Multiple GWAS have since been performed and repeatedly confirmed the H1 haplotype as a risk locus in replication studies, in different populations, and in meta-analyses (Do et al. 2011; Edwards et al. 2010; Hamza et al. 2010; Hill-Burns et al. 2014; Lill et al. 2012; Nalls et al. 2011; Nalls et al. 2014; Spencer et al. 2011; Vacic et al. 2014). The latest mega meta-analysis reported an odds ratio of 1.3 for the association of H1 haplotype with PD (Nalls et al. 2014). Here, GWAS have demonstrated consistently polymorphisms in the H1 haplotype confer increased risks for PD, a common complex disorder that is mostly sporadic in nature. The association of the H1 haplotype with FTD and AD however, has not been replicated in GWAS, suggesting that previous association studies performed prior to GWAS have identified the H1 haplotype as a false positive result. GWAS also identified the H1 haplotype as a high risk locus for PSP and CBD with odds ratios of 5.5 and 3.7, respectively, agreeing with previous reports that the H1 haplotype is over-represented in PSP and CBD (Hoglinger et al. 2011; Kouri et al. 2015; Pittman et al. 2005).

Given that most DNA sequences in the human genome are non-coding in nature, it is not surprising that the majority of GWAS risk SNPs identified map to non-coding regions (Abecasis et al. 2010; Frazer et al. 2009; Freedman et al. 2011; Schaub et al. 2012). Together with the recent advancements in post-GWAS interpretation methods, increasing evidence has pointed towards the enrichment and the functionality of GWAS risk SNPs and their associated SNPs in non-coding regulatory elements such as epigenetic markers, transcription factor binding sites, DNase I hypersensitive sites, RNA splicing and gene expression (Cowper-Sal lari et al. 2012; Maurano et al. 2012; Pomerantz et al. 2009; Riaz et al. 2012; Schaub et al. 2012). All of the above highlight the importance of understanding the functional polymorphisms within large expanses of LD which is often challenging due to the difficulty of working with large genomic regions in models of disease and the potential subtle effects of functional polymorphisms.

1.3.7 Section Summary

The H1 and H2 *MAPT* haplotypes locate on chromosomal region 17q21.31. The haplotypes are characterised by a 900 kb chromosomal inversion, a 238 bp insertion/deletion at *MAPT* intron 9, and various haplotype-tagging SNPs. GWAS identified common polymorphisms in the H1 haplotype to associate with increased susceptibility for PD, PSP and CBD. In the next section, several aspects of PD, PSP and CBD will be introduced, including the clinical presentations, neuropathology and disease genetics.

1.4 Parkinson's Disease

1.4.1 Parkinson's Disease

Parkinson's disease (PD) was first described by James Parkinson, a surgeon from London, in 1817 in his monograph "*An Essay on the Shaking Palsy*", detailing six cases where patients exhibit signs of paralysis agitans or shaking palsy. The shaking palsy was define as "*involuntary tremulous motion, with lessened muscular power, in parts not in action and even when supported; with a propensity to bend the trunk forwards, and to pass from a walking to a running pace: the senses and intellects being uninjured.*" The movement disorder James Parkinson described is known today as PD. The prevalence of PD is approximately 180 in 100,000 worldwide with an average age of onset at 60 (Quinn et al. 2009). Two centuries on from James Parkinson's monograph, a lot has been discovered regarding PD disease pathogenesis and genetics. However, there is currently still no disease modifying treatment available, and the cause of PD is still unknown.

1.4.2 Clinical Features of Parkinson's Disease

PD is a heterogeneous disorder both clinically, pathologically and genetically (Figure 4). The clinical diagnostic criteria which was first elaborated in 1988 by the UK Parkinson's Disease Society Brain Bank, is currently still the most widely used clinical diagnostic criteria for PD (Quinn et al. 2009). The diagnosis of PD includes the cardinal features of parkinsonism. Bradykinesia is an obligatory sign, and is defined as the slowness of initiation of voluntary movement with progressive reduction in the speed and amplitude of repetitive actions. At least one or more signs of parkinsonism including muscle rigidity, resting tremor of 4 – 6 Hz and postural instability not caused by visual, vestibular, cerebellar or proprioceptive dysfunction should also be present. Supportive diagnostic criteria for PD include asymmetric onset, persistent asymmetry of signs, excellent response to levodopa treatment, levodopa responsive for 5 years or more, progressive worsening of symptoms, resting tremor, levodopa induced chorea, and a clinical course of a decade or above. For the definitive diagnosis of PD, at least 3 or more of the supportive diagnostic criteria is required (Quinn et al. 2009). Despite classically considered as a movement disorder, PD is also characterised by a

wide range non-motor features, which are often described by patients to be more disabling and disruptive than the obvious motor signs. Preclinical symptoms can include constipation, olfactory impairment, rapid eye movement sleep behaviour disorder and dysphagia can precede long before the motor signs are observed. Non-motor features that occur after onset can include depression, anxiety or panic, insomnia, day time sleepiness, fatigue, impotence, psychosis and hallucinations, pain, urinary urgency and frequency, drenching sweats and sialorrhoea or excessive drooling (Lees et al. 2009).



Figure 4. An illustration of a Parkinson's patient.
This diagram illustrates the characteristic flexed posture of a Parkinson's patient. The frozen "hangdog" facial feature is also typical of Parkinson's patient. Notice the thumb appears to be pilling against the index finger, illustrating the pill-rolling tremor, a sign of parkinsonism. Figure adapted from (Lees et al. 2009).

1.4.3 Treatments for Parkinson's Disease

As mentioned before, PD is currently incurable. However, symptomatic treatments are available from disease onset. The most potent method available for treating PD is the use of levodopa combined with peripheral decarboxylase inhibitors such as carbidopa or benserazide. Levodopa is the precursor of dopamine which can cross the blood brain barrier and is synthesised into dopamine once when it reaches the brain, providing stimulation for the dopaminergic receptors in

the striatum. Peripheral decarboxylase inhibitors prevents levodopa from being metabolised before it reaches the brain. Marked improvement on the motor symptoms should be observed at early stages of the disease. However, long-term levodopa syndrome will develop in almost all the patients 10 years into the treatment where akinesia occurs when levodopa wears off, and dyskinesia occurs when levodopa takes effect. Other drugs are also available for augmenting levodopa treatment, such as catechol-o-methyl transferase inhibitors for extending the half-life of levodopa and monoamine oxidase B inhibitors to potentiate levodopa side effects (Quinn et al. 2009). Patients with advanced PD may exhibit dyskinesias or troublesome tremors, which cannot be treated by drugs, can undergo surgery for deep brain stimulation (DBS) electrode implants that delivers high-frequency impulses. DBS mainly targets several brain regions in the basal ganglia, including the thalamus, subthalamic nucleus and globus pallidus internus (Quinn et al. 2009).

1.4.4 Neuropathological Features of Parkinson's Disease

PD is characterised pathologically by the selective death of dopaminergic, neuromelanin-containing neurons in the substantia nigra pars compacta region. In addition, more widespread neuronal pathology in other brain regions also occur, including the locus coeruleus, serotonergic raphe nuclei, dopaminergic mesolimbic pathways, cholinergic nucleus basalis of Meynert, cerebral cortex, hypothalamus, olfactory tract and sympathetic ganglia (Lees et al. 2009; Quinn et al. 2009). At the microscopic level, PD is an α-synucleinopathy, characterised by the formation of neuronal Lewy body and pale body in the cytoplasm and Lewy neurite in the processes (Dickson 2012). Lewy bodies found in the highly affected brainstem region and in other less vulnerable regions such as the cortex and the amygdala, have different morphologies and can be divided broadly into the classical and cortical Lewy bodies (Dickson 2012).Classical Lewy bodies present as spherical structures of 8 – 30 μm, with a hyaline core surrounded by a pale halo. Cortical Lewy bodies lack the hyaline core and the halo, and are often more pale than classical Lewy bodies (Dickson 2012; Lees et al. 2009). Abnormally aggregated α-synuclein forms the main component of Lewy bodies and Lewy neurite (Lees et al. 2009).

1.4.5 Genetics of Parkinson's Disease

PD is a largely sporadic disorder with only approximately 10% of familial cases (C. Klein and Westenberger 2012). The complex genetic architecture of PD encompasses a broad range of risk factors, including low frequency high risk mutations, moderate frequency moderate risk mutations and high frequency low risk common variations (van der Brug et al. 2015). Mendelian mutations in *SNCA*, *LRRK2*, *VPS35*, *EIF4G1*, *DNAJC13*, *PARK2*, *PINK1*, *DJ-1*, *DNAJC6*, *ATP13A2*, *FBXO7*, *PLA2G6*, *ATP6AP2*, and *SYNJ1* have been identified in linkage analysis to underlie familial forms of PD (Lin and Farrer 2014). Lysosomal pathways such as endosomal trafficking, and autophagic dysfunction are implicated by *VPS35* and *ATP13A2* mutations. Mitophagy dysfunction are implicated in mutations in *PARK2*, *PINK1* and *FBXO7* where the loss of function of the proteins caused by the mutation led to the disruption of mitochondrial recycling process (van der Brug et al. 2015).

Missense mutations in *LRRK2* (leucine-rich repeat kinase 2) is the most common cause of familial PD, and the clinical presentations are almost identical to that observed in idiopathic cases. *LRRK2* is both a kinase and a Ras GTPase, with roles implicated in organising actin cytoskeleton (Marin et al. 2008). The kinase, GTPase and C-terminal domains of *LRRK2* appear to interact with each other and function accordingly, however, the complex roles of each domain have yet to be fully understood (Gilsbach and Kortholt 2014). *SNCA* (α-synuclein) locus duplication and triplication, and missense mutations, lead to PD with dementia and Lewy body pathology (O. A. Ross et al. 2008). α-synuclein participates in various pre-synaptic processes, hence its pathological aggregation and the various mutations identified indicate both loss and gain of function mechanisms in disease pathogenesis (Burre et al. 2010; Olanow and Prusiner 2009). In addition, the prion-like spreading of α-synuclein may also underlie disease pathogenesis (H. Braak et al. 2003).

GBA (glucocerebrosidase) mutations have been found to confer moderate risks for PD without full penetrance. While individuals carrying homozygous *GBA* mutations will develop Gaucher's disease, heterozygous carriers will develop a five-fold increase in the risk for PD instead of developing Gaucher's (Sidransky and Lopez 2012). Interestingly, GBA and α-synuclein have been shown to

interact with each other, resulting in decreased lysosomal function and increased α-synuclein aggregation, indicating lysosomal dysfunctional as a major pathogenic mechanism in PD (Mazzulli et al. 2011). Interestingly, unlike the Mendelian mutations mentioned above, the penetrance of the *LRRK2* mutation G2019S appears to be dependent on the ethnic background and the risks vary from approximately 22% in Ashkenazi Jews in America to 80% in Arab-Berbers in Tunisia and Ashkenazi Jews in Israel (Clark et al. 2006; Hentati et al. 2014; Ozelius et al. 2006).

GWAS in PD have identified multiple low risk loci that contribute to sporadic PD, including *SNCA*, *MAPT*, *LRRK2*, *GBA*, *GAK*, *BST1*, *LAMP3*, *HLA-DRB5*, *STK39* and *SYT11* (van der Brug et al. 2015). *SNCA* and *MAPT* have consistently been identified as the most significant risk loci in PD GWAS (Lin and Farrer 2014). The identification of *SNCA* mutations in linkage analysis and polymorphisms in the *SNCA* locus by GWAS, together with the pathological aggregation of α-synuclein as a defining neuropathology, strongly places *SNCA* at the centre of PD pathogenesis. The mechanisms of how polymorphisms in the *MAPT* haplotypes contribute to disease pathogenesis however, is currently unknown. Test of epistatic interaction between *SNCA* and *MAPT* loci did not identify interactions between the loci, indicating they may contribute to PD progression via different mechanisms (Simon-Sanchez et al. 2009).

1.4.6 Progressive Supranuclear Palsy and Corticobasal Degeneration

PSP and CBD are rare 4R tauopathies that share parkinsonism as one of the features during disease progression. PSP and CBD are sometimes referred to as “Parkinson’s-plus” disorders as patients are often misdiagnosed as PD cases due to the parkinsonian features. Basal ganglia dysfunction, cortical deafferentation and loss of ascending projections from subcortical structures are observed in PSP and CBD, which result in dementia and neuropsychiatric features (Obeso et al. 2014). Their clinical presentations can be heterogeneous but each has a distinctive set of distinguishing features that could aid in the differential diagnosis between PSP, CBD and PD (Kouri et al. 2011; Ludolph et al. 2009; Obeso et al. 2014). Genetically PSP, CBD and PD all share the increased susceptibility by the

common H1 *MAPT* haplotypic polymorphisms (Hoglinger et al. 2011; Kouri et al. 2015; Nalls et al. 2014).

PSP is a sporadic disorder and the prevalence is approximately 5 in 100,000 with an average age of onset of 63. The mean survival is 7 years (Quinn et al. 2009). Motor features include rigidity, akinesia, backward falls, dysarthria, dysphagia, blepharospasm, severe postural instability, motor recklessness, sitting *en bloc* (Quinn et al. 2009). Prominent eye features are a typical sign of PSP, and can include frontalis over-activity that gives rise the typical wide-eye stare, levator inhibition or apraxia of eyelid opening and supranuclear gaze palsy. Supranuclear gaze palsy is an obligatory sign for PSP diagnosis, and is also an exclusion criteria for PD (Quinn et al. 2009). Non-motor features include emotional incontinence, memory deficit, cognitive impairment, sialorrhoea, bradyphrenia and executive deficits such as perseverative responses, environmental dependency and utilisation behaviours. The typical applause sign is a characteristic of the executive deficits observed in PSP patients where when patients are asked to clap three times as fast as possible after the clinician, patients often make additional claps or even begin an applause (Quinn et al. 2009; Rossor et al. 2009). Familial cases are rare, however, genetic mutations in *MAPT* and *LRRK2* have been identified to cause familial PSP (Im et al. 2015).

CBD is a sporadic disorder with a prevalence rarer than PSP. While the exact prevalence is unknown, crude estimations placed the number of CBD cases at about 4.9 – 7.3 in 100,000. Average age of onset is 63. The average survival is 7 – 8 years (Kouri et al. 2011; Quinn et al. 2009). Motor features include akinesia, rigidity, fixed dystonia, myoclonus, tremor and apraxia. Alien limb phenomenon may associate with CBD, where patients do not have control over the use of the limb, with the alien limb carrying out actions opposing the actions of the limb under control (Quinn et al. 2009). CBD can be distinguished from PSP by its early features of prominent cortical deficits such as bucco-facial apraxia, impaired speech, and parietal signs. Similar to PD, there is currently no disease modifying treatment available for both PSP and CBD.

1.4.7 Section Summary

PD, PSP and CBD are clinically, pathologically and genetically heterogeneous neurodegenerative disorders. They are currently incurable although symptomatic managements are possible. PD, PSP and CBD are largely sporadic in nature, however, common genetic variations have been identified to contribute towards increased disease susceptibility.

1.5 H1/H2 Haplotype-Specific Exon 3 Expression and Neurodegenerative Diseases

1.5.1 Haplotype-Specific Exon 3 Expression in Humans

The identification of two principal genetic H1 and H2 haplotypes at the *MAPT* locus, and the subsequent association of the H1 haplotype with PSP (odds-ratio 5.5), CBD (odds-ratio 3.7) and PD (odds-ratio ~1.7) in multiple GWAS and meta analyses placed *MAPT* among the most important genetic risk loci in neurodegeneration (Hoglinger et al. 2011; Kouri et al. 2015; Nalls et al. 2014; Simon-Sanchez et al. 2009). Recent post-GWAS analyses have pointed towards the enrichment and the functionality of GWAS risk SNPs and their associated SNPs in non-coding regulatory elements such as epigenetic markers, transcription factor binding sites, DNase I hypersensitive sites (representing active chromatin structures), RNA splicing and gene expression (Cowper-Sal lari et al. 2012; Maurano et al. 2012; Pomerantz et al. 2009; Riaz et al. 2012; Schaub et al. 2012).

The functions of non-coding GWAS risk and their associated SNPs could explain the contributions of the risk SNPs to disease predisposition and their complex combinatorial effects in diseases, highlighting the importance to understand their functional mechanisms in the genome. Two main mechanisms could be under the influence of H1/H2 haplotypes: gene expression and alternative splicing. Previous evidence has suggested that the expression levels of *MAPT* are increased from the H1 allele (Myers et al. 2007). However, more recent studies indicate that this is likely an artefact created by the haplotype-specific polymorphisms, in which the probes were initially mainly designed for the H1 haplotype (Trabzuni et al. 2012). The tau N-terminal domain is of particular interest here because its normal functions and its contributions to neurodegeneration were not well studied.

The H2 *MAPT* haplotype was previously shown to express exon 3-containing transcripts two-fold higher than the H1 haplotype in post-mortem human brain and in a neuronal cell line (Caffrey et al. 2008). This two-fold increase of exon 3 expression by the under-represented H2 allele was

proposed to be protective against PD, PSP and CBD (Caffrey et al. 2008). While mutations that cause frontotemporal dementia and parkinsonism linked to chromosome 17 lead to an increase of exon 10 expression, no mutation has been identified to alter exon 3 inclusion (Goedert et al. 2012b). The H1/H2 haplotype-specific differences observed in exon 3 splicing are therefore likely to arise from the haplotype-specific polymorphisms that are different between the H1 and H2 alleles.

1.5.2 *MAPT* Exon 3 and Neurodegenerative Diseases

Evidence that an altered ratio of *MAPT* isoforms can have functional consequences first came from mutations identified in the *MAPT* gene. In 1998, several studies found mutations within the *MAPT* gene in families with frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17) (Hutton et al. 1998; Poorkaj et al. 1998; Spillantini et al. 1998b; Spillantini et al. 1998a). To date, at least 39 different *MAPT* mutations have been identified to cause FTDP-17 (F. Liu and Gong 2008). The effects of the mutations on tau can be classified into two main groups. Coding changes caused by missense mutations or deletions present in exons 1, 9, 10, 11, 12 and 13 affect the interaction between tau proteins and microtubules, resulting in reduced microtubule assembly and increased aggregation of mutated tau proteins into filaments (Buee et al. 2000). Non-coding changes caused by silent mutations in exons 10 and 11 or intronic mutations in introns 9 and 10 affect the alternative splicing of exon 10, resulting in an increase in exon 10 inclusion and an increase in the normal 1:1 ratio of 4R (exon 10+) to 3R (exon 10-) tau in the adult human brain (F. Liu and Gong 2008). The two groups of mutations both predominantly associate with the formation filamentous tau inclusions composed mainly of the 4R isoforms, demonstrating that the misregulation of tau exon 10 alone is sufficient to cause neurodegeneration (F. Liu and Gong 2008). More importantly, the silent and intronic mutations demonstrate that an imbalance in the normal 1:1 4R:3R tau ratio alone with no changes in the amino acid sequence is enough lead to pathological aggregation of 4R tau.

Neuropathological features of PSP and CBD provide further insights into how increased exon 10 expression and decreased exon 3 expression by the H1 allele may lead to genetic predisposition by

common polymorphisms in the H1 haplotype. PSP and CBD are 4R tauopathies characterised by the formation of insoluble aggregated tau proteins predominantly composed of the 4R isoforms (Arai et al. 2001; Dickson et al. 2011; Gibb et al. 2004). While 4R:3R tau expression in the normal adult human brain is maintained at approximately a 1:1 ratio, gene expression analyses have found 4R:3R transcript ratio to be two to three-fold higher in vulnerable brain regions in PSP and CBD (C. B. Chambers et al. 1999; Goedert and Jakes 1990; Hong et al. 1998; Kosik et al. 1989a; W. K. Liu et al. 2001; Takanashi et al. 2002). The H1-associated risk to PSP and CBD may thus partially be contributed by the increased exon 10 expression (Caffrey et al. 2006). Insoluble tau aggregates in PSP and CBD were found to lack 2N (exon 3+) isoforms in neuropathological studies (Arai et al. 2001; Feany et al. 1995; Gibb et al. 2004; Ksiezak-Reding et al. 1994; Terada et al. 2005). *In vitro* fibrillisation assays using human recombinant tau showed that different tau isoforms intrinsically have different aggregation propensities (Zhong et al. 2012). In the presence of Thiazine red inducer, 4R tau isoforms have much higher tendency to nucleate, nucleation rate and aggregation propensity, while 2N tau isoforms antagonise the aggregation propensity by lowering the efficiency of monomer aggregation (Zhong et al. 2012). Interestingly, the efficacy of 2N tau to inhibit tau aggregation is slightly reduced if 4R tau is present, while 4R tau has higher aggregation propensity than 3R tau (Zhong et al. 2012).

H1 polymorphisms in the extended LD block at the *MAPT* locus on chromosome 17q21.31 are a leading genetic risk factor for PD, PSP and CBD. In addition, tau proteins mediate various pathological processes in neurodegeneration. Understanding the mechanisms of how H1/H2-specific SNPs contribute to the H1-associated risk or H2-associated protection against neurodegeneration through decreasing and increasing exon 3 expression, respectively, will provide important insights into diseases mechanisms.

1.6 Aims of This Thesis

The main aims of this thesis are to investigate whether H1/H2-specific non-coding polymorphisms regulate *MAPT* exon 3 splicing and the underlying mechanisms.

1. Chapter 3 – *MAPT* H1/H2 haplotype genomic expression PAC vectors

   To identify haplotype-specific SNPs that may affect the allele-specific expression of *MAPT* exon 3-containing transcripts and to generate haplotype-hybrid genomic vectors carrying the complete *MAPT* locus.

2. Chapter 4 – Expression of pMAPT genomic vectors in SK-N-F1 neuroblastoma cells

   To study the effects of candidate SNPs using the haplotype-hybrid genomic expression vectors.

3. Chapter 5 – RNA–protein interaction studies on H1 and H2 rs17651213 and rs1800547 sequences

   To determine the effects of H1/H2 functional SNPs on RNA-protein interactions and to identify their interacting splice factors.

4. Chapter 6 – Investigating the allele-specific regulation of *MAPT* exon 3 expression

   To study the allele-specific regulation of *MAPT* exon 3 inclusion by splice factors interacting with H1/H2 functional SNPs.

CHAPTER 2

MATERIALS AND METHODS

2.1 Chapter 3 Methods

Chemicals were obtained from Sigma Aldrich unless otherwise stated.

2.1.1 Sequence Comparison of pMAPT-H1 and pMAPT-H2 Vectors

Genomic coordinates of the H1 and H2 *MAPT* vectors sequences were calculated by aligning the sequence of the vectors with the *MAPT* sequence obtained from the UCSC genome browser (https://genome.ucsc.edu/cgi-bin/hgGateway) GRCh38/hg38 assembly using APE – A Plasmid Editor (http://biologylabs.utah.edu/jorgensen/wayned/ape/). SNP data was downloaded from the Variation Viewer (http://www.ncbi.nlm.nih.gov/variation/view/), a National Center for Biotechnology Information (NCBI) tool. Only SNPs with a minor allele frequency of >=0.05 were included in the download. The downloaded SNP data were viewed in Microsoft Excel. To determine whether a SNP is different between the H1 and H2 *MAPT* vectors, the vectors were aligned using APE – A Plasmid Editor. SNPs present were manually compared to the downloaded SNP data in Microsoft Excel and the percentage of SNPs captured by the H1 and H2 vectors were calculated.

2.1.2 Splice Factor Binding Site Analysis

H1 and H2 sequences containing SNP rs117124984, rs118087478, rs77875796, rs1800547 and rs17651213 were analysed using (i) SFmap (http://sfmap.technion.ac.il/) with the WR scoring function with medium stringency; (ii) SpliceAid2 (http://193.206.120.249/splicing_tissue.html) with the default settings; (iii) ESEfinder (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home) with the default settings. Results obtained were manually compared and combined.

2.1.3 Analysis of Linkage Disequilibrium Structure

Haplotype tagging SNPs were obtained from (Simon-Sanchez et al. 2009). SNP data of the CEU population (Utah residents with Northern and Western European ancestry from the CEPH collection) was obtained from the HapMap database (HapMap_rel27) via HapMart (http://hapmap.ncbi.nlm.nih.gov/biomart/martview/). SNP data was exported in the .xml format. SNP data was then converted to .info and .ped files using SNP tools for Microsoft Excel (http://www.bioinformatics.org/snp-tools-excel/). Haploview 4.2 (http://www.broadinstitute.org /haploview/haploview) was used to analyse the data in the .info (locus information file) and .ped (pedigree data) files. Linkage disequilibrium heat plot was generated by Haploview.

2.1.4 Mini Preps

Bacterial colonies were cultured in 1.5 mL Luria Broth (LB) (Fisher) containing the appropriate antibiotics. The cultures were incubated at 37^{o}C overnight at 225 rpm. Cell pellets were then obtained by centrifugation at 5000 rpm for 10 minutes in a bench-top centrifuge. The supernatant was aspirated and the pellets were resuspended in STET buffer [8% sucrose, 5% Triton X-100, 50 mM EDTA and 50 mM Tris-HCl (pH 8)] by pipetting the pellet up and down. Bacterial cells were then lysed in 200 µL lysis buffer (1% sodium dodecyl sulfate, 0.2 M NaOH) for no more than 5 minutes. Lysis was terminated by the addition of 150 µL 7.5M ammonium acetate. The reaction mixture was placed on ice for five minutes and were centrifuged at 13000 rpm for 20 mins at 4^{o}C. The supernatant containing the plasmid DNA was then transferred to a new microcentrifuge tube. Plasmid DNA was precipitated by the addition of 250 µL room temperature isopropanol. The reaction mixture was mixed by inversion and DNA was pelleted by centrifugation at 10000 rpm for 8 minutes. The supernatant was aspirated and 200 µL 70% ethanol was added to wash the DNA pellet and the reaction mixture was centrifuged at 10000 rpm for 8 minutes. The 70% ethanol supernatant was aspirated and the DNA pellet was air dried for 5 – 10 minutes. DNA was resuspended in 50 µL TE buffer (5 µg mL^{-1} RNase A, 10 mM Tris-HCl, 1 mM EDTA, pH 8).

2.1.5 Maxi Preps

Bacterial colonies were cultured in 1.5 mL LB (Fisher) containing the appropriate antibiotics. The cultures were incubated at 37°C for 8 hours at 225 rpm. The bacterial cultures were then transferred to 250 mL LB containing the appropriate antibiotic and were incubated at 37°C overnight at 225 rpm. The following buffers (P1, P2, P3, QBT, QC and QF) used were from QIAGEN Plasmid Maxi Kit. The cultures were then centrifuged at 6000 rpm for 15 minutes at 4°C. The supernatant was discarded and each cell pellet was resuspended in 15 mL P1 buffer by incubation at 37°C at 225 rpm for 10 minutes. 15 mL of P2 buffer was then incubated with the cell + P1 mixture for cell lysis for 5 minutes with gentle swirling every minute. 15 mL of ice cold P3 buffer was then added to the mixture and mixed immediately through inversion of the container for neutralising the lysis buffer. The mixture was then kept on ice for 20 minutes and then centrifuged at 9000 rpm for 35 minutes at 4°C. The supernatant was transferred onto QIAGEN-tip 500 column, through KimWipe tissue as a filter, which was pre-equilibrated with 15 mL QBT buffer. The filtered supernatant was allowed to run through the column and the column was then washed twice with 30 mL QC buffer. DNA was eluted using 15 mL QF solution pre-heated at 50°C. 10.5 mL isopropanol was added to the eluted DNA for DNA precipitation and was mixed gently by inversion of the container. The mixture was centrifuged at 14000 rpm for 30 minutes at 4°C to pellet the DNA. The supernatant was discarded and 3.5 mL 70% ethanol was added to the DNA pellet. The pellet was centrifuged again at 14000 rpm for 30 minutes at 4°C. The supernatant was discarded and the DNA pellet was air dried for 15 minutes. DNA was resuspended with 250 µL TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8) by incubation at 4°C overnight without shaking.

2.1.6 Generating Electrocompetent Cells

Bacterial colonies were incubated in 1.5 mL LB containing the appropriate antibiotics at 37°C overnight at 225 rpm. The bacterial cultures were transferred to 100 mL LB containing the appropriate antibiotics and was incubated at 37°C for 1.5 hours. The optical density of the cultures at 600 nm (OD 600 nm) was measured until OD 600 nm reaches 0.35 – 0.36. The cultures were then incubated on ice in the cold room (4°C) for 40 minutes. Bacteria was pelleted at 5000 rpm for 15

minutes at 4°C and supernatant was discarded. To wash pellets, 100 mL ice cold autoclaved 10% glycerol was added to resuspend the pellets. The mixture was centrifuged at 6000 rpm and the supernatant glycerol was discarded. The glycerol wash step was repeated two more times, with centrifugation speed at 7000 rpm and 7500 rpm in the subsequent steps. After the final wash, the supernatant was discarded and the pellet was resuspended in the remaining glycerol in the container in the cold room (4°C). Electrocompetent bacteria was aliquoted in pre-chilled tubes and was snap frozen on dry ice and stored at -80°C.

2.1.7 Generating Electrocompetent Cells with Induction of Homologous Recombination Enzymes

DH10B bacterial colonies were incubated in 1.5 mL LB containing the appropriate antibiotics at 30°C overnight at 225 rpm. The bacterial cultures were transferred to 100 mL LB containing the appropriate antibiotics and was incubated at 37°C for 1 hour. The optical density of the cultures at 600 nm (OD 600 nm) was measured until OD 600 nm reaches 0.10 – 0.13. 1.5 mL 10% L-arabinose was added to the cultures and the cultures were incubated at 37°C until OD 600 nm reaches 0.30 – 0.35. The cultures were then incubated on ice in the cold room (4°C) for 40 minutes. Bacteria was pelleted at 5000 rpm for 15 minutes at 4°C and supernatant was discarded. To wash pellets, 100 mL ice cold autoclaved 10% glycerol was added to resuspend the pellets. The mixture was centrifuged at 6000 rpm and the supernatant glycerol was discarded. The glycerol wash step was repeated two more times, with centrifugation speed at 7000 rpm and 7500 rpm in the subsequent steps. After the final wash, the supernatant was discarded and the pellet was resuspended in the remaining glycerol in the container in the cold room (4°C). Electrocompetent bacteria was aliquoted in pre-chilled tubes and was snap frozen on dry ice and stored at -80°C.

2.1.8 Electroporation

Electrocompetent cells, DNA template and electroporation cuvettes (BioRad) were pre-chilled on ice. 50 – 500 ng PAC/BAC DNA was added to electrocompetent cells thawed on ice. Cells and the DNA were mixed and were transferred to the electroporation cuvette. Electroporation was performed under 35 µF (capacitance), 200 Ω (resistance) and 1.8 kV (voltage) using BioRad

GenePulser XCell. Immediately following electroporation, 450 µL of S.O.C. medium (Invitrogen) was added to the cells and the cells were transferred to an incubation tube for incubation at 37°C for 1 hour at 225 rpm. Cells were then plated onto LB agar (Fisher) plates containing the appropriate antibiotics and were incubated at 37°C overnight.

2.1.9 Exchanging H1/H2 SNPs rs117124984, rs118087478 & rs77875796 on pMAPT Vectors

Primer pairs and sequences in Table 1.

1st PCR for cloning rpsl/chl cassette flanked by intron 2 sequences.

Plasmid containing rpsl/chl cassette was first digested by SpeI enzyme (New England Biolabs) for linearisation and purified by gel purification (QIAquick Gel Extraction Kit, Qiagen) according to the manufacturer's instruction. Int2_rpslchl_F and Int2_rpslchl_R were used to clone the rpsl/chl cassette flanked by intron 2 sequences neighbouring rs117124984, rs118087478 & rs77875796 using Ranger polymerase (Bioline). Reaction volumes: 5X buffer 5 µL, DNA template (SpeI digested rpsl/chl plasmid) 40 ng, 10 µM Int3_rpslchl_F 1 µL, 10 µM Int3_rpslchl_R 1 µL, polymerase 0.5 µL, water 16.5 µL. PCR condition: 95°C 1 min, (98°C 10 sec, 65°C 2 mins, 72°C 7 mins) x 35 cycles. PCR product from 6 reactions was pulled together and was purified by QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions.

Replacing rs117124984, rs118087478 & rs77875796 with rpsl/chl cassette.

pRed/ET plasmid encoding enzymes for homologous recombination was electroporated into electrocompetent DH10β (Invitrogen) cells (Section 2.1.6) containing either H1 or H2 pMAPT wildtype vectors to generate electrocompetent cells with induction of homologous recombination enzymes (Section 2.1.7). 6 µL PCR product was electroporated into DH10β cells carrying either pMAPT H1 or H2 wildtype vectors with homologous recombination enzymes induced. Bacterial colonies were screened based on streptomycin sensitivity and kanamycin and chloramphenicol resistance on LB agar plates.

2nd PCR for cloning H1 and H2 rs117124984, rs118087478 & rs77875796.

Ex3SNPs_Int_F and Ex3SNPs_Int_R were used to clone SNPs rs117124984, rs118087478 and rs77875796 using pMAPT H1 and H2 wildtype vectors as templates by KAPA Hifi polymerase (KAPA Biosystems). Reaction volumes: 5X buffer 5 µL, 10 mM dNTP mix 0.75 µL, KAPA HiFi DNA polymerase (1 U/µL) 0.5 µL, Ex3SNPs_Int_F (10 µM) 0.75 µL, Ex3SNPs_Int_R (10 µM) 0.75 µL, water 16.25 µL, 40 ng/µL DNA template (H1/H2 pMAPT maxi prep) 1 µL. PCR condition: 95°C 5 mins, (98°C 20 sec, 75°C 15 sec, 72°C 30 sec) x 30 cycles, 72°C 5 mins. PCR product was pooled form 6 reactions and was purified by QIAquick PCR purification kit (Qiagen), digested with DpnI enzyme (New England Biolabs) and purified again, according to the manufacturer's instructions.

Replacing rpsl/chl cassette with rs117124984, rs118087478 & rs77875796.

Recombination protocol was as above. Bacterial colonies were screened by colony PCR with primers Ex3_SNP_A and Ex3_SNP_B using AmpliTaq Gold DNA polymerase (Applied Biosystems). Reaction volumes: 10X buffer 2 µL, 25 mM $MgCl_2$ 2 µL, 10 mM dNTP mix 0.5 µL, water 8.5 µL, 1 µM forward & reverse primers 5 µL, AmpliTaq Gold 0.1 µL. PCR condition: 95°C 15 mins, (95°C 30 sec, 57°C 30 sec, 72°C 30 sec) X 35 cycles, 72°C 10 mins.

Confirmation of vectors.

Colonies with the correct band size from colony PCR were mini prepped screened with NotI and SalI restriction digests (New England Biolabs) and analysed by pulsed-field gel electrophoresis. Colonies carrying vectors with the correct band sizes from pulsed-field gel analysis were maxi prepped. The vectors were re-electroporated and maxi prepped again.

DNA sequencing

PCR for sequencing was performed using Ex3_SNP_A and Ex3_SNP_B using AmpliTaq Gold DNA polymerase (Applied Biosystems) and Pfu Turbo DNA polymerase (Agilent). Reaction volumes: 10X buffer 10 µL, 25 mM $MgCl_2$ 2 µL, 10 mM dNTP mix 0.5 µL, 1 µM forward & reverse priemrs 5 µL, AmpliTaq Gold 0.08 µL, Pfu Turbo 0.02 µL, water 8.5 µL, 20 – 100 ng DNA template 2 µL. PCR condition: 95°C 15 mins, (95°C 30 sec, 57°C 30 sec, 72°C 30 sec) X 35 cycles, 72°C 10 mins. PCR product was cleaned up using: FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher) 1 µL,

exonuclease I (New England Biolabs) 0.1 µL, water 0.9 µL. PCR clean up reaction was incubated at 37°C for 30 minutes, and then at 80°C for 20 minutes. Following PCR clean up, the PCR product was subjected to BigDye Terminator v3.1 (Applied Biosystems) sequencing reaction: BigDye 0.5 µL, 5X Sequencing Buffer 1.75 µL, 10 µM forward primer 1 µL, 10 µM reverse primer 1 µL, DNA template from PCR clean up 5 µL, water 1.75 µL. BigDye reaction condition: 96°C 1 min, (96°C 10 sec, 50°C 5 sec, 60°C 4 mins) X 30 cycles. Following BigDye sequencing reaction, 2 µL 125 mM EDTA, 2 µL 3 M sodium acetate, and then 50 µL 100% ethanol were added to each reaction. The mixtures were mixed by inversion of the container and were then centrifuged at 3000 x g at 4°C for 30 minutes. The supernatant was removed by centrifugation at 185 x g for 10 seconds. 70 µL 70% ethanol was then added to each reaction followed by centrifugation at 1650 x g at 4°C for 15 minutes. The supernatant was removed by centrifugation at 185 x g for 1 minute. The DNA was allowed to air dry for 5 minutes and was stored at -20°C.

Table 1. Primers for generating H1-*2-2-2* & H2-*1-1-1*

Primer name	5' – 3' sequence
Int2_rpslchl_F	CGGGCTTTCCTGGTGGATGAGCTGGGCGGTTCATGAGCCAGAACCACTCAGCAGC ATGTCAGTGTGCTTCCTGGGGAGGCCTGGTGATGATGGCGGGATC
Int2_rpslchl_R	CTTTGCTTCTTGCGCAGGAGCAATGCAGGGAGGAGGAGTGAGGAGGGGATCAGC CATATGCCAGAAAGCAGCCCTGAATTACGCCCCGCCCTGCCACTCA
Ex3SNPs_Int_F	CACTGCAGCGTTTACACAGG
Ex3SNPs_Int_R	ACCTTAGAGGGGCCAGGAC
Ex3_SNP_A	GCCTCACAGCAAAACCATAG
Ex3_SNP_B	GTTTTCTGGGGCAGGTGGTG

2.1.10 Exchanging H1/H2 SNP rs17651213 on pMAPT Vectors

Primer pairs and sequences in Table 2.

1st PCR for cloning rpsl/chl cassette flanked by intron 3 sequences.

Intron 3 + rpsl/chl F and Intron 3 + rpsl/chl R were used to clone the rpsl/chl cassette flanked by intron 3 sequences neighbouring rs17651213 using Expand Long PCR Mix (Roche). Reaction volumes: 10X buffer 2.5 µL, DNA template (H1 or H2 maxi preps) 40 ng, 10 µM Intron 3 + rpsl/chl 1 µL, 10 µM Int3_rpslchl_R 1 µL, polymerase 0.375 µL, 10 mM dNTPs 0.875 µL, water 15.75 µL. PCR condition: 92°C 2 mins, (92°C 10 sec, 70°C 2 mins, 72°C 4 mins) x 35 cycles, 72°C 6 mins. PCR product

from 6 reactions was pulled together and was purified by QIAquick PCR purification kit (Qiagen), followed by DpnI digests (New England Biolabs) and then PCR purified again according to the manufacturer's instructions.

Replacing rs17651213 with rpsl/chl cassette.

pRed/ET plasmid encoding enzymes for homologous recombination was electroporated into electrocompetent DH10β (Invitrogen) cells (Section 2.1.6) containing either H1 or H2 pMAPT wildtype vectors to generate electrocompetent cells with induction of homologous recombination enzymes (Section 2.1.7). 6 µL PCR product was electroporated into DH10β cells carrying either pMAPT H1 or H2 wildtype vectors with homologous recombination enzymes induced. Bacterial colonies were screened based on streptomycin sensitivity and kanamycin and chloramphenicol resistance on LB agar plates.

2nd PCR for cloning H1 and H2 rs17651213.

Rs17651213 F and rs17651213R were used to clone SNP rs17651213 using pMAPT H1 and H2 wildtype vectors as templates by AmpliTaq Gold polymerase (Applied Biosystems). Reaction volumes: 10X buffer 2 µL, 25 mM $MgCl_2$ 2 µL, 10 mM dNTP mix 0.5 µL, AmpliTaq Gold 0.1 µL, rs17651213 F (1 µM) 2.5 µL, rs17651213 R (1 µM) 2.5 µL, water 9.4 µL, 40 ng/µL DNA template (H1/H2 pMAPT maxi prep) 1 µL. PCR condition: 95°C 15 mins, (95°C 30 sec, 60°C 30 sec, 72°C 30 sec) x 25 cycles, 72°C 30 sec. PCR product was pooled from 6 reactions and was purified by QIAquick PCR purification kit (Qiagen), digested with DpnI enzyme (New England Biolabs) and purified again, according to the manufacturer's instructions.

Replacing rpsl/chl cassette with rs17651213.

Recombination protocol was as above. Bacterial colonies were screened by colony PCR with primers rs17651213 F and rs17651213 R using AmpliTaq Gold DNA polymerase (Applied Biosystems). Reaction volumes: identical as above. PCR condition: identical as above.

Confirmation of vectors.

As described as above (Section 2.1.9)

DNA sequencing

PCR for sequencing was performed using DNA sequencing F and DNA sequencing R using AmpliTaq Gold DNA polymerase (Applied Biosystems) and Pfu Turbo DNA polymerase (Agilent). Sequencing identical as above (Section 2.1.9)

Table 2. Primers for generating H1-*1-2* & H2-*2-1*

Primer Name	Primer Sequence 5' – 3'
Intron 3 + rpsl/chl F	AGCCCCAGAGACCCCCAGGCAGTCAAGGCCCTGCTGGGTGCCCCAGCTGA CCTGTGACAGAAGTGAGGGAGCTTTGGGCCTGGTGATGATGGCGGGATC
Intron 3 + rpsl/chl R	AAGGCATCGAGCTACTCACAACCCAAGATTCCCAGGAGCCAGAATCCACCC ATGTTCCTGCCCCACAGGAGGATAAACTTACGCCCCGCCCTGCCACTCA
rs17651213 F	CAGAGACCCCCAGGCAGT
rs17651213 R	CACCAAGGCATCGAGCTACT
DNA sequencing F	TGTGACAGCACCCTTAGTGG
DNA sequencing R	GAGGCAAGATGAGGAGCATC

2.1.11 Retrofitting pMAPT Vectors with pH-FRT-HY Plasmid

1 – 2 µg of pMAPT plasmids were incubated with 50 ng pH-FRT-HY plasmid with Cre recombinase (Novagen) for 1 hour at 37^{o}C, followed by 5 mins of recombinase inactivation at 70^{o}C. The reaction mixture was dialysed using 0.22 micron filter (Milipore) for 3 hours at room temperature. The reaction mixture was electroporated into DH10β cells. Plasmids were then isolated and analysed by restriction digests.

2.1.12 Other Primers used to modify pMAPT Vectors

Primers previously used to modify the pMAPT vectors by other lab members are in Table 3.

Table 3. Primers previously used to modify pMAPT vectors.

Primer name	Primer Sequence 5' – 3'
Primers for cloning pMAPT vectors	
pCYPAC2 25-mer R/ tau 5' homology	ttaagtgaaaatgtacagattgattattttcacctggtttctgttagattatcttAAAATCATTTAA TTGGTGGTGCTGC
pCYPAC2 25-mer L/ tau 3' homology	agatagaaaatatcatacagctgacttcactagagagaaagtgcatcaactgcttATTGACCCG GAACCCTTAATATAAC
Generating pMAPT with HA-tag sequence in exon 5	
Exon 5 + rpsl/chl F	TGAACAGTGAAAATGGAGTGTGACAAGCATTCTTATTTTATATTTTATCAGCT CGCATGGTCAGTAAAAGCAAAGACGGCCTGGTGATGATGGCGGGATC
Exon 5 + rpsl/chl R	CTCTGTAAACTTGACCAGCTGCAGAGCTCCGTGGCATCGTCAGCTTACCTTG GCTTTTTTGTCATCGCTTCCAGTCCCTTACGCCCCGCCCTGCCACTCA
Ex5 HAtag F	GTGAACAGTGAAAATGGAGTGTGACAAGCATTCTTATTTTATATTTTATCAGC TCGCATGGTCAGTAAAAGCAAAGACTACCCTTATGATGTTCCAGATT
Ex5 HAtag R	CTCTGTAAACTTGACCAGCTGCAGAGCTCCGTGGCATCGTCAGCTTACCTTG GCTTTTTTGTCATCGCTTCCAGTCCCTGCGTAATCTGGAACATCATAA
Colony PCR F	ACCCAGTGTGCTGTGTTGA
Colony PCR R	CTGCTCTTCCCTGGGGTAAG
Generating pMAPT rs1800547 hybrid vectors	
Ex3 + rpsl/chl F	AAGCAGGCTGCCGCGCAGCCCCACACGGAGATCCCAGAAGGAACCACAGGT GAGGGTGGCCTGGTGATGATGGCGGGATC
Ex3 + rpsl/chl R	GTCACAGGTCAGCTGGGGCACCCAGCAGGGCCTTGACTGCCTGGGGGTCTC TGGGGCTTTACGCCCCGCCCTGCCACTCA
rs1800547 F	CTTCAGGGCTGCTTTCTGGC
rs1800547 R	GCAAAGCTCCCTCACTTCTG
Colony PCR F	GCCTCACAGCAAAACCATAG
Colony PCR R	CACCACCTGCCCCAGAAAAC
Generating pMAPT rs1800547 and rs17651213 hybrid vectors	
H1 Intron 3 + rpsl/chl F	GGATGAGGGAGCTCCCGGCAAGCAGGCTGCCGCGCAGCCCCACACGGAGA TCCCAGAAGGAACCACAGGTGAGGGTGAGCCCCAGAGACCCCCAGGCAGT
H1 Intron 3 + rpsl/chl R	GAGGCAAGATGAGGAGCATCTAGAAGTCCAGAAGGCCCTAAATGCTCTGAG AGGCTGGCAGGCAGCCAGGGAGGTAACTGAGCACCAAGGCATCGAGCTA
H2 Intron 3 + rpsl/chl F	GGATGAGGGAGCTCCCGGCAAGCAGGCTGCCGCGCAGCCCCACACGGAGA TCCCAGAAGGAACCACAGGTGAGGGTAAGCCCCAGAGACCCCCAGGCAGT
H2 Intron 3 + rpsl/chl R	GAGGCAAGATGAGGAGCATCTAGAAGTCCAGAAGGCCCTAAATGCTCTGAG AGGCTGGCAGGCAGCCAGGGAGGTAACTGGGCACCAAGGCATCGAGCTA
rs1800547 and rs17651213 F	CTTCAGGGCTGCTTTCTGG
rs1800547 and rs17651213 R	GCACCAAGGCATCGAGCT
Colony PCR F	ATATGGCTGATCCCCTCCTC
Colony PCR R	GAAGGCCCTAAATGCTCTGA
DNA sequencing F	GCCTCACAGCAAAACCATAG
DNA sequencing R	CACCACCTGCCCCAGAAAAC

2.2 Chapter 4 Methods

All reagents were from ThermoFisher Scientific unless otherwise stated.

2.2.1 Primer Design

All primers were designed using the online tool Primer 3 web version 4.0.0 (http://primer3.ut.ee/). Genomic DNA sequences for primers design were obtained from the UCSC genome browser (https://genome.ucsc.edu/cgi-bin/hgGateway). All primers were purchased from either Eurofins Genomics or Integrated DNA Technologies.

2.2.2 Cell Culture

BE(2)-M17 was cultured in DMEM (Sigma Aldrich), 10% foetal bovine serum, 50 U/mL penicillin and 50 ug/mL streptomycin, 2 mM L-glutamine and 1X minimum essential medium non-essential amino acids. SK-N-MC was cultured in EMEM (Lonza), 10% foetal bovine serum, 50 U/mL penicillin and 50 ug/mL streptomycin, 2 mM L-glutamine, 1X minimum essential medium non-essential amino acid and 1 mM sodium pyruvate. SK-N-F1 was cultured in DMEM, 10% foetal bovine serum albumin, 50 U/mL penicillin and 50 ug/mL streptomycin, 4 mM L-glutamine and 1X minimum essential medium non-essential amino acid. SH-SY5Y was cultured in DMEM/F12, 10% foetal bovine serum albumin, 50 U/mL penicillin and 50 ug/mL streptomycin and 2 mM L-glutamine. Vero 2-2 and G16.9 cells were cultured as described (Lufino et al. 2011). All cells were incubated at 37^{o}C and 5% CO_2.

2.2.3 Transfections of SK-N-F1 Neuroblastoma Cells

Except during optimisation steps where concentrations of different reagents have been varied (detailed in the results section), all transfections were performed as follow. 7 – 10 x 10^5 cells were seeded per 6-well well. Transfection was carried out when cell confluence reached ~75 %. 6.25 µg of each pMAPT vector DNA was incubated for 10 minutes at room temperature with 6.25 µL of PLUS™ reagent in Opti-MEM® reduced serum medium (Life Technologies). 15.6 µL of Lipofetamine® LTX reagent diluted in Opti-MEM® medium, was then incubated with the DNA mixture for a further 30 minutes, forming the transfection mix. Cells were washed with Opti-MEM® medium and were then incubated with the transfection mix for 4 hours at 37^{o}C, 5 % CO_2. Cells were

washed with OptiMEM® medium and were then further incubated in culturing medium for 48 hours before proceeding to RNA extraction.

2.2.4 RNA Extraction and cDNA Synthesis

SK-N-F1 cells were harvested in TRIzol® reagent (1 mL per 6-well well, 750 mL per 12-well well). Chloroform (Fisher) (200 µL per 1 mL Trizol) was added to TRIzol containing cell lysates and was vigorously shaken for 15 seconds to mix thoroughly. The organic and aqueous phases were then separated by centrifugation at 4°C for 10 mins at 10000 rpm. The top aqueous was transferred to a new microcentrifuge tube and total RNA was precipitated by the addition of 70% ethanol at an equal volume with the aqueous phase. RNA purification was then carried out using the RNeasy Mini Kit (Qiagen). The precipitated RNA was run through the RNeasy columns at 10000 rpm for 1 minute. The columns were washed using 350 µL RW1 buffer and centrifuged at 13000 rpm for 1 minute. The columns were then treated with RNase-Free DNase I (Qiagen) for 15 minutes. 350 µL RW1 buffer was then added the columns and were centrifuged at 10000 rpm for 1 minute. The columns were then washed twice with 500 µL RPE buffer at 10000 rpm for 2 minutes. The columns were further spun dry for 2 minutes to remove remaining buffer. 30 µL RNase free water was added to the columns and RNA was eluted in new microcentrifuge tubes. RNA concentration was determined using the ND-1000 nanodrop spectrophotometer.

cDNA was generated using either SuperScript® III Reverse Transcriptase or SuperScript® VILO Master Mix using 1 – 5 µg RNA, according to the manufacturer's protocol.

2.2.5 qPCR for Total and Exon 3 Expression by pMAPT vectors

Except during optimisation steps where concentrations of different reagents have been varied as detailed in the results section, all qPCR reactions were performed as follow. Reaction volumes: 2X SYBR Green PCR Master 12.5 µL, 1 µM forward primer 1.25 µL, 1 µM reverse primer 1.25 µL, cDNA (1:10 dilution) 10 µL. qPCR reactions were performed using the StepOnePlus System using the standard ~2 hours cycling protocol for SYBR Green reagents with the inclusion of melt curve analysis. Serial dilution cDNA standard curves using transgenic *MAPT* exon 3 primers transcripts

were generated with a starting cDNA volume of 10 µL 1:2.5 cDNA per well followed by 3-fold dilutions at 5 dilution points. Serial dilution cDNA standard curves using transgenic total *MAPT* primers were generated with a starting cDNA volume of 10 µL 1:3.3 cDNA per well followed by 5-fold dilutions at 5 dilution points.

Table 4. Primers for measuring total and exon 3 expression from pMAPT vectors

Primer Name	Primer Sequence 5' – 3'
Transgenic Total MAPT F	GAAGACGAAGCTGCTGGTCA
Transgenic Total MAPT R	GCTTCCAGTCCCTGCGTAAT
Transgenic MAPT Exon 3 F	CAGAAGGAACCACAGCTGAAGA
Transgenic MAPT Exon 3 R	GGAACATCATAAGGGTAGTCTTTGC

2.2.6 qPCR analysis for transgenic *MAPT* expression

A relative standard curve method was used where Ct values were plotted against log DNA concentrations. The equations of the lines of best fit (y = mx + c) were used to calculate the relative amounts of total and exon 3 transcripts present using Microsoft Excel. *MAPT* exon 3 expression was expressed as the amount of exon 3/total *MAPT* expression.

2.2.7 HSV-1 Packaging Protocol

Packaging of pMAPT vectors was as described (Lufino et al. 2011). Titration and X-gal staining was performed as described (Lufino et al. 2011).

Transduction was performed as follow: 3 x 10^5 cells were seeded the day before in 12-well wells. Amplicons in DMEM (10 % FBS, 1 % pen/strep, 4 mM L-glutamine) were applied onto cells. The cells were spun down at 1500 xg for 20 minutes, then at 700 xg for 45 minutes. Media was replaced after 6 hours and cells were allowed to express the constructs.

2.3 Chapter 5 Methods

All chemicals were from Sigma Aldrich unless otherwise stated. All RNA oligonucleotides were purchased from Integrated DNA Technologies. All reagents for RNA EMSA, RNA 3' Biotinylation and RNA affinity pull down experiments were purchased from Pierce.

2.3.1 Nuclear Lysate Extraction

SK-N-F1 nuclear lysate was extracted using a protocol by Prof J. Knight (University of Oxford). All reagents and cell pellets were kept on ice during the extraction process, and all centrifugation steps were performed at 4^{o}C unless otherwise stated. SK-N-F1 cells were grown in 15 cm dishes for 48 hours and were washed twice with ice cold PBS before harvested by gentle scraping. Cells were pelleted by centrifugation at 1500 rpm for 3 minutes and the supernatant was discarded. Cell pellets were resuspended in lysis buffer A [10 mM HEPES, 10 mM KCl (Ambion), 0.1 mM EDTA, 0.1 mM EGTA, Halt protease and phosphatase inhibitors cocktail (ThermoFisher) and 0.67 % Igepal and was left to swell for 15 minutes. 67 µL 10% Igepal was added and was mixed thoroughly by pipetting up and down for 1 minute. The mixture was further left to swell for 5 minutes. The mixture was then centrifuged for 1 minute at 13000 x g. The supernatant contains the cytoplasmic layer and was aliquoted and snap frozen using dry ice. The remaining nuclear pellet was washed with 200 µL lysis buffer A and was centrifuged for 1 minute at 13000 x g. The supernatant was discarded and 500 uL lysis buffer B (20 mM HEPES, 400 mM KCl, 1 mM EDTA, 1mM EGTA and Halt protease and phosphatase inhibitors cocktail) was added to the nuclear pellet. The mixture was vortexed slowly at 4^{o}C in the cold room for 60 minutes and was centrifuged at 13000 x g for 5 minutes. The supernatant contains the nuclear enriched lysate and was aliquoted and snap frozen using dry ice.

2.3.2 RNA Biotinylation

RNA biotinylation was performed using Pierce RNA 3' End Biotinylation Kit. RNA oligonucleotide probe sequences: H1 (5' CCA CAG GUG AGG GU**A** AGC CCC AGA GAC CCC 3') and H2 (5' CCA CAG GUG AGG GU**G** AGC CCC AGA GAC CCC 3'). Reaction condition: 10X RNA ligase buffer 3 µL, 40 U/µL RNase inhibitor 1 µL, 1.67 µM (50 pmol) unlabelled RNA oligonucleotides, 1 – 4 nmol biotinylated

cytidine bisphosphate, 20 U/µL T4 RNA ligase, 30% polyethylene glycol 15 µL and RNase free water to make up to 30 µL reaction volume. Reactions were incubated at 16^{o}C overnight. To isolate RNA oligonucleotide, 70 µL RNase free water and 100 µL chloroform:isoamyl alcohol were added to the reaction mixture and were briefly vortexed before centrifuged at high speed in a bench-top microcentrifuged to separate phases. The top aqueous layer containing RNA was transferred to a new microcentrifuge tube. 10 µL 5 M NaCl, 1 µL glycogen, 300 µL 100% ethanol were added to the aqueous layer and RNA was allowed to precipitate for over 1 hour at -20^{o}C. The mixture was centrifuged at >13000 x g at 4^{o}C. The supernatant was carefully discarded. The RNA pellet was washed with 300 µL ice cold 70% ethanol. The ethanol was carefully removed and the RNA pellet was allowed to air dry for ~5 minutes. RNA was resuspended in 20 µL RNase free water.

Biotinylation efficiency was determined according to the manufacturer's protocol using the standard provided in the kit.

2.3.3 RNA Electrophoretic Mobility Shift Assay (RNA EMSA)

RNA EMSA was perform using LightShift Chemiluminescent RNA EMSA kit that includes all the RNA-protein binding reaction buffers and components except RNA oligonucleotides and protein extract. RNA oligonucleotide containing rs17651213 SNP sequence: H1 (5' UGA GGG AGC UUU GC**G** UGU UUA UCC UCC UGU 3') and H2 (5' UGA GGG AGC UUU GC**A** UGU UUA UCC UCC UGU 3'). RNA probes contain the same sequence with a 15 atom triethyleneglycol linker biotin at the 3' end of the oligonucleotide (Integrated DNA Technologies). RNA oligonucleotide containing rs1800547 SNP sequence: H1 (5' CCA CAG GUG AGG GU**A** AGC CCC AGA GAC CCC 3') and H2 (5' CCA CAG GUG AGG GU**G** AGC CCC AGA GAC CCC 3'). RNA probes contain the same sequence with a biotinylated cytidine bisphosphate attached to the 3' end of the oligonucleotide.

Reaction volumes:

Except during optimisation steps where concentrations of different reagents have been varied as detailed in the results section, all RNA EMSA reactions were performed as follows. All reagents were added in the order as stated below.

For rs1800547 probe: RNase free water to make up 20 µL reaction volumes, 1X RNA EMSA buffer, 0.25 µg/µL tRNA (H1 probe) or 0.1 µg/µL tRNA (H2 probe), 0.25 µg/µL SK-N-F1 nuclear extract, 15 nM biotinylated H1 or 4 nM H2 RNA probe, unlabelled RNA oligonucleotide (varying as detailed in results section). For rs17651213 probe: RNase free water to make up 20 µL reaction volumes, 1X RNA EMSA buffer, 0.1 µg/µL tRNA, 4 µg SK-N-F1 nuclear extract, unlabelled RNA oligonucleotide (varying as detailed in results section) and 1 nM biotinylated RNA probe. The reaction mixtures were incubated at room temperature for 20 – 30 minutes before proceeding to electrophoresis.

RNA EMSA electrophoresis and transfer:

5 µL 5X loading buffer was added to each reaction mixture. 20 µL of each reaction mixture was loaded onto a 6% polyacrylamide gel made in 0.5X TBE buffer (pre-electrophoresed at 100 V for 30 – 60 minutes) and was electrophoresed at 100 V for 50 minutes in 0.5X TBE buffer. The RNA was then transferred onto positively charged nylon membrane (Sigma) using the Trans-Blot Turbo Transfer System (BioRad), together with the membrane and whatman filter paper pre-soaked in 0.5X TBE buffer for at least 10 minutes. Transfer was performed at 0.2 A for 18 minutes. The membrane was then allowed to absorb excess buffer (but not dried) and RNA was UV-crosslinked at 120 mJ/cm^2 with 254 nm bulbs for 60 seconds using the UV Stratalinker 1800 (Stratagene). The membrane was then washed and blocked using buffers provided in the LightShift Chemiluminescent RNA EMSA kit and the RNA bands were visualised by chemiluminescent according the manufacturer's protocol.

2.3.4 RNA-Affinity Pull-Down

RNA pull-down of RNA binding proteins was performed using the magnetic RNA-protein pull-down kit (Pierce) according to the manufacturer's instruction. Briefly, streptavidin coated magnetic beads were washed in wash buffers. 50 pmol RNA oligonucleotide probes labelled with 3' biotin-tetra-ethyleneglycol spanning rs17651213 either H1-G (5'TGA GGG AGC TTT GCG TGT TTA TCC TCC TGT3') or H2-A (5'TGA GGG AGC TTT GCA TGT TTA TCC TCC TGT3') (Integrated DNA Technology) were allowed to bind to the streptavidin coated magnetic beads and were then allowed to incubate with

40 ug SK-N-F1 nuclear enriched lysates. Unbound proteins were washed off. The magnetic beads together with the RNA bait and bound proteins were snap frozen and stored at -20^{o}C until further analysis.

2.3.5 Mass Spectrometry

Samples were analysed on an Ultimate 3000 RSLCnano HPLC (Dionex, Camberley, UK) system run in direct injection mode coupled to a QExactive Orbitrap mass spectrometer (Thermo Electron, Hemel Hempstead, UK). Samples were resolved on a 25cm by 75 micron inner diameter picotip analytical column (New Objective, Woburn, MA, USA) which was packed in-house with ProntoSIL 120-3 C18 Ace-EPS phase, 1.9um bead (Bischoff Chromatography, Germany). The system was operated at a flow-rate of 250nL min^{-1}. A 120 min gradient was used to separate the peptides. The mass spectrometer was operated in a "Top 20" data dependent acquisition mode. Precursor scans were performed in the orbitrap at a resolving power of 70,000, from which the twenty most intense precursor ions were selected by the quadrupole and fragmented by HCD at a normalised collision energy of 30%. The quadrupole isolation window was set at *1.6 m/z.* Charge state +1 ions and undetermined charge state ions were rejected from selection for fragmentation. Dynamic exclusion was enabled for 27s. Data were converted from .RAW to .MGF using ProteoWizard (M. C. Chambers et al. 2012).

2.3.6 Mass Spectrometry Data Analysis

Mass spectrometry results in .MGF format were imported and processed with Progenesis QI for Proteomics software (Nonlinear Dynamics A Waters Company). The H1 probe sample data were selected as the alignment reference for all the replicates in the alignments of ion density maps. Automatic alignment was first performed followed by manual reviewing of the alignments. Peptide ions were filtered according to charge. Peptide ions with a charge between 2 and 4 were included, the rest were excluded. Peptide abundance normalisation was performed automatically by the software. A between-subject design was set for analysed runs. Analysed data were exported as .MGF files. Protein identification was performed using MS/MS ions search on the Computation

Biology Research Group (Oxford University) Mascot Server (Matrix Science). Peptide ions were searched against the UPR_HomoSapiens database with fixed modications set for carbamidomethyl at cysteine residues and variable modifications set for oxidation at methionine residues. Completed protein identifications were imported back into Progenesis QI for matching to their corresponding detected peptide ions. Proteins were quantified by Progenesis QI where a relative quantitation method using non-conflicting peptides was chosen for all the results. The final data were exported as .CSV files. To identify splicing factors interacting with the RNA probes, the exported data in .CSV from Progenesis QI were sorted using Microsoft Excel. Proteins identified with a spectral count and unique peptides of below two were excluded from the data. Proteins that were not present in all three experimental replicates further excluded from the data. The remaining candidate proteins were matched to a list of 71 experimentally validated splicing factors obtained from the SpliceAid-F database (http://srv00.ibbe.cnr.it/SpliceAidF/help.php) (Giulietti et al. 2013). The abundance of each splicing factor interacting with the H1 and H2 RNA probes were compared.

2.3.7 Western Blot

10 µg of proteins from nuclear lysate extraction (Section 2.3.1) were denatured in sample buffer (6X: 12% SDS, 30% β-mercaptoethanol, 60% glycerol, 0.012% bromophenol blue, 375 mM Tris pH 6.8) at 95^{o}C for 10 minutes. Following denaturation, proteins were loaded onto a 12% SDS-PAGE gel with a 4% SDS-PAGE stacking gel. Electrophoresis was performed in running buffer (10X: 144 g glycine, 30.3 g Tris-base, 10 g SDS, water to make up to 1 L) at 80 V for 5 – 10 minutes followed by 200 V for ~45 minutes. Proteins were transferred using the Trans-Blot Turbo Transfer System (BioRad) onto polyvinylidene fluoride membrane contained in the Trans-Blot Turbo PVDF Transfer Packs (BioRad) at 2.5 A for 7 minutes. The membrane was then washed twice in TBS-Tween [100 mL 10X TBS (200 mL 1 M Tris pH 7.5, 80 g NaCl, water to make up to 1 L), 1 mL Tween-20 and 900 mL water] for a few minutes. The membrane was then blocked in 5% milk in TBS-Tween for 1 hour at room temperature with gentle rocking. Following blocking, the membrane was incubated in antibody diluted in 1% TBS-Tween overnight at 4^{o}C with gentle rocking. The membrane was then

washed twice with TBS-Tween for 10 minutes, and was incubated with 1:5000 HRP conjugated secondary antibody (BioRad) in 1% TBS-Tween at room temperature for 1 hour with gentle rocking. The membrane was then washed three times with TBS-Tween. Protein bands were then visualised by chemiluminescent.

Anti-GAPDH mouse monoclonal antibody, 1:1000 for Western blot, Abcam ab9484;

Anti-Histone H3 (D1H2) rabbit monoclonal antibody, 1:2000 for Western blot, Cell Signalling 4499.

2.4 Chapter 6 Methods

All qPCR reagents were from Life Technologies unless otherwise stated. All chemicals were from Sigma Aldrich. All primers were from Integrated DNA Technologies. Cell culture methods (Section 2.2.2), RNA extraction and cDNA synthesis (Section 2.2.4) methods were as described above.

2.4.1 Allele-Specific Assay Probes and Primers Design

Primers, H1 (FAM labelled) and H2 (VIC labelled) specific Taqman probes spanning rs17650901 in exon 1 were designed using both Primer Express 3.0 (Applied Biosystems) and Primer 3 (Koressaar and Remm 2007; Untergasser et al. 2012). Probes designed satisfied all the following guidelines from Applied Biosystems: probe length 13 – 25 bp, melting temperature 65 – 67, GC content (%) 30 – 80, 5' end cannot be a G, [SNP site] should locate in [NNNNN]NN3', avoid 4 or more consecutive Gs, avoid GGG3' and GGAG3', avoid 6 or more consecutive As, avoid CC or more in the middle, 2nd G at 5' end should be avoided for FAM probe.

2.4.2 Allele-Specific qPCR Assay

Primers and probes sequences in Table 5.

Except during optimisation steps where concentrations of different reagents have been varied as detailed in the results section, all allele-specific Taqman qPCR reactions were performed as follow. qPCR was performed using Taqman Gene Expression Master Mix. Reaction condition: 2X Taqman Gene Expression Master Mix 10 µL, 10 µM forward primer 1.8 µL, 10 µM reverse primer 1.8 µL, 40X custom Taqman probes (H1+H2) 2.5 µL, cDNA template 1 – 3 µL, and water to make up to 20 µL. qPCR reactions were performed using the StepOnePlus System using the standard cycling protocol for Taqman reagents. All reactions were performed in triplicate.

Standard curve was generated using genomic primers where the delta Ct values between VIC (H1) and FAM (H2) signals were plotted against the $\log_2$ ratios of H1:H2 transcripts by mixing 50 pg of H1 and H2 pMAPT vectors in the ratios 8:1, 4:1, 2:1, 1:1, 1:2, 1:4 and 1:8.

The equation of regression line obtained from the standard curve $\log_2$ (H1/H2) = -1.194 × delta Ct was used calculate the ratios of H1:H2 transcripts.

Table 5. Allele-specific qPCR assay primers and probes.

Primers/probes Name	Sequence 5' – 3'
Exon 3 F	CCTCGCCTCTGTCGACTATC
Exon 3 R	GTCACATCTTCCGCTGTTGG
Genomic F	TACGTCCCAGCGTGATCTTC
Genomic R	CCCCAACACTCCTCAGAACT
Total MAPT F	TACGTCCCAGCGTGATCTTC
Total MAPT R	CCCCAACACTCCTCAGAACT
H1 probe (FAM-NFQ MGB labelled)	CTGGTTCAAAGTTC
H2 probe (VIC-NFQ MGB labelled)	TGGTTCAAAGCTCAC

2.4.3 qPCR for Measuring Endogenous Gene Expression

Primer sequences in Table 6.

Primers were designed using primer 3 except for hnRNP M where primer sequence was obtained from PrimerBank (https://pga.mgh.harvard.edu/primerbank/). Gene sequences were otabined from UCSC genome browser (https://genome.ucsc.edu/cgi-bin/hgGateway). Except during optimisation steps where concentrations of different reagents have been varied as detailed in the results section, all qPCR reactions were performed as follows. Reaction volumes: 2X Fast SYBR Green PCR Master 10 µL, 4 µM forward primer 1 µL, 4 µM reverse primer 1 µL, water 3 µL, cDNA (1:10 dilution) 5 µL. For primers measuring total MAPT expression, 0.5 µL 4 µM forward and reverse primers were used and water was adjusted accordingly. qPCR reactions were performed using the StepOnePlus System using the fast cycling protocol for SYBR Green reagents with the inclusion of melt curve analysis. All qPCR reactions were performed in triplicate. Serial dilution cDNA standard curves were generated with a starting cDNA volume of 5 µL 1:3.3 cDNA per well followed by 5-fold dilutions at 5 dilution points.

Table 6. Primers for expression of endogenous genes.

Primer Name	Primer Sequence 5' – 3'
MAPT Exon 3 F	CAGCCCCACACGGAGATC
MAPT Exon 3 R	GTGTCTCCAATGCCTGCTTC
Total MAPT F	CTGAGCCCCGCCAGGAGTTC
Total MAPT R	CCCCCTGATCTTTCCTGTCCCCC
GAPDH F	GGTCTCCTCTGACTTCAACA
GAPDH R	AGCCAAATTCGTTGTCATAC
HPRT F	GCCAGACTTTGTTGGATTTG
HPRT R	CTCTCATCTTAGGCTTTGTATTTTG
ACTB F	AGCGCGGCTACAGCTTCA
ACTB R	CGTAGCACAGCTTCTCCTTAATGTC
hnRNP F F	ACTGCCAGGAGGTACATTGG
hnRNP F R	CTGAGGTCTCTCCCGAACAG
hnRNP Q F	TGGGAAACTGGAACGAGTGA
hnRNP Q R	GATCTGGTGGCTTGGCAAAA
PTBP2 F	CTCGGTTCTTGTGAGCGAAG
PTBP2 R	GCAACCTCAGTGACGATTCC
hnRNP A2B1 F	CTGAGGTTGATGCTGCCATG
hnRNP A2B1 R	CTACAGCACGTTTTGGCTCA
SFPQ F	CGGACTCGGAGGGGTTTAAA
SFPQ R	GTTTTCTCTCCAGGCCTCCT
TARDBP F	GGCTGGGGAAATCTGGTGTA
TARDBP R	TCGGATGTTTTCTGGACTGC
hnRNP M F	CTCTTAATGGACGCTGAAGGAAA
hnRNP M R	CGCTCAGACTATGCTTGTTTAGG
ELAVL1 F	TCGGGATAAAGTAGCAGGACA
ELAVL1 R	GATCGCTCTCTCTGCATCCT
hnRNP A0 F	CCTCAATGTGCAGACGAGTG
hnRNP A0 R	AGTCCGTCAGAGTCCCAAAG

2.4.4 qPCR Data Analysis

For measuring splice factor expression levels and total *MAPT* expression, the $2^{-(\text{delta delta Ct})}$ method (K. J. Livak and Schmittgen 2001) was used where gene expression is internally normalised against the geometric means of three endogenous control genes *GAPDH*, *HPRT1* and *ACTB* and externally normalised to the corresponding untransfected controls (Vandesompele et al. 2002). For measuring *MAPT* exon 3 expression, the $2^{-(\text{delta delta Ct})}$ method was used as described above except gene expression is internally normalised against total *MAPT* expression.

2.4.5 siRNA Knockdown

All siRNAs were purchased from Ambion (Table 7). Except during optimisation steps where different transfection reagents and different reagent concentrations have been used as detailed in the results section, all siRNA knockdown experiments were performed as follow. Cells were seeded in 12-well well the day before transfection at 2 x 10^5 (SK-N-F1) and 1 x 10^5 (SK-N-MC) densities. To transfect, 50 µM siRNA was diluted to 10 µM in 50 µL Opti-MEM. 3 µL Lipofectamine RNAiMAX was diluted separately in 50 µL Opti-MEM. Diluted Lipofectamine RNAiMAX was added to the diluted siRNA drop by drop and was incubated at room temperature undisturbed for 5 – 10 minutes to allow for siRNA-lipid complex formation. Meanwhile, cells were washed once to twice using their corresponding media. The siRNA-lipid complex was then added directly onto cells in their corresponding media. Media was topped up 48 hours after transfection to prevent wells from drying, and RNA was extracted from the cells 96 hours post transfection.

Table 7. Ambion silencer select pre-designed siRNA IDs.

Gene Name	Ambion Silencer Select Pre-Designed siRNA ID
Negative Control #1	N/A
TARDBP	S23831
PTBP2	S33791
SFPQ	S224606
hnRNP A2B1	S6715
SYNCRIP (hnRNP Q)	S20562
hnRNP M	S9259
hnRNP A0	S21544
ELAVL1	S4610
hnRNP F	S6725

2.4.6 Camptothecin Treatments

Camptothecin (Sigma) and was reconstituted in DMSO to 20 mM concentration. 1 x 10^6 cells were seeded in 6-well well the day before camptothecin treatment. Camptothecin was directly added to cells at different final concentrations as detailed in the results section, and was incubated with cells for 24 hours before RNA extraction.

2.4.7 RNA EMSA Supershift Assays

RNA EMSA supershift assays were perform using LightShift Chemiluminescent RNA EMSA kit. RNA oligonucleotide probes used were identical as described above (Section 2.3.3)

Reaction volumes:

Except during optimisation steps where concentrations of different reagents have been varied as detailed in the results section, all RNA EMSA supershift reactions were performed as follows. All reagents were added in the order as stated below.

For rs1800547 probe: RNase free water to make up 20 µL reaction volumes, 1X RNA EMSA buffer, 0.1 µg/µL tRNA, 4 µg SK-N-F1 nuclear extract, 200 U Protector RNase Inhibitor (Roche), 5 µg antibody, biotinylated H1 or 4 nM H2 RNA probe. For rs17651213 probe: RNase free water to make up 20 µL reaction volumes, 1X RNA EMSA buffer, 0.1 µg/µL tRNA, 4 µg SK-N-F1 nuclear extract, 200 U Protector RNase Inhibitor (Roche), 5 µg antibody, 1 nM biotinylated RNA probe.

Antibody was allowed to incubate with the nuclear lysate in the RNA EMSA binding condition for 30 minutes at 25^{o}C before biotinylated probes were added. Biotinylated probes were then added and the reaction mixtures were further incubated at 25^{o}C for 15 minutes before electrophoresis.

Anti-hnRNP F antibodies: Abcam ab6095 and ImmuQuest IQ208

Anti-hnRNP Q antibody: Abcam ab10687

Anti-GAPDH antibody: Abcam ab9484

RNA EMSA electrophoresis and transfer:

Same as that described above (Section 2.3.3).

CHAPTER 3

MAPT H1/H2 Haplotype Genomic Expression PAC Vectors

3.1 Introduction

3.1.1 Genomic Expression PAC Vectors

Bacterial and P1-derived artificial chromosomes (BAC and PAC) are plasmid backbones originally designed as cloning systems for constructing high-resolution physical map of the human genome where large fragments of DNA of over 100 kb can be incorporated into BAC/PAC and cloned in *Escherichia coli* (*E. coli*) (Ioannou et al. 1994; Shizuya et al. 1992). BAC and PAC vectors encode regulatory genes to maintain the vectors at one to two copies per *E. coli*, allowing the vectors high transgene capacity (Ioannou et al. 1994; Shizuya et al. 1992). The cloning of the human genome fragments using BAC/PAC vectors and subsequent DNA sequencing have led to the completion of generating the first complete human genome sequence in the human genome project (International Human Genome Sequencing Consortium 2004).

With increasing understanding of genetic elements as sequence data were being produced from the human genome project, BAC/PAC vectors have evolved into useful tools for genetic studies. The high transgene capacity of BAC/PAC vectors has been used for developing gene delivery systems where the BAC/PAC vectors carry the transgene and viral coding sequences, forming amplicons, which are maintained in the cell population as an episome but can be packaged into infectious virions when viral proteins are present (Lufino et al. 2008; Macnab and Whitehouse 2009). Several viral-based elements have been exploited to incorporate BAC/PAC vectors, including Epstein-Barr virus (J. Black and Vos 2002; Coren and Sternberg 2001; Wade-Martins et al. 1999), herpes simplex virus type 1 (Lufino et al. 2007; Saeki et al. 2001; Wade-Martins et al. 2001) and cytomegalovirus (Borst and Messerle 2003). Numerous studies in both cultured cell lines and

animal models have gradually emerged, illustrating the functional advantages of transgene expression from BAC/PAC vectors carrying whole genomic loci, including correct promoter usage and alternative splicing, physiologically relevant gene and protein expression, maintenance of long-term expression, correct tissue specificity and developmental expression pattern (Alon et al. 2009; Antoch et al. 1997; Callow et al. 1994; Cavelier et al. 2001; Chandler et al. 2007; Deal et al. 2006; Hibbitt et al. 2007; Hibbitt et al. 2010; R. Inoue et al. 2004; van den Boogaard et al. 2014; Wade-Martins et al. 2000; Wade-Martins et al. 2003; W. Xing et al. 2004; L. Xing et al. 2013).

3.1.2 *Cis*-Acting Sequence Elements in Alternative Splicing

Genetic variation has long been recognised to have allele-specific effects on gene expression and splicing processes (Emilsson et al. 2008; Hull et al. 2007; Nembaware et al. 2004; Pickrell et al. 2010; Stranger et al. 2007). In order to understand and identify the sequence variants that may contribute to the haplotype-specific splicing of exon 3-containing transcripts previously observed in post-mortem brain tissue and in a cultured neuroblastoma cell line (Caffrey et al. 2008), we employed in this study PAC vectors carrying the H1 and H2 *MAPT* gene to capture genetic polymorphisms present in the *MAPT* haplotypes and to ensure their physiological relevance in gene expression and splicing. We performed a detailed sequence analysis of the H1 and H2 *MAPT* gene sequences in our PAC vectors to determine the percentage of sequence variants captured by the vectors.

The alternative splicing of pre-mRNA transcripts is an extremely complex process. It is highly context specific and is tightly controlled by RNA-binding proteins (X. D. Fu and Ares 2014; Matlin et al. 2005; Z. Wang and Burge 2008). The outcome of inclusion or exclusion of an exon depends on whether the spliceosome can successfully recognise and pair up the correct 5' and 3' splice sites that are often thousands of nucleotides apart due to the long intron lengths (International Human Genome Sequencing Consortium 2004; L. P. Lim and Burge 2001). However, 5' and 3' splice site sequences are poorly conserved and generally contain insufficient information for defining exon/intron boundaries (Berget 1995; L. P. Lim and Burge 2001; Z. Wang and Burge 2008). Splice site recognition and pairing is therefore strongly regulated by the multiple exonic and intronic cis-

regulatory enhancer and silencer elements (Berget 1995; D. L. Black 1995, 2003; C. W. Smith and Valcarcel 2000). It is also important to note that both proximal and distant intronic regulatory elements have both been shown to be important contributors to splicing outcomes (Lovci et al. 2013; X. H. Zhang et al. 2005).

3.1.3 Aims of Chapter 3

In order to understand how SNPs may alter splicing regulatory elements, several H1/H2 candidate SNPs were analysed and compared for trans-acting splicing factor binding motifs. For functional studies of H1/H2 *MAPT* polymorphisms, we generated haplotype-hybrid H1/H2 *MAPT* vectors where several candidate SNP sequences were exchanged between the H1/H2 haplotype backgrounds. The hybrid vectors were generated using an advanced BAC/PAC modification strategy by homologous recombination (J. Wang et al. 2006; Y. Zhang et al. 1998; Y. Zhang et al. 2000). We aim here to strategically select candidate SNPs that may affect *MAPT* exon 3 inclusion by performing computational sequence analyses and to generate PAC *MAPT* haplotype-hybrid expression vectors for functional studies of the candidate SNPs in a cell culture model.

3.2 Results

3.2.1 Comparative Sequence Analysis of the Genomic *MAPT* H1/H2 Haplotype PAC Vectors

To investigate the haplotype-specific polymorphisms that may yield a difference in the expression of exon 3-containing *MAPT* transcripts, two PAC vectors containing the whole *MAPT* gene (143 kb) of either the H1 (pMAPT-H1) or H2 (pMAPT-H2) haplotypes (Figure 5A) were employed in our study. Five BACs and PACs were previously identified from the BACPAC resources to contain the entire *MAPT* gene. The five BAC/PAC vectors were previously genotyped at six haplotype tagging SNPs and at a 238 bp deletion in intron 9 that unambiguously define the H1/H2 haplotypes (Table 8) (performed by Tara M. Caffrey) (Baker et al. 1999; Pittman et al. 2005). Four vectors were identified as the H1 haplotype and one as H2 (Table 8). The pMAPT-H1 vector used in this study originates from PAC61D06 (Table 8) which has previously cloned into a BAC backbone (Peruzzi et al. 2009). The pMAPT-H2 vector originates from BAC769P22 (BACPAC resources ID RP11-769P22; accession BX544879) (Table 8). In order to engineer the *MAPT* gene such that the H1 and H2 *MAPT* genes start and end at the same sequence, the H1 and H2 vectors in their BAC backbones were cloned into PAC plasmids (by Richard Wade-Martins and Franziska Denk) (Table 3). Primer pairs containing homology sequences flanking the *MAPT* gene at its 5' and 3' untranslated regions, together with sequences of the PAC plasmid (Table 3), were previously used to generate the pMAPT-H1 and –H2 vectors.

Table 8. Genotyping five BAC/PAC vectors spanning the whole *MAPT* locus identified four H1 and one H2 alleles.

BAC/PAC ID	Genotype	ΔInt9	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
PAC24i13	H1	-	G	A	G	T	G	T
BAC669E14	H1	-	G	G	A	C	A	T
BAC769P22	H2	+	A	G	G	C	G	C
BAC2524N8	H1	-	G	A	G	C	G	T
PAC61D06	H1	-	G	A	G	T	G	T

Five BAC/PAC vector spanning the whole *MAPT* locus were genotyped at six haplotype-tagging SNPs and the intron 9 238 bp deletion (ΔInt9) that segregates the H1/H2 haplotypes (Baker et al. 1999; Pittman et al. 2005). Genotyping identified 4 H1 and 1 H2 *MAPT* genes contained in the BAC/PAC vectors. (SNP1: rs14676967; SNP2: rs242557; SNP3: rs3785883; SNP4: rs2471738; SNP 5: rs7521; SNP6: rs9468)

The pMAPT-H1 and pMAPT-H2 vectors contain equivalent up and downstream regulatory regions: 7.7 kb promoter sequences, all introns and exons, and 5.5 kb of downstream sequences sufficient to contain 2 polyadenylation sequences (Figure 5A). The inclusion of all the elements of the *MAPT* gene is crucial for maintaining its physiological relevance for our studies on the effects of haplotype-specific polymorphisms on alternative splicing (Peruzzi et al. 2009). The start and end coordinates of the *MAPT* gene sequence correspond to chromosome 17:45886670-46029648 on the Genome Reference Consortium GRCh38 human genome assembly (Figure 5B).

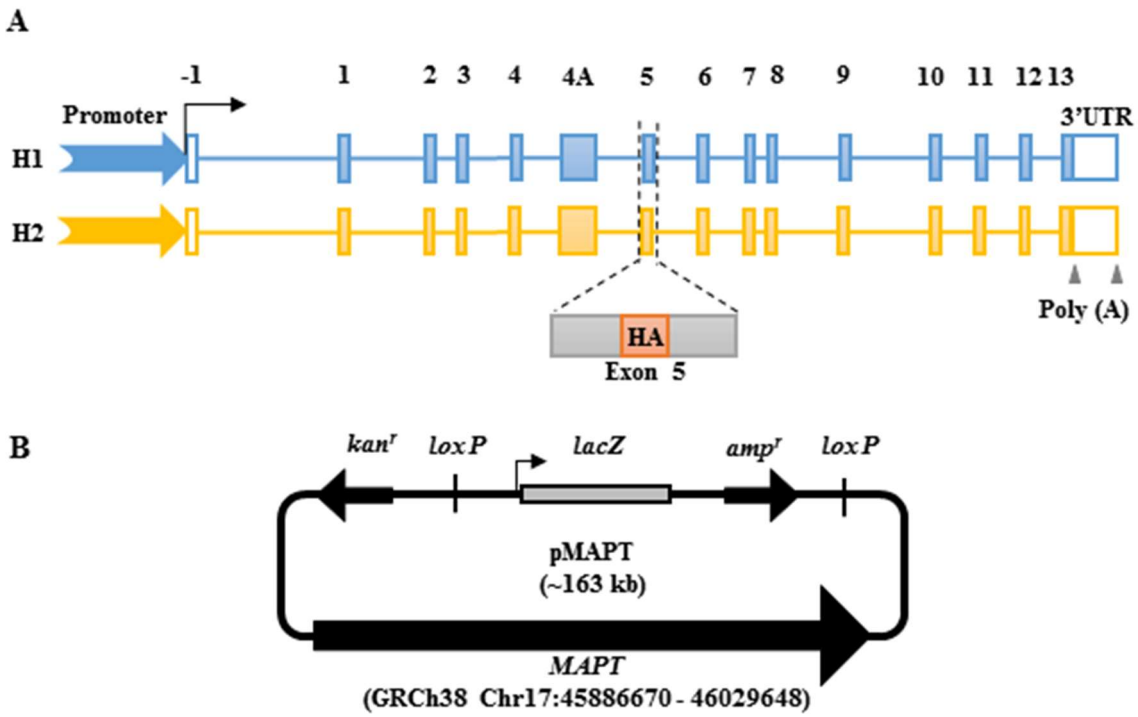


Figure 5. pMAPT-H1 and -H2 genomic expression PAC vectors.

(A) Schematic representation of the H1 and H2 *MAPT* gene in the PAC vectors containing 7.7 kb promoter sequences, all introns and exons, 5.5 kb 3' untranslated region (3'UTR). (Poly (A): polyadenylation signal, black arrow: transcription start, HA: haemagglutinin tag). The H1 and H2 genes start and end at the same position, and the sequences are the same except at positions where haplotypic variations occur. Exons numbers are indicated above. Coloured and empty boxes represent translated and untranslated regions, respectively.

(B) Schematic representation of the pMAPT vector used in this study. The *MAPT* gene is approximately 143 kb and correspond to chromosome 17:45886670 – 46029648 on the Genome Reference Consortium GRCh38 human genome assembly. (lacZ: β-galactosidase gene, kan: kanamycin resistance, amp: ampicillin resistance). The above figures are not drawn to scale.

The coverage of common SNPs registered on the dbSNP b144 database (1000 Genomes minor allele frequency >= 0.05) by the pMAPT-H1 and pMAPT-H2 vectors was first determined by a detailed analysis on the 143 kb *MAPT* region covered by the vectors (Figure 6). A total of 655 common SNPs were identified between the H1 and H2 vectors, capturing over 86% of the common SNP diversity as registered on the dbSNP database in the corresponding *MAPT* region. Of the 655 SNPs, approximately 60% of the common SNPs are present in the upstream 7.7 kb promoter region and 67.7 kb 5' untranslated intron -1 region (Figure 6). The 59.5 kb intronic regions between coding exons harbour about 38% of the common SNPs. The remaining 2% common SNPs locate in the 5.5 kb 3' untranslated and downstream regions, and in coding exons (Figure 6).

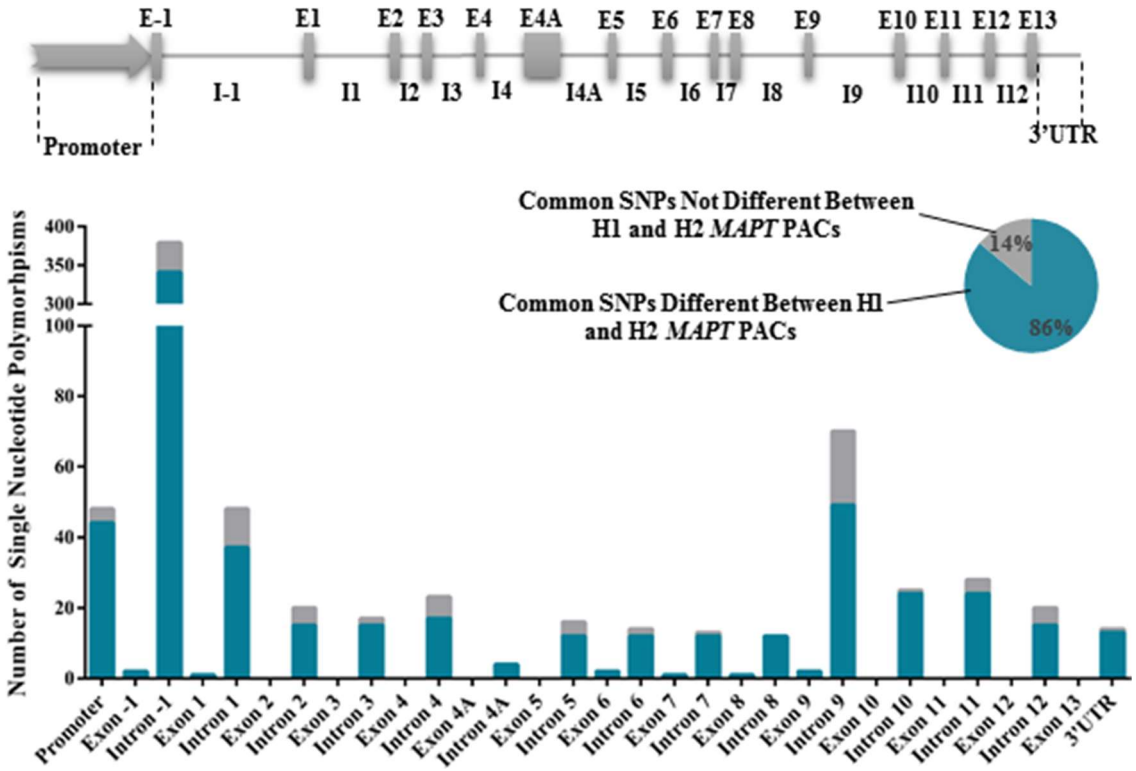


Figure 6. Comparative sequence analysis of SNPs present on the H1 and H2 *MAPT* gene.
Common SNPs identified between the H1 and H2 *MAPT* gene. SNP data are obtained from dbSNP b144 database. Only SNPs with a minor allele frequency of 5% or above as reported in the 1000 Genomes project are included for the comparison. Overall, 86% of the common SNPs are identified to be different between the H1 and H2 *MAPT* gene. A schematic representation of the *MAPT* gene structure, with exon (E) and intron (I) numbers, promoter and 3' untranslated region (3'UTR) labelled, is shown above the graph.

The polymorphisms occur predominately in the non-coding regions though nine variants were identified in exons (Figure 6). Of these, SNP rs17650901 upstream of the translation start site in exon 1, and two synonymous SNPs rs1052553 and rs17652121 in exon 9 have been described previously (Baker et al. 1999). Two SNPs rs11575895 and rs62056779 are present in the untranslated exon -1. Two additional synonymous SNPs occur in exons 7 (rs1052551) and 8 (rs62063845). Exon 6 contains two non-synonymous missense variants that result in serine to proline (rs10445337) and tyrosine to histidine (rs2258689) changes. The amino acid sequence of tau was analysed using SNAP (http://rostlab.org/services/snap/) (Hecht et al. 2015), a program that makes predictions regarding the functionality of mutated proteins. The amino acid changes resulted from rs10445337 and rs2258689 in exon 6 were both predicted to have a neutral effect on protein function.

3.2.2 Analysis of Splice Factor Binding Sites at the Exon 3 Splice Sites *In Silico*

To identify candidate polymorphisms that may be responsible for the increased inclusion of exon 3 by the H2 allele, the H1 and H2 intronic sequences upstream and downstream of exon 3 were aligned using the programme ApE-A-Plasmid Editor. In total, 30 SNPs were found to be different between the pMAPT-H1 and –H2 vectors, in the upstream and downstream introns of exon 3 (Figure 6). To determine if the H1/H2 SNPs present in our pMAPT-H1 and –H2 vectors locate in splicing regulatory sequences, a 500 nucleotide region adjacent to the exon 3 and exon 4 5' splice sites were analysed using SVM-BP finder for the most probable branch point and polypyrimidine tract sequences (Corvelo et al. 2010). SNPs were not identified in the splice sites, predicted branch points or the predicted polypyrimidine tract sequences. SNPs closest to the exon-intron boundaries have previously been associated with splicing patterns (Hull et al. 2007). Therefore 5 SNP sequences (rs117124984, rs118087478, rs77875796, rs1800547 and rs17651213) falling within 200 bp of exon 3 splice sites were analysed using SFmap, SpliceAid2 and ESEfinder to predict the binding sites of common splice factors (Figure 8) (Paz et al. 2010; Piva et al. 2012; P. J. Smith et al. 2006). It was found that the 5 SNPs of H1 or H2 haplotypes represent different potential splice factor binding sites (Figure 7). Moreover, splice factor binding sites common to both haplotypes were also identified (Figure 7). The results here suggest that altered splice factor binding sequence by SNPs could be a potential mechanism of how haplotype-specific SNPs regulate the allele-specific inclusion of *MAPT* exon 3. Due to the proximity of the 5 SNPs (rs117124984, rs118087478, rs77875796, rs1800547 and rs17651213) to the exon 3 splice sites and the putative haplotype-specific binding of splice factors, these variants were selected as potential functional variants for further study in our cell culture model.

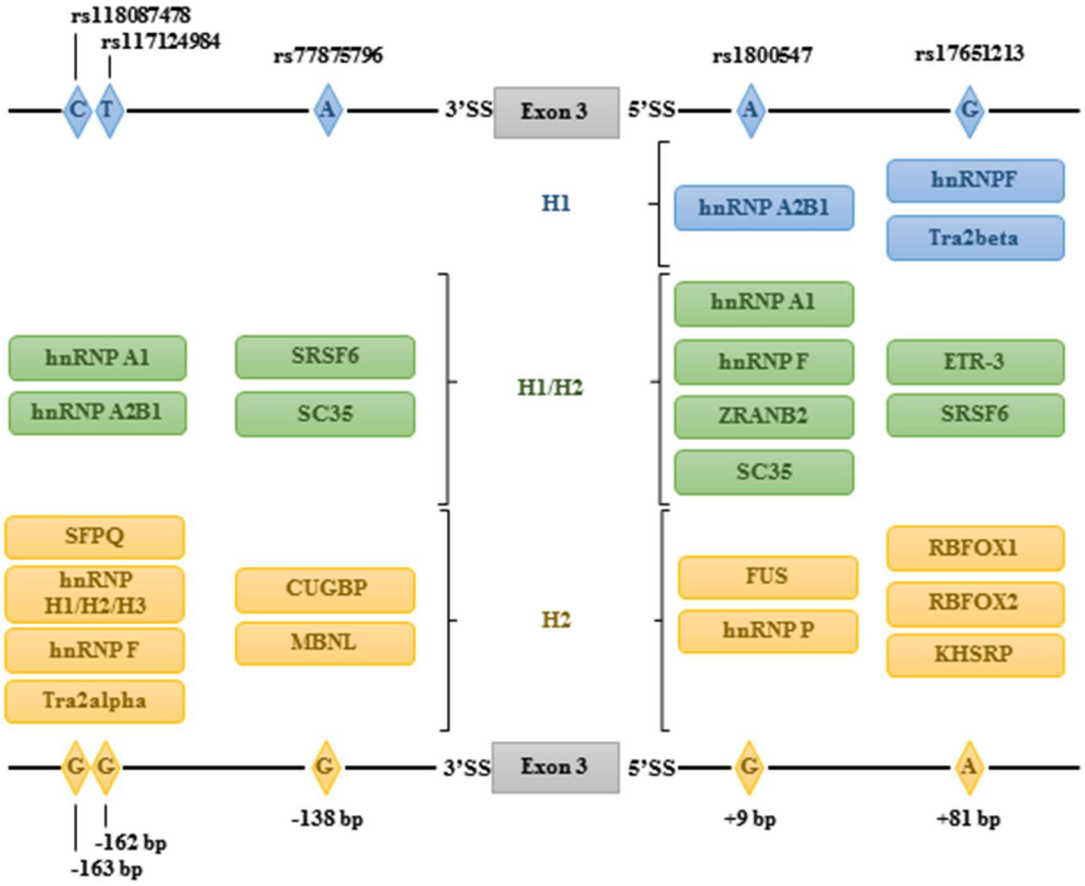


Figure 7. Splice factor binding sites prediction on five candidate SNP sequences.

Splice factor binding site prediction at the SNP sequences using three different online tools: SFmap, SpliceAid2 and ESE. The five candidate SNPs share common splice factor binding sites between the H1 and H2 alleles as indicated in green boxes. The five candidate SNPs feature splice factor binding sites unique to either the H1 or H2 allele, indicated in blue or yellow boxes, respectively. SNP IDs are indicated above the SNPs, and their distance relative to the 3' or 5' splice sites (3'SS or 5'SS) are indicated below the SNPs.

3.2.3 Candidate SNPs rs1800547 and rs17651213 are in Strong Linkage Disequibrium with the H1/H2 Haplotype Structure

To determine whether the five candidate SNPs (rs117124984, rs118087478, rs77875796, rs1800547 and rs17651213) in close proximity to the exon 3 5' and 3' splice sites (Figure 7) are H1/H2 haplotype-specific, linkage analysis between the candidate SNPs and the published H1/H2 haplotype-tagging SNPs was performed (Simon-Sanchez et al. 2009). Briefly, genotype data from individuals in the CEU population (Utah residents with Northern and Western European ancestry from the CEPH collection) were obtained from the international HapMap project (The International HapMap Consortium 2003) and LD analysis was performed using Haploview (Barrett et al. 2005). D' value is a measure of LD between two genetic markers. In general, D' = 1 indicates complete LD, while D' < 1 indicates incomplete LD. Only D' values near 1 are a reliable indicator of LD. D' can range from 0 to 1 (The International HapMap Consortium 2003). Logarithm of the odds (LOD) score is a statistical estimate of whether two markers are likely linked. In general, LOD ≥ 3 indicates that two markers are likely linked together on the chromosome.

Haploview generates the LD plot of D'/LOD values, where the intensity of the box colour is proportional to the strength of the LD property (Barrett et al. 2005). Bright red (D' = 1, LOD ≥ 2) indicates strong evidence of LD, light red (D' < 1, LOD ≥ 2) indicates uninformative evidence of LD, and white (D' < 1, LOD < 2) indicate strong evidence of recombination (Barrett et al. 2005; The International HapMap Consortium 2003). Two candidate SNPs, rs1800547 and rs17651213, downstream of the 5' splice site of exon 3, were found to be in strong LD with the H1/H2 haplotype structure (Figure 8). In addition, the two SNPs were in strong LD with a GWAS significant SNP rs17649553 (Figure 8) from a recent large-scale meta-analysis of PD GWAS (Nalls et al. 2014). There was no available genotype data for rs117124984, rs118087478 and rs77875796 from the HapMap project. The five candidate SNPs are common variants with minor allele frequencies of over 5 % as identified in the 1000 Genomes project (1000 Genomes Project Consortium et al. 2012). Since all the five common SNPs are captured by our H1 and H2 *MAPT* PAC vectors, they were therefore

selected for generating haplotype-hybrid *MAPT* vectors where the SNPs were exchanged between the H1/H2 haplotype backgrounds.

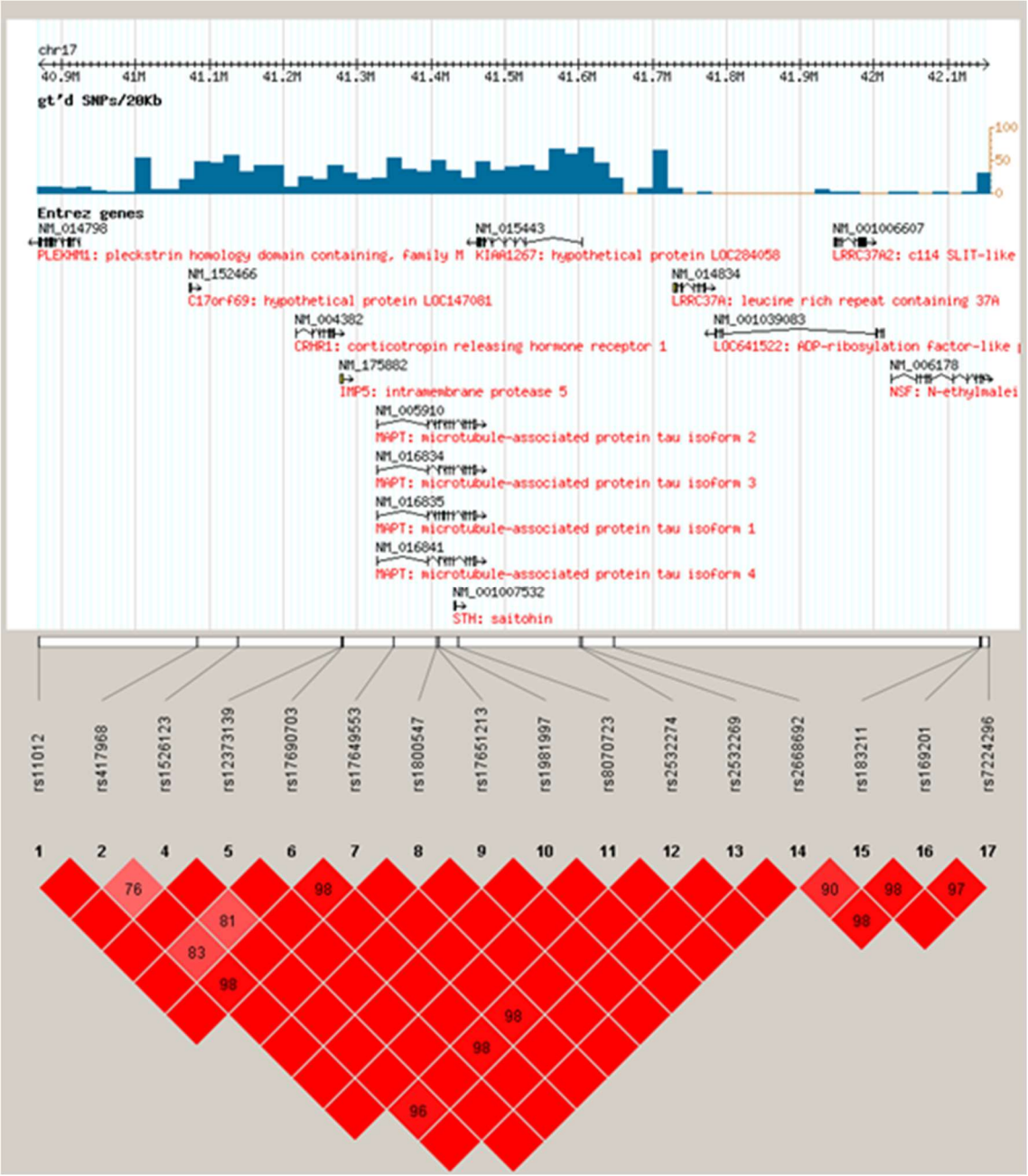


Figure 8. rs1800547 and rs17651213 are in linkage disequilibrium with the H1/H2 haplotype structure.

Linkage analysis performed using CEU population genotype data from the international HapMap project by the programme Haploview. The linkage disequilibrium plot shows rs1800547 and rs17651213 are in linkage disequilibrium with the H1/H2 haplotype structure as represented by haplotype tagging SNPs. Genomic coordinates on the top. gt'd SNPs/20kb: genotyped SNPs per 20 kb region. Gene names and transcripts are indicated above the linkage disequilibrium plot.

3.2.4 The Homologous Recombination Strategy by Gene Bridges for SNPs Modification

To assess the potential functional effects of the five candidate intronic SNPs (rs117124984, rs118087478, rs77875796, rs1800547 and rs17651213) on *MAPT* exon 3 expression, a selection-counter selection method by Gene Bridges (Figure 9) (J. Wang et al. 2006) was used to modify the 5 SNP sequences on the pMAPT-H1 and pMAPT-H2 vectors. The selection-counter selection method uses a homologous recombination strategy (Figure 9). First, the pRedE/T plasmid encoding RecE exonuclease and RecT single strand binding proteins was introduced into bacterial colonies carrying the PAC vectors (Figure). Second, the expression of RecE and RecT was induced upon a temperature shift from 30^{o}C to 37^{o}C with the addition of L-arabinose (Figure 9). The pRedE/T plasmid contains a temperature sensitive origin of replication and is not retained after the temperature shift. Third, a PCR product carrying the selection-counter selection cassette encoding streptomycin sensitive and chloramphenicol resistant genes (rpsl/chl), flanked by homology sequences to the PAC, was introduced into the bacteria (Figure 9). Fourth, the rpsl/chl cassette with homology sequences replaces the desired sequence to be modified, mediated by RecE and RecT proteins (Figure 9). RecE digests the 5' ends of the PCR product while RecT proteins bind to the overhanging single strand DNA. RecT directs the PCR product towards the homologous sequence on the PAC, invades and displaces the double stranded DNA for integration. The replacement of the PAC sequence by the PCR product is completed *in vivo* by the DNA replication machinery of the bacteria. Lastly, this process is repeated again to replace the rpsl/chl cassette with the desired DNA sequence to be modified on the PAC (Figure 9). Using this recombination strategy, our laboratory has previously engineered a haemagglutinin (HA) tag in frame in exon 5 (by Anne-Laure Bechy) (Figure 5, Table 3). The HA tag allows for the differentiation of transgenic and endogenous *MAPT* transcripts in cell culture models. This method is employed here to generate H1/H2 *MAPT* haplotype-hybrid genomic expression vectors.

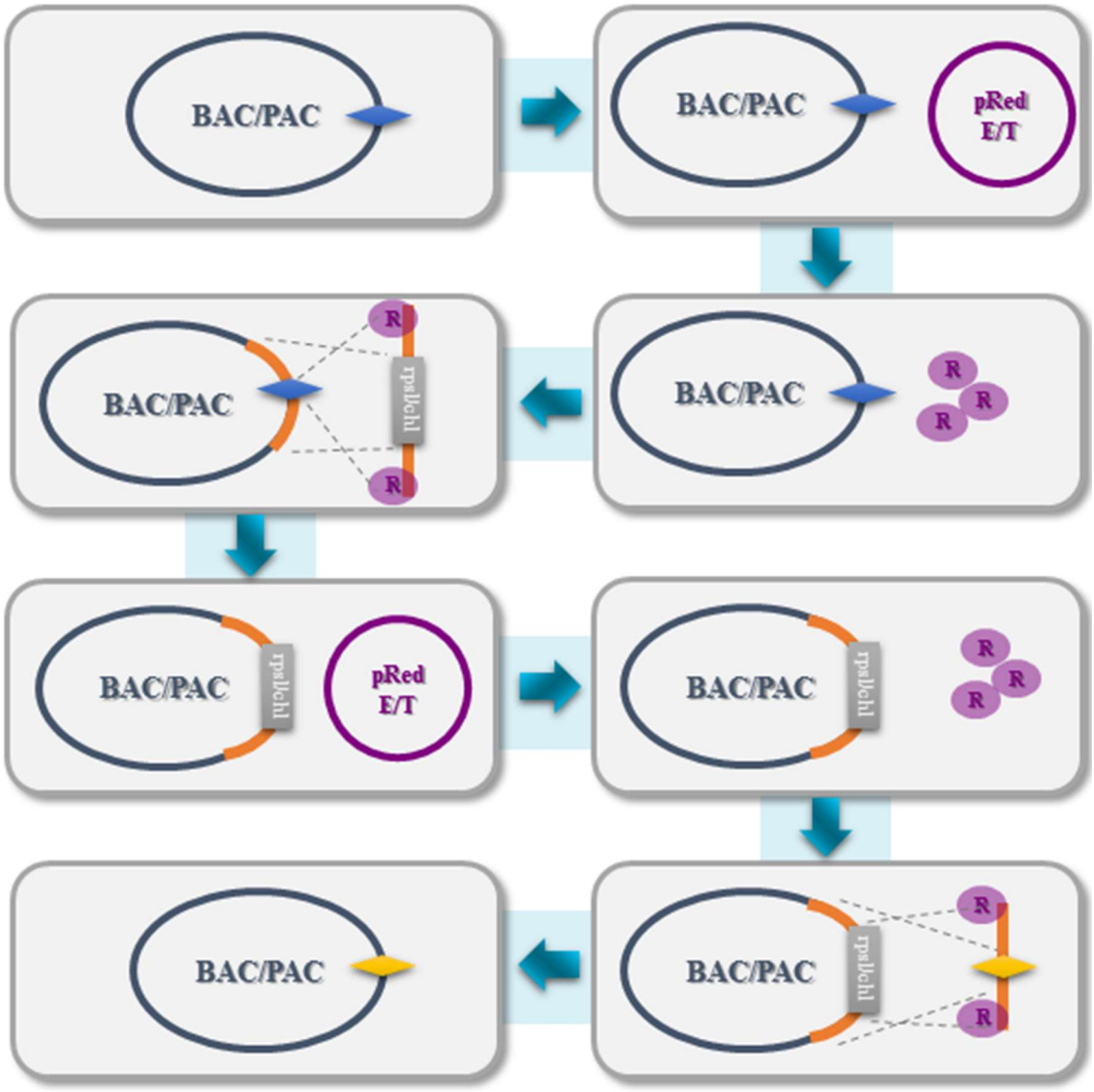


Figure 9. The selection counter-selection method for PAC modifications.

The selection counter-selection method by Gene Bridges for PAC modifications. PAC vectors are modified in DH10β *E. coli*. The sequence to be replaced (blue diamond) was first replaced by a selection cassette that confers streptomycin sensitivity and chloramphenicol resistance (rpsl/chl) with flanking homology sequences (orange). Homologous recombination of homology sequences was mediated by RecE and RecT exonuclease and single stranded binding proteins, illustrated as R in the diagram. The selection cassette was then replaced by the desired sequence (yellow diamond). pRed/ET plasmid encoding RecE and RecT were cloned at 30ºC. RecE and RecT expression was induced by a temperature shift from 30ºC to 37ºC with the addition of L-arabinose. pRed/ET plasmid itself was not retained at 37ºC due to a temperature sensitive origin of replication. PCR products are plasmids were introduced into cells by electroporation.

Using the homologous recombination strategy described above, our laboratory has previously generated the vectors with rs1800547 exchanged (H1-*2-1* and H2-*1-2*) (by Anne-Laure Bechy), and with rs1800547 and rs17651213 exchanged (H1-*2-2* and H2-*1-1*) (by Maria Gavriliouk) from the pMAPT-H1 and –H2 wildtype vectors (Figure 10, Table 3). The aim of this chapter is to produce two more pairs of vectors with rs17651213 (H1-*1-2* and H2-*2-1*), and with rs117124984, rs118087478 and rs77875796 (H1-*2-2-2* and H2-*1-1-1*) exchanged (Figure 10).

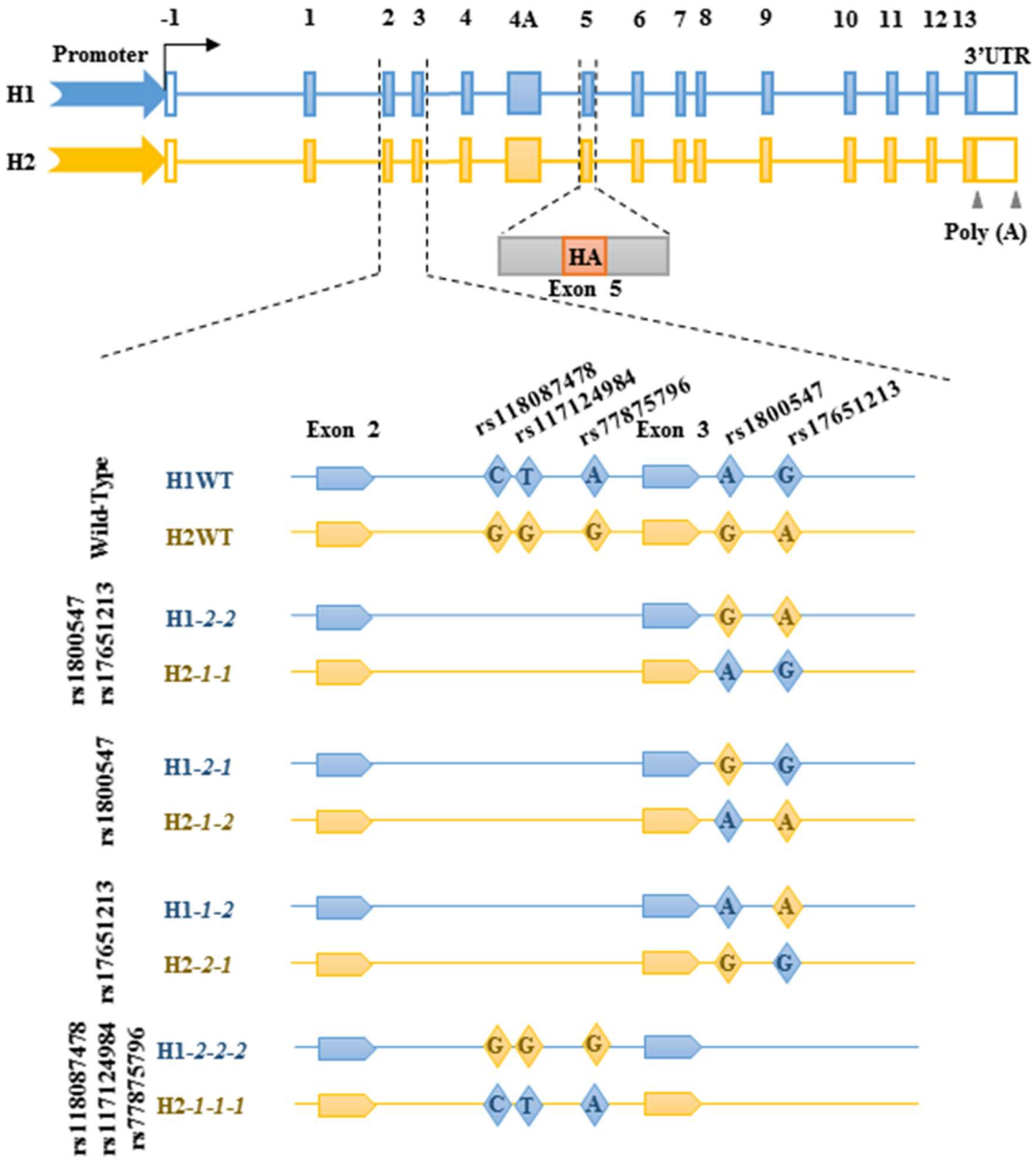


Figure 10. *MAPT* haplotype-hybrid genomic expression vectors.

A schematic representation of the five pairs of *MAPT* haplotype-hybrid vectors generated from the H1 and H2 wild-type (H1WT and H2WT) vectors. SNPs rs118087478, rs117124984, rs77875796, rs1800547 and rs17651213 are exchanged in different combinations between the haplotype backgrounds. SNPs are represented in diamonds with their alleles indicated.

3.2.5 Generation of Haplotype-Hybrid Vectors H1-2-2-2, H2-1-1-1, H1-1-2 and H2-2-1

To exchange SNPs on the H1 and H2 wild type backgrounds, pRedE/T plasmid was first introduced into bacterial colonies carrying either H1 or H2 vectors wild-type. Successful incorporation of pRedE/T plasmid into colonies was confirmed by EcoRI digestion of the plasmid (Figure 11A and Figure 12A). We next PCR amplified the rpsl/chl selection-counter selection cassette with long homology arm primers and checked that the amplicon size was correct by gel electrophoresis (Figure 11B and Figure 12B). The rpsl/chl cassette with homology sequences was then introduced into bacterial colonies expressing RecE and RecT proteins encoded by the pRedE/T plasmid. We screened for successful incorporation of the rpsl/chl cassette by streaking colonies on plates containing chloramphenicol and both chloramphenicol and streptomycin. Colonies that are resistant to chloramphenicol but sensitive to streptomycin indicate that the rpsl/chl cassette has been incorporated successfully (Figure 11C and Figure 12C). We then PCR amplified the SNP sequences to be exchanged and again, PCR products were analysed by gel electrophoresis (Figure 11D and Figure 12D). Using the same homologous recombination process, we replaced the rpsl/chl cassette with PCR products containing our desired SNP sequences. The elimination of the rpsl/chl cassette was checked using colony PCR (Figure 11E and Figure 12E) by primers spanning the region of homologous recombination. Finally, we checked for PAC integrity by digesting PAC vectors with NotI and SalI restriction enzymes followed by pulsed-field gel electrophoresis (Figure 11F and Figure 12F).

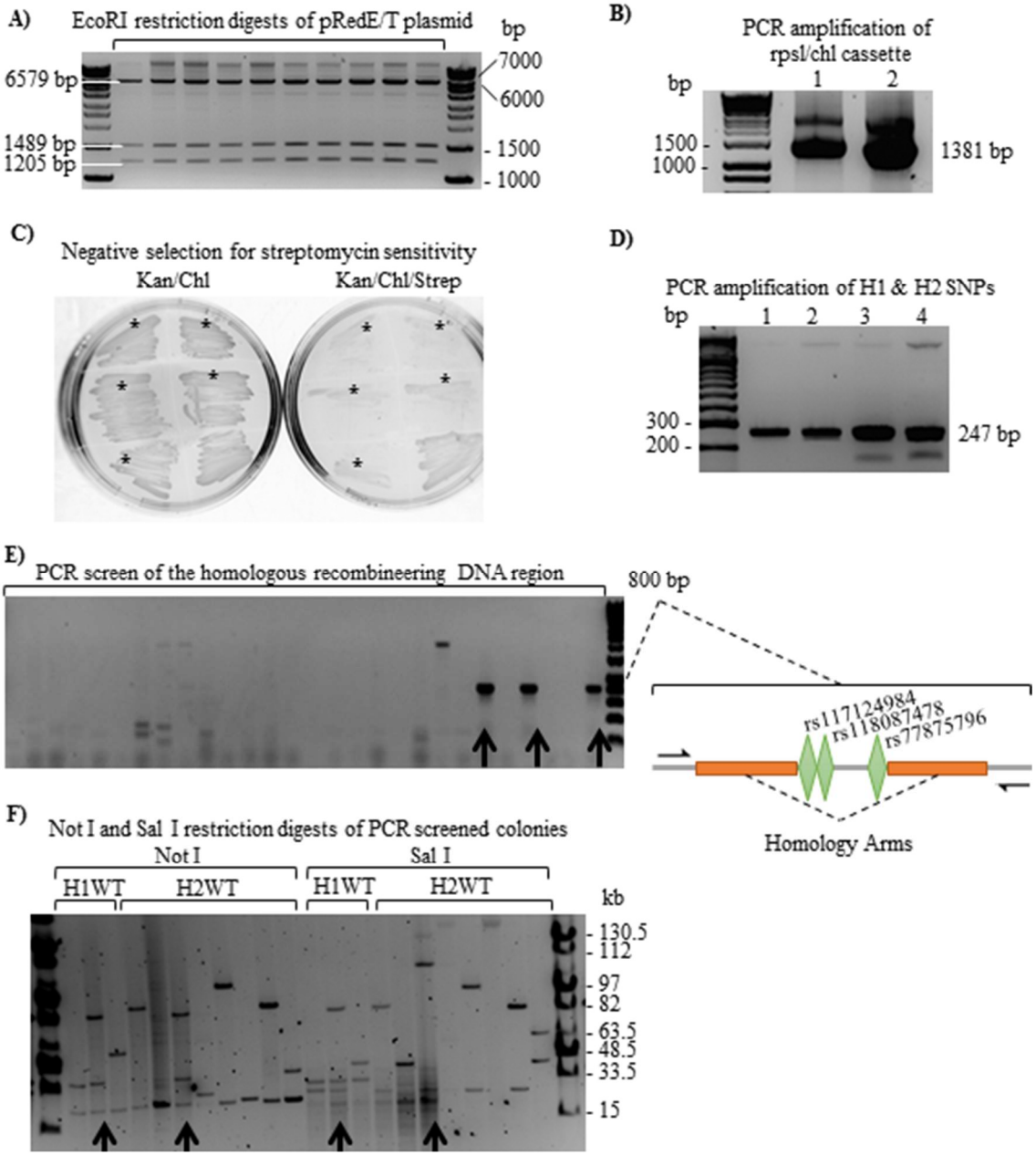


Figure 11. Generating *MAPT* haplotype-hybrid vectors H1-*2-2-2* and H2-*1-1-1*.

(A) pRedE/T plasmid electroporated into *E. coli* carrying the pMAPT-H1WT and –H2WT vectors were screened using EcoRI restriction digests. Correct band sizes confirmed successful electroporation. **(B)** rpsl/chl cassette amplified by primers with homology sequence flanking the SNPs. PCR products before (lane 1) and after (lane 2) PCR purification. **(C)** Kanamycin (Kan) and chloramphenicol (Chl) resistant, and streptomycin (Strep) sensitive colonies (stars) indicate the integration of rpsl/chl cassette into pMAPT vectors. **(D)** SNPs amplified by primers with homology sequence flanking the SNPs. PCR products for the H1 and H2 SNPs before (lanes 1 & 2) and after (lanes 3 & 4) DpnI restriction digests to remove the parental DNA template. **(E)** Left panel: colonies with rpsl/cassette removed as shown by correct PCR product size (arrows). Right panel: PCR primer locations. **(F)** Colonies (arrows) carrying intact modified haplotype-hybrid vectors.

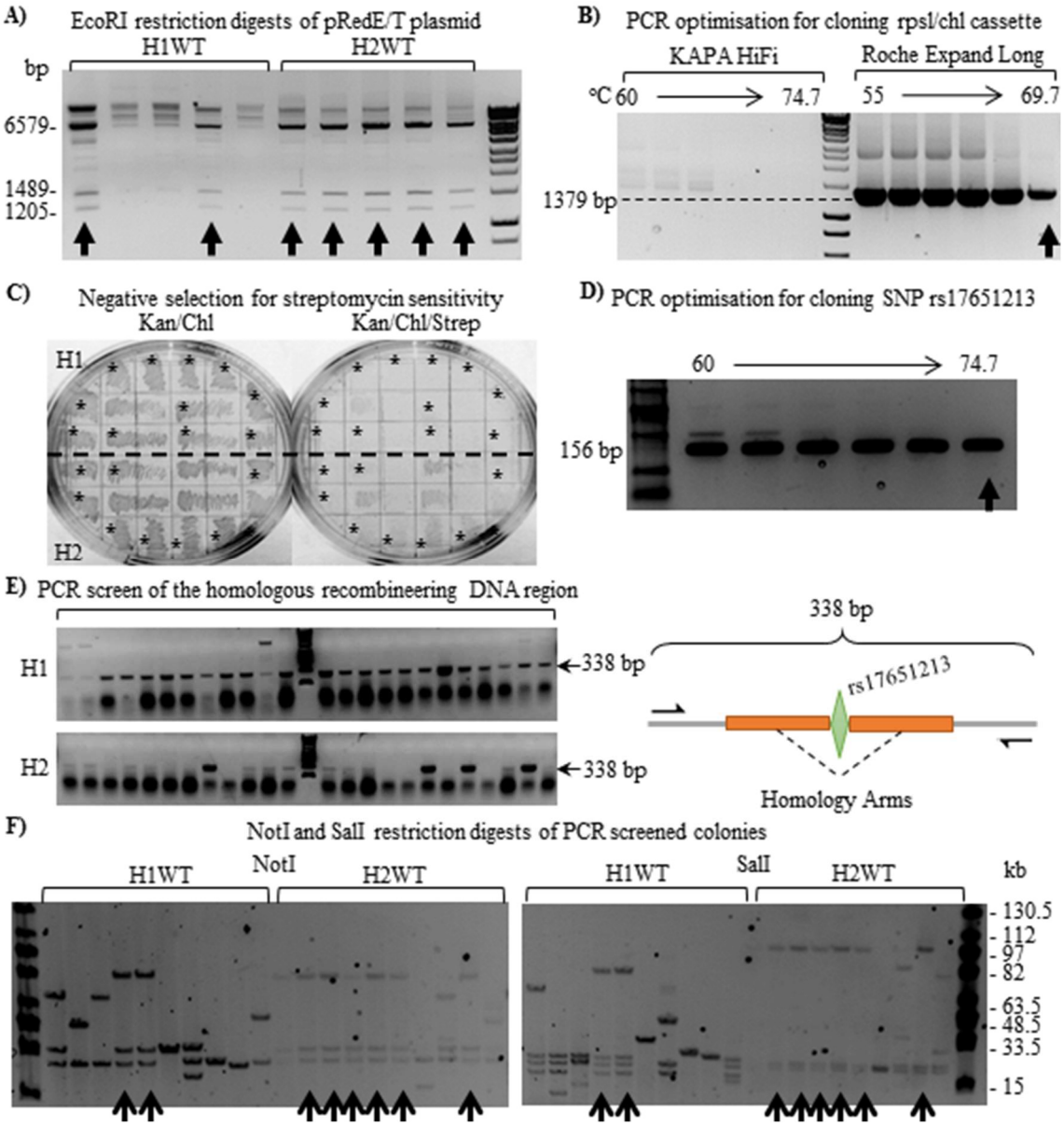


Figure 12. Generating MAPT haplotype-hybrid vectors H1-*1-2* and H2-*2-1*.

(A) pRedE/T plasmid electroporated into *E. coli* carrying pMAPT-H1WT or –H2WT vectors were screened using EcoRI restriction digests. Correct band sizes confirmed successful electroporation (arrows). **(B)** PCR optimisation for rpsl/chl cassette by primers with homology sequence flanking SNP rs17651213. KAPA HiFi PCR kit and Roche Expand Long Template PCR System were used with increasing annealing temperatures. PCR condition with specific PCR product (arrow). **(C)** Kanamycin (Kan) and chloramphenicol (Chl) resistant, and streptomycin (Strep) sensitive colonies (stars) indicate the integration of rpsl/chl cassette into pMAPT vectors. **(D)** PCR optimisation with increasing annealing temperatures for SNP rs17651213 by primers with homology sequence flanking rs17651213. PCR condition with specific PCR product (arrow). **(E)** Left: PCR screen for colonies with rpsl/chl cassette removed produce 338 bp PCR product (arrows). Right: PCR primer locations. **(F)** Colonies (arrows) carrying intact modified haplotype-hybrid vectors.

PAC vectors of the correct band sizes at NotI SalI restriction digests were then maxi-preped and re-electroporated into *E. coli* for reducing the chances of two different vectors being carried in the same *E. coli* colony. Vector integrity was once again examined using NotI and SalI restriction digests (Figure 13A-B and Figure 14A-B). The vectors were subsequently sequenced to confirm that the SNPs have been engineered successfully with no other mutations present in the region of homologous recombination (Figure 13C and Figure 14C).

Here, four pairs of pMAPT haplotype-hybrid vectors were successfully generated: H1-*2-2-2* and H2-*1-1-1* (with SNPs rs117124984, rs118087478 and rs77875796 exchanged), and H1-*1-2* and H2-*2-1* (with SNPs rs17651213 exchanged) (Figure 10). Vector pairs previously generated by our laboratory include: H1WT and H2WT (wildtype), H1-*2-1* and H2-*1-2* (with SNP rs1800547 exchanged), and H1-*2-2* and H2-*1-1* (with SNPs rs1899547 and rs17651213 exchanged) (Figure 10). All the vectors will be used in our cell culture model for investigating the haplotype-specific effects of intronic SNPs on *MAPT* exon 3 inclusion. In the next section, a β-galactosidase gene carried in the plasmid pH-FRT-Hy (Lufino et al. 2013), were introduced into our pMAPT vectors by Cre-mediated lox P recombination.

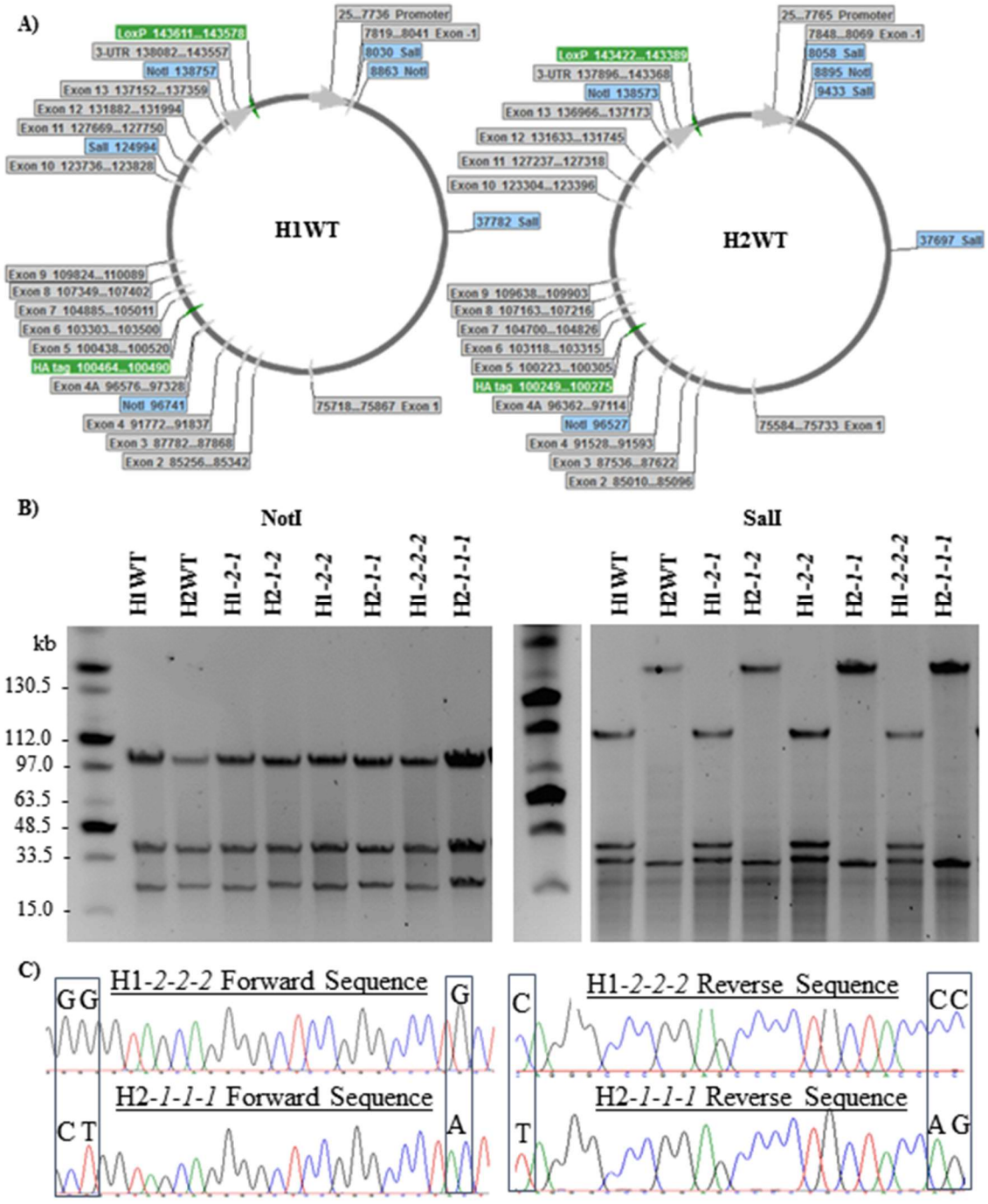


Figure 13. Confirmation of H1-*2-2-2* and H2-*1-1-1* *MAPT* haplotype-hybrid vectors.

(A) Restriction map for NotI and SalI enzymes on the H1 and H2 vectors. *MAPT* gene promoter, 3' UTR, exons, lox P site and the HA tag are indicated in boxes. Numbers represent their relative location on the vectors. **(B)** Restriction digests of H1-*2-2-2* and H2-*1-1-1* vectors by NotI and SalI enzymes. Correct band sizes indicate intact vectors. **(C)** Sequencing chromatograms confirming SNPs rs118087478, rs117124984 and rs77875796 exchanged on the H1 and H2 backgrounds (sequence highlighted in boxes). Sequencing was carried out both in the forward and reverse directions.

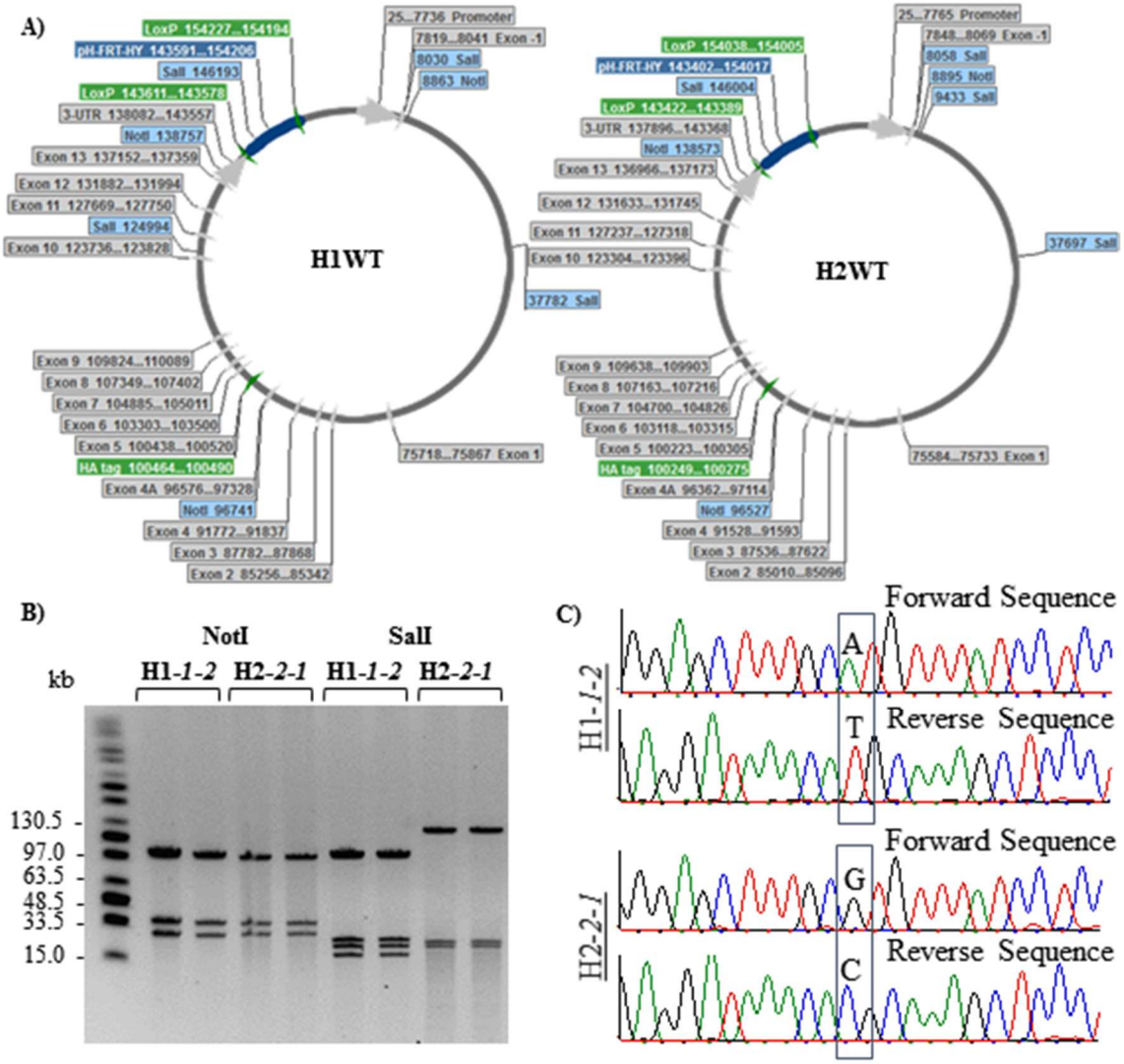


Figure 14. Confirmation of H1-*1-2* and H2-*2-1* *MAPT* haplotype-hybrid vectors.

(A) Restriction map for NotI and SalI enzymes on the H1 and H2 vectors. *MAPT* gene promoter, 3' UTR, exons, lox P site and the HA tag are indicated in boxes. Numbers represent their relative location on the vectors. H1-*1-2* and H2-*2-1* vectors are generated from wildtype vectors previously retrofitted with pH-FRT-Hy plasmids (Lufino et al. 2013). **(B)** Restriction digests of H1-*1-2* and H2-*2-1* vectors by NotI and SalI enzymes. Correct band sizes indicate intact vectors. **(C)** Sequencing chromatograms confirming rs17651213 H1-G and H2-A alleles are exchanged on the H1 and H2 backgrounds (sequence highlighted in boxes). Sequencing was carried out both in the forward and reverse directions.

3.2.6 Retrofitting pMAPT Vectors with pH-FRT-Hy Plasmid

All the pMAPT vectors used in this study were retrofitted with the pH-FRT-Hy plasmid that contains a β-galactosidase (*lacZ*) reporter gene (Lufino et al. 2013). The *lacZ* reporter gene can be used for assessing gene delivery efficiencies in our cell culture model. Both the pH-FRT-Hy plasmid and the pMAPT vectors carry a lox P sequence (Figure 15). *In vitro* Cre-mediated lox P recombination was used to combine pH-FRT-Hy and pMAPT vectors. The recombination reactions were performed with different ratios of pH-FRT-HY to pMAPT ratios. The optimal recombination ratio was found to be 1 µg of pH-FRT-Hy plasmid to 50 ng pMAPT vector. Following Cre-mediated recombination reactions, the resulting vectors were electroporated into DH10β *E. coli* cells. Successful recombination of the pH-FRT-Hy plasmid and the pMAPT vector was determined by restriction digest of the recombined pMAPT vector using NotI and SalI enzymes, analysed by pulsed-field gel electrophoresis. Intact vectors with correct band sizes were then re-electroporated into *E. coli* cells for minimising the chances of two different vectors being carried in the same *E. coli* colony. Again, NotI and SalI restriction digests were used to confirm vectors (Figure 16).

All the haplotype-hybrid vectors generated were retrofitted with pH-FRT-Hy plasmid except for H1-*1-2* and H2-*2-1* (Figure 14) where they were generated from wildtype vectors that were already retrofitted. Together, the pMAPT vectors with the *lacZ* reporter gene will be used in our expression studies in our SK-N-F1 neuroblastoma cell culture model.

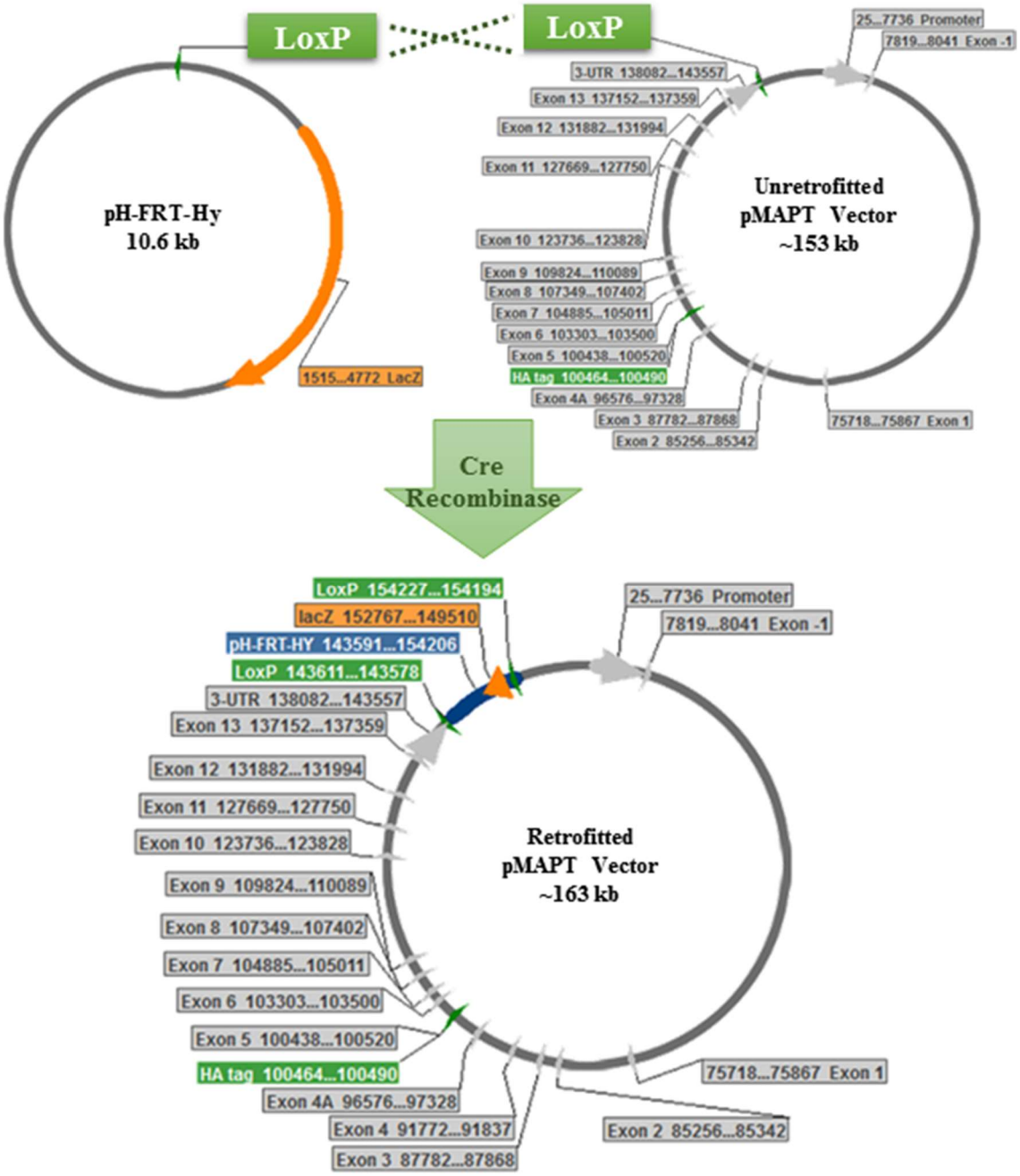


Figure 15. Retrofitting pMAPT vectors with pH-FRT-Hy plasmid.

pMAPT vectors are retrofitted with pH-FRT-Hy plasmid via Cre-mediated lox P recombination. pH-FRT-Hy carries a β-galactosidase gene (lacZ). β-galactosidase gene acts as a reporter gene for assessing pMAPT vector transfection/transduction efficiencies in cell culture models.

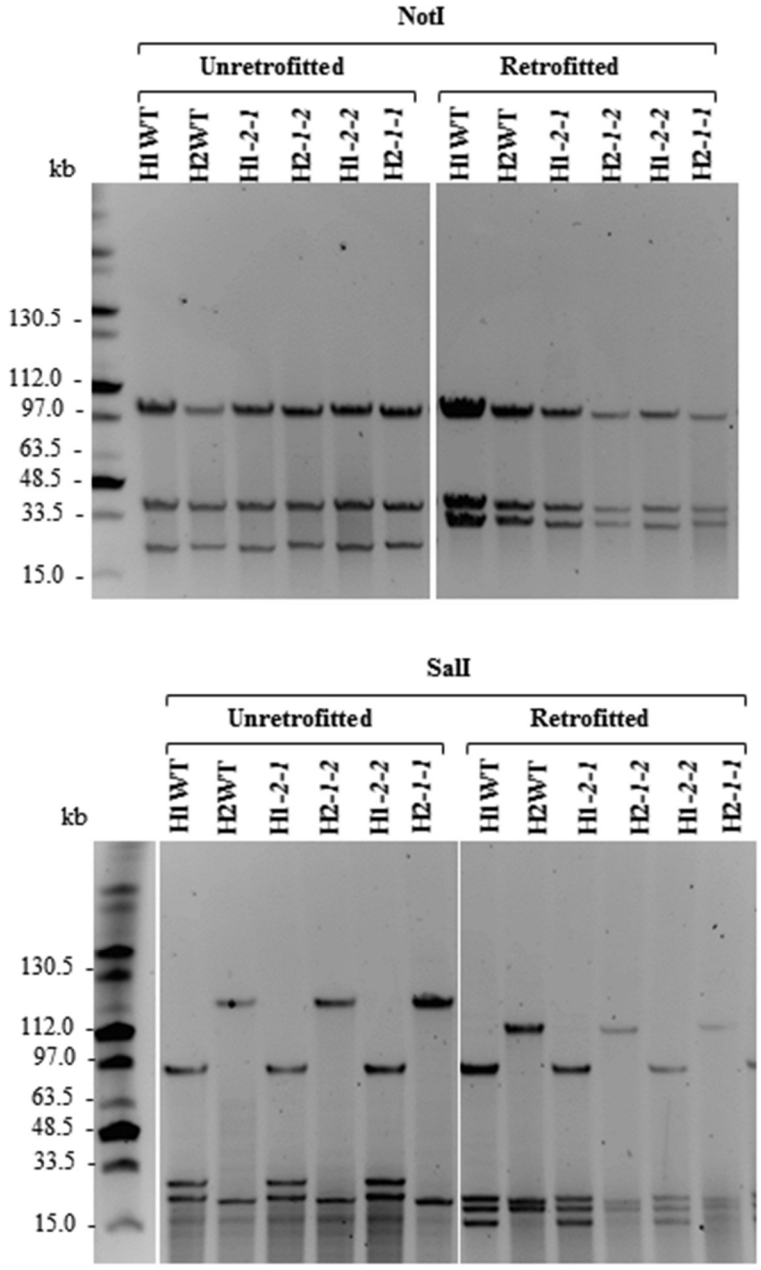


Figure 16. Restriction digests of retrofitted and unretrofitted vectors by NotI and SalI.

Vectors retrofitted are digested alongside unretrofitted vectors using NotI and SalI enzymes, followed by pulsed-field gel electrophoresis. Correct band sizes indicate the retrofitting reaction has occurred successfully and that the vectors are intact.

3.3 Discussion

3.3.1 Generating Haplotype-Hybrid pMAPT Vectors by Engineering Candidate SNPs

The main aims in this first results chapter of the thesis are to i) select candidate SNPs that may affect *MAPT* exon 3 inclusion by performing computational sequence analyses and ii) to generate PAC *MAPT* haplotype-hybrid expression vectors by engineering SNP sequences through homologous recombination for studying the functional effects of the candidate SNPs. Five candidate SNPs that are proximal to the exon 3 3' (rs117124984, rs118087478 and rs77875796) and 5' (rs1800547 and rs17651213) splice sites have been selected to be studied in greater details on their effects on exon 3 inclusion. Haplotype-hybrid PAC vectors were successfully generated where the five candidate SNPs were exchanged between the H1/H2 haplotype backgrounds (Figure 10). PAC modification with single base pair accuracy has allowed us to generate haplotype-hybrid vectors for expression studies in cell culture models.

3.3.2 BAC/PAC Vectors as a Tool for Alternative Splicing Studies

To my knowledge, this is the first study to use genomic expression PAC vectors to employ genomic expression vectors to study alternative splicing changes that may be caused by SNPs. Traditionally, the effects of cis-regulatory enhancer and silencer elements on splicing are largely identified and defined through mini-gene constructs (D. L. Black 2003; Cartegni et al. 2002; X. D. Fu and Ares 2014). Mini-genes are short versions of a complete gene, often containing only selected and incomplete sections of exons and introns, and the cis-regulatory sequence elements of interests, which can be expressed in cell culture models or studied in *in vitro* splicing assays (Cooper 2005). Gene expression by mini-genes is not controlled by the physiological promoter sequences and cannot accommodate downstream polyadenylation signals. In order to understand the functional significance of SNPs on alternative splicing in their genomic context, it is thus essential to use whole genomic locus vectors where all the haplotype-specific polymorphisms can be captured and manipulated. The H1 and H2 *MAPT* genomic vectors were shown here to capture over 86 % sequence diversity in the *MAPT* gene, demonstrating their physiological relevance for studying

allele-specific exon 3 splicing. The preservation of genomic promoter sequence, polyadenylation signals, 3' untranslated sequences and all the introns in our H1 and H2 *MAPT* PAC vectors further provide the physiological genomic context for interrogating the effects of SNPs on the haplotype-specific inclusion of *MAPT* exon 3.

3.3.2 Alternative Splicing Beyond *Cis*-Acting Regulatory Elements

Alternative splicing is known to couple with other gene expression and pre-mRNA processing machineries, including transcription, mRNA export, 5' capping, 3' polyadenylation, non-sense mediated decay and pre-mRNA degradation (Maniatis and Reed 2002; Montes et al. 2012). It is also regulated by various epigenetic components such as chromatin structure, DNA methylation, and histone modifications (Luco et al. 2011; Zhou et al. 2012). Therefore besides understanding cis-regulatory sequence elements which play an important role in determining the outcome of exon inclusion, future studies could focus on exploring how single nucleotide polymorphisms may modulate splicing activity indirectly via splicing coupled processes and epigenetic components. Chromatin immunoprecipitation (ChIP) based methods will be invaluable for studying how SNPs directly affect various processes and elements linked to splicing. For example, the haploChIP method demonstrated haplotype-specific polymerase loading at the LTA gene, correlating with the haplotype-specific mRNA expression levels (Knight et al. 2003). More recent studies that utilise genome wide data sets generated from various high throughput ChIP based methods further linked genetic variants to haplotype-specific DNA methylation, nucleosome occupancy as assessed by DNase I hypersensitivity, CTCF (a chromatin insulator protein) binding, histone modifications and RNA polymerase II occupancy (Almen et al. 2012; Bell et al. 2010; Maynard et al. 2008; McDaniell et al. 2010; Prendergast et al. 2012; Verlaan et al. 2009).

Computational motif analyses provide informative data that suggest possible mechanisms of haplotype-specific exon 3 inclusion by the five candidate SNPs as their H1 and H2 alleles were predicted to bind different splice factors. We postulate that H1/H2 SNPs may modulate the haplotype-specific inclusion of exon 3 by generating new splice factor binding sites and/or by

affecting the strengths of their corresponding regulatory motifs, two mechanisms that are not mutually exclusive. Previous studies showed poor correlation between computational predictions and actual RNA-protein interactions (Cowper-Sal lari et al. 2012; Goren et al. 2006; Rimoldi et al. 2013). It is therefore important to validate *in vitro* how the five candidate SNPs affect their interactions with splice factors. Nevertheless, *in silico* predictions here showed SNPs sequences can locate within splice factor binding consensus motifs, which could be completely eliminated by a change in a single nucleotide sequence.

CHAPTER 4

Expression of pMAPT Genomic Vectors in SK-N-F1 Neuroblastoma Cells

4.1 Introduction

4.1.1 Exon 3 Splicing Mechanisms

This chapter focuses on studying the effects of five candidates SNPs (rs117124984, rs118087478, rs77875796, rs1800547 and rs17651213) on the haplotype-specific expression of *MAPT* exon 3 using H1/H2 genomic expression vectors. The *MAPT* gene contains sixteen exons and six tau isoforms are expressed in the human central nervous system via the alternative splicing of exons 2, 3 and 10 (Andreadis et al. 1992; Andreadis et al. 1993). Exon 3-containing isoforms are the least expressed tau isoforms in the adult human brain, comprising only approximately 10 % of the total tau expression (Boutajangout et al. 2004; Hong et al. 1998). The most expressed tau isoforms contain exon 2, representing more than 50 % of the total tau expressed (Boutajangout et al. 2004; Hong et al. 1998). Exons 2 and 3 belong to the rare "incremental combinatorial" category of alternative exons where exon 3 never appears independently of exon 2 (Andreadis et al. 1995). This category of splicing has been documented in the amyloid precursor protein gene and in the c-src gene (D. L. Black 1991, 1992; Golde et al. 1990). While most efforts on *MAPT* gene splicing focused on exon 10, a few studies that have examined exon 3 splicing have provided insights about cis-regulatory elements in this much less understood region of *MAPT*. Experiments using mini-gene expression constructs have demonstrated that the low inclusion rate of exon 3 is a result primarily of an intrinsically suboptimal branch point sequence upstream of the 3' splice site, but not due to the competition from exon 2 (Andreadis et al. 1995). Mutagenesis of exon 3 sequence in mini-gene

constructs unveiled the presence of two exonic splice silencer and two exonic splice silencer sequences (Arikan et al. 2002).

Diseases that associate with splice factor misregulation and altered *MAPT* splicing have shed light on the trans-acting splice factors involved in exon 3 inclusion. Myotonic dystrophy type I (DM1) is caused by an expanded CTG repeat at the 3' untranslated region of the myotonic dystrophy protein kinase gene (Brook et al. 1992). The expression of the expanded CTG repeats leads to the sequestration of muscleblind-like factors (MBNL) in RNA foci (Fardaei et al. 2001). MBNL and CUG-BP and ETR-3-like factors (CELF) are splicing factors that have antagonistic actions and their expressions are both affected in DM1 brains (Ho et al. 2004; O. Leroy et al. 2006). Exons 2 and 3 were found to be preferentially excluded in post-mortem brains of DM1 patients (H. Jiang et al. 2004; Sergeant et al. 2001; Vermersch et al. 1996), MBNL and CELF are thus implicated in regulating the splicing of *MAPT* exons 2 and 3. Indeed, experiments expressing different splice factors in T98 cells identified ETR-3, a member the CELF family, to promote the skipping of exons 2 and 3 (O. Leroy et al. 2006). Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) represent a clinicopathological spectrum of neurological disorders that are clinically, pathologically and genetically heterogeneous. A subtype of FTD/ALS cases are characterised by the abnormal aggregation of the splice factor fused in sarcoma (FUS) (Hallier et al. 1998; Josephs et al. 2011; Kwiatkowski et al. 2009; Vance et al. 2009). The loss of nuclear localisation of FUS has been linked to the pathogenesis of FTD/ALS (Dormann et al. 2010; Kwiatkowski et al. 2009; Mackenzie et al. 2010; Vance et al. 2009). FUS was found to bind *MAPT* transcripts both in human brains and in rat hippocampal neurons (Nakaya et al. 2013; Orozco et al. 2012). Interestingly, the knockdown of FUS in rat hippocampal neurons led to an increase in exon 3-containing transcript expression, suggesting FUS regulates the exclusion of exon 3 (Orozco et al. 2012).

4.1.2 Aims of Chapter 4

This chapter aims to identify the SNP(s) responsible for the observed two-fold increase in exon 3 expression by the H2 allele compared to the H1, the effects of five candidate SNPs (rs117124984,

rs118087478, rs77875796, rs1800547 and rs17651213) on exon 3 inclusion were investigated by expressing H1/H2 PAC *MAPT* haplotype-hybrid genomic expression vectors in SK-N-F1 neuroblastoma cells. A qPCR assay for quantifying the expression of exon 3-containing transcripts from the transgene was developed.

4.2 Results

4.2.1 Designing, Optimising and Validating a qPCR Assay for Studying pMAPT Vectors Expression

To study *MAPT* exon 3 expression in SK-N-F1 neuroblastoma cells using our pMAPT genomic vectors, a relative standard curve method was employed to determine the levels of transgenic exon 3-containing transcripts normalising to total transgenic *MAPT* transcripts. The transcript levels were determined by the PCR amplification threshold cycle (Ct) values relative to a standard dilution series. In order to perform qPCR measurements using SYBR Green reagents, a qPCR assay using the HA tag present in exon 5 (Figure 17) that will allow us to differentiate between transgenic and endogenous *MAPT* transcripts was first designed. The reverse primers span the HA tag and exon 5 (constitutive exon) sequence, with forward primers locating in exon 3 (alternative exon) and exon 4 (constitutive exon) to measure exon 3 and total *MAPT* expression, respectively (Figure 17).

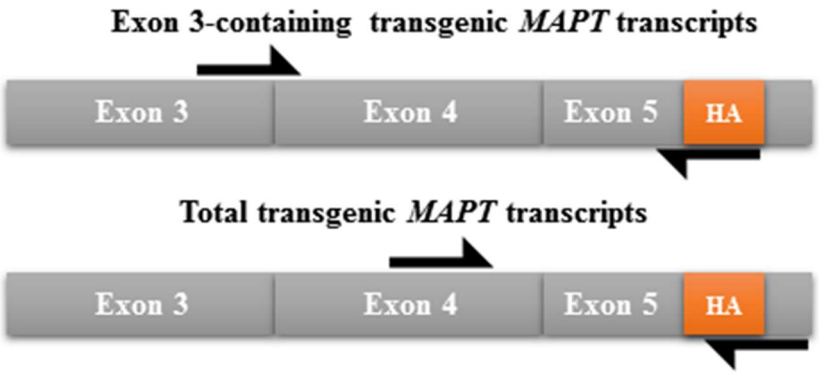


Figure 17. qPCR assay design for measuring transgenic *MAPT* expression.
pMAPT vectors carry an in-frame haemagglutinin (HA) tag sequence in the constitutive exon 5, allowing for transgenic transcripts to be differentiated from endogenous *MAPT* transcripts. Reverse primers were designed to span the HA tag sequence in exon 5. Forward primers locate in exon 3 and exon 4 (constitutive exon) for amplifying exon 3-containing transcripts and total *MAPT* transcripts, respectively.

To ensure the assay's validity, several selection criteria were used for primer pair selection including primer specificity, their capability to differentiate between endogenous and transgenic transcript, and primer efficiency. Specificities of primer pairs were determined by melting curves from the amplification products of qPCRs using cDNA from SH-SY5Y cells expressing pMAPT-H1WT and –H2WT vectors. A single peak observed in the melting curve, which correspond to the release in fluorescence detected by the qPCR machine as temperature increases, indicates the dissociation of a single species of DNA duplex in the PCR product (Figure 18).

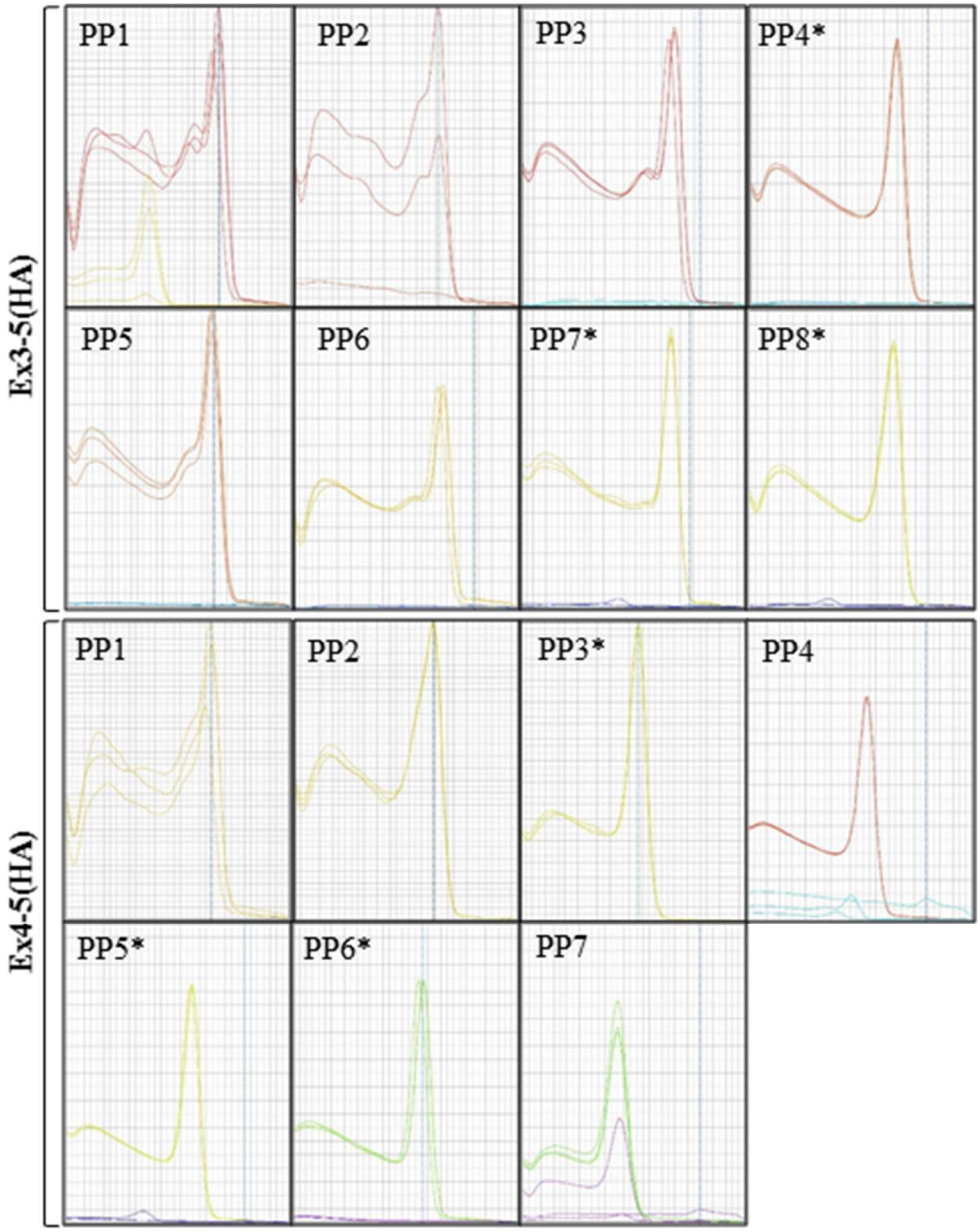


Figure 18. Primer pair specificities assessed by melting curve.

qPCR performed using cDNA from SH-SY5Y cells expressing pMAPT-H1WT and –H2 with primers for amplifying exon 3-containing (Ex3-5(HA)) and total (Ex4-5(HA)) transgenic *MAPT* transcripts. Primer pairs (PP) with a single melting curve peak (starred) indicate specific PCR product was produced in the amplification reaction.

Primer pairs that produce specific PCR products were then analysed for their ability to differentiate between endogenous transcripts and transgenic transcripts carrying the HA tag by qPCRs using cDNA from untransfected BE(2)-M17, SK-N-MC neuroblastoma, and HEK293 cells. For example, primers that are able to differentiate between endogenous and transgenic transcripts should not exhibit any amplification as assessed by the amplification plots of the qPCRs (Figure 19).

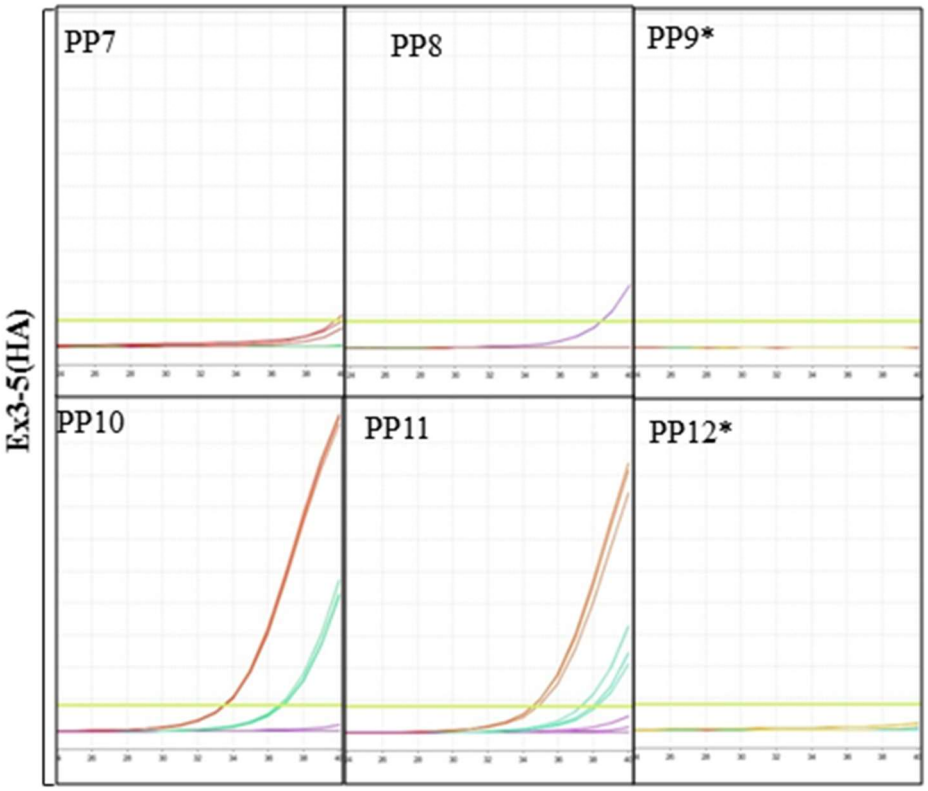


Figure 19. Amplification plots of qPCRs using cDNA from untransfected cells.
qPCRs using cDNA from untransfected BE(2)-M17, SK-N-MC and HEK293 cells as template. Primer pairs (PP) that do not exhibit any amplification (starred) indicate that they can differentiate between endogenous and transgenic *MAPT* transcripts. qPCR cycles 24 – 40 (end of cycling programme) are shown.

Our preliminary data indicated low gene delivery efficiency by transfection reagents of the pMAPT vectors in cultured cells due to the large vector size of approximately 160 kb (Figure 20). The expression of transgenic exon 3-containing transcripts were also found to be very low, which could in part due to exon 3-containing isoforms being the least expressed *MAPT* isoforms (Boutajangout et al. 2004; Hong et al. 1998) and in part due to the low gene delivery efficiency, rendering detecting exon 3 transcripts by qPCR difficult. The low gene delivery efficiency and the intrinsically low inclusion rate of *MAPT* exon 3 could both contribute to difficulties in measuring transgenic exon 3-containing transcripts.

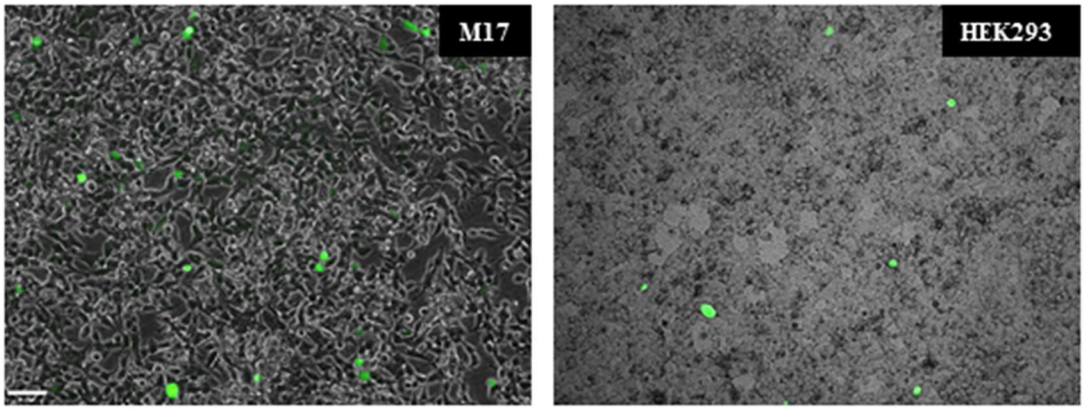


Figure 20. Low transfection efficiency in BE(2)-M17 and HEK293 cells by a large PAC vector. BE(2)-M17 and HEK293 were transfected with a vector of approximately 160 kb containing a *MAPT* gene and an EGFP gene. The vectors were expressed for 48 hours before the images were taken. Green cells represent transfected cells. The transfection efficiencies for both cell lines using the large vector were < 1%. Error bar: 35 μm.

It may therefore be advantageous to determine the optimal primer concentration and ratio of forward to reverse primers in order to increase the chances of detecting transgenic exon 3 transcripts. For example, when detecting transgenic exon 3 expression, the forward primer will have a much higher chance to binding to the template due to the presence of both endogenous and transgenic *MAPT* transcripts compared to the reverse primer which interacts with the HA-tag sequence. Using primer pairs amplifying transgenic exon 3-containing or total *MAPT* transcripts, the general effects of primer concentrations and different forward to reverse primer ratios on qPCR

signals were investigated. A range of forward to reverse primer ratios and concentrations were tested: 18:1, 18:6, 18:18, 6:1, 6:6, 6:18, 1:1, 1:6 and 1:18 (Figure 21). The intensity of the fluorescent signal was assessed using the melting curves of the PCR products and are compared to the signal from a reactions using water as template. In general, the higher the primer concentrations, the higher the fluorescent signals detected (Figure 21). However, increasing primer concentrations also increase primer dimer formation in water controls as observed as a peak signal at a lower temperature (Figure 21). The optimal primer concentration and primer ratio is therefore a compromise between assay sensitivity and amplification specificity. When selecting candidate primer pairs, the effects of varying forward and reverse primer concentrations were also taken into account in order to identify an optimal condition for measuring exon 3-containing transcripts.

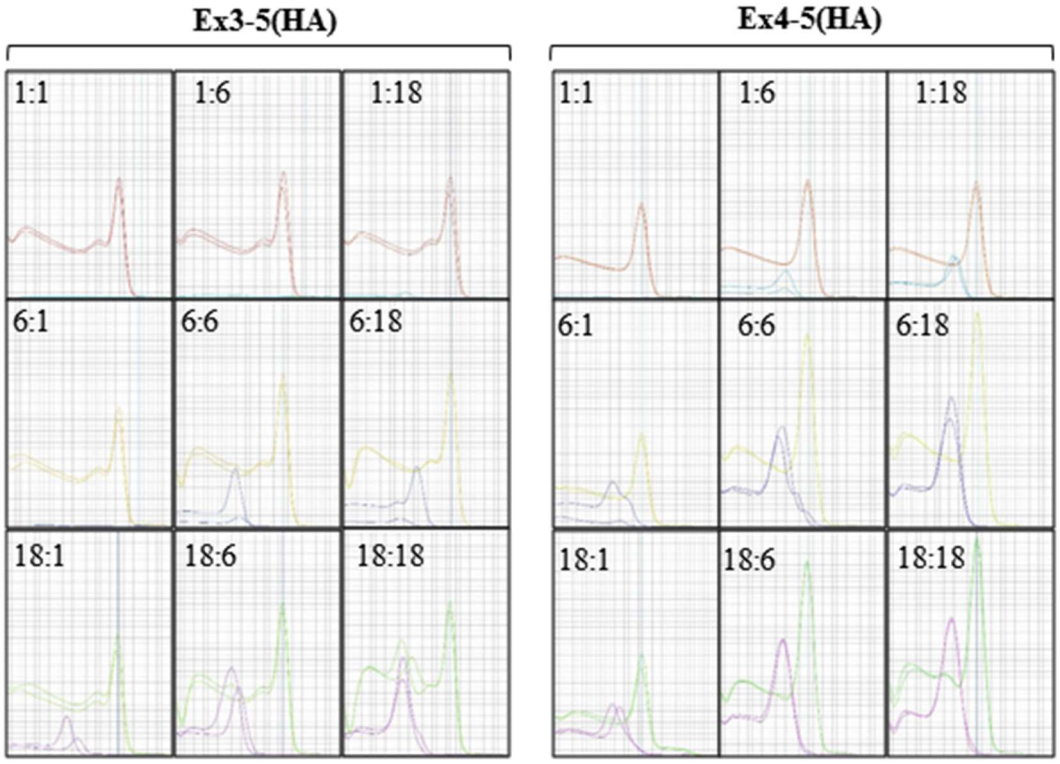


Figure 21. Determining the effects of varying forward to reverse primer ratios.

qPCRs were performed using different ratios of forward to reverse primers and their effects on amplification were assessed through the melting curves of the PCR products. In general, increasing primer concentrations provide higher assay sensitivity as reflected by increased fluorescent signals, but also increases the chances of non-specific amplification in water controls as indicated by a smaller peak to the left. In the two pairs of primers tested, a 1:6 concentration showed the best result.

Two candidate primer pairs (Table 4) were ultimately identified after the above selection process, one for measuring exon 3-containing transcripts and one for total *MAPT* expression, which satisfy with all the above selection criteria. Primer specificities and efficiencies were assessed in order to validate the primers. Both primer pairs produce specific PCR products as shown in their melting curves with correct amplicon sizes as verified by gel electrophoresis (Figure 22A). Primer pair efficiency (E) was determined from the slope of the standard curve by the equation $E = (10^{-1/slope}-1) \times 100$ (Rutledge and Cote 2003). High primer efficiencies of near 100 % were consistently achieved for both primer pairs (Figure 22B). For our assay, a forward to reverse primer ratio of 1:4 and 1:1 for measuring exon 3-containing and total *MAPT* transcripts, respectively, was found to provide the best sensitivity, consistency and specificity in qPCR triplicates, particularly at low cDNA concentrations when producing the standard curves. The primer pairs also do not detect endogenous *MAPT* transcripts. In summary, a qPCR assay to measure exon 3 expression using the pMAPT genomic haplotype-hybrid vectors in SK-N-F1 cells has been designed, optimised and validated.

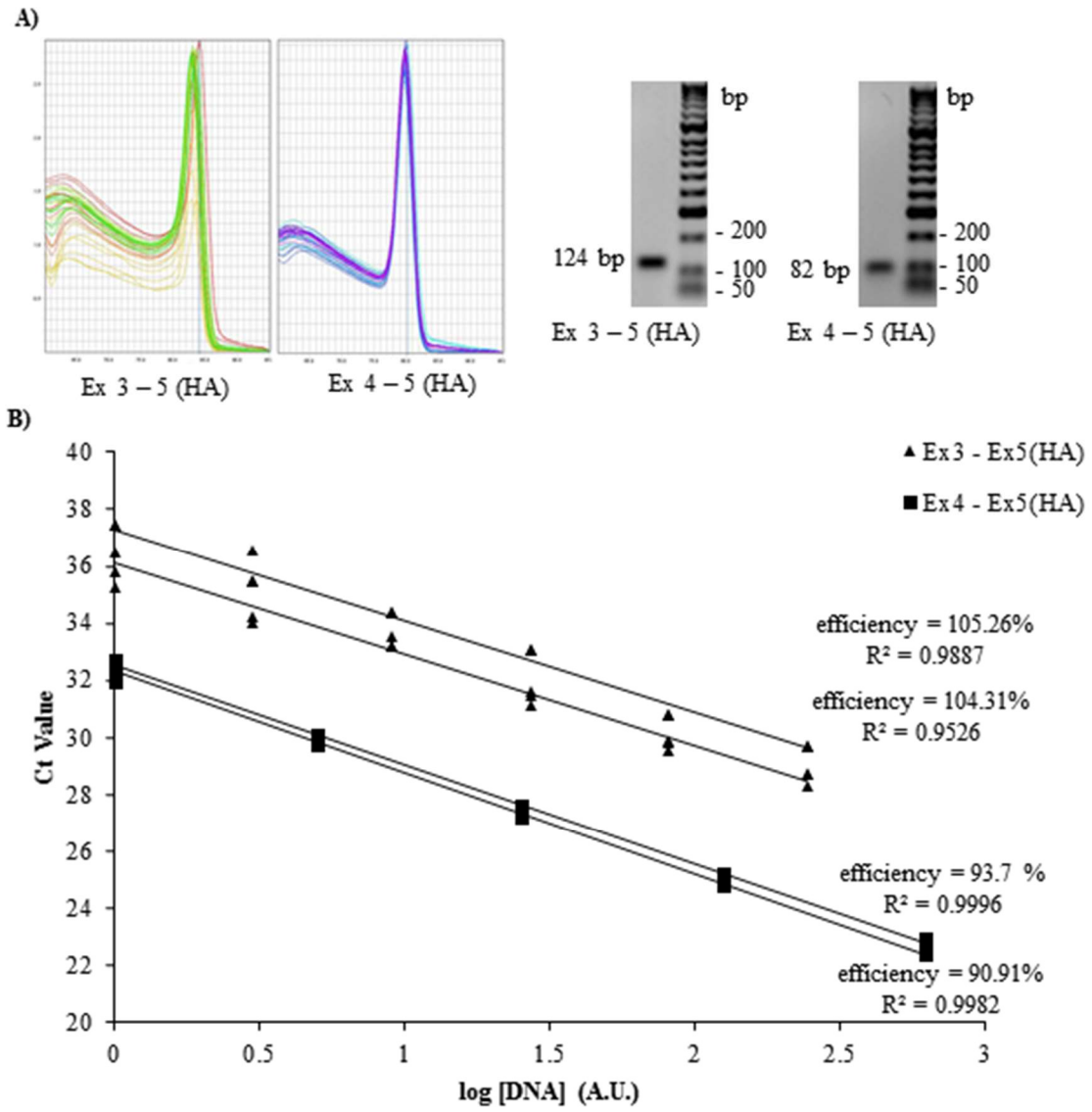


Figure 22. qPCR assay validations for selected candidate primer pairs.

(A) Primer specificities determined by a single peak observed in the melting curves of the qPCR products. A single peak indicates a specific product is being produced. Correct amplicon sizes were verified by agarose gel electrophoresis. **(B)** Primer efficiencies determined by cDNA dilution series standard curves for the two candidate primer pairs. Both primer pairs showed high efficiencies of near 100%.

4.2.2 Delivery of pMAPT Genomic Vectors by HSV-1 Amplicons

Conventional chemical methods for introducing the large vectors into cultured cell lines are inefficient (Figure 20). This section of the project therefore explores the use of a helper virus-free system for producing herpes simplex virus-type I (HSV-1) amplicons (Saeki et al. 2001) (Figure 23A) as a means of delivering the complete *MAPT* locus into cultured cells. The helper virus-free HSV-1 system (Saeki et al. 2001) has several advantages for gene delivery: a high transgene capacity (~150 kb) (Saeki et al. 2001); a broad range of cell tropism (Sena-Esteves et al. 2000; Wade-Martins et al. 2003); a limited cytotoxicity and immunogenicity due to a lack of viral coding sequences (Fraefel et al. 1996; Saeki et al. 2001); and a high amplicon titre production (up to 10^9 transduction unit (TU)/ mL) without the generation of replication-competent helper virus (Saeki et al. 2001; Saeki et al. 2003).

The production of HSV-1 amplicons involves a two-step protocol for amplicon production *in vivo* and amplicon concentration *in vitro* (Lufino et al. 2011). Briefly, HSV-1 amplicons were produced by co-transfecting three plasmids into Vero 2-2 cells: pMAPT vector; fHSVΔpacΔ27 0++; and pEBHICP27 (Figure 23A). The pH-FRT-Hy plasmid, previously retrofitted with pMAPT vectors, contain the HSV-1 origin of replication (*oriS*), a *lacZ* gene under the control of the HSV-1 immediate-early (IE) 4/5 (HSV-1 IE 4/5) promoter, and the HSV-1 packaging signal (Figure 23B) (Lufino et al. 2013). The fHSVΔpacΔ27 0++ (~178 kb) is an oversized plasmid that cannot be packaged into HSV-1 amplicons to prevent the production of replication competent virus. It contains the HSV-1 genome but lacks the packaging signal and an essential viral gene ICP27 as additional safety features (Figure 23A) (Lufino et al. 2011; Saeki et al. 2001). The essential viral gene ICP27 is supplied in trans by pEBHICP27 and by the Vero 2-2 cells (Figure 23A) (Lufino et al. 2011; Saeki et al. 2001). Amplicons were harvested following their production in Vero 2-2 cells, and concentrated by ultracentrifugation. The *lacZ* gene was used for as amplicon titration (Figure 23B) (Lufino et al. 2011) and we consistently obtained amplicons with titres between $3 \times 10^6 – 1.5 \times 10^7$ TU/ mL.

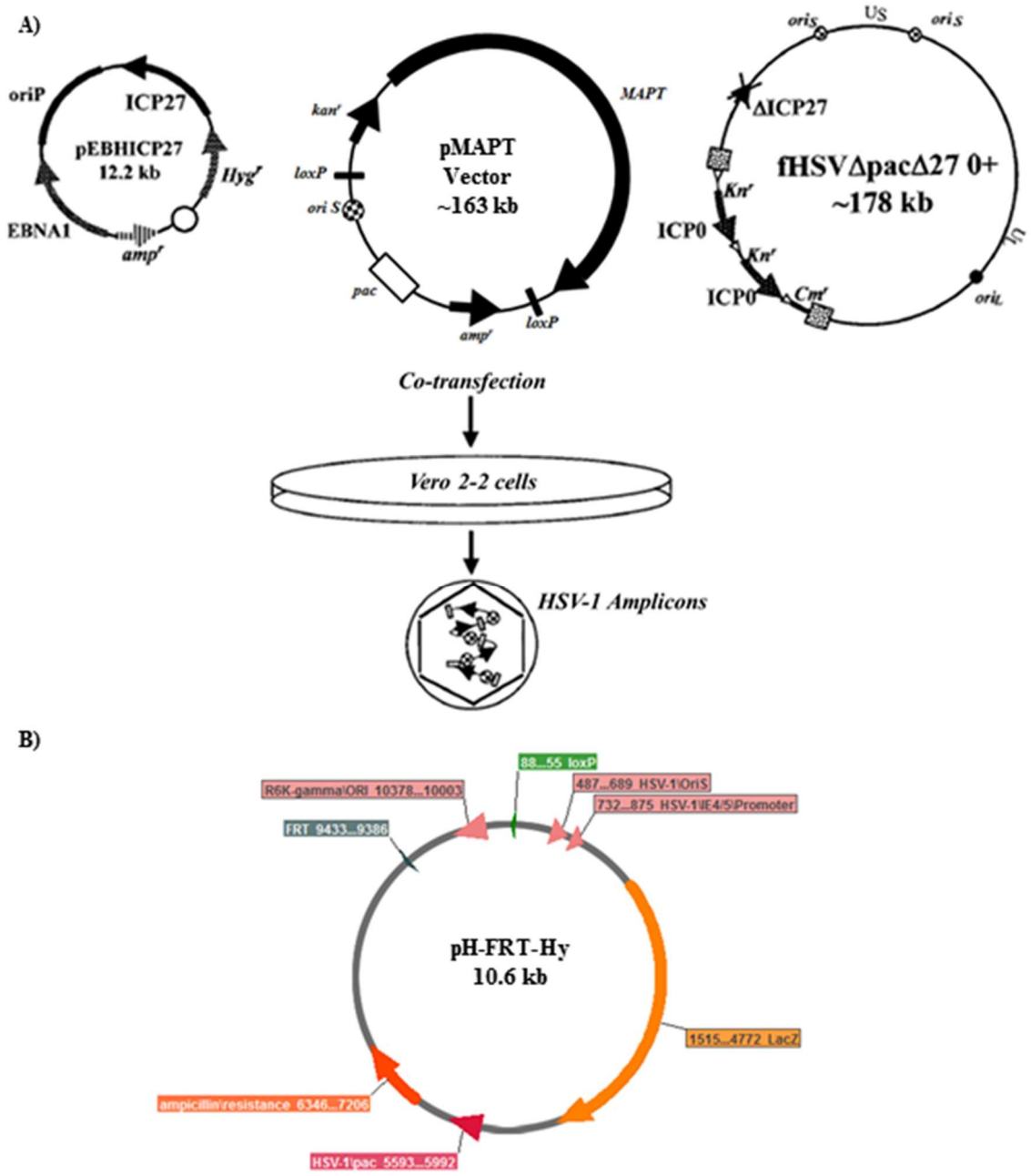


Figure 23. HSV-1 amplicon production for infectious delivery of pMAPT vectors.

(A) HSV-1 elements present on the pH-FRT-Hy plasmid. The incorporation of the pH-FRT-Hy plasmid into pMAPT vectors allows for the pMAPT vectors to be packaged as HSV-1 amplicons and subsequent infectious gene delivery by HSV-1 particles. (B) The helper virus-free packaging system. The packaging of amplicons involves the co-transfection of the pMAPT vector, the HSV-1 genome (fHSVΔpacΔ27 0++) and an HSV-1 essential gene ICP27 supplied in trans by pEBHICP27 in Vero 2-2 cells. Figure modified from (Saeki et al. 2003).

The amplicons produced were applied to deliver pMAPT gene into 4 different neuronal cell lines SH-SY5Y, BE(2)-M17, SK-N-MC and SK-N-F1. The *lacZ* gene was used for assessing transduction efficiencies and high transduction efficiencies can be achieved in all the tested cell lines (Figure 24). In general, an > 90 % transduction efficiency was consistently achieved with a virus to cell ratio, or multiplicity of infection (MOI), of 1 (Figure 24). Transduction efficiencies of ~ 30% – 50% were consistently achieved with a virus to cell ratio of 0.5 (Figure 24). HSV-1 amplicons thus provide an efficient and robust way of delivery large genomic vectors into different neuronal cell culture models for gene expression studies.

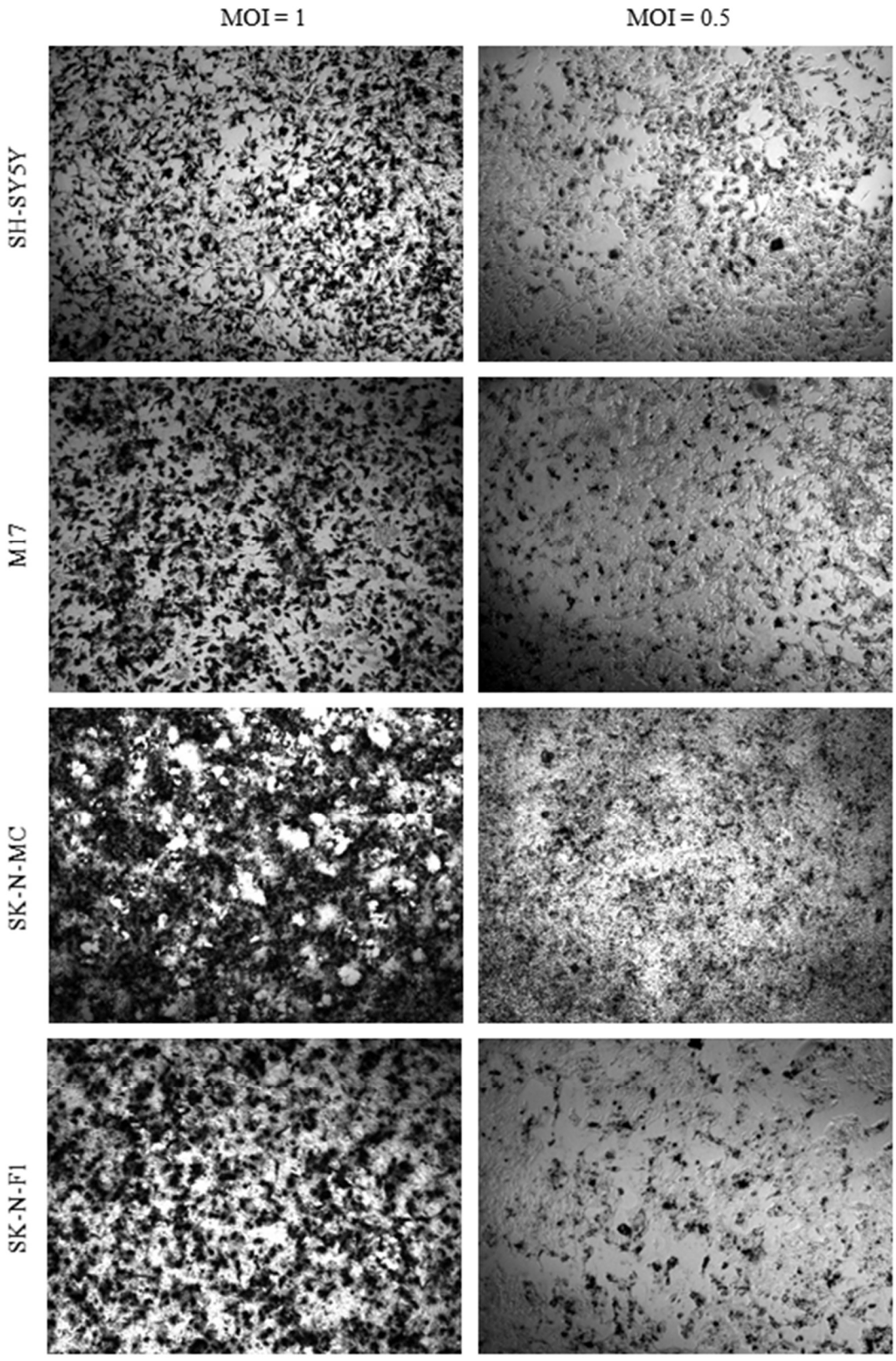


Figure 24. Infectious delivery of pMAPT vectors in cultured neuroblastoma cells.

X-gal staining for lacZ expression (cells stained in black) for assessing transduction efficiency by HSV-1 amplicons. In general, > 90 % efficiency was achieved by an MOI (multiplicity of infection) of 1. About 30 – 50 % efficiency was achieved by an MOI of 0.5.

To measure transgenic *MAPT* expression, 4 cell lines (SH-SY5Y, BE(2)-M17, SK-N-MC and SK-N-F1) were transduced using HSV-1 amplicons carrying either pMAPT-H1 or pMAPT-H2 vectors. Using the qPCR assay described above (Figure 17), the expression levels of transgenic exon 3 and total *MAPT* expression in the transduced cells were measured. In general, the higher the Ct (threshold cycle) value measured by qPCR, the lower the transcript quantity (Higuchi et al. 1993). Using an MOI of 1, high transduction efficiencies were achieved and transgenic *MAPT* was successfully expressed in all 4 cell lines (Figure 25A). It should be of note that compared to qPCR measurements, agarose gel electrophoresis is not a quantitative method for measuring the relative levels of transcripts expressed since the visibility of the PCR product band is dependent on the amount of DNA, loading dye and DNA ladder loaded and the exposure time of the image taken. Indeed, despite the high transduction efficiencies, expression of total transgenic *MAPT* transcripts from pMAPT vectors delivered by HSV-1 amplicons were in fact very low as reflected in their high measured Ct values (> 30) in qPCR. The expression of exon 3-containing transcripts were not detectable by qPCR due to the low total transgene expression levels. To ensure that the cell lines express *MAPT* under both transduced and untransduced conditions, endogenous *MAPT* expression was measured by RT-PCR (Figure 25B). Since the *MAPT* expression levels were not sufficient for studying exon 3 splicing, the use of different transfection reagents to deliver pMAPT vectors was next investigated.

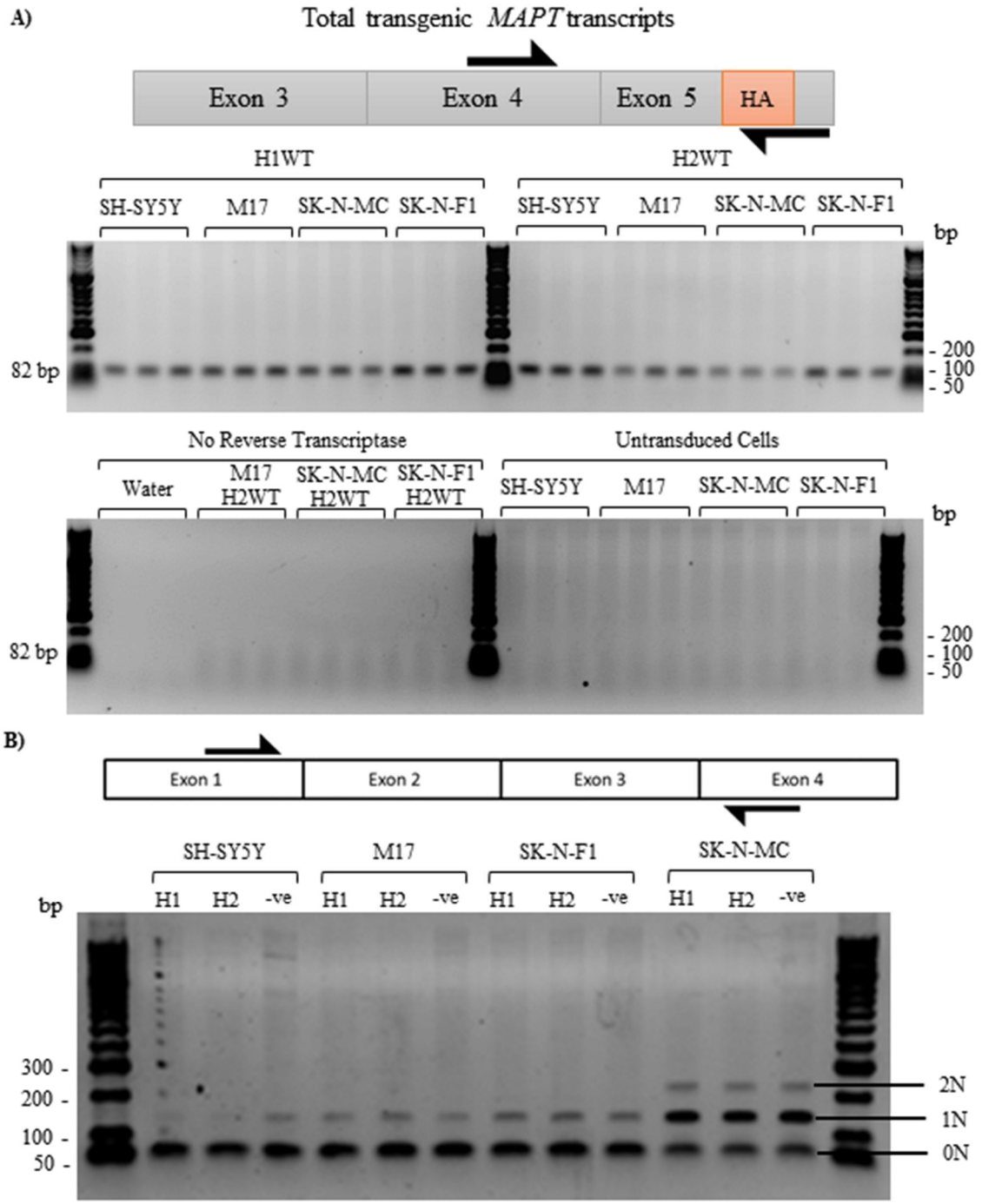


Figure 25. Expression of pMAPT-H1WT and -H2WT vectors in HSV-1 transduced cells.

(A) qPCR performed to measure total transgenic *MAPT* expression in neuroblastoma cells transduced with pMAPT-H1WT and –H2WT vectors. qPCRs were performed in triplicate and amplification products (82 bp) were analysed by agarose gel electrophoresis. **(B)** RT-PCR performed to confirming the expression of endogenous *MAPT* transcripts in the four neuroblastoma cell lines. Primers locate in exons 1 and 4. The three N-terminal *MAPT* isoforms 2N (exon 2+3+), 1N (exon 2+3-) and 0N (exon 2-3-) are detected simultaneously.

4.2.3 Delivery of pMAPT Genomic Expression Vectors by Transfection Reagents

Three commercially available transfection reagents (Lipofectamine, Lipofectamine LTX and X-tremeGene HP) were tested for delivering pMAPT vectors into SK-N-MC and SK-N-F1 cells. SK-N-MC and SK-N-F1 were previously genotyped by our laboratory to be heterozygous for H1 and H2 *MAPT* and were therefore chosen to be used in our experiments for their relevant genotypic background. Transfection reactions were optimised by varying the ratios of transfection reagent to DNA (1:1, 1:2, 1:2.5, 1:3, 1:3.5, 1:4 and 1:5) and by varying the amounts of DNA used ranging from 1 to 5 µg. *lacZ* expression was again used as a reporter to assess transfection efficiencies. pMAPT vectors were delivered successfully into SK-N-F1 cells using the three transfection reagents (Figure 26). In general, the optimal transfection efficiency was less than 1% using the three transfection reagents (Figure 26). Since the transfection efficiencies were similar, Lipofectamine LTX was chosen to be used in our experiments due to the availability of the reagent in our laboratory. Transfections of pMAPT vectors into SK-N-MC were unsuccessful. pMAPT vectors were successfully introduced into SK-N-F1 cells using commercially available transfection reagents. SK-N-F1 cells will therefore be used in our following experiments to study transgenic *MAPT* exon 3 expression.

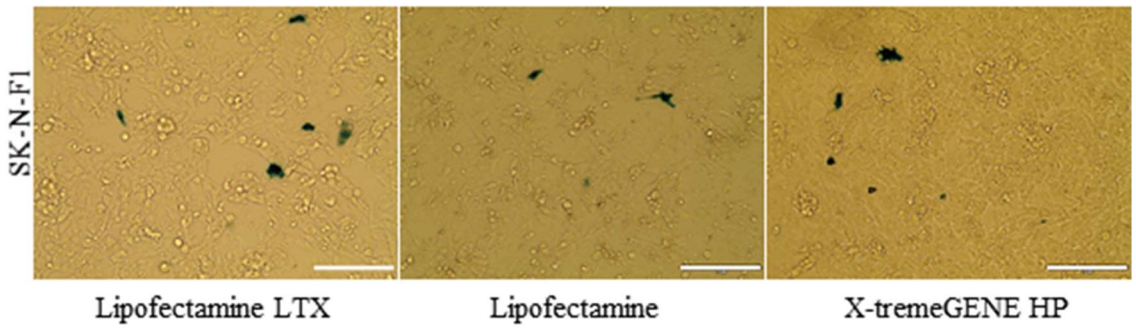


Figure 26. pMAPT vectors delivery using transfection reagents.

pMAPT vectors delivered into SK-N-F1 cells using three different commercially available transfection reagents. lacZ expression was used for assessing transfected cells (stained in blue) by x-gal staining. The transfection efficiencies using different transfection reagents were all < 1 %. Bar indicates 200 µm.

4.2.4 pMAPT–H1WT and –H2WT Vectors Recapitulate Human Physiological Exon 3 Expression

To study the effects of rs117124984, rs118087478, rs77875796, rs1800547 and rs17651213 on *MAPT* exon 3 expression, pMAPT H1 and H2 hybrid vectors (Figure 27) were expressed in SK-N-F1 cells. Briefly, SK-N-F1 cells were allowed to express the pMAPT H1 and H2 hybrid vectors for 48 hours before cells were harvested. qPCR (Figure 17) was performed to measure the amount of transgenic *MAPT* transcripts. Transgenic *MAPT* exon 3 expression is expressed as the amount of transgenic exon 3-containing transcripts relative to that of total transgenic *MAPT* using the relative standard curve method (K. Livak 1997 & 2001).

To determine in our SK-N-F1 cell culture model if the pMAPT-H1WT and –H2WT vectors recapitulate human physiological expression where the H2 allele expresses two-fold higher exon 3-containing transcripts when compared to H1 (Caffrey et al. 2008), the expression exon 3-containing transcripts from our H1 and H2 wild-type vectors in SK-N-F1 neuroblastoma cells was first assessed (Figure 27). Exon 3 expression from the H2 wild-type vector (H2WT 0.81 ± 0.16) was found to be 1.8 fold higher when compared to the H1 wild-type vector (H1WT 0.46 ± 0.12) (Figure 27), indicating that our genomic haplotype pMAPT wild-type vectors are representative of the physiological *MAPT* exon 3 expression (Caffrey et al. 2008; Trabzuni et al. 2012) in our SK-N-F1 cell culture model.

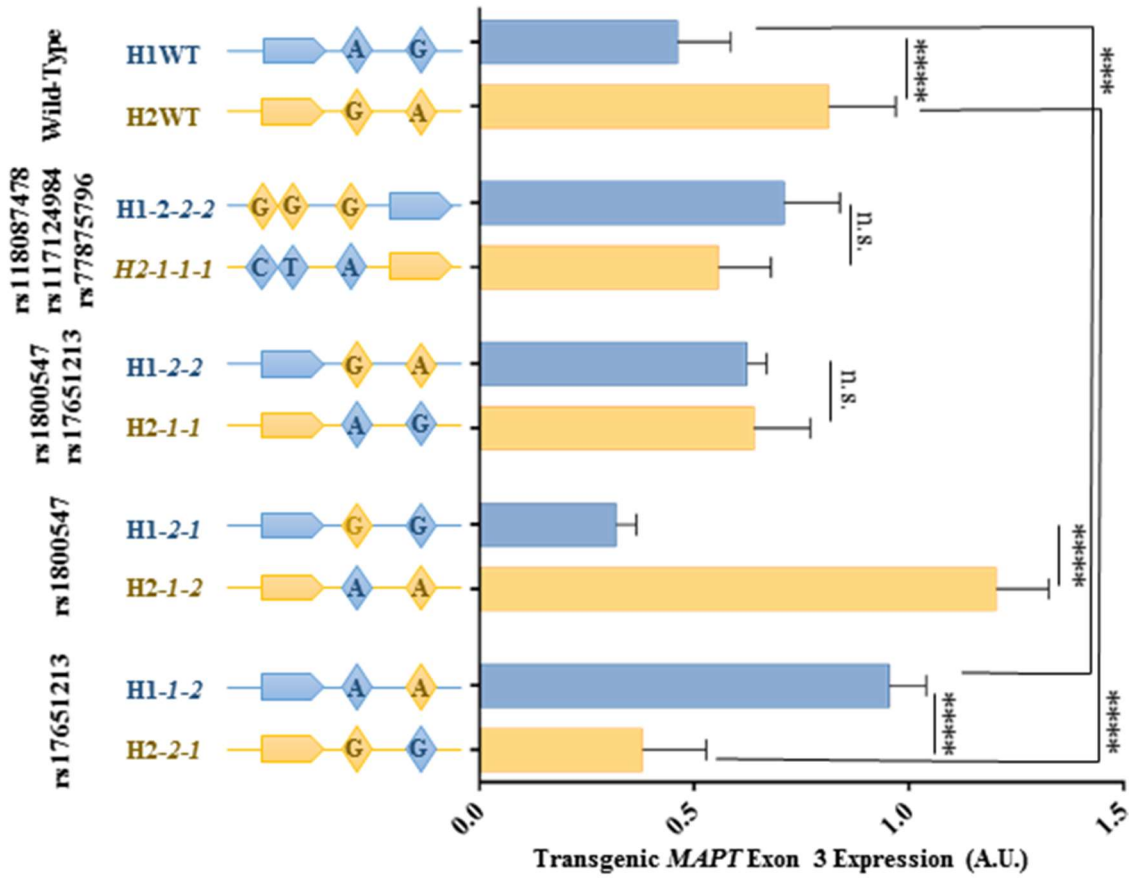


Figure 27. rs17651213 regulates the haplotype-specific inclusion of *MAPT* exon 3.

Left panel: a schematic representation illustrating the combinations of SNPs exchanged between the H1 and H2 pMAPT vectors. Pentagons indicate exon 3. Right panel: the corresponding pMAPT exon 3 expression in SK-N-F1 neuroblastoma cells as measured by qPCR. Exon 3 expression from the H2WT (0.81 ± 0.16, n=8) allele was 1.8 fold higher when compared to the H1WT (0.46 ± 0.12, n=8) allele. No difference in exon 3 expression was observed between the H1-*2-2-2* (0.71 ± 0.13, n=4) and H2-*1-1-1* (0.56 ± 0.12, n=4), and H1-*2-2* (0.62 ± 0.05, n=4) and H2-*1-1* (0.64 ± 0.13, n=4) vectors. Exon 3 expression was 3.8 fold higher from the H2-*1-2* (1.20 ± 0.12, n=4) vector when compared to the H1-*2-1* (0.32 ± 0.05, n=4) vector. Exon 3 expression was 2.5 fold higher from the H1-*1-2* (0.95 ± 0.09, n=4) vector when compared to the H2-*2-1* (0.38 ± 0.15, n=4) vector. Statistical significance was determined using one-way ANOVA followed by Bonferroni correction (*** $p<0.001$, **** $p<0.0001$). Error bars represent standard deviation.

4.2.5 Candidate SNPs Near the 5' and 3' Splice Sites are Functional and Modulate Exon 3 Inclusion

In order to determine if the candidate haplotype-specific SNPs near the 5' and 3' splice sites of exon 3 are functional, haplotype-hybrid vectors were expressed in the SK-N-F1 cell culture model (Figure 27). The H1 and H2 vector pair with SNPs rs117124984, rs118087478 and rs77875796 exchanged between the haplotype backgrounds were first expressed. It was found that exon 3 transcripts are now expressed to the same level between the pair of vectors (H1-*2-2-2* 0.71 ± 0.13 and H2-*1-1-1* 0.56 ± 0.12), suggesting the three SNPs near the exon 3 3' splice site overall are functional and regulate exon 3 expression (Figure 27).

Next, the vector pair H1-*2-2* and H2-*1-1* with both SNPs rs1800547 and rs17651213 exchanged between the haplotype backgrounds were expressed and again it was found that exon 3 is now expressed to the same level between the pair of vectors (H1-*2-2* 0.62 ± 0.05 and H2-*1-1* 0.64 ± 0.13). The data suggests that rs1800547 and rs17651213 near the exon 3 5' splice site overall are functional and regulate exon 3 expression (Figure 27).

4.2.6 rs17651213 Regulates the Haplotype-Specific Inclusion of *MAPT* Exon 3

SNPs rs1800547 and rs17651213 locating within 100 bp from the exon 3 5' splice site were next focused on as they are closest to the splice site and may have the strongest effects on splicing. To study the effects of the SNPs rs1800547 and rs17651213 on exon 3 expression individually, the vector pair pMAPT-H1-2-1 and pMAPT-H2-1-2 with only rs1800547 exchanged between the haplotype backgrounds was first expressed in SK-N-F1 cells. SNP rs1800547 situates 9 bp downstream of the exon 3 5' splice site and is expected to have a strong effect on affecting the inclusion of exon 3 due to is close proximity. Indeed, exchanging rs1800547 led to a 3.8 fold greater expression of exon 3 from the H2-*1-2* vector when compared to the expression of H1-*2-1* vector (H2-*1-2* 1.2 ± 0.12 and H1-*2-1* 0.32 ± 0.05; $p < 0.0001$), further exacerbating the inclusion pattern of exon 3 from the H2 allele when compared to the wild-type expression (Figure 27). Lastly, expressing the vector pair with only rs17651213 exchanged between the haplotype backgrounds

resulted in a 2.5 fold higher exon 3 expression from the H1-*1-2* vector compared to the expression from H2-*2-1* vector (H1-*1-2* 0.95 ± 0.09 and H2-*2-1* 0.38 ± 0.15; $p < 0.0001$) (Figure 27).

All together, the expressions of exon 3-containing transcripts from different genomic-hybrid pMAPT vectors in SK-N-F1 neuroblastoma cells indicate that rs1800547 and rs17651213 are both functional SNPs that regulate exon 3 expression. The reversal of the wild-type exon 3 expression profiles by vectors H1-*1-2* and H2-*2-1* (Figure 27) strongly suggests that rs17651213 confers the H1/H2 haplotype-specific expression of *MAPT* exon 3, we therefore further investigated the *trans*-acting splicing factors that interact with rs17651213 SNP sequences.

4.3 Discussion

4.3.1 rs1800547 and rs17651213 Regulate the Allele-Specific Splicing of Exon 3

This chapter aims to identify the SNP(s) responsible for the observed two-fold increase in exon 3 expression by the H2 allele compared to the H1 using genome expression vectors. A functional variant rs17651213 was identified to drive the two-fold increase in exon 3 expression by the H2 allele compared to H1 that was previously observed in both cultured neuronal cells and in post-mortem brain tissue (Caffrey et al. 2008). Multiple GWAS and subsequent meta-analyses repeatedly identify the association between H1 and H2 haplotype-specific polymorphisms with risk and protection, respectively, in PD, PSP and CBD (Hoglinger et al. 2011; Kouri et al. 2015; Nalls et al. 2014). Identifying H1/H2 polymorphisms that lead to splicing changes and dissecting the underlying mechanisms are therefore of great interest to better understand the possible mechanisms of disease association by the haplotypes. By expressing different haplotype-hybrid *MAPT* genomic locus vectors in SK-N-F1 neuroblastoma cells, a functional intronic variant rs17651213 which, when exchanged between the H1 and H2 haplotypes, was identified to regulate the H1/H2 haplotype-specific alternative splicing of exon 3. It was also shown here that our H1 and H2 wild-type genomic vectors recapitulate the physiological two-fold higher expression of exon 3-containing transcripts by the H2 allele compared to the H1, demonstrating the biological relevance of capturing the sequence diversity in splicing studies. Additionally, our pMAPT H1 vector carries the rs242557-G allele which is a marker for the H1c *MAPT* sub-haplotype that is highly associated with CBD and PSP (Pittman et al. 2005; Rademakers et al. 2005), therefore carries variants in LD with this highly relevant marker.

4.3.2 Gene Delivery by HSV-1

Attempts were made here to deliver the *MAPT* PAC vectors using the helper-virus free HSV-1 amplicon system in order to overcome the low transfection efficiencies of large PAC vectors in cell lines by conventional chemical methods (Lufino et al. 2011; Saeki et al. 2001). Despite the high transgene delivery efficiency of over 90 % achieved, the gene expression levels of the transgene

remained lower than transfected cells with less than 1 % transfection efficiency. Several possible mechanisms could lead to the unexpectedly low levels of transgenic *MAPT* expression in the transduced cells as measured by qPCR. HSV-1 is known to have various cytopathic effects upon the infection of cells (Aubert and Blaho 1999; Fraefel et al. 1996). Further purification of the HSV-1 particles produced could be important in reducing the possible toxic effects.

The theoretical packaging capacity of HSV-1 is the size of the HSV-1 genome (approximately 152 kb). During HSV-1 particles production, the amplicon vectors carrying the replication of origin are replicated as head-to-tail concatemers at the cleavage/packaging signal sequence and are packaged into HSV-1 viral particles as linear DNA molecules of up to approximately 150 kb (Saeki et al. 2003). The linear amplicon DNA is recircularised upon entry into cell nucleus after cells are being transduced with the viral particles (Wade-Martins et al. 2001). The recircularised vector can be shuttled between human cells and bacteria, allowing for the rate of intact and unrearranged recircularization efficiency to be assessed using a plasmid rescue assay (Lufino et al. 2007; Wade-Martins et al. 2001). Larger vectors of 156 kb and 157 kb have been shown they can be packaged efficiently, and can be delivered intact with the efficiencies of 83 % and 67 % respectively (Lufino et al. 2007; Peruzzi et al. 2009). However, the reduced intact delivery efficiency with increasing vector size would imply that the 163 kb *MAPT* vector may simply have exceeded the capacity of the HSV-1 particle.

The successful expression of the lacZ gene which locates near the cleavage/packaging signal and the origin of replication, and the low expression of transgenic *MAPT* in the transduced cells, indicate that the amplicon vector may be recircularised but with rearrangement of sequences likely in the 143 kb *MAPT* gene which is not essential to cleavage/packaging of the vector. A reduction in the amplicon size may be necessary for increasing the efficiencies of packaging and intact delivery of the vectors. However, since chemical reagents for transfections are available as an alternative method for introducing our PAC vectors into cultured cells, determining the mechanisms of

unsuccessful transgene expression in HSV-1 transduced cells would have deviated from our main goal of this project in understanding exon 3 splicing, and was therefore not pursued in this study.

4.3.3 rs1800547, rs17651213 and Splicing Regulatory Elements

The functional intronic SNPs rs17651213 and rs1800547 locate 9 bp and 86 bp downstream of the 5' splice site of exon 3. Since both H1 and H2 alleles of the two SNPs regulate exon 3 inclusion, they are likely to locate in intronic splice enhancer (ISE) or silencer (ISS) cis-acting elements. ISE and ISS are often binding sites for trans-acting factors such as heterogeneous ribonucleoproteins (hnRNPs) or serine/arginine-rich (SR) proteins. In comparison to the splicing of constitutive exons where splice sites are efficiently recognised, ISS and ISE play a much more important role in regulating splice site recognition and pairing by the core spliceosome components U1 and U2 small nuclear ribonucleoproteins (snRNPs) at the 5' and 3' splice sites in the splicing of alternative exons (X. D. Fu and Ares 2014). The effects of splicing regulatory elements are highly context and position specific (X. D. Fu and Ares 2014; Z. Wang and Burge 2008). For example, the strengths of neighbouring 3' and 5' splice sites will influence likelihood of each other being recognised by U1 and U2 snRNPs, thereby altering splicing outcomes (Kuo et al. 1991; Robberson et al. 1990). Intronic cis-regulatory elements could have both enhancer and silencer effects depending on their surrounding sequences. For example, deletion experiments of hnRNP F binding motif in the fibrinogen gamma-chain gene pseudoexon showed a hnRNP F silencer element could have enhancing effect if the neighbouring hnRNP F binding elements are absent (Rimoldi et al. 2013). Whether if the H1 and H2 rs17651213 and rs1800547 sequences represent ISS or ISE, will therefore require more experiments to be determined.

An approach to determine whether if the H1 and H2 rs1800547 and rs17651213 SNPs locate in ISS or ISE would be to generate H1 and H2 PAC *MAPT* vectors with serial deletions on sequences around rs1800547 and rs17651213. The H1 and H2 vectors could then be expressed in cell culture models and the effects of the deletions on exon 3 inclusion could pinpoint the nature of the sequence elements surrounding the functional SNPs. Sequence modifications in PAC by

homologous recombination is accurate and precise. However, to generate a large number of PAC vectors with different deletions, a high throughput approach will be required. Instead of generating mutations on PAC vectors, the clustered, regularly interspaced, short palindromic repeat-associated Cas9 endonuclease (CRISPR-Cas9) genome editing systems could provide an alternative solution. The CRISPR-Cas9 systems have the capacity to generate a large number of immortalised human cell lines, each carrying a random deletion, insertion or mutation in a specified genomic region of interest (Sander and Joung 2014). By measuring how the indels and mutations affect endogenous *MAPT* exon 3 inclusion in the engineered cell lines using qPCR, the nature of the splicing regulatory sequence elements could be deduced. The CRISPR-Cas9 method has been used to modify genomic sequence in *E. coli* cells carrying a plasmid. Thus theoretically a PAC vector carried by *E. coli* could also be modified in this manner. However, this method is only possible in a recombineering *E.coli* strain which may not be suitable for carrying large PAC vectors, complicated by the use of even more antibiotic-resistant genes for the plasmids bearing the CRISPR array and the Cas9 endonuclease.

CHAPTER 5

RNA-Protein Interaction Studies on H1 and H2 rs17651213 and rs1800547 Sequences

5.1 Introduction

5.1.1 Splicing Regulation by RNA-Binding Proteins

The splicing of human pre-mRNA transcripts is directly controlled by a large number of RNA-binding proteins. Splicing occurs as a two-step transesterification process where the 2'hydroxyl group of an adenosine residue in a branch point sequence first attacks the phosphodiester bond at the 5' splice site, followed by the attack of the 3' splice site by the 3'-hydroxyl group of the upstream exon, resulting in exon ligation and intron excision. The stepwise assembly of several small nuclear ribonucleoprotein (snRNP) complexes (named U1, U2, U4/6 and U5) form different spliceosome intermediates that catalyse and coordinating the spatial and temporal joining of two exons (Wahl et al. 2009). Studies of the spliceosome components at their different stages of assembly and function suggest that approximately 170 proteins in total are associated with the spliceosome (Wahl et al. 2009). Each spliceosome intermediate species is associated with approximately 125 proteins with a high level of exchange of the protein components from one stage to the next during splicing, demonstrating the complexity and multiple levels of regulation of the process (Wahl et al. 2009).

Most 5' and 3' splice site consensus sequences are relatively degenerate, with numerous decoy splice sites locating in introns that are often over thousands of bp in length (Matera and Wang 2014). Decoy splice sites are sequences that resemble authentic splice site sequences but are rarely successfully recognised and utilised by the spliceosome (Sun and Chasin 2000). Since the splicing of human transcripts occur with high accuracy, the degenerate nature of splice sites suggest that they

do not contain enough information for spliceosome assembly (Matera and Wang 2014; Wahl et al. 2009). Indeed, *cis*-acting regulatory sequence elements play a major role in augmenting splice site recognition by spliceosomal complexes. The difference between constitutive and alternative exons therefore lies in the relative strengths of the splice site signals, in which alternative exons generally have weaker consensus splice site sequence than constitutive exons (X. D. Fu and Ares 2014). Depending on the location and the mode of action, *cis*-acting regulatory elements can be classified as exonic or intronic splicing enhancers (ESEs or ISEs) or silencers (ESSs or ISSs), and are binding sites for *trans*-acting splicing factors.

Serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) are two major groups of *trans*-acting splice factors and are classically regarded as promoters and repressors, respectively, for exon inclusions (X. D. Fu and Ares 2014). In vitro biochemistry studies have provided insights into the general mechanisms of how their binding towards exonic and intronic splice regulatory elements regulate splicing. SR protein family promotes exon inclusions by interacting with U1 snRNP and U2 auxiliary factor during the assembly of spliceosome complexes for splice site recognitions (Busch and Hertel 2012). hnRNPs repress exon inclusions by physically blocking snRNPs from being recruited, preventing the formation of spliceosome through their multimerisation on exons or by looping out the exon (Busch and Hertel 2012). The actual mechanisms of splicing regulation by SR proteins and hnRNPs are now known to be much more complicated and sophisticated than the classical view of enhancer and silencer proteins as revealed by recent genome-wide analyses of SR proteins and hnRNPs mediated splicing events (Huelga et al. 2012; Pandit et al. 2013).

Crosslinking and immunoprecipitation followed by high-throughput sequencing (CLIP-seq) technology has allowed for genome-wide identification of *in vivo* splice factor binding sites. While the study of global splicing events mediated by splice factors is now made possible with microarrays for splicing detection. Using a combination of the two methods, studies have demonstrated that the effects of SR proteins and hnRNPs are highly context specific and they regulate both enhancing

and silencing events, defying their tradition unidirectional roles as being enhancer and silencer, respectively, in splicing (Huelga et al. 2012; Pandit et al. 2013). Binding motifs for SR proteins and hnRNPs were found to be degenerate, again defying the classical view that they have a defined set of consensus motifs as identified by *in vitro* RNA-affinity purification methods and by SELEX (systematic evolution of ligands by exponential enrichment) experiments (Huelga et al. 2012; Pandit et al. 2013). Through studying the effects of the of individual splice factors, it is now understood that splicing control and outcome are the overall complex combinatorial effects of splice factors through both their competition for *cis*-acting elements and cooperative interactions with each other and with other splice factor families (X. D. Fu and Ares 2014; Huelga et al. 2012; Pandit et al. 2013).

5.1.2 Biochemical Studies on RNA-Protein Interaction

Biochemistry techniques have been indispensable in shaping our understanding of the roles of RNA-RNA, RNA-protein and protein-protein interactions in alternative splicing. For example, small nuclear RNAs (snRNAs) at the catalytic centre of the spliceosomal complexes catalysis the transesterification of pre-mRNA molecules; RNA-binding splice factors such as SR proteins and hnRNPs directly interact with pre-mRNA molecules to regulate splicing outcomes; and numerous proteins interact with the snRNP complexes and splice factors to augment the splicing process (Fica et al. 2013; Y. Lee and Rio 2015; Wahl et al. 2009). Various in vitro RNA-binding assays have been used to demonstrate RNA-protein interaction between splice factors and *cis*-acting enhancer/silencer elements, and to the study effects of different binding site sequences on splicing. Experimentally validated binding site sequences are often used by reference tools, such as SFmap and SpliceAid2, for splice factor motif analysis and binding predictions (Akerman et al. 2009; Giulietti et al. 2013; Paz et al. 2010; Piva et al. 2012).

The SELEX procedure has been particularly useful in screening for candidate RNA sequences with the highest affinities amongst a pool of randomised RNA sequences (Tuerk and Gold 1990). Evolutionary selection of variations followed by their subsequent replication and propagation

forms the basis of the SELEX procedure (Tuerk and Gold 1990). In this procedure, a pool of RNA molecules containing a region with randomised sequences are allowed to bind to a protein of interest immobilised to an anchor, such as sepharose beads or nitrocellulose filters. Unbound RNA species are washed off and the bound RNA molecules are reverse transcribed and amplified by PCR. The PCR products are then transcribed in vitro and the process is repeated again several times until the dominant or "winner" RNA species are identified (Tuerk and Gold 1990). The binding properties of the winner sequences towards the protein of interest and the functional outcome on splicing are often verified and further investigated using various methods such as electrophoretic mobility shift assays (EMSAs), UV-crosslinking, Western blotting, in vitro splicing assays, surface plasmon resonance, siRNA knockdown, RNA-affinity pull-down, to confirm binding with the protein and to study reaction binding kinetics (Burd and Dreyfuss 1994; Cavaloc et al. 1999; Marquis et al. 2006; Ponthier et al. 2006; Schaal and Maniatis 1999).

Today's understanding of pre-mRNA splicing has been a huge effort of the above biochemical studies on RNA-protein interaction characterisations. Most techniques mentioned above involve the use of cell nuclear extract as a source of splice factor proteins. While RNA-protein interaction *in vitro* is indicative of *in vivo* events, it should be noted that the interaction observed may be artificially induced due to the various buffer and binding reaction conditions, which may not occur in vivo (Mili and Steitz 2004). To be able to identify physiologically relevant RNA-protein binding sites *in vivo*, recent studies have evolved to focusing on studying genome-wide RNA-protein interactions and alternative splicing patterns. Using CLIP-seq, studies have been able to generate genome-wide maps of splice factor binding sites on mRNA transcripts (Huelga et al. 2012; Pandit et al. 2013; Shankarling et al. 2014). Consensus motifs for splice factors are often degenerate and context specific. SELEX experiments cannot account for the degenerative nature of motifs as they select for ligands with higher affinities with the protein depending on the reaction condition, as illustrated in by different motifs deduced from different SELEX experiments (Cavaloc et al. 1999; H. X. Liu et al. 2000; Schaal and Maniatis 1999; Tacke and Manley 1995). Novel motif discovery, the

refinement and the validation of known motifs have been made possible by the mapping of genome-wide binding sites (Anko et al. 2010; Anko et al. 2012; Huelga et al. 2012).

Using sequence motifs deduced from CLIP-seq data, it is also possible to identify regulatory regions where the motifs are enriched in, providing functional insights into their role in splicing regulation (Ray et al. 2013). For example, binding site clusters derived from CLIP-seq data of six different hnRNP proteins in over ten thousand genes showed they locate predominantly to intronic regions 500 nucleotides from an exon (Huelga et al. 2012). Enrichment of binding sites in 3' untranslated regions was found for hnRNP A1, hnRNP A2/B1, hnRNP F and hnRNP U while hnRNP M sites demonstrate a bias towards the 5' end of genes (Huelga et al. 2012). Context specificity and the difficulty in motif prediction is likely explained by the formation of RNA secondary and tertiary structures *in vivo*. Indeed, it has been shown that RNA structures play a crucial role in alternative splicing (Buratti et al. 2004). Several methods have emerged recently using chemical reagents to probe for RNA structures in living cells, coupling with RNA sequencing, to generate genome-wide maps of RNA structures in vivo (Ding et al. 2014; Rouskin et al. 2014; Spitale et al. 2013; Talkish et al. 2014). Future studies in alternative splicing are likely to integrate various genome-wide data such as splice factor binding, RNA structures, epigenetic maps and genetic variations to unravel the complex picture in alternative splicing regulation.

5.1.3 Aims of Chapter 5

This chapter aims to understand the mechanisms behind the splicing regulation by rs17651213, by studying how a single base change caused by the SNP affect RNA-protein interactions. Electrophoretic mobility shift assay (EMSA) was first described as a sensitive and robust method to study DNA-protein binding equilibrium and kinetics (Fried and Crothers 1981; Garner and Revzin 1981). The separation of free DNA and DNA-protein complexes in non-denaturing polyacrylamide gel by electrophoresis forms the basis of EMSA (Fried and Crothers 1981; Garner and Revzin 1981). EMSA has been subsequently successfully used to study RNA-protein interactions (Lowary and Uhlenbeck 1987; Peltz and Ross 1987). RNA-EMSA was performed here to study the effects of SNPs

rs1800547 and rs17651213 on RNA-protein interactions using biotinylated H1 and H2 RNA oligonucleotide probes spanning the SNPs and SK-N-F1 nuclear lysate (section 2.3.2). RNA-affinity pull-down approach followed by mass spectrometry analysis was further employed to study candidate splice factor bindings towards rs17651213 H1 and H2 sequences.

5.2 Results

5.2.1 SK-N-F1 Nuclear Lysate Extraction

To obtain nuclear enriched lysate from SK-N-F1 cells for studying how rs17651213 and rs1800547 SNP sequences affect RNA-protein interactions, we first modified and tested two centrifugation methods of nuclear lysate extraction. The methods were modified from a protocol by Prof. J. Knight's laboratory (University of Oxford) and a published protocol (Ye et al. 1994). Stepwise separation of the cytoplasmic fraction from the nuclear fraction by centrifugation to obtain cell nuclei forms the basis of the centrifugation method (Schreiber et al. 1989). The centrifugation method involves two major steps. The cytoplasmic membrane is first lysed by a hypotonic buffer aided by a detergent. The nuclei fraction is then separated from the cytoplasmic fraction by centrifugation and is then incubated with a hypertonic buffer for collapsing the nuclear membrane aided by vigorous mechanical disruption such as vortexing. The nuclear enriched lysate is then separation from the unlysed nuclei again by centrifugation. Using histone H3 and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as nuclear and cytoplasmic markers (Dimauro et al. 2012), we performed Western blot using anti-histone H3 and anti-GAPDH antibodies to determine the purity of the nuclear and cytoplasmic fractions obtained (Figure 28). Using the two protocols, very little contamination of GAPDH in the nuclear fraction and histone H3 in the cytoplasmic fraction was found (Figure 28), suggesting that the nuclear and cytoplasmic fractions were obtained with relatively high purity. SK-N-F1 nuclear enriched lysates was successfully generated here for RNA-protein interaction studies. Since similar results were obtained using both protocols, nuclear extraction for was performed using the protocol from Prof J. Knight's laboratory as we have received more detailed technical advices regarding reagents and methodology.

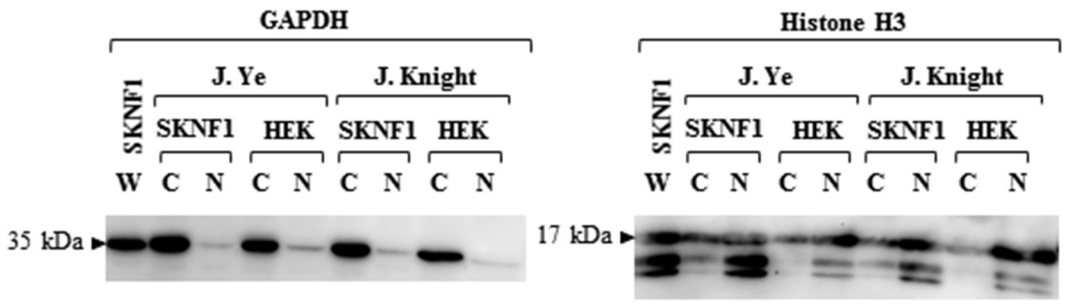
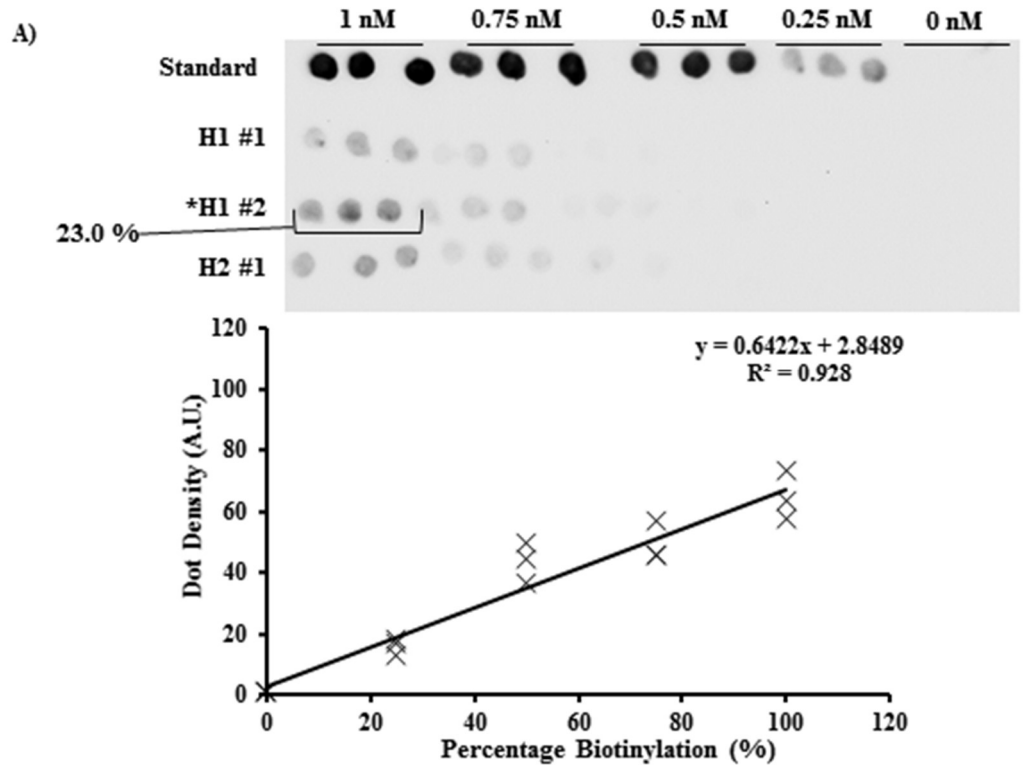


Figure 28. Confirmation of nuclear enriched cell lysate fractions by Western blot.
Two nuclear lysate extraction protocols (J. Ye and J. Knight) have been used for separating nuclear (N) and cytoplasmic (C) fractions of SK-N-F1 and HEK293 cell extracts. GAPDH (35 kDa) and histone H3 (17 kDa) were used as cytoplasmic and nuclear marker, respectively. SK-N-F1 whole cell extract (W) was used as a positive control.

5.2.2 Biotinylation of rs1800547 RNA Oligonucleotides for RNA-EMSA Studies

RNA oligonucleotide probes of 30 bases in length that span the H1 (5' CCA CAG GUG AGG GU**A** AGC CCC AGA GAC CCC 3') or H2 (5' CCA CAG GUG AGG GU**G** AGC CCC AGA GAC CCC 3') rs1800547 SNP sequences were designed. To produce biotinylated RNA oligonucleotide probes for chemiluminescent detection, biotinylated cytidine bis-phosphate was ligated to the 3' end unlabelled RNA oligonucleotides by T4 RNA ligase using the RNA 3' End Biotinylation Kit (Pierce). Biotinylation efficiency was determined by comparing the chemiluminescent intensity of the biotinylated RNA oligonucleotides to a standard dilution of biotinylated RNA control (Figure 29). Since different RNA sequences will lead to different RNA structures, different biotinylation conditions may be required to achieve the optimal biotinylation efficiencies. Several biotinylation reaction conditions were assessed for the H1 and H2 rs1800547 RNA probes and it was found that the optimal biotinylation efficiencies achieved for the H1 and H2 probes were 23 % (Figure 29A) and 56 % (Figure 29B), respectively. The biotinylated probe spanning rs1800547 SNP H1 and H2 sequences were in our RNA-EMSA.


A)
1 nM
0.75 nM
0.5 nM
0.25 nM
0 nM
Standard
H1 #1
*H1 #2
23.0 %
H2 #1
y = 0.6422x + 2.8489
R² = 0.928
Dot Density (A.U.)
Percentage Biotinylation (%)
0
20
40
60
80
100
120

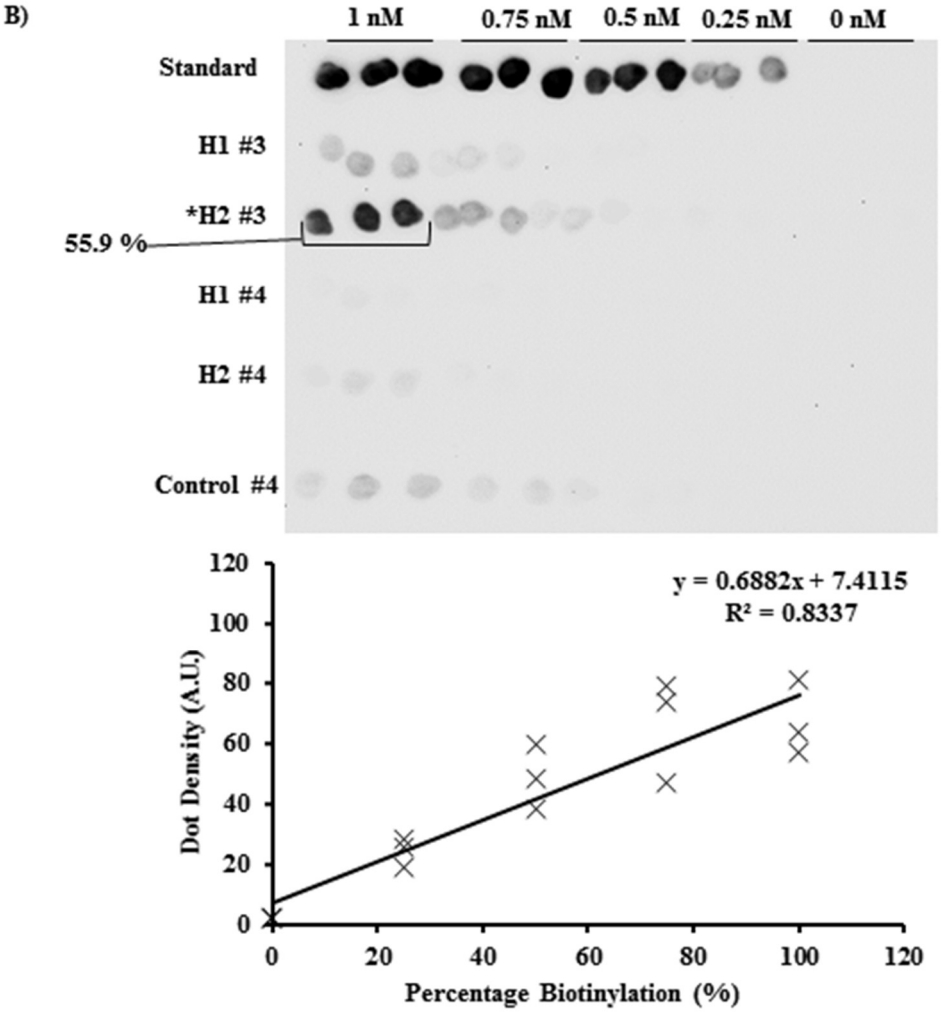

B)
1 nM
0.75 nM
0.5 nM
0.25 nM
0 nM
Standard
H1 #3
*H2 #3
55.9 %
H1 #4
H2 #4
Control #4
y = 0.6882x + 7.4115
R² = 0.8337
Dot Density (A.U.)
Percentage Biotinylation (%)
0
20
40
60
80
100
120

Figure 29. Estimation of biotinylation efficiency for rs1800547 H1 and H2 probes.

Standard dilution curve for estimating biotinylation efficiency was generated by dot blotting a biotinylated RNA control at 1 nM, 0.75 nM, 0.5 nM, 0.25 nM and 0 nM concentrations. The dot density of each dilution was assessed by Image J and was plotted against the theoretical biotinylation percentage where a curve of best fit was then generated. Each biotinylated RNA sample was then diluted to 1 nM concentration and were spotted at 1 nM, 0.5 nM and 0.25 nM concentrations. The biotinylation efficiency of each sample was determined by the average dot density of the non-saturated spot replicates using the standard curve equation. All dilution points were spotted in triplicate. Several biotinylation conditions (#1 - #4) of the H1 and H2 probes were assessed. **(A)** The optimal biotinylation condition (starred) for the H1 probe with an efficiency of 23.0 %. **(B)** The optimal biotinylation condition (starred) for the H2 probe with an efficiency of 55.9 %.

5.2.3 Optimising RNA-EMSA Binding Reaction Conditions Using rs1800547 RNA Probes

To optimise RNA-protein binding reaction conditions, different amounts of SK-N-F1 nuclear lysate were first titrated with increasing concentrations of biotin-labelled probe (Figure 30A). Optimal resolution of gel shift was determined when the shifted bands representing RNA-protein complexes are clearly visible (Figure 30A). 5 µg of SK-N-F1 nuclear lysate together with 10 and 4 nM of H1 and H2 biotin-labelled rs1800547 probes, respectively, was found to show RNA-protein complexes with the best resolution and clarity. The optimal H1 and H2 probe concentrations differ (Figure 30A) due to the difference in their biotinylation efficiencies (Figure 29). The amounts of tRNA in the binding reactions were next titrated. tRNA acts as non-specific competitor to sequester RNA binding proteins that interact non-specifically with the RNA probe (D. L. Black et al. 2002). 2 µg tRNA per reaction was found to show an optimal effect on eliminating smears on the gel without compromising the visibility of the specific bands (Figure 30B).

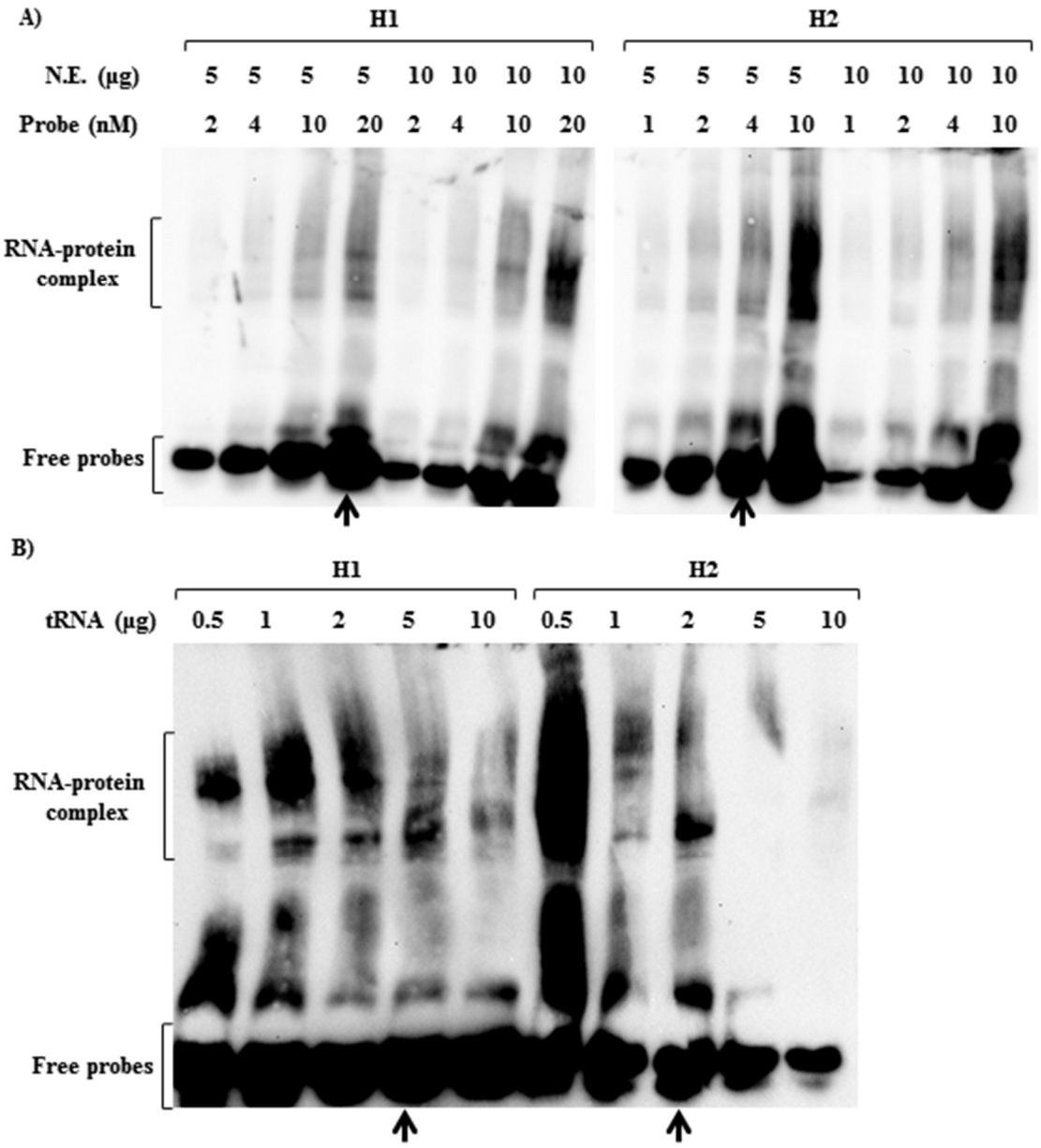


Figure 30. RNA-EMSA optimisations for rs1800547 H1 and H2 probes.

(A) Different amounts of H1 and H2 RNA probes were incubated with different amounts of nuclear extract (N.E.). RNA complex formed with the best clarity (arrows) represent the optimal binding conditions. **(B)** Titration of tRNA (non-specific competitor) to determine the optimal amounts of tRNA required (arrows) for reducing non-specific binding but not completely sequestering all relevant proteins.

5.2.4 rs1800547-A/G Sequences Alter RNA-Protein Interactions

To determine how rs1800547 SNP sequences impact on RNA-protein interactions, RNA-EMSA was performed using SK-N-F1 nuclear lysate and H1 and H2 biotin-labelled RNA probes that span rs1800547 SNP sequences. Both H1 and H2 probes led to the formation of RNA-protein complexes (Figure 31A and B, lane 2). Different gel mobility of the H1 and H2 RNA probes was observed(Figure 31A and B, lane 2), suggesting that the proteins present in the H1 and H2 RNA-protein complexes differ.

Competition experiments, where increasing amounts (10-fold to 100-fold molar excess) of either H1 or H2 unlabelled rs1800547 oligonucleotides were allowed to compete with the labelled probes for protein binding (Figure 31A and B), were performed to determine the binding specificities of the RNA-protein complexes formed by the H1 and H2 rs1800547 probes. The H2 RNA-protein complexes were nearly completely competed away by 100-fold excess H2 competitor oligonucleotides (Figure 31B, lanes 3-6), while 100-fold excess H1 competitor did not successfully compete with the H2 probe (Figure 31B, lanes 7-10). The competition of the H2 RNA-protein complexes by only H2 but not H1 competitor indicates that proteins interacting with the H2 probe was specifically caused by a base change from A to G in the oligonucleotide sequence.

H1 specific RNA-protein interaction was not observed in our competition experiment (Figure 31A). H1 (Figure 31A, lanes 3-6) and H2 (Figure 31, lanes 7-10) competitor oligonucleotides were both able to compete with the H1 probe for the formation of H1 RNA-protein complexes with no observable difference in their competition strengths. The results indicate that the binding of proteins to H1 labelled probe is not specific to the H1 allele *per se*.

Here, H1 and H2 rs1800547 SNP sequences was found to lead to differential RNA-protein complex formation. Next, the effects of rs17651213 on RNA-protein complex formation were investigated using RNA-EMSA.

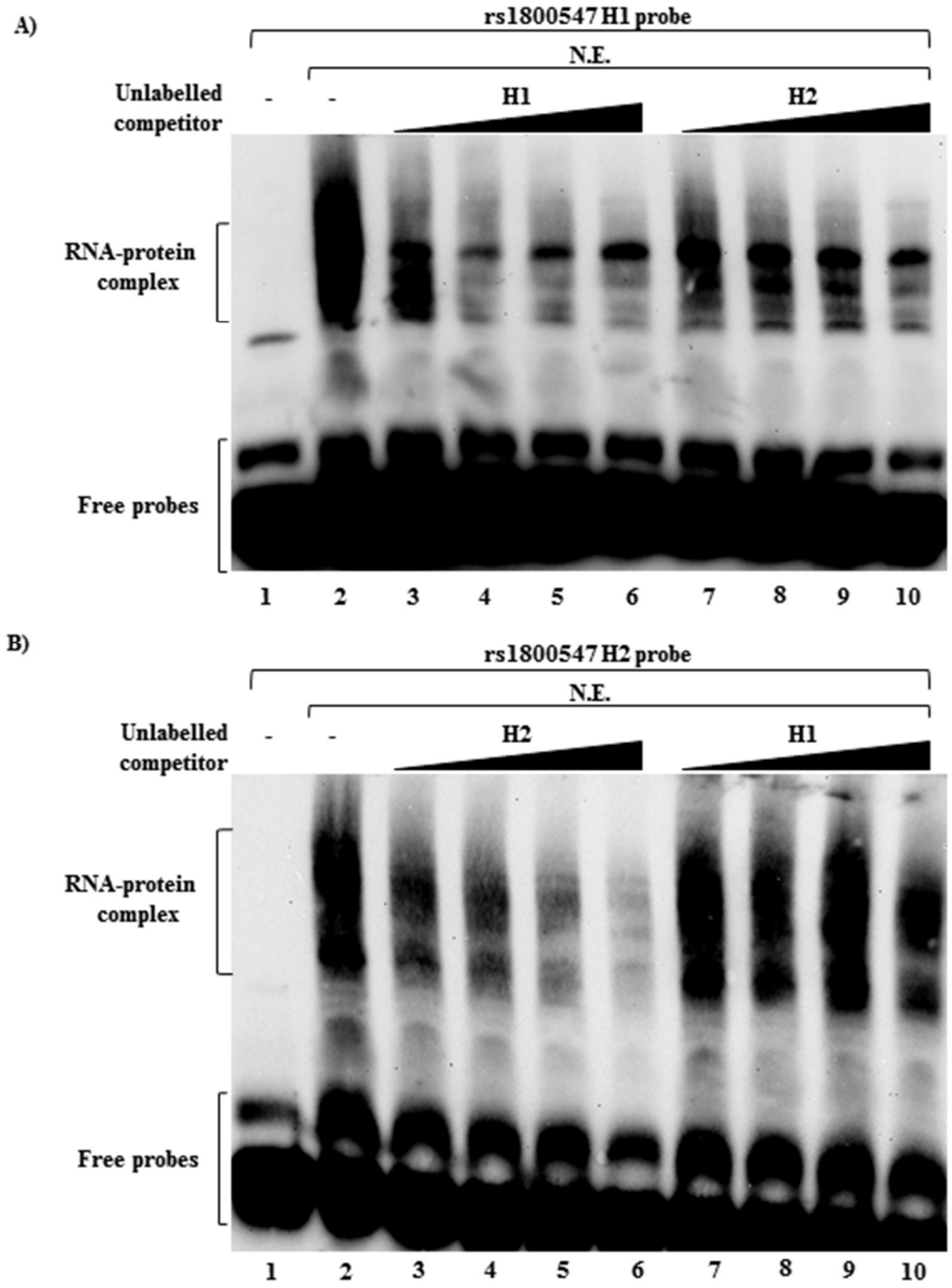


Figure 31. RNA-EMSA analysis demonstrates differential RNA-protein complex formation at the H1 and H2 variants of SNP rs1800547.

(A) RNA-protein complex formation by the H1 rs1800547 probe (lane 2). The H1 (lanes 3-6) and H2 (lanes 7-10) unlabelled competitor RNA oligonucleotides displace protein complex formation at the H1 rs1800547 sequence. **(B)** RNA-protein complex formation by the H2 rs1800547 probe (lane 2). The H2 (lanes 3-6) but not H1 (lanes 7-10) unlabelled competitor RNA displace protein complex formation at the H2 rs1800547 sequence. Unlabelled competitor probes were incubated 10-, 20-, 40- and 100-fold concentration in excess. N.E.: SK-N-F1 nuclear extract.

5.2.5 Optimise RNA-EMSA Binding Reaction Condition Using rs17651213 RNA Probes

To study how rs17651213 SNP sequence affect RNA-protein, 30 base pairs H1 (5' UGA GGG AGC UUU GC**G** UGU UUA UCC UCC UGU 3') and H2 (5' UGA GGG AGC UUU GC**A** UGU UUA UCC UCC UGU 3') probes were again used in our RNA-EMSA assays. To optimising binding reaction condition, increasing amounts of SK-N-F1 nuclear extract was first titrated against 1 nM H1 and H2 biotin-labelled RNA probes (Figure 32A). 5 µg nuclear lysate was found to show RNA-protein complexes with the best resolution and clarity (Figure 32A). Next, increasing amounts of tRNA were titrated to determine the amounts of non-specific competitor required in the binding reaction condition (Figure 32B). 2 µg tRNA per reaction was found to show an optimal effect on eliminating smears on the gel without compromising the visibility of the specific bands (Figure 32B).

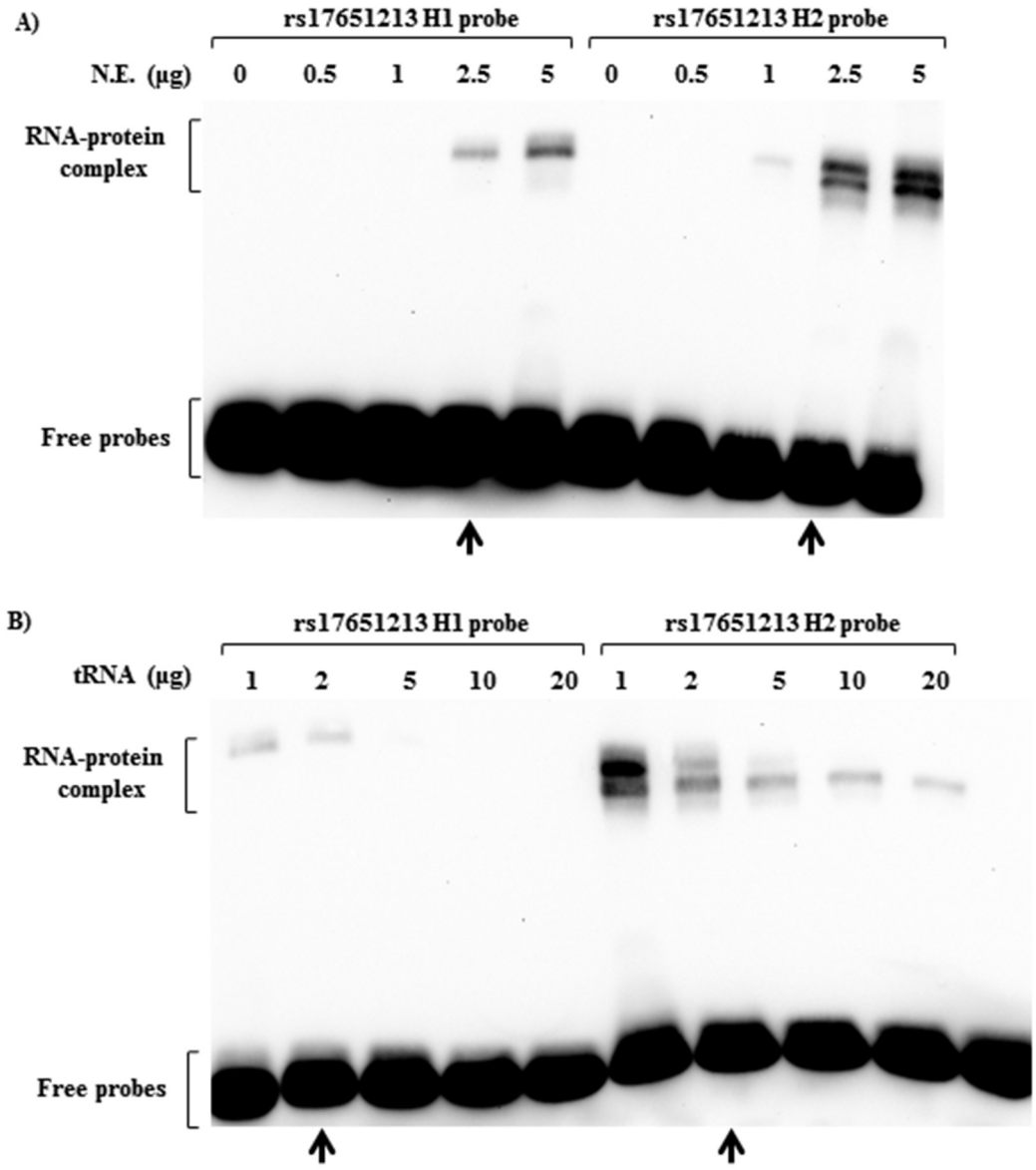


Figure 32. RNA-EMSA optimisations for rs17651213 H1 and H2 probes.

(A) H1 and H2 RNA probes were incubated with different amounts of nuclear extract (N.E.). RNA complex formed with the best clarity (arrows) represent the optimal binding conditions. **(B)** Titration of tRNA (non-specific competitor) to determine the optimal amounts of tRNA required (arrows) for reducing non-specific binding but not completely sequestering all relevant proteins.

5.2.6 rs17651213-G/A Sequences Alter RNA-Protein Interactions

To determine how rs17651213 SNP sequences affect RNA-protein interactions, we performed RNA-EMSA using SK-N-F1 nuclear lysate and H1 and H2 biotin labelled probes that span rs17651213 SNP sequences. Both H1 and H2 probes led to the formation of RNA-protein complexes (Figure 33A and B, lane 2). The H1 and H2 RNA probes showed different gel mobility (Figure 33A and B, lane 2), suggesting that the proteins present in the H1 and H2 RNA-protein complexes differ.

Competition experiments were next performed, where increasing amounts (200- and 500-fold molar excess) of either H1 or H2 unlabelled competitor RNA oligonucleotides were allowed to compete with the labelled probes for protein interaction, to determine the binding specificities of SK-N-F1 nuclear proteins toward the H1 and H2 probes (Figure 33). The competition of the H2 RNA-protein complex by only H2 (Figure 33B, lanes 3-4) but not H1 (Figure 29B, lanes 5-6) competitor indicates that components of the H2 RNA-protein complex was formed specifically due to a single base change from G to A in the oligonucleotide sequence. Unlabelled H1 and H2 rs1800547 competitor oligonucleotides (Figure 33B, lanes 7-10) showed similar competition effects as the H1 rs17651213 competitor, further indicating that the H2 RNA-protein complex formed was specific to the H2 sequences.

H1 specific RNA-protein interaction was not observed in our competition experiment (Figure 33A). Both H1 and H2 competitor oligonucleotides were able to compete away the formation of H1 RNA-protein complex with no observable differences in their competition strengths (Figure 33A, lanes 3-6). Unlabelled H1 and H2 rs1800547 competitor oligonucleotides showed similar competition effects as the H1 and H2 rs17651213 competitors (Figure 33A, lanes 7-10). The results indicate that the binding of proteins to H1 labelled probe is not specific to the H1 allele *per se*.

The results here indicate that H1 and H2 17651213 SNP sequences led to differential RNA-protein complex formation. The next section focuses on SNP rs17651213 to determine the identity of interacting proteins by performing RNA-affinity pull-down assays followed by mass spectrometry analysis.

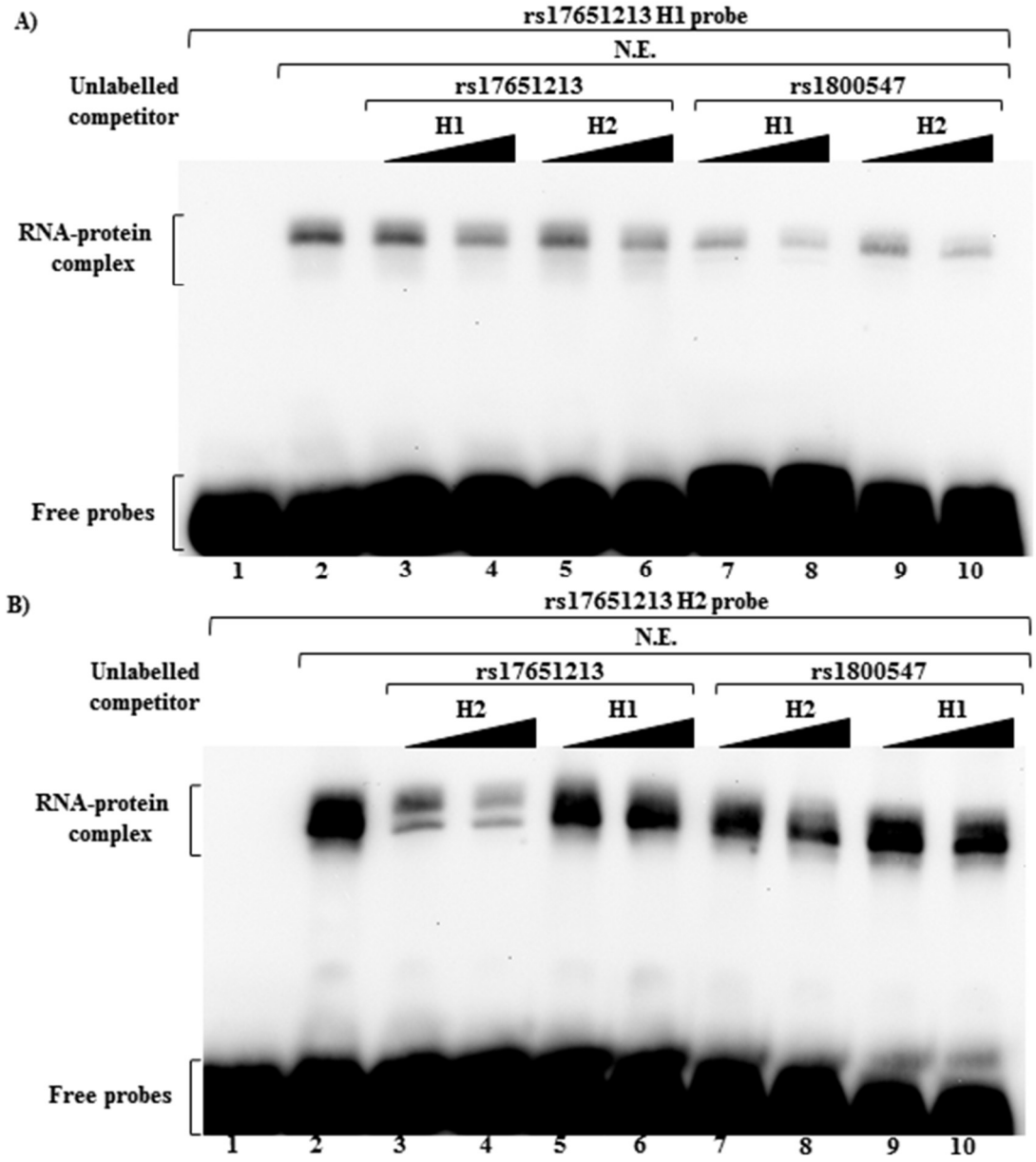


Figure 33. RNA-EMSA analysis demonstrates differential RNA-protein complex formation at the H1 and H2 variants of SNP rs17651213.

(A) RNA-protein complex formation by the H1 rs17651213 allele (lane 2). The H1 (lanes 3-4) and H2 (lanes 5-6) unlabelled competitor RNA probes displace protein complex formation at the H1 rs17651213 sequence. rs1800547 H1 (lanes 7-8) and H2 (lanes 9-10) unlabelled competitor showed similar strength as the rs17651213 competitors in competing with the H1 probe for complex formation. **(B)** RNA-protein complex formation by the H2 rs17651213 allele (lane2). The H2 (lanes 3-4) but not the H1 (lanes 5-6) unlabelled competitor RNA probe displaces protein complex formation at the H2 rs17651213 sequence. rs1800547 H1 (lanes 7-8) and H2 (lanes 9-10) unlabelled competitor showed similar strength as the rs17651213 H1 competitor, but not the H2, in competing with the H2 probe for complex formation.

5.2.7 Identifying Splice Factors Interacting With rs17651213 H1 and H2 SNP Sequences

RNA-EMSA provided us with a powerful method to visualise the overall effects of a single base change in the RNA oligonucleotide sequences on RNA-protein interactions. However, we wish to ask what splice factors interact with the SNP sequences. To address this question without pre-existing assumptions in the interacting partners of rs17651213 sequences, RNA-affinity pull-down of SK-N-F1 nuclear proteins was performed using the rs17651213 H1 and H2 RNA probes. Briefly, SK-N-F1 nuclear proteins were allowed to bind to biotin labelled H1 or H2 RNA baits immobilised on streptavidin coated magnetic beads (Pierce) (Figure 34A). Unbound proteins were washed off and the resulting samples were subjected to label free tandem mass spectrometry (LC-MS/MS) analysis (Figure 34A) (Mass spectrometry by Svenja Hester). Mass spectrometry signals were first analysed using the software Progenesis QI for Proteomics (Figure 34B and C). A much higher level of similarities was observed between the peptide ion compositions identified between the H1 and H2 samples than that between the beads only and H1/H2 samples as observed in their ion density maps (Figure 34B) and also in the overall peptide ion abundances (Figure 34C). This indicates that most proteins identified were a result of their interactions with the RNA baits but not with the beads. The produced peptide data were then analysed through the Mascot Server where peptide sequences were mapped onto proteins, generating protein identities (Figure 34A). The proteins identified were further analysed by the Progenesis software where proteins abundance were calculated according to (Figure 34A).

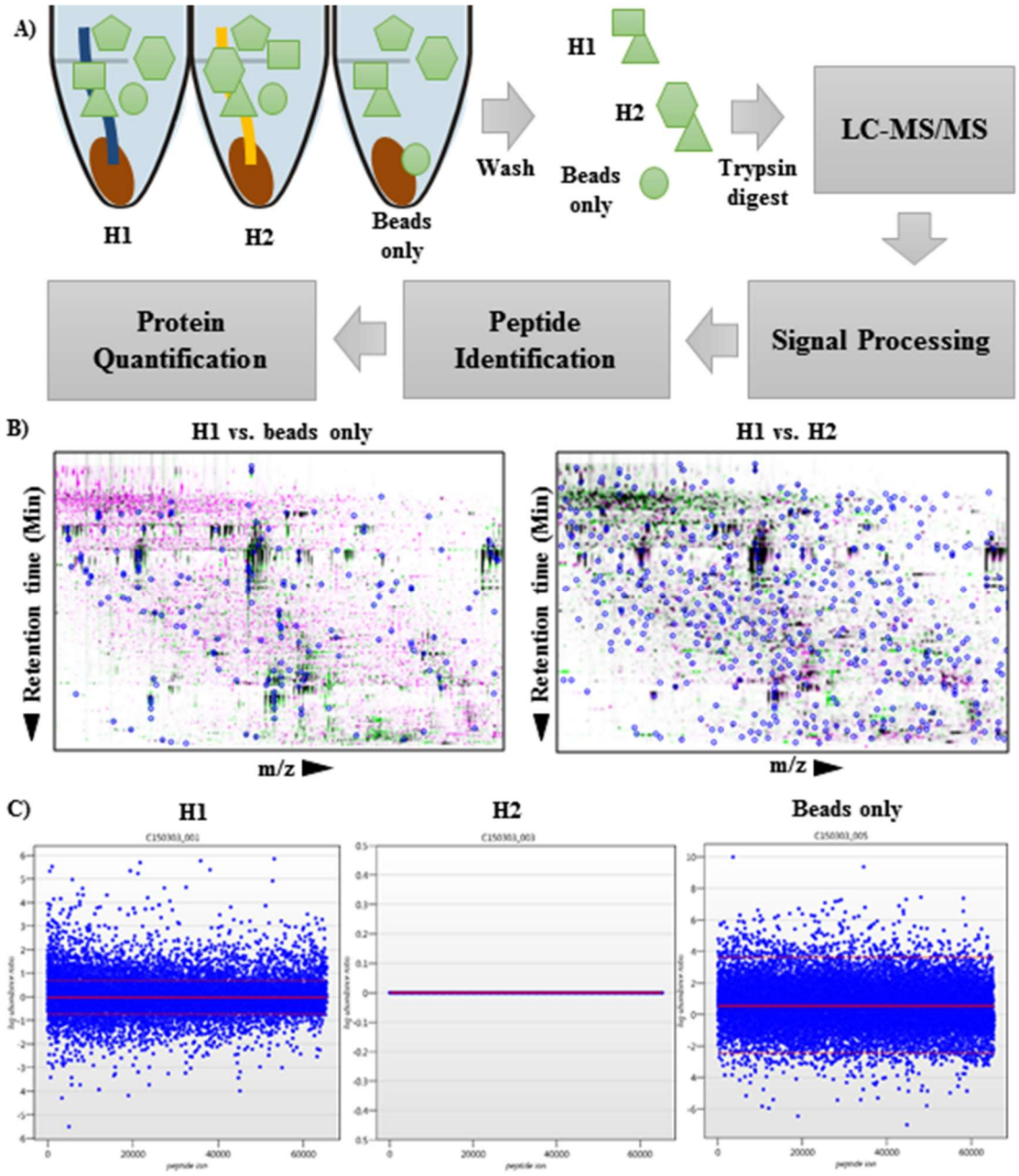


Figure 34. RNA-affinity pull-down of SK-N-F1 nuclear proteins interacting with rs17651213 H1 and H2 alleles.

(A) A schematic representation of the RNA-affinity pull-down workflow for identifying SK-N-F1 proteins that interact with rs17651213 H1/H2 alleles. Biotinylated probes were attached to streptavidin coated magnetic beads (Pierce) where SK-N-F1 nuclear extracts were allowed to bind to the probes. A control with no RNA was also set up for mass spectrometry analysis. Unbound proteins were then washed off and the remaining proteins were subjected to trypsin digestion followed by liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. Peptide ion peak signals were analysed (Progenesis QI) and the peptide sequences were mapped to their corresponding proteins (Mascot Server). Lastly, mass spectrometry data was generated where proteins abundance were quantified (Progenesis QI). **(B)** Ion intensity maps of peptides detected by

the mass spectrometer in the H1, H2 and beads only samples. The maps show the mass to charge ratio of the peptide ion peaks detected (x-axis) versus the retention time of the peptides through LC (y-axis). The intensity of the peaks (z-axis) is not visible in this two dimension plot. Peaks were aligned by the Progenesis software such that their intensity could be compared and quantified. Peptide peaks identified between H1 and H2 samples showed a much higher degree of similarity in their relative compositions, as observed by overlapping peaks, as compared to that between the H1 and beads only samples where most peptide peaks are only present in the H1 samples (pink signals). **(C)** Progenesis QI normalises the peptide ions by automatically selecting one of the runs, in this case the H2 samples, that is the least different from the other runs in the data set to be a reference for normalisation. Blue dots represent individual peptide ions. The central red line represent the normalisation factor as calculated by the mean of the log abundance ratios of the peptide ions that fall with the "robust estimated limits" (dotted red lines). The Progenesis software sets the limits such that peptide ions lying outside of the limits are considered outliers and will not affect the normalisation process. The H1 and H2 samples showed a much smaller difference, as shown by the log abundance ratios of the "robust estimation limits", when compared to that between H2 and beads only samples.

Proteins identified to be of low abundance or not identified in experimental replicates were excluded from the results (Figure 35A). Our results were compared to a list of 71 known human splice factors obtained from the SpliceAidF database (Giulietti et al. 2013) to identify splicing factors that interact with rs17651213 sequences (Table 9). Nine splice factors were identified to show preferential bindings of over 20 % to either H1 or H2 alleles. The splice factors are SFPQ, TARDBP, hnRNP M, ELAVL1, hnRNP A0, SYNCRIP, hnRNP A2B1, PTBP2 and hnRNP F (Table 9). The number of peptides which the Progenesis software used for quantifying each of the nine splice factors were further examined and they were found to be highly consistent between experimental replicates (Figure 35B). The differences in the relative abundance of splice factors identified interacting with the H1 and H2 rs17651213 SNP sequences indicate the SNP alleles alter the strengths of binding sequence for the splice factors.

In this chapter, differential effects on RNA-protein interactions were shown to be caused by the H1 and H2 rs1800547 and rs17651213 SNP alleles using RNA-EMSAs. Nine candidate splice factors interacting with SNP rs17651213 H1 and H2 alleles with differential binding strengths were further identified. To understand the mechanisms behind splicing regulation that may be caused by the SNPs, functional studies were performed to identify the potential allele-specific effects of the nine candidate splice factors on *MAPT* exon 3 splicing in SK-N-F1 neuroblastoma cell culture model in the next chapter.

A

Mass Spectrometry Data Analysis

Data Inclusion Criteria	Protein number
High abundance	~1200
↓ Abundance quantified with high confidence	~800
↓ Present in all three replicates	505
↓ Matching known human splice factors on SpliceAidF database	18
↓ Preferential binding of 20 % or above to either H1 or H2 probe	9

B

Protein Abundance Quantification of 9 Candidate Splice Factors

Splice Factors	Peptide Count
SFPQ	89.7 (3.9)
TARDBP	12.0 (1.4)
hnRNP M	110.7 (7.9)
ELAVL1	15.0 (0.0)
hnRNP A0	14.7 (0.9)
hnRNP Q	23.3. (1.7)
hnRNP A2B1	34.3 (2.5)
PTBP2	18.0 (1.4)
hnRNP F	16.0 (3.7)

Figure 35. Mass spectrometry data analysis to identify rs17651213 interacting splice factors. **(A)** Mass spectrometry protein data inclusion criteria for identifying splice factors that preferentially bind to either the H1 or H2 rs17651213 probe. **(B)** The peptide count for protein abundance quantification of the nine candidate splice factors. Standard deviation values in brackets.

Table 9. Identifying candidate splice factors interacting with rs17651213-G/A alleles.

UniProtKB	Gene Symbol	Description	Ratio H1:H2 (±S.D.)
P23246	SFPQ	Splice Factor Proline/Glutamine-Rich	2.41 (0.50)
Q13148	TARDBP	TAR DNA Binding Protein	2.17 (0.08)
P52272-2	hnRNP M	heterogeneous nuclear ribonucleoprotein M	2.05 (0.32)
Q15717-2	ELAVL1	Embryonic Lethal Abnormal Vision Drosophila-Like 1	1.96 (0.10)
Q13151	hnRNP A0	heterogeneous nuclear ribonucleoprotein A0	1.71 (0.21)
O60506-2	hnRNP Q	heterogeneous nuclear ribonucleoprotein Q	1.65 (0.51)
P22626	hnRNP A2B1*	heterogeneous nuclear ribonucleoprotein A2/B1	1.59 (0.26)
Q9UKA9-2	PTBP2	Polyrimidine Tract Binding Protein 2	1.30 (0.21)
P52597	hnRNP F*	heterogeneous nuclear ribonucleoprotein F	1.21 (0.14)
H3BPE7	FUS	Fused in Sarcoma	1.15 (0.08)
O75533	SF3B1	Splicing Factor 3b Subunit 1	1.13 (0.17)
O75525-2	KHDRBS3	KH Domain-Containing RNA Binding Signal Transduction Associated 3	1.13 (0.26)
Q96EP5-2	DAZAP1	Deleted in Azzospermia-Associated Protein 1	1.09 (0.19)
Q07666	KHDRBS1	KH Domain-Containing RNA Binding Signal Transduction Associated 1	1.09 (0.11)
P49756	RBM25	RNA Binding Motif Protein 25	0.94 (0.10)
P84103	SRSF3*	Splicing Factor 3b Subunit 1	0.94 (0.09)
Q15365	PCBP1	Poly (rC) Binidng Protein 1	0.89 (0.12)
P14866	hnRNP L	heterogeneous nuclear ribonucleoprotein L	0.83 (0.06)

H1 and H2 rs17651213 RNA-pull down of SK-N-F1 nuclear proteins followed by mass spectrometry identification. The values of the mean abundance ratio of proteins interacting with H1 : H2 probes were reported with the standard deviation. * Proteins predicted to bind the H1 and/or H2 probe sequences by SFmap, ESE and SpliceAid2. Nine splice factors (highlighted in grey) exhibit preferential bindings of 20 % or above to either H1 or H2 alleles.

5.3 Discussion

5.3.1 rs1800547 and rs17651213 SNP Sequences Alter RNA-Protein Interactions

This chapter aims to study how a single nucleotide difference caused by SNP rs17651213 affect RNA-protein interactions using biochemical approaches. Haplotype-specific differences were identified in the RNA-protein complexes formed by rs17651213 and rs1800547 H1 and H2 sequences through RNA-EMSA, suggesting that a change in SNP sequence can create/abolish binding sites for splice factors. Candidate splice factors that interact with rs17651213 SNP sequence were further identified for functional studies using an RNA-affinity pull-down approach followed by mass spectrometry analysis. The relative abundance of splice factors interacting with the H1 and H2 rs17651213 sequences were found to be different between the alleles, suggesting that a change in SNP sequence can alter the sequence strength as ligands for splice factors. The results indicate that altered RNA-protein interactions may form the functional basis for exon 3 splicing where different splice factors may exhibit different binding affinities for the cis-acting elements as a result of H1/H2 rs17651213.

5.3.2 RNA-EMSA in Studying RNA-Protein Interactions

In RNA-EMSA experiments, a short RNA oligonucleotide probe length of 30 nucleotides was used to minimise non-specific interactions of proteins to the RNA regions that not are bound by specific proteins (D. L. Black et al. 2002). Shorter oligonucleotide probes also allow for higher resolution of protein complexes visible on the gel shift due to a higher mobility of the probes upon their interaction with proteins (Min et al. 1995). However, it is important to note that RNA forms secondary and tertiary structures *in vivo*, and using short RNA probe lengths *in vitro* may not permit the proper RNA structures to be formed that would reflect *in vivo* folding. Nevertheless, consensus sequence motifs of SR proteins and hnRNPs deduced in genome-wide CLIP-seq experiments are consistent with those discovered using the SELEX procedure where short RNA baits of approximately 20 nucleotides are used for motif discovery, indicating that short RNA length may

already be sufficient to preserve the RNA folding for specific RNA-protein interactions (Huelga et al. 2012; Pandit et al. 2013).

A number of steps were performed to optimise the RNA-EMSA binding reaction condition, including varying the amounts of SK-N-F1 nuclear extract, RNA probe and non-specific competitor tRNA. It is important to perform these optimisation steps to increase the specificity of the RNA-protein complexes formed. Unlike transcription factors, RNA binding proteins are relatively abundant and are generally high in concentrations in mammalian cell nuclear extracts (Dreyfuss et al. 1993; Mittal et al. 2009). Since all nucleic acid binding proteins exhibit some degree of non-specificity, it is therefore essential to include non-specific competitors to prevent the RNA probe from being sequestered by non-specific RNA-protein interaction as non-specific interactions will obscure specific RNA-protein complex formed to be observed (D. L. Black et al. 2002; Lane et al. 1992). Different types of competitors have different properties to sequester proteins. Commercially available tRNA were applied here as the non-specific competitor in our experiments. Other commonly used non-specific competitor molecules include anionic heparin polymer and synthetic polynucleotides such as poly(C) or poly(U) (D. L. Black et al. 2002; Lane et al. 1992). It may be necessary to determine an optimal combination of non-specific competitors to include in the binding reactions. Since our aim here was to observe differences in RNA-protein binding between the H1 and H2 alleles but not to characterise the composition or the stoichiometry of the RNA-protein complexes, we did not test other non-specific competitors. Other components to consider in EMSA binding conditions that will affect he formation of RNA-protein complexes include the presence of cofactors such as cyclic AMP or ATP, monovalent and divalent salt concentrations, pH levels, neutral solutes such as sucrose or glycerol, non-ionic detergents and inhibitors for protease, nuclease and phosphatase (Hellman and Fried 2007).

RNA-EMSA is an incredibly useful tool not only to detect RNA-protein complex formation, but to characterise the binding properties of proteins towards a specific RNA sequence. For example, cooperative binding of proteins can be inferred from EMSA experiments by titrating the protein of

interest in increasing amounts (Senear and Brenowitz 1991; J. Wang and Bell 1994). If the protein is binding non-cooperatively to two neighbouring binding sites, the sites will be filled up sequentially until saturation, hence two bands will be observed as protein concentration increases (J. Wang and Bell 1994). If the protein binds cooperatively, two neighbouring binding sites will be filled simultaneously, hence only one band will be observed as protein concentration increases (J. Wang and Bell 1994). Interestingly, the Michaelis-Menten model of reaction kinetics could also be applied in RNA-protein interactions where the dissociation constant K_d of the binding of a protein to an RNA binding site can be determined using EMSA. Binding kinetics can be defined by the equation $K_d = [P]^n[R]/[P_nR]$ where [P] represent the total protein concentration, [R] and [PR] represent the unbound and bound RNA fractions, and n represent the number of proteins present per RNA-protein complex (D. L. Black et al. 2002; J. Wang and Bell 1994). A splice factor that interacts with a SNP sequence can be investigated using this methodology, providing insights into how SNP alleles affect the binding kinetics or the mode of binding of the splice factor.

5.3.3 Quantitative Proteomics in RNA-Binding Protein Identification

In the previous chapter, a list of splice factors predicted to bind to rs17651213 H1 and H2 sequences by computational analyses were described. However, as discussed previously that computational predictions correlate poorly with actual biological interactions, while RNA-EMSA does not provide the identity of interacting proteins. An RNA-centric affinity purification approach was therefore applied here using H1 and H2 rs17651213 RNA probes, to capture interacting proteins using SK-N-F1 nuclear extract. Proteins captured by the rs17651213 sequences are identified by quantitative mass spectrometry. RNA-affinity approaches provide a relatively unbiased means of screening for nucleic acid-protein interactions as the need for pre-existing knowledge on proteins is unnecessary (McHugh et al. 2014; Tacheny et al. 2013), while label-free quantification of peptides provides a flexible method for comparing protein abundance in different samples (Nahnsen et al. 2013). Many variations of this approach has been used for identifying RNA-binding proteins. In general the experiments involve two main parts. First, affinity purification of RNA-binding proteins are

performed (McHugh et al. 2014; Tacheny et al. 2013). Second, mass spectrometry is used to determine identity of the RNA-binding proteins (McHugh et al. 2014). Two major strategies have been developed to capture *in vivo* and *in vitro* assembled RNA-binding proteins (McHugh et al. 2014; Tacheny et al. 2013). Affinity purification of *in vivo* and *in vitro* assembled RNA-binding protein complexes both have their pros and cons. Briefly, the crosslinking of *in vivo* assembled RNA-protein complexes allows for stringent wash conditions to be used during purification, thereby reducing non-specific RNA-protein and protein-protein interactions that will interfere with downstream mass spectrometry analysis (Tacheny et al. 2013). The purifying of RNA-protein complexes assembled *in vitro* does not require the overexpression of the RNA sequence of interest using genetically modified constructs, but one could argue the RNA-protein complexes may not be in their physiological states (Tacheny et al. 2013).

To capture *in vivo* assembled RNA-binding proteins, tagged RNA baits expressed from RNA constructs are often used. The protein complexes bound to the RNA bait assembled *in vivo* are then crosslinked by short wavelength UV light or formaldehyde. The cells are lysed and the RNA-protein complexes formed can be recovered using the RNA tag present on the RNA bait. The secondary structures of RNA molecules have been utilised for their recovery and purification via RNA-RNA and RNA-protein interactions. An example of affinity purification via RNA-RNA interactions is the use of antisense RNA probes to recover MBII-52 small nucleolar RNA (snoRNA) in mouse brain cells (Soeno et al. 2010). The antisense RNA probe was biotin and was immobilised to streptavidin agarose beads. To capture MBII-52 snoRNA-protein complexes, nuclear lysates were incubated with the antisense RNA probe which hybridises with the MBII-52 loop structure that was previously assessed by RNase H digestion assays to be an accessible moiety (Soeno et al. 2010). The author reported approximately 7.3% of the total MBII-52 snoRNA-protein complexes were captured by the antisense RNA, and they subsequently identified and confirmed nucleolin and ELAVL1 as novel interacting partners of MBII-52 (Soeno et al. 2010). An example of affinity purification via RNA-protein interactions is a method termed MS2-BioTRAP which elegantly employed the

bacteriophage MS2 coat protein to capture *in vivo* stem-loop tagged RNA-protein complexes (Tsai et al. 2011). This method utilises the high affinity of MS2 coat protein to the MS2 stem-loop structure, with the dissociation constant at the nanomolar range (Hook et al. 2005; Keryer-Bibens et al. 2008; F. Lim et al. 1994). The MS2 coat protein was labelled with biotin and was immobilised streptavidin coated magnetic beads. The author employed MS2 stem loop tagged internal ribosome entry site (IRES) sequence of the lymphoid enhancer factor-1 gene and identified 36 IRES binding proteins (Tsai et al. 2011).

A method for affinity purification of *in vitro* assembled proteins binding H1 and H2 sequences rs17651213 was adopted here. *In vitro* methods in general involve using a synthetic tagged RNA bait that is first immobilised on a support. Proteins obtained from lysed cells are then captured by the RNA bait on the support. As the RNA bait does not have to be expressed in cells before being attached to a solid support, a diverse range of methods have been developed to tether RNA baits. Chemical modifications of RNA bait can be used to covalently link RNA molecules to adipic acid dihydrazide agarose beads or cyanogen-activated sepharose beads (Miniard et al. 2010; Patel et al. 2005; Talukdar et al. 2011). The RNA bait can also be modified chemically to contain different moieties such that they can be tethered non-covalently to a solid support, such as biotin moiety and streptavidin coated beads (Harris et al. 2006; A. F. Ross et al. 1997). Other strategies include incorporating natural sequences, such as polyadenylation, into the RNA bait synthesised, which could be recovered by oligo(dT)-sepharose beads (Bose et al. 2006). One should be aware that different ways of immobilising RNA baits to a solid support may affect protein binding in different ways. For example, chemical modifications may alter RNA structures, while biotin labels may cause steric hindrance to protein binding (Tacheny et al. 2013; Walker et al. 2008).

Since mass spectrometry is a sensitive method to detect the presence of proteins, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) separation of captured proteins are often performed to reduce sample complexity. Protein bands of interest are excised from the gel before being subjected to mass spectrometry analysis. However, SDS-PAGE raises the question of

sensitivity due to poor staining of proteins. Gel free studies offer the solution to SDS-PAGE by using liquid chromatographies where protein sample complexities are reduced before being analysed by the mass spectrometer. Typical affinity purification followed by SDS-PAGE experiments analysed by mass spectrometry identify between 20 and 250 proteins, while gel free studies identify approximately 500 proteins (D. Jiang et al. 2008; Ranish et al. 2003; Ren et al. 2012; Talukdar et al. 2011; Tsai et al. 2011). Our experiments captured over 1000 proteins using RNA baits just 30 nucleotides in length. It is therefore important in future experiments to reduce sample complexity, especially during the affinity purification process. Blocking or pre-clearing the solid support for RNA baits can be performed using tRNA, bovine serum albumin, heparin polymer or salmon sperm DNA (Goraczniak and Gunderson 2008; Mehta and Driscoll 1998). Alternatively the cell lysate can be pre-cleared by incubating with the solid support for RNA baits itself or with tRNA to sequester non-specific RNA-binding proteins (Iioka et al. 2011; Patel et al. 2005; Ward et al. 2011). Just as in RNA-EMSA, non-specific RNA competitor species, such as tRNA and anionic heparin polymer, can be added to the binding reaction together with the cell lysate to reduce the chances of non-specific RNA-binding proteins sequestering the RNA baits (Butter et al. 2009; Peddigari et al. 2013; Rattenbacher and Bohjanen 2012). However, the quantities of non-specific RNA competitor to be included in a reaction should be carefully determined by titration as it is possible that they may sequester the proteins of interest, thereby reducing the sensitivity of the experiment.

The affinity purification experiment here was originally designed as a screening step for possible in vitro splice factor candidates that interact with rs17651213 but not as a validation step. Further optimisations for the binding reaction conditions were therefore not carried out despite the need for sample complexity reduction due to time constraints. More importantly, the list of candidate splice factors will be functionally validated in our next chapter by siRNA knockdown experiments and also by RNA-EMSA. The results here provide an initial screen for candidate splice factors that may play a role in regulating *MAPT* exon 3 inclusion via rs17651213 H1 and H2 alleles.

CHAPTER 6

Investigating the Allele-Specific Regulation of *MAPT* Exon 3 Expression

6.1 Introduction

6.1.1 Understanding the Functional Biological Basis of Non-Coding Polymorphisms

To date, almost 19,000 lead SNPs have been described in GWAS to associate with over 1,500 human traits and diseases, and over 95% of the lead SNPs locate in non-coding regions (Burdett et al. ; Welter et al. 2014). Since GWAS were designed to detect low penetrance risk-associated variations (Bush and Moore 2012), it is unlikely that the lead SNPs themselves or nearby candidate genes to be the sole contributors to genetic predisposition. Variants that locate in the linkage disequilibrium (LD) architecture as captured by the lead SNPs are now increasingly recognised as putative causal polymorphisms that cumulatively account for the phenotypes according to the biological functions they modulate (X. Zhang et al. 2014). Understanding the biological functions of non-coding polymorphisms has therefore become increasingly important in the post-GWAS era.

The characterisation of allele-specific effects has been traditionally focused on gene expression levels. Genomic imprinting and X-chromosome inactivation in cells of extraembryonic lineages are a classic example of allelic gene expression where gene expression is determined by their parental origin, mediated by epigenetic mechanisms (Jaenisch and Bird 2003). The silencing of the X-chromosome during development is mediated by the expression and coating of the chromosome to be inactivated by the non-coding *Xist* RNA originating from the X inactivation centre (Xic), resulting in the formation of stable heterochromatin structures for the entire chromosome that are maintained during cell divisions (Knight 2004; Plath et al. 2002). The downstream expression of the non-coding RNA *Tsix* antisense to *Xist* in the active X-chromosome that remains prevents its

inactivation by *Xist* (Plath et al. 2002). The maintenance of X inactivation is dependent on the continuous and allele-specific activity of the protein Dnmt1 that maintains the methylation of the imprinted genes (Jaenisch and Bird 2003). While the example here illustrates that epigenetic mechanisms can have drastic and chromosome wide effects on genetic loci, recent evidence from a global study on the allelic expression in a human cell line suggests imprinted genes account for just 0.1% of the variation observed in allele-specific expression (Ge et al. 2009).

The advancement of next generation sequencing (NGS) methods not only allows for greater understanding of the human transcriptome, but also how variations contribute toward allelic differences in transcript expression levels as a quantitative trait through the generation of genome-wide allelic expression maps (Ge et al. 2009; Lappalainen et al. 2013; Pickrell et al. 2010). Indeed, widespread independent allele-specific gene expression events, or expressed quantitative trait loci (eQTL), have been identified to associate with polymorphisms different between alleles (Ge et al. 2009; Lappalainen et al. 2013; Pickrell et al. 2010). Through analysing the enrichment of eQTLs in regulatory elements, the author found a high level of eQTL enrichment in non-coding elements such as transcription factor binding sites, DNase I hypersensitive sites, transcription repressor CTCF sites, and the chromatin state of promoters (Lappalainen et al. 2013). Future studies to integrate GWAS data and eQTL maps will help understand and establish whether if the risk associated polymorphisms are associated with a particular gene expression pattern, and whether if such association forms a causal relation with certain disease traits or phenotypes. The enrichment of allelic traits in these regulatory elements helps explain how polymorphisms could act to regulate gene expression levels, through the functional non-coding elements in the genome the polymorphisms locate in.

The study of quantitative allelic gene expression levels sets a model for quantitative traits associated with genetic polymorphisms. In addition to eQTL, genetic variations are associated with several other cellular physiological traits, such as tissue or cell type specific allelic gene expression, alternative splicing patterns, chromatin loop formation, microRNA functions and long non-coding

RNA functions (Brest et al. 2011; Dimas et al. 2009; Ishii et al. 2006; Kwan et al. 2008; Shen et al. 1999; Visser et al. 2012). Importantly, the widespread allele-specific alternative splicing identified to associate with SNPs demonstrate the need to understand the effects of functional polymorphisms beyond gene expression (Kwan et al. 2008). More recently, studies have emerged to integrate GWAS data, LD structures and NGS data of functional elements, to study the enrichment of disease risk-associated SNPs in non-coding regulatory elements and to understand how genetic polymorphisms contribute to disease phenotype, cell type specific vulnerability, and the underlying molecular mechanisms in common complex human diseases (X. Zhang et al. 2014). In this final chapter, we aim to understand the mechanisms behind the H1/H2 regulation of *MAPT* exon 3 inclusion by the risk associated SNP rs17651213 in SK-N-F1 neuroblastoma cells. Several studies have previously employed a quantitative approach using Taqman based allele-specific qPCR methods to measure allelic imbalance in gene expressions (Arai et al. 2011; X. Chen et al. 2008b; Harmon et al. 2010; H. S. Lo et al. 2003; Martinez-Canto et al. 2012; A. J. Smith et al. 2012; Snow et al. 2015).

6.1.2 Co-Transcriptional Splicing Mechanisms

The coupling of transcription and pre-mRNA splicing is an important regulatory mechanism. Chromatin immunoprecipitation experiments have demonstrated the direct interaction between spliceosomal components and both RNA polymerase II (RNAPII) and the chromatin scaffold (Listerman et al. 2006). Kinetic experiments showed that pre-mRNA splicing occurs almost instantly after the relevant introns and exons have been produce, before the synthesis of the whole transcript has been completed (Singh and Padgett 2009). Recent RNA-seq experiments have further demonstrated that co-transcriptional splicing is a global mechanism (Ameur et al. 2011; Nojima et al. 2015). The elongation rate of RNAPII through a gene is non-uniform, and it is known to pause at various gene features including promoter regions, at exon-intron junctions and at polyadenylation sites (Kwak et al. 2013; Veloso et al. 2014). The pausing of RNAPII, which affects elongation rate along a gene, is functional and regulate alternative splicing. Experiments using chemical

compounds that modulate RNAPII elongation speed, have demonstrated slowing down or speeding up RNAPII lead to promoting exon inclusions and exclusions, providing evidence that RNAPII elongation rate regulate splicing outcome (de la Mata et al. 2003; Dujardin et al. 2014; Singh and Padgett 2009). Since several hundred SNPs exist in the *MAPT* gene, we are interested to study i) if RNAPII elongation rate regulate *MAPT* exon 3 splicing and ii) if altering RNAPII elongation rate may have different effects on the inclusion of H1 and H2 *MAPT* exon 3-containing transcripts.

6.1.3 Aims of Chapter 6

In this chapter, a Taqman based allele-specific qPCR assay was developed to measure H1/H2 *MAPT* exon 3-containing transcript isoforms. The allele-specific effects on *MAPT* exon 3 inclusion by knocking down candidate splice factors (Table 9) using siRNA in SK-N-F1 cells were studied with the Taqman based qPCR assay. The interaction of splice factors interaction with rs17651213 H1 and H2 SNP sequences was studied through RNA-EMSA supershift assays. Lastly, the developed allele-specific qPCR assay was further applied to study the effects of RNAPII elongation rate on *MAPT* exon 3 splicing.

6.2 Results

6.2.1 Developing and Validating a Quantitative *MAPT* Allele-Specific Expression qPCR Assay

A Taqman based allele-specific qPCR assay was designed in order to quantitatively measure the H1 to H2 *MAPT* exon 3 transcript ratios in our SK-N-F1 cell culture model. The H1 and H2 specific Taqman probes, labelled with FAM and VIC fluorophores respectively, were designed to span a haplotype-specific SNP rs17650901 that is present in the untranslated region of exon 1 (Figure 36). Primers were designed such that we could measure H1 and H2 genomic DNA, total *MAPT* transcripts and exon 3-containing transcripts by qPCR (Figure 36).

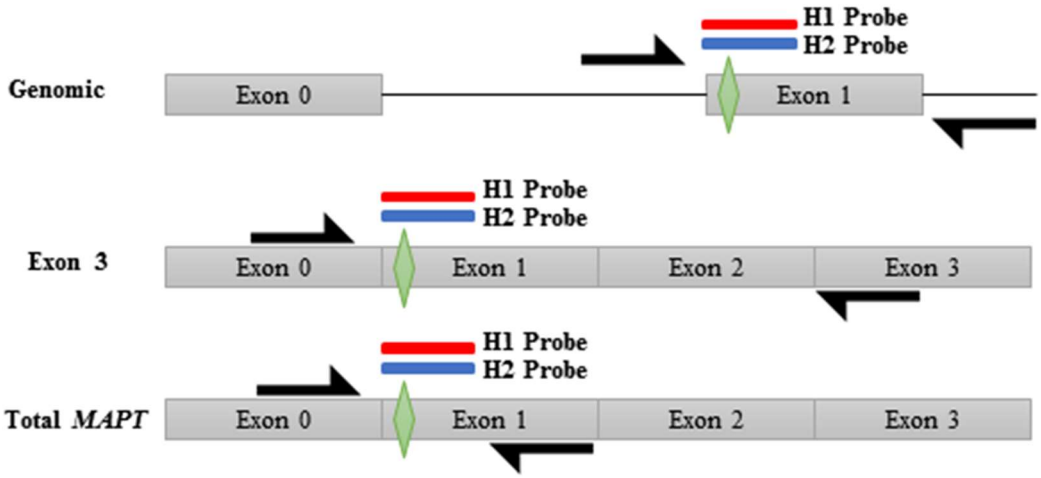


Figure 36. Allele-specific Taqman assay probes and primers design.

A schematic representation of the allele-specific Taqman assay design showing probes and primers locations on *MAPT* genomic DNA and mRNA transcripts. The H1 and H2 probes span SNP rs17650901 (green diamond), allowing for the detection of H1 and H2 transcripts by FAM and VIC fluorescence, respectively. Primers locating in different positions allows for the quantification of genomic DNA, exon 3-containing transcripts and total *MAPT* transcripts.

Several optimisation steps were performed to ensure the validity of the assay. To identify the optimal amount of primers to use in a reaction, primer concentrations were titrated from 50 to 900 nM. 900 nM of primers was found to produce the highest amounts of amplification products with specific product amplifications (Figure 37). The primer pairs (PP) for amplifying genomic DNA, total *MAPT* transcripts and exon 3-containing transcripts (PP2 – PP4) produce PCR products with the correct amplicon sizes (Figure 37). To choose between PP2 – PP4 exon 3 primers, qualitative assessment of their PCR product band intensities indicate PP3 and PP4 to produce a higher amount of PCR products compared to PP2 (Figure 37). As different primer sequences may interact with the Taqman probes differently, the fluorescent signals generated by the Taqman probes during amplifications using PP3 and PP4 were compared and it was found that PP4 amplifications produce a higher amount of signal as measured by the qPCR machine. Here, the optimal primer concentrations for prime pairs that produce specific PCR products have been identified for the allele-specific Taqman assay.

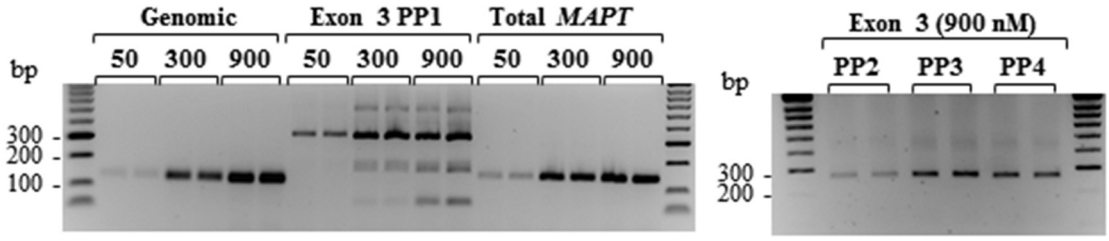


Figure 37. Selecting primer pairs for allele-specific Taqman assay.

Primer pairs (PP) amplifying genomic *MAPT* DNA, total and exon 3-containing *MAPT* transcripts. Primer concentrations (50, 300 & 900 nM) are indicated above the lanes. Except for exon 3 PP1, all primer pairs showed specific amplification and correct amplicon sizes.

To understand the effects of Taqman probe concentrations on the fluorescent signals, the amplification of H1/H2 heterozygous genomic DNA using different probe concentrations ranging from 0.5X to 2X as well as the recommend 1X concentration was assessed (Figure 38A). Increasing probe concentrations in general increases the fluorescent signals detected as observed by a decrease in the Ct values (Figure 38A). To determine the probe specificities in detecting the H1 and H2 alleles, we measured homozygous H1/H1 or H2/H2 genomic DNA with probes of 2X and 5X concentrations. Probe specificity was found to increase with probe concentrations (Figure 38B). 5X probe concentration provides the best separation of H1 and H2 signals (Figure 38B) in our homozygous samples, indicating that the H1 and H2 probes can differentiate between H1 and H2 alleles respectively with minimal non-specific cross interactions with the opposing alleles.

Lastly, to determine if the H1 and H2 probe signals can remain consistent over a range of DNA template concentrations, a serial dilution standard curve was generated using heterozygous H1/H2 genomic DNA as template. The amplification efficiency of the genomic primer pair as measured by the H1 and H2 probe fluorescence were similar between the H1 and H2 probes (H1 96.42 % and H2 94.6 %), indicating the H1 and H2 probe signals can remain consistent over a range of template concentrations (Figure 38C).

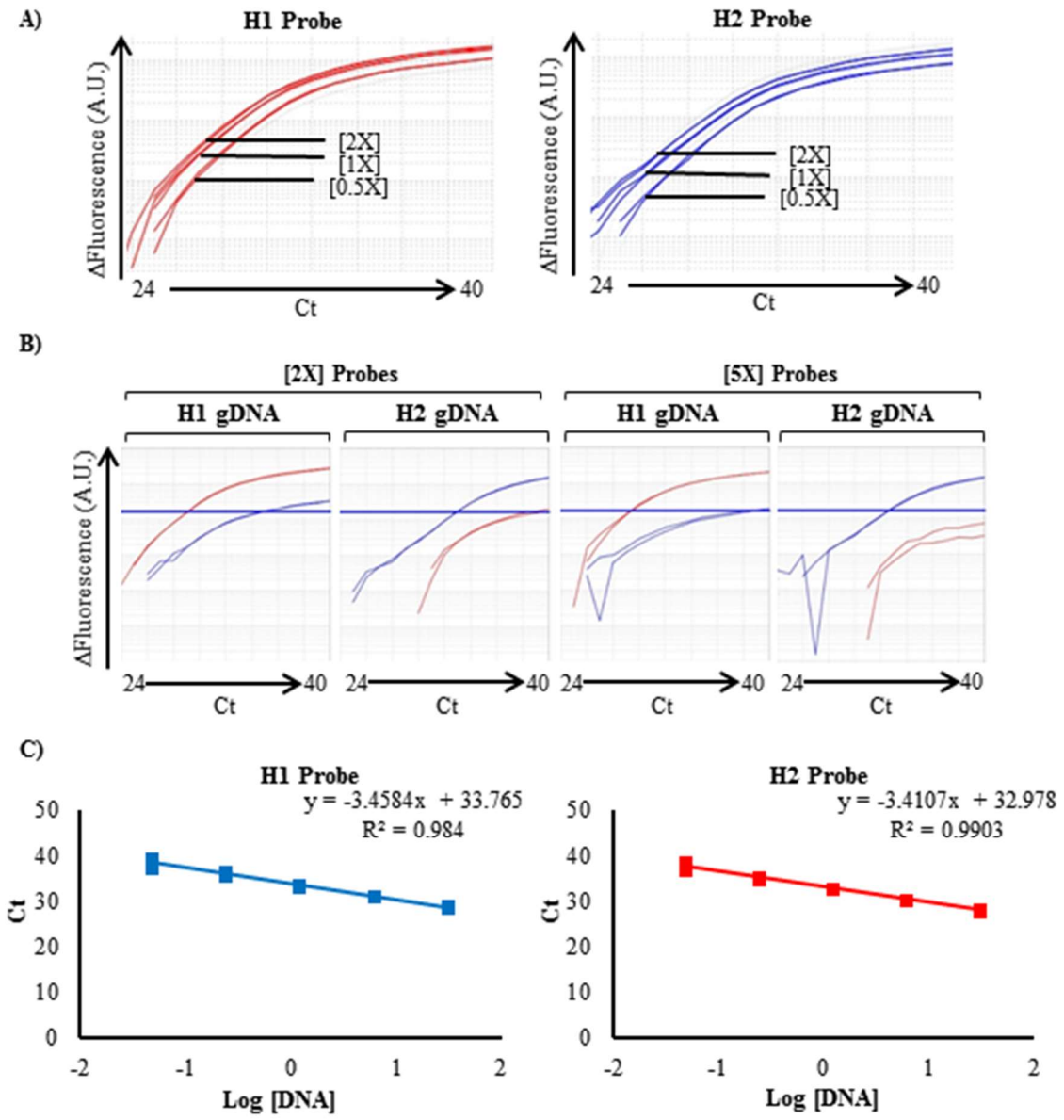


Figure 38. Optimisations of the *MAPT* allele-specific Taqman qPCR assay.

(A) Titration of Taqman H1 and H2 probe concentrations from 0.5X to 2X. Increasing probe concentration increases the fluorescent signals detected. **(B)** Measurement of the H1/H1 or H2/H2 homozygous genomic DNA by the H1 (red) and H2 (blue) probes at 2X and 5X concentrations. Increased probe concentration from 2X to 5X led to the separation of H1 (FAM) and H2 (VIC) probe signals, indicating the cross binding of H1 probe to the H2 allele, and vice versa, is minimised. The horizontal line indicate the identical threshold Ct value **(C)** Serial dilution standard curves generated using H1/H2 heterozygous genomic DNA. The primer efficiencies as detected by the H1 and H2 probes were calculated by the slope of the standard curves. The H1 and H2 probes indicate the genomic primer pair amplification efficiency to be at 96.4% and 94.6% respectively.

The relation between qPCR fluorescent signals and threshold cycle (Ct) values have been previously established to quantitatively describe the relative ratios of transcript levels between two alleles (X. Chen et al. 2008b; H. S. Lo et al. 2003). This approach which describes a linear relation between $\log_2$ allelic DNA ratio and their dCt values ($Ct_{(FAM)} - Ct_{(VIC)}$) (X. Chen et al. 2008b), was adapted here to establish a quantitative linear relation between the amounts of H1 to H2 *MAPT* transcripts, and their corresponding qPCR fluorescent probe signals from FAM (H1) and VIC (H2) fluorophores. By measuring H1 and H2 pMAPT vectors mixed at various ratios (8:1, 4:1, 2:1, 1:1, 1:2, 1:4, 1:8) (Figure 39A), an allele-specific standard curve was generated where the delta Ct values were plotted against the $\log_2$ H1:H2 transcript ratio (Figure 39B). With increasing ratios of H2:H1 DNA concentrations, the VIC fluorescent signal (detecting H2 pMAPT vector) shifts towards a lower Ct value (Figure 39A), while the FAM fluorescent signal (detecting H1 pMAPT vector) shifts towards a higher Ct value (Figure 39A). By regression analysis (Microsoft Excel), the linear line of best-fit for the data points on the standard curve has an r^2 value of 0.9979, where $\log_2$ (H1/H2) = (-1.194 x dCt) + 0.5615 ($p < 0.0001$) (Figure 39B), strongly indicating a linear relationship between $\log_2$ H1:H2 ratios and dCt values. Since a linear relationship is observed here and the slope is constant regardless of the y-intercept, the equation of $\log_2$ (H1/H2) = -1.194 x dCt was employed for the calculations for the $\log_2$ ratios of H1:H2 transcripts from their dCt values.

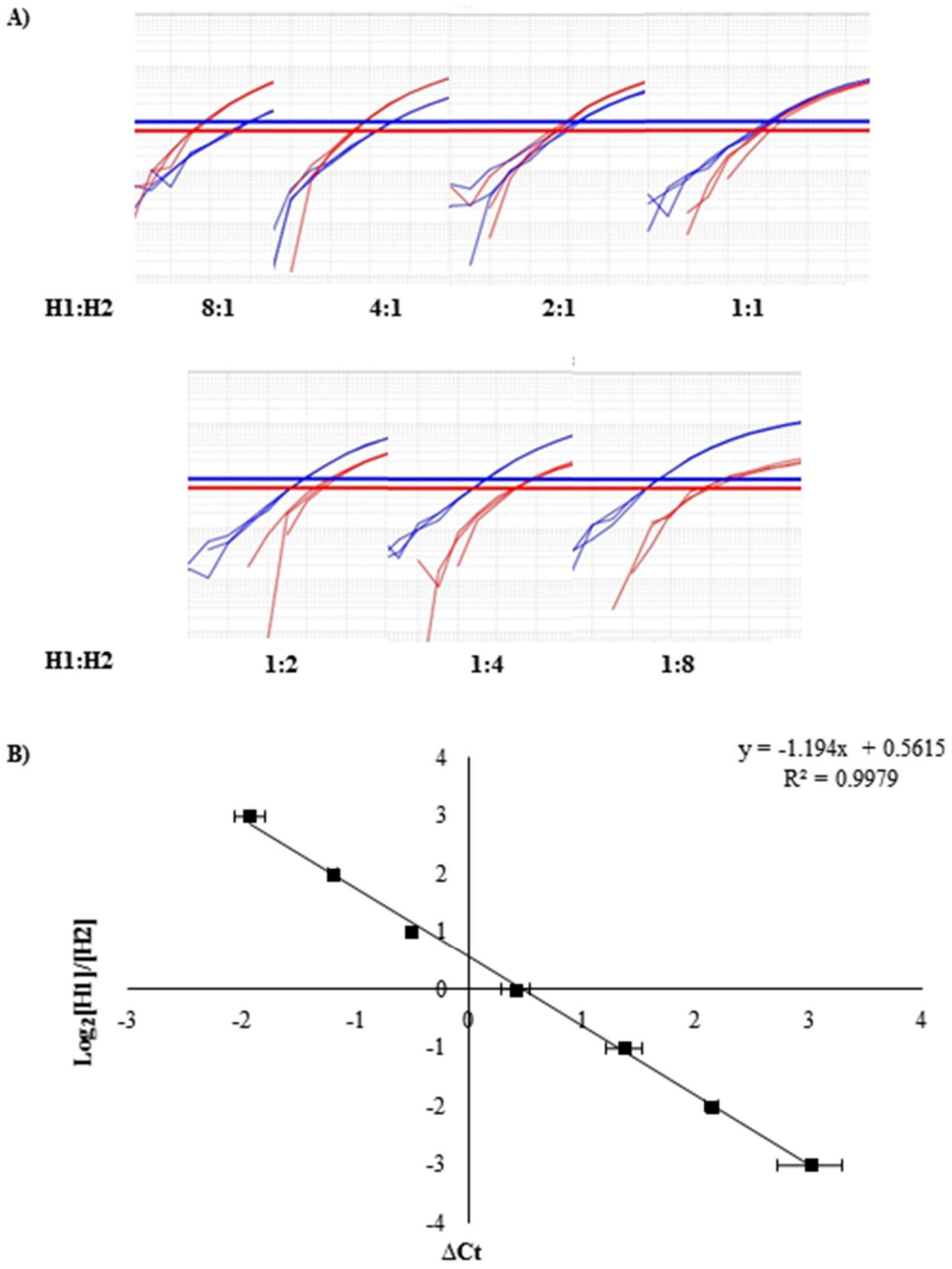


Figure 39. *MAPT* allele-specific expression assay standard curve

(A) pMAPT H1 and H2 DNA standards were mixed in the ratios 8:1, 4:1, 2:1, 1:1, 1:2, 1:4 and 1:8. The H1 probe fluorescent signal (red) shifts from the left to the right with decreasing H1 DNA concentration, and vice versa for the H2 probe (blue) signal. **(B)** A standard curve generated from the pMAPT DNA standards. The H1:H2 ratios are reflected on the standard curve by the delta Ct values between the H1 and H2 signals. The equation for the line of best-fit describes the linear relation between delta Ct values and the $\log_2$ H1:H2 DNA ratios, and can be used for calculating H1:H2 DNA ratios from measured H1 and H2 Ct values in different samples.

6.2.2 Optimisations of qPCR Primers for Measuring Gene Expression Changes

In order to measure the expression levels of different genes in SK-N-F1 cells, the $2^{(-\text{delta delta Ct})}$ method (K. J. Livak and Schmittgen 2001; Schmittgen and Livak 2008) was employed. Briefly, the Ct values for a target gene was normalised to that of a housekeeping gene. This delta Ct value from a sample was then normalised to the delta Ct value of a control sample. The untransfected control has been chosen as our control sample in our following siRNA knockdown experiments. Gene expression levels are expressed as the mean fold change ($2^{-\text{delta delta Ct}}$) values. To increase the robustness in measuring the fold change in gene expression of our target genes, a methodology which takes into account the variability of endogenous control genes was used through performing normalisations to the geometric mean of the levels of multiple housekeeping genes (GAPDH, ACTB and HPRT1) (Vandesompele et al. 2002). Unless otherwise stated, all genes were normalised to the geometric mean of the three housekeeping genes in the following experiments using the $2^{(-\text{delta delta Ct})}$ method. The untransfected cells were used as the control sample, unless otherwise stated, in all our gene expression measurements using the $2^{(-\text{delta delta Ct})}$ method.

To ensure the validity of using the $2^{(-\text{delta delta Ct})}$ method for gene expression, primer efficiencies of all our qPCR primer pairs (Table 6) were ensured to be similar to the controls as determined by constructing serial dilution standard curves (Figure 40). Primer pair efficiency (E) is determined from the slope of the standard curve by the equation $E = (10^{-1/\text{slope}}-1) \times 100$ (Figure 40) (Rutledge and Cote 2003). Primer specificities were determined from the melting curves where a single melting peak of the qPCR products indicate specific amplification (Figure 41A). Amplicon sizes were confirmed by agarose gel electrophoresis (Figure 41B). The primer pairs used here satisfy with the requirements to measure gene expression changes with the $2^{(-\text{delta delta Ct})}$ method (K. J. Livak and Schmittgen 2001). Next, gene expression changes in splicing factors, total and exon 3-containing *MAPT* transcripts, were measured using qPCR.

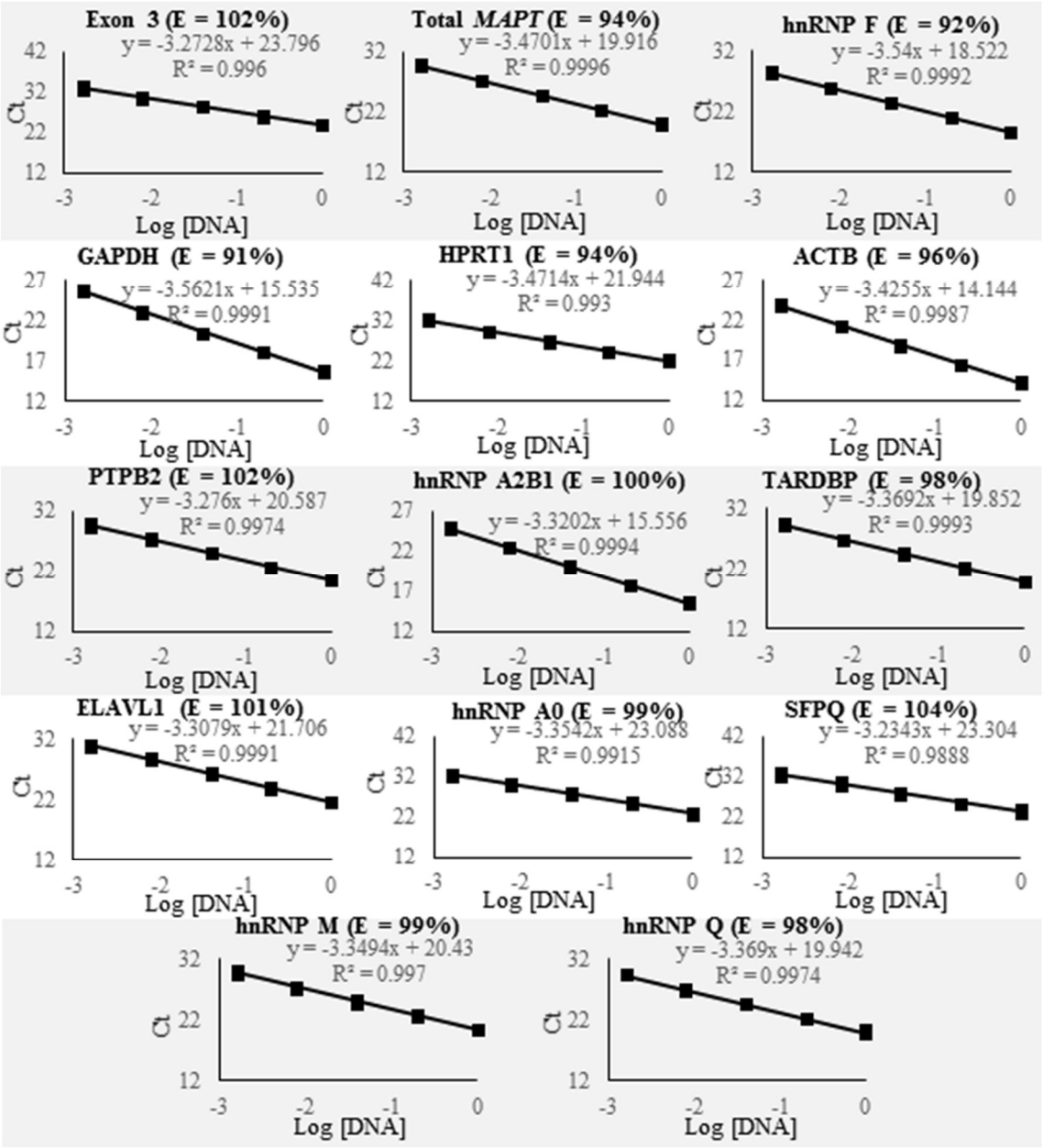


Figure 40. qPCR primer efficiencies determined by serial dilution standard curves

Serial dilution standard curves generated using different primer pairs to measure the gene expression levels of total and exon 3-containing *MAPT* transcripts, hnRNP F, GAPDH, HPRT1, ACTB, PTBP2, hnRNP A2B1, TARDBP, ELAVL1, hnRNP A0, SFPQ, hnRNP M and hnRNP Q. All the primer pairs have high amplification efficiencies of approximately 100% and can be used in the 2 (-delta delta Ct) method for measuring gene expression changes. The efficiencies (E) were determined from the slope of the standard curves by the equation $E = (10^{-1/slope}-1) \times 100$.

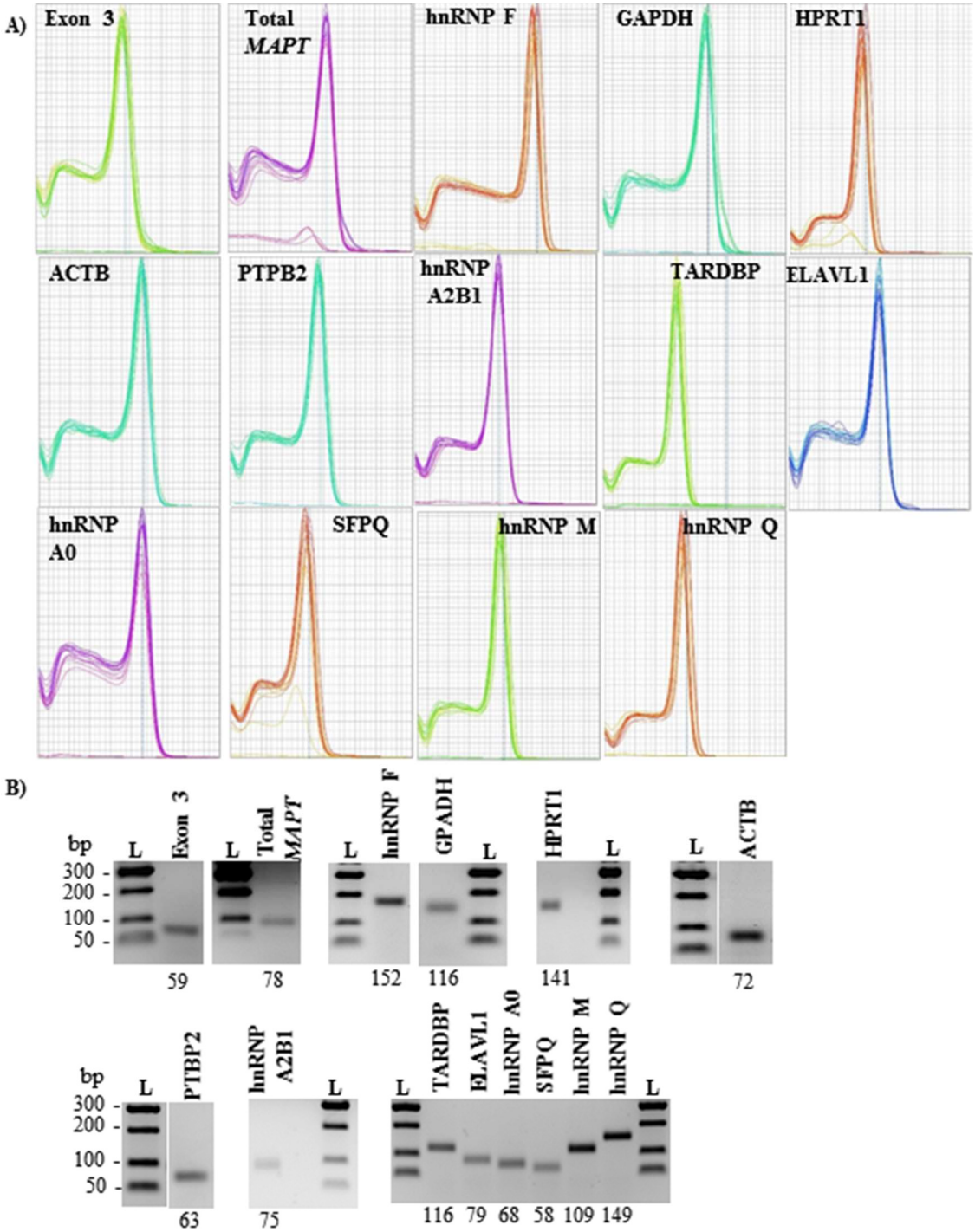


Figure 41. qPCR primer specificities assessed by melting curve and gel electrophoresis

(A) Melting curves of different qPCR primer pairs measuring total and exon 3-containing *MAPT* transcripts, and different splicing factors as indicated by their gene names. A single melting peak indicates specific PCR amplification. **(B)** PCR product amplicon sizes were assessed by gel electrophoresis. Primer pairs were indicated above the lanes and the amplicon size in bp below the lanes. Ladder (L).

6.2.3 Optimisations of siRNA delivery condition in SK-N-F1 cells

To assess siRNA delivery efficiencies by two different commercially transfection reagents Lipofectamine 2000 and Lipofectamine RNAiMAX, we used siRNAs against two candidate splice factors previously identified PTBP2 and hnRNP A2B1 (Table 9) in SK-N-F1 and SK-N-MC cells. The siRNA delivery efficiencies are then inferred from the corresponding gene knockdown efficiencies. We first transfected siRNAs against PTBP2 and hnRNP A2B1 by Lipofectamine 2000 using a range of siRNA concentrations (10 - 200 nM) in both SK-N-F1 and SK-N-MC cells. Gene expression levels of PTBP2 and hnRNP A2B1 were measured by qPCR 48 hours post transfection. The $2^{(-\text{delta delta Ct})}$ method was used to determine gene knockdown levels. The knockdown efficiencies for PTBP2 and hnRNP A2B1, at the various siRNA concentrations used, were approximately 40 % and 60 % respectively (Figure 42A). Next, Lipofectamine RNAiMAX (Invitrogen) was used to deliver 10 nM siRNA in SK-N-F1 and SK-N-MC cells. Gene knockdown efficiencies for PTBP2 and hnRNP A2B1 were approximately 60 % and 85 % respectively (Figure 42B). The results here indicate that siRNA delivery in SK-N-F1 and SK-N-MC cells is more effective using Lipofectamine RNAiMAX when compared to Lipofectamine 2000 as reflected in the higher gene knockdown levels. All the following siRNA knockdown experiments were therefore performed using Lipofectamine RNAiMAX.

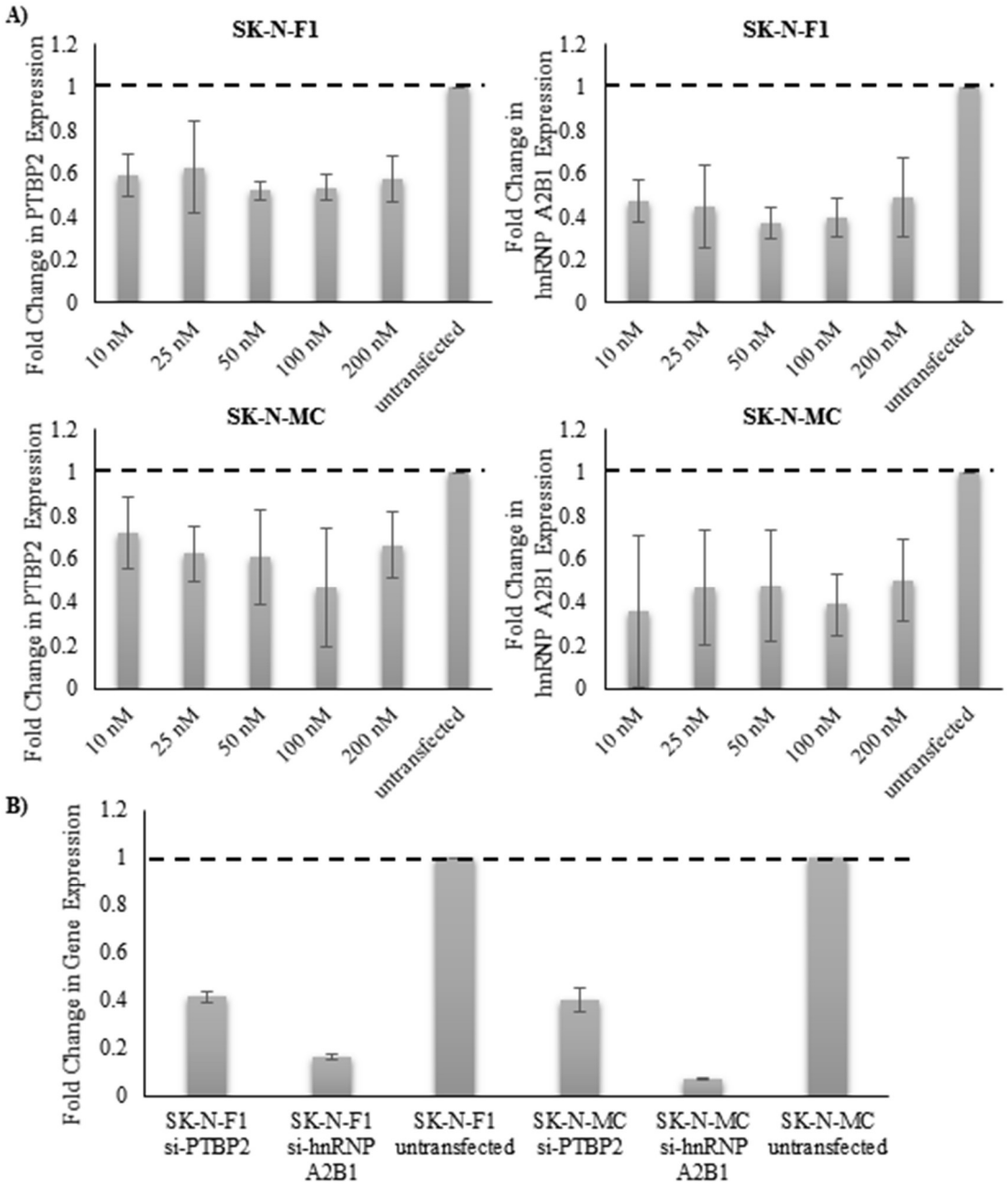


Figure 42. Lipofectamine RNAiMAX delivers siRNA more efficiently than Lipofectamine 2000

(A) Gene knockdown using 10 – 200 nM siRNAs by Lipofectamine 2000 transfection reagent in SK-N-F1 and SK-N-MC cells. Gene knockdown levels were expressed as the fold change in gene expression levels compared to the untransfected control. PTBP2 and hnRNP A2B1 knockdown levels were approximately 40% and 60%, respectively. Error bars represent standard deviation (n=3). **(B)** Gene knockdown using 10 nM siRNAs by Lipofectamine RNAiMAX transfection reagent in SK-N-F1 and SK-N-MC cells. Gene knockdown levels were expressed as the fold change in gene expression levels compared to the untransfected control. PTBP2 and hnRNP A2B1 knockdown levels were approximately 60% - 85%, respectively. Error bars represent standard deviation (n=2).

Cellular mRNA levels have previously been shown to significantly correlate with protein levels but not with protein half-lives (Schwanhäusser et al. 2011), while many proteins generally have a half-life of approximately 2 days (Sandoval et al. 2013). We wish to ensure sufficient time is allowed for protein levels to equilibrate after a reduction in their mRNA levels by increasing the time of gene knockdown from 2 to 4 days. To determine if gene expression levels will remain knocked down 4 days post transfection, we performed knockdown experiments using siRNA against PTBP2, hnRNP A2B1 and hnRNP F in SK-N-F1 and SK-N-MC cells. Gene expression levels remained knocked down 4 days post transfection when compared to 2 days (Figure 43). Hence, all knockdown experiments were performed with a 4 days siRNA incubation period to allow for sufficient time for the protein levels to equilibrate.

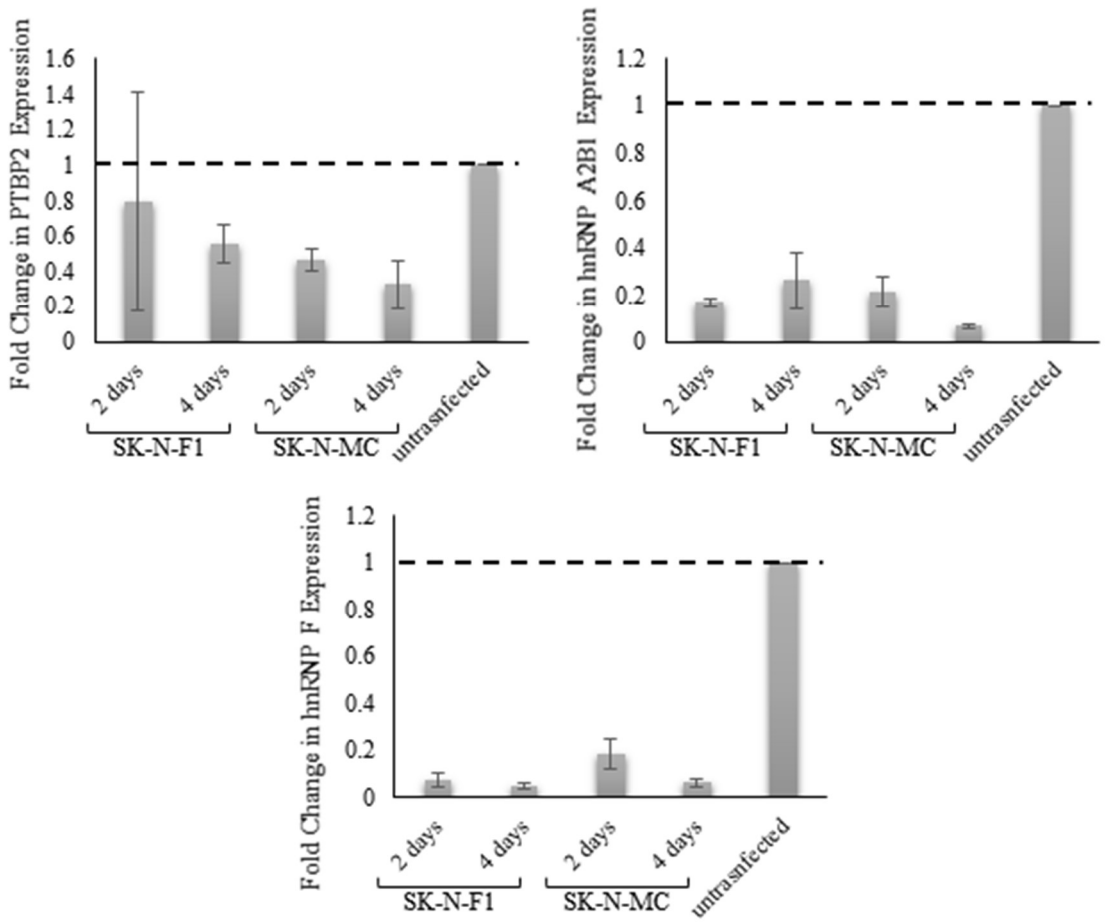


Figure 43. Gene expression remains knocked down 4 days post transfection.

Gene knockdown using siRNA against PTBP2, hnRNP A2B1 and hnRNP F for 2 and 4 days. Gene expression remained knocked down 4 days post transfection compared to 2. Error bars = S.D. (n=3).

6.2.4 siRNA Knockdown of Splice Factors in SK-N-F1 Neuroblastoma Cells

Through RNA-protein pull-down experiments (Section 5.2.7, Figure 34), nine splicing factors that interact with the H1 and H2 rs17651213 sequences with different affinities were identified (Table 9). To study how the nine splicing factors (SFPQ, TARDBP, hnRNP M, ELAVL1, hnRNP A0, hnRNP Q, hnRNP A2B1, PTBP2 and hnRNP F) regulate the inclusion of exon 3 from the H1 and H2 *MAPT* alleles, gene knockdown of the splicing factors in SK-N-F1 cells was performed using commercially available siRNAs. The fold change in gene expression was determined by the $2^{(-\text{delta delta Ct})}$ method using the optimised qPCR assays (Figure 40 & Figure 41). Gene knockdown levels of over 80 % were achieved for all the splicing factors except for PTBP2 and hnRNP M where the knockdown efficiencies are approximately 60 % (Figure 44). The results here showed high gene knockdown levels of the splice factors were achieved in SK-N-F1 cells. We next investigated the effects of splice factor knockdown on the 1) allele-specific expression of total *MAPT* and *MAPT* exon 3; and 2) overall expression of total *MAPT* and *MAPT* exon 3.

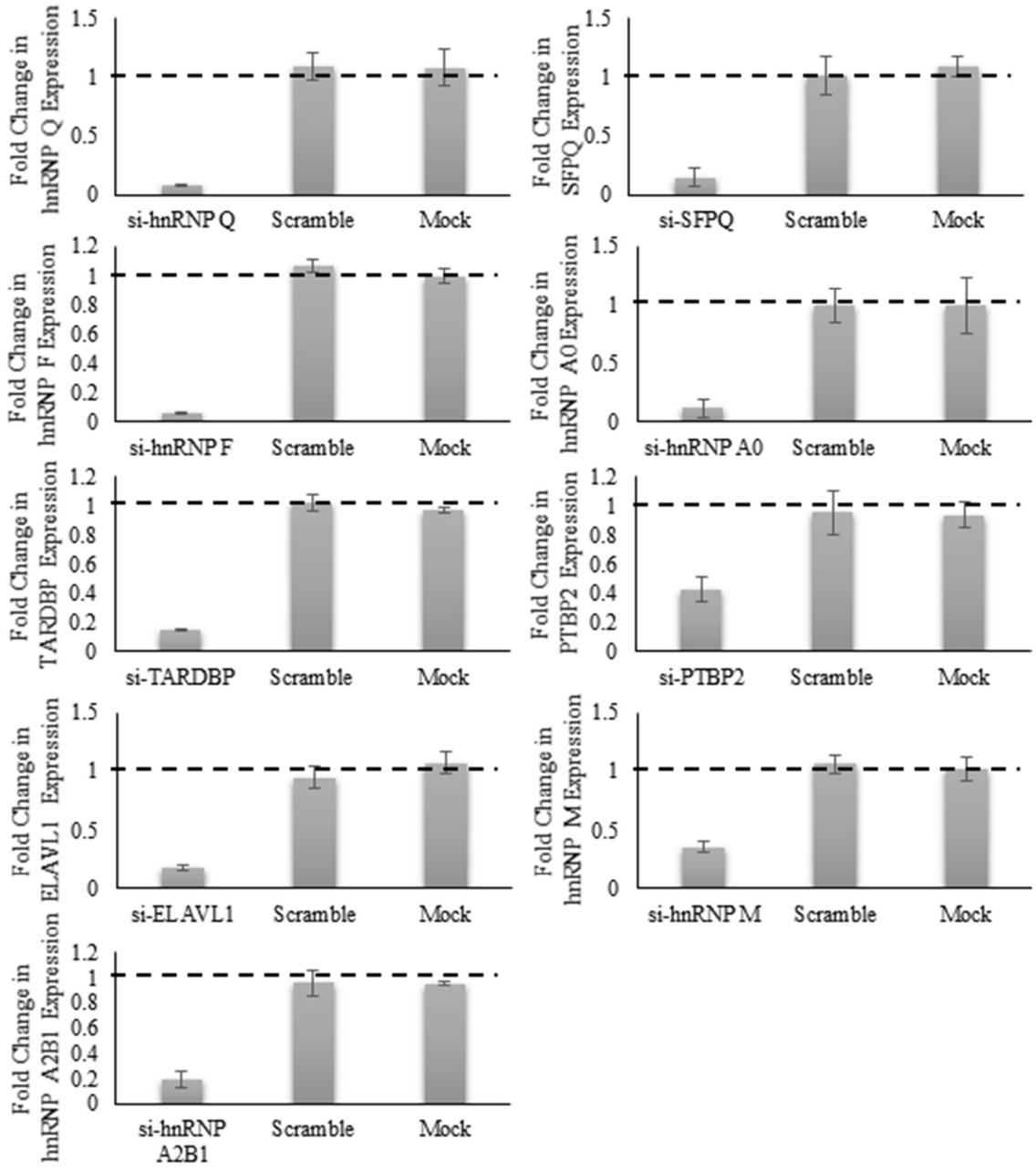


Figure 44. High gene knockdown efficiencies of the candidate splice factors in SK-N-F1 cells. siRNA knockdown of nine candidate splice factors in SK-N-F1 cells. Gene expression levels in cells treated with siRNAs against splice factors, scramble siRNA (negative control) and in the mock transfections were assessed using the 2(-delta delta Ct) method. Gene knockdown levels of over 80% were achieved in cells treated with si-hnRNP Q, si-SFPQ, si-hnRNP F, si-hnRNP A0, si-TARDBP, si-ELAVL1 and si-hnRNP A2B1. Gene knockdown levels of approximately 60% were achieved in cells treated with si-PTBP2 and si-hnRNP M. Error bars represent standard deviation (n=3).

6.2.5 hnRNP Q and hnRNP F Regulate the Allele-Specific Expression of *MAPT* Exon 3

Using the quantitative allele-specific expression qPCR assay developed (Figure 36 - Figure 39), the H1:H2 *MAPT* exon 3 and H1:H2 total *MAPT* transcript ratios (Figure 45) were measured in the splice factor knockdown experiments. Gene expression in the samples were normalised to the untransfected controls and were expressed as the normalised mean fold change in the H1:H2 transcript ratios. Here, a significant increase in the H1:H2 exon 3 transcript ratios was observed when hnRNP Q (H1:H2 1.28 ± 0.26, $p < 0.05$) and hnRNP F (H1:H2 1.47 ± 0.19, $p < 0.0001$) were knocked down compared to the mock transfection control (H1:H2 0.96 ± 0.08) (Figure 45A). To ensure that the change in exon 3 H1:H2 transcript ratios observed is not a result from a change in total *MAPT* ratios, the H1:H2 total *MAPT* transcript ratios were also measured using the allele-specific assay developed (Figure 45B). Significant changes in the total H1:H2 *MAPT* transcript ratios were not observed when hnRNP Q and hnRNP F were being knocked down. Here, the results showed that the silencing of hnRNP Q and hnRNP F led to an increase in the H1:H2 *MAPT* exon 3 transcript ratio, indicating that hnRNP Q and hnRNP F play a role in promoting exon 3 inclusion in the H1 allele, and/or decreasing exon 3 inclusion in the H2 allele.

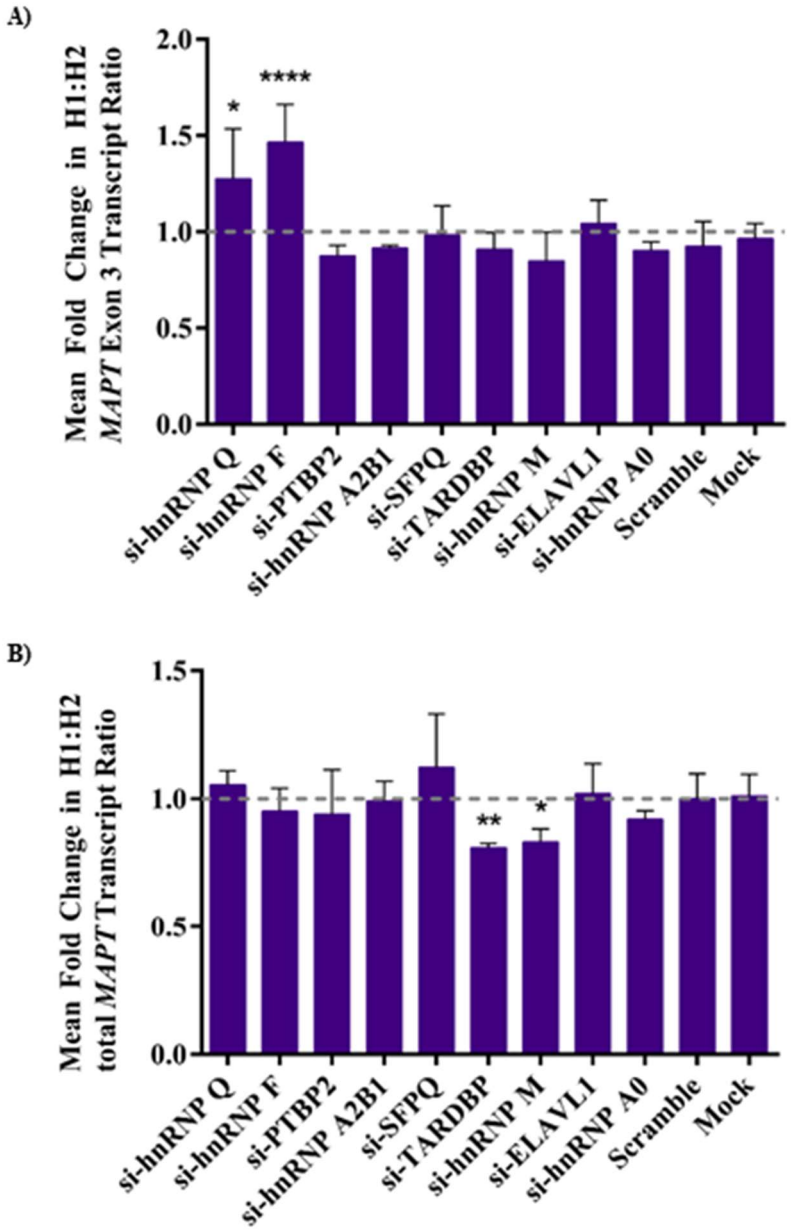


Figure 45. hnRNP Q and hnRNP F regulate the haplotype-specific inclusion of *MAPT* exon 3

(A) Mean fold change in the normalised H1:H2 *MAPT* exon 3 transcript ratios in SK-N-F1 cells transfected with siRNA against different splice factors (n=3-6), scramble siRNA (n=8) and mock transfections (n=8). hnRNP F (H1:H2 1.47 ± 0.19, $p<0.0001$), hnRNP Q (H1:H2 1.28 ± 0.26, $p<0.05$). Statistical significance was determined by comparing each column to the mock control using one way ANOVA with Bonferroni correction. Error bars represent standard deviation. **(B)** Mean fold change in the normalised H1:H2 total *MAPT* transcript ratios in SK-N-F1 cells transfected with siRNA against different splice factors (n=3-6), scramble siRNA (n=8) and mock transfections (n=8). Columns were compared to the mock control by one way ANOVA with Bonferroni correction (**$p<0.01$, *$p<0.05$). Error bars represent standard deviation.

6.2.6 hnRNP F Regulates Both Exon 3 Inclusion and Total *MAPT* Expression

The splice factors hnRNP F and hnRNP Q were identified in the previous section as allele-specific regulators of *MAPT* exon 3 inclusion (Figure 45). In order to further study the effects of their knockdown on the overall expression of total *MAPT* and *MAPT* exon 3 expression, *MAPT* exon 3 (Figure 46A) and total *MAPT* expression (Figure 46B) were measured by qPCR using the 2(-delta delta Ct) method. To prevent the caveat of observing exon 3 changes as a result of changes in total *MAPT* levels, exon 3-containing transcript levels were normalised to total *MAPT* levels instead of the geometric mean of the levels of multiple housekeeping genes.

The knockdown of SFPQ led to a 50 % reduction in total *MAPT* level when compared to the mock control (si-SFPQ 0.46 ± 0.05; mock 0.91 ± 0.09, $p < 0.0001$) (Figure 46B), while exon 3 inclusion level has not been altered (si-SFPQ 1.14 ± 0.18; mock 1.14 ± 0.2) (Figure 46A). Similarly, the knockdown of hnRNP M led to a 2-fold increase in exon 3 inclusion compared to the mock control (si-hnRNP M 2.16 ± 0.73; mock 1.14 ± 0.2, $p < 0.0001$) (Figure 46A), while total *MAPT* level remains unchanged (si-hnRNP M 1.01 ± 0.05; mock 0.91 ± 0.09) (Figure 46B). The two examples here illustrate that a change in total *MAPT* level is an independent event from a change in exon 3 inclusion level. The gene normalisation strategy for measuring exon 3-containing transcripts allows us to account for changes in the total *MAPT* level, preventing a false positive observation of changes in exon 3 inclusion as a result of changes in the total *MAPT* expression.

The silencing of hnRNP F was found leading to a reduction in exon 3 transcript levels when compared to the mock transfection control (si-hnRNP F 0.54 ± 0.03; mock 1.14 ± 0.2, $p < 0.05$) (Figure 46A). A reduction in total *MAPT* expression was also observed when hnRNP F was silenced, as compared to the mock transfection control (si-hnRNP F 0.65 ± 0.02; mock 0.91 ± 0.09, $p < 0.05$) (Figure 46B). The reduction in total *MAPT* expression could be caused by a reduction in exon 3 levels. However, exon 3 is a minor isoform that makes up less than 10% of the total transcripts and is unlikely to explain the 30% reduction in total *MAPT* level observed here. The results here showed hnRNP F regulates both the expression of total *MAPT* and the inclusion exon 3-containing

transcripts in SK-N-F1 cells. Interestingly, significant changes in exon 3 inclusion and total *MAPT* expression were not observed when hnRNP Q was knockdown.

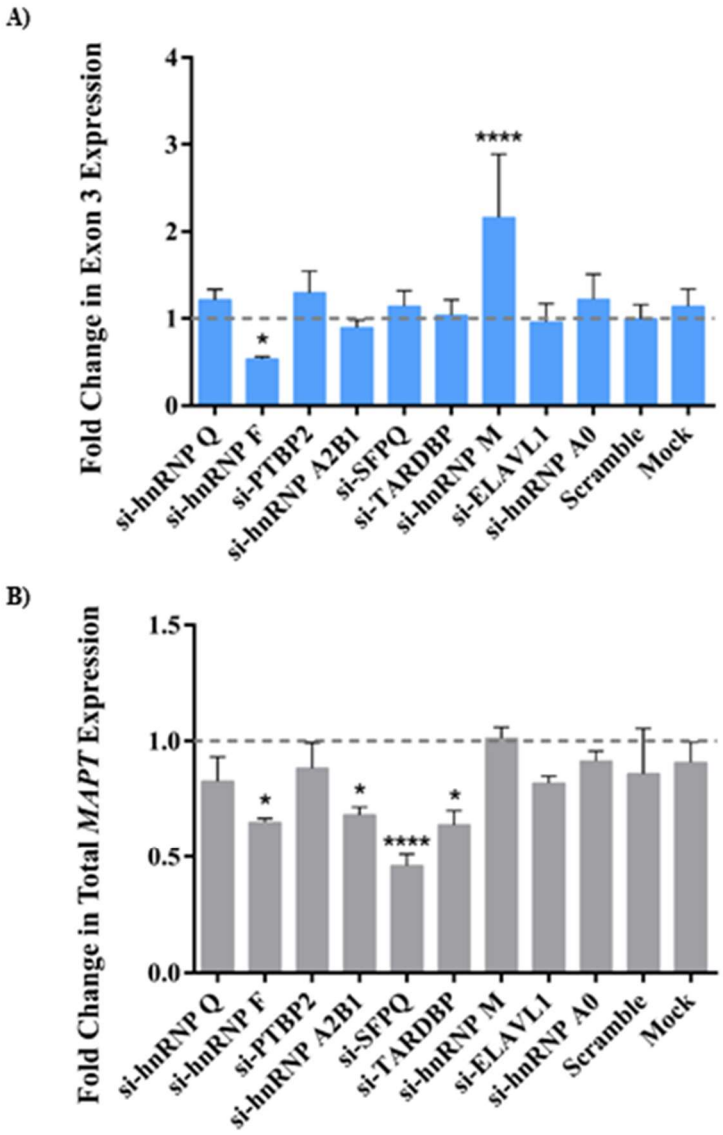


Figure 46. Splice factor knockdown on total *MAPT* and exon 3 expression.

(A) *MAPT* exon 3 levels normalised to total *MAPT* expression in SK-N-F1 cells. Gene knockdown of hnRNP F and hnRNP M led to a 2-fold reduction (si-hnRNP F 0.54 ± 0.03, $p<0.05$) and increase (si-hnRNP M 2.16 ± 0.73, $p<0.0001$) in exon 3 levels when compared to the mock control. Error bars indicate standard deviation (n=3-8). **(B)** Total *MAPT* expression normalised to endogenous control genes in SK-N-F1 cells. Gene knockdown of hnRNP F (si-hnRNP F 0.83 ± 0.11, $p<0.05$), hnRNP A2B1 (si-hnRNP A2B1 0.68 ± 0.03, $p<0.05$), SFPQ (si-SFPQ 0.46 ± 0.05, $p<0.0001$) and TARDBP (si-TARDBP 0.64 ± 0.06, $p<0.05$) led to a reduction in total *MAPT* expression as compared to the mock control. Error bars indicate standard deviation (n=3-8).

6.2.7 hnRNP Q and hnRNP F Interact with rs1800547 and rs17651213 Respectively

To provide mechanistic insights into how hnRNP Q and hnRNP F regulate exon 3 inclusion, RNA-EMSA supershift assays were performed to determine their interactions with rs1800547 and rs17651213 sequences. In supershift assays, specific antibodies against hnRNP Q and hnRNP F were incubated with SK-N-F1 nuclear lysate the addition of RNA probes in the binding reactions. Due to the extra mass of the antibodies, the interaction of hnRNP Q and hnRNP F with the RNA probes will be observed a further shift upwards (supershift) as compare to the RNA-protein complex observed without the antibodies. Here, the biotin labelled RNA probes that span H1 and H2 rs1800547 and rs17651213 sequences and the SK-N-F1 nuclear lysate were used.

Using rs17651213 H1 probe and anti-hnRNP Q antibody, several optimisation steps were performed for the supershift binding reaction conditions. Different amounts of tRNA (non-specific competitor) were titrated in the binding reactions to determining their effects on the supershift. Different tRNA concentrations affected the profiles of the RNA-protein complex formed (Figure 47A, lanes 1, 3, 4, 6, 7, 9, 10 & 12), but not the true supershift signal (Figure 47A, lanes 2, 5, 8 & 11). It is also evident that a significant amount of RNase activity is present in the hnRNP Q antibody as observed by the disappearance of the RNA probe signal (Figure 47A, lanes 2, 5, 8 & 11). To minimise RNase activity, the effects of RiboLock RNase inhibitor (ThermoFisher), RNaseOUT recombinant ribonuclease inhibitor (ThermoFisher) and Protector RNase inhibitor (Roche) in the binding reactions were determined (Figure 47B, lanes 2-5). Comparatively, Protector RNase inhibitor showed the most potent RNase inhibition activity at the tested condition as observed by the least amount of RNA-protein complex formed by the degraded RNA species (Figure 47B, lane 4). Increasing amounts (1 µL, 3 µL and 5 µL) of Protector RNase inhibitor were next titrated to determine the optimal concentration for maximum RNase inhibition activity (Figure 47C). The formation of RNA-protein complex by degraded RNA species was minimal in the presence of 5 µL of RNase inhibitor without affecting the true supershift signal (Figure 47C, lane 5). It was also found that 5 µg antibody was necessary for the supershift to be observed as lower amounts of antibody

did not lead to a supershift (Figure 47B, lanes 2-5; Figure 47D, lanes 2-4). The optimal RNA-EMSA supershift binding reaction condition was determined here for studying the binding of hnRNP Q and hnRNP F to rs17651213 and rs1800547 H1 and H2 sequences.

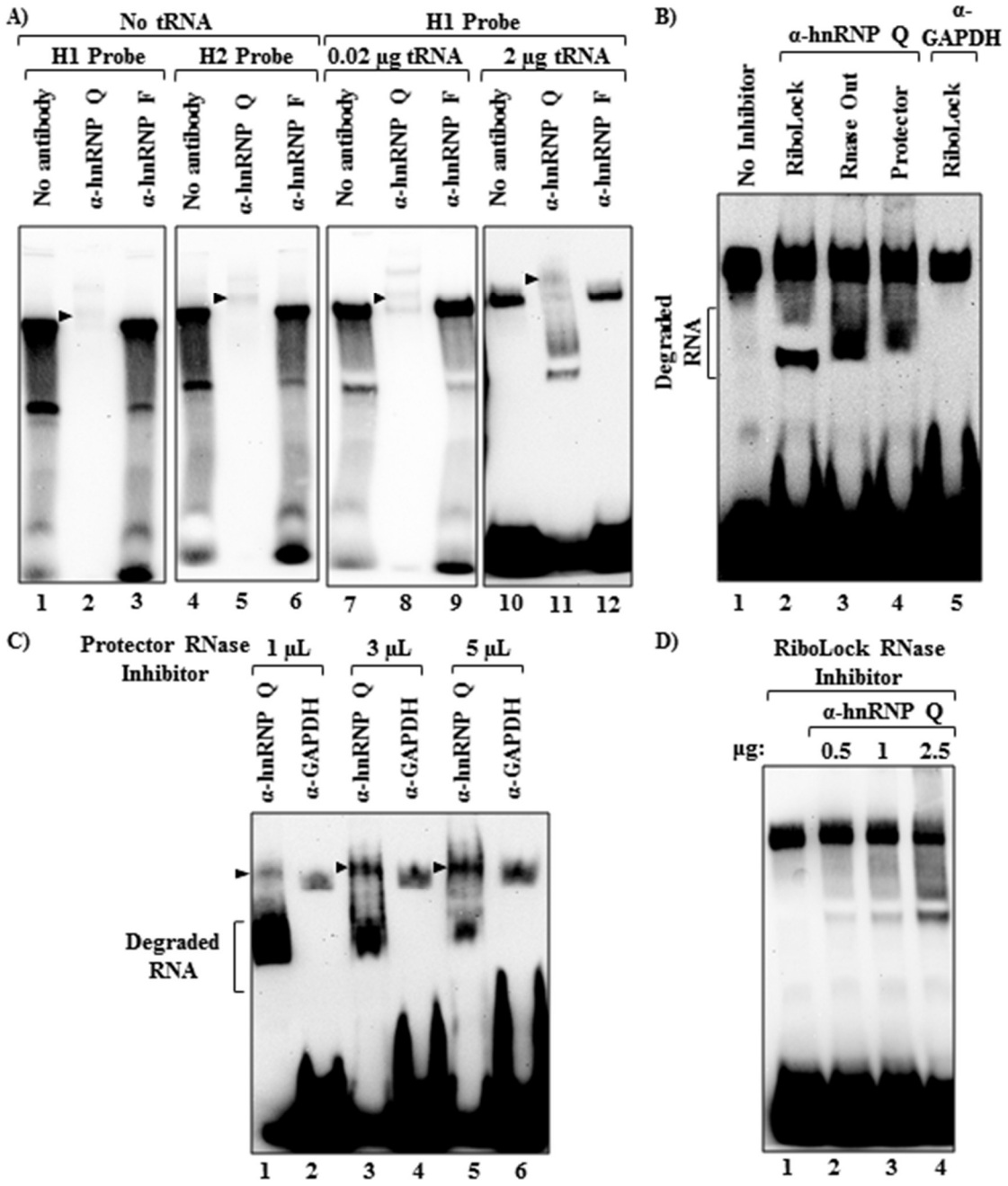


Figure 47. Optimisations of RNA-EMSA supershift binding reaction condition with rs17651213 H1 and H2 probes.

(A) Titrating tRNA concentrations in supershift reactions. No tRNA (lanes 1-6), 0.02 µg (lanes 7-9) and 2 µg tRNA (lanes 10-12). tRNA concentrations affected the profiles of RNA-protein complex but not the true supershift signal (black arrow) (lanes 2, 5, 8 & 11). 5 µg antibody per lane. **(B)** Determining the effects of RiboLock, RNase Out and Protector RNase inhibitors. Anti-GAPDH antibody as a positive control as it contains minimal amount of RNase activity. 2.5 µg antibody per lane. **(C)** Titrating Protector RNase inhibitor concentrations from 1 µL to 5 µL. Black arrows indicate the supershift. 5 µg antibody per lane. **(D)** Titrating antibody concentrations from 0.5 µg to 2.5 µg.

Following binding reaction condition optimisation, the interaction of hnRNP Q and hnRNP F, two splice factors that showed allele-specific regulation of exon 3 inclusion (Figure 45), with rs17651213 H1 and H2 sequences were studied by RNA-EMSA supershift assays. hnRNP Q was found to interact with rs17651213 H1 and H2 probe sequences (Figure 48A, lanes 2 & 6) but not hnRNP F (Figure 48A, lanes 3 & 7). In one of the previous chapters, hnRNP F was predicted in computational analysis to interact with rs1800547 H1 and H2 sequences (Figure 7). To understand how hnRNP F may affect the allele-specific splicing of exon 3 (Figure 45), the interaction of hnRNP Q and hnRNP F with rs1800547 H1 and H2 sequences were studied using RNA-EMSA. It was found that rs1800547 interacts with hnRNP F (Figure 48B, lanes 2 & 5) but not hnRNP Q (Figure 48B, lanes 3 & 6). The results here showed that hnRNP Q and hnRNP F interact with rs17651213 and rs1800547 respectively.

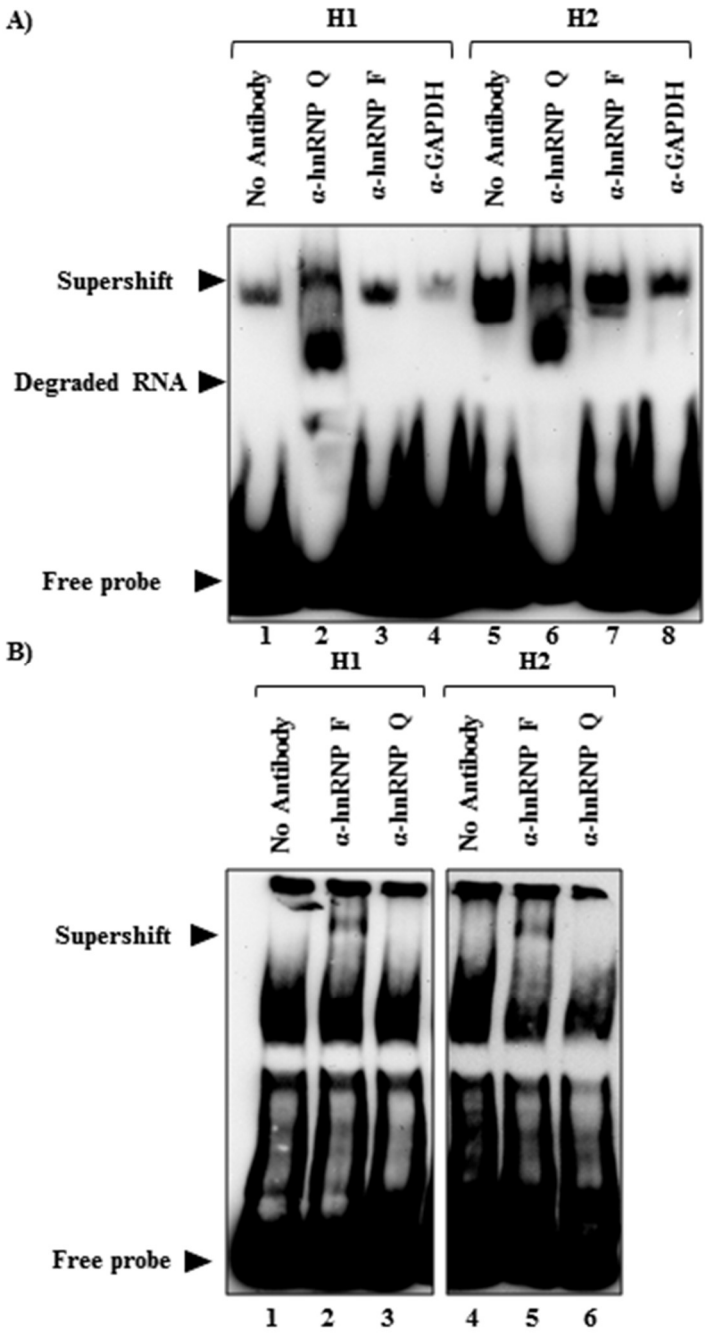


Figure 48. hnRNP Q and hnRNP F interact with SNPs rs17651213 and rs1800547 respectively. RNA-EMSA supershift assay to confirm the binding of hnRNP Q and hnRNP F to rs17651213 and rs1800547 H1 and H2 SNPs. (A) The binding of hnRNP Q to rs17651213 H1 and H2 sequences (lanes 2&6) are observed as a supershift by the anti-hnRNP Q antibody as compared to no antibody controls (lanes 1&5). Anti-GAPDH and anti-hnRNP F are IgG controls (lanes 4&8). (B) The binding of hnRNP F to rs1800547 H1 and H2 sequences (lanes 2&5) are observed as a supershift by the anti-hnRNP F antibody as compared to no antibody controls (lanes 1&4). Anti-hnRNP Q as IgG control (lanes 3&6).

6.2.8 Investigating the Allele-Specific Effects of RNA Polymerase II Elongation on *MAPT* Exon 3 Inclusion

To study if RNAPII elongation rate regulate *MAPT* exon 3 splicing in an allele-specific fashion, the compound camptothecin was employed to alter RNAPII elongation speed. Camptothecin is a non-competitive reversible inhibitor of topoisomerase I, which results in the slowing down RNAPII elongation rate from approximately 3.8 kb min^{-1} to 1 kb min^{-1} (Capranico et al. 2007; Darzacq et al. 2007; Singh and Padgett 2009). Camptothecin was applied to BE(2)-M17, SK-N-F1 and SK-N-MC neuroblastoma cells to determine its effects on *MAPT* exon 3 expression. *MAPT* exon 3 inclusion levels was determined by qPCR and was expressed as the fold change in exon 3 expression (Figure 49) using the 2(-delta delta Ct) method as previously described (Figure 46). Camptothecin was found to reduce the levels of *MAPT* exon 3 inclusion in all three cell lines, suggesting that RNAPII elongation rate plays a role in regulating the inclusion of exon 3 (Figure 49A).

The ratios of H1:H2 exon 3 and total *MAPT* transcripts were next measured using the allele-specific assay developed (Figure 36 - Figure 39). Allele-specific *MAPT* expression was normalised to untreated cells and was expressed as the mean fold change in H1:H2 *MAPT* transcript ratio (Figure 49B). A significant change in the H1:H2 ratios of both exon 3-containing and total *MAPT* transcript in camptothecin treated cells was not observed when compared to cells treated with the vector DMSO. However, more experiments will be required to dissect the effects and the underlying mechanisms of RNAPII elongation kinetics on regulating the inclusion of exon 3 from the H1 and H2 alleles. The preliminary data here indicates that the splicing of *MAPT* exon 3 is coupled to RNAPII elongation rate and that RNAPII elongation rate appears to have similar effects on exon 3 inclusion on both H1 and H2 alleles.

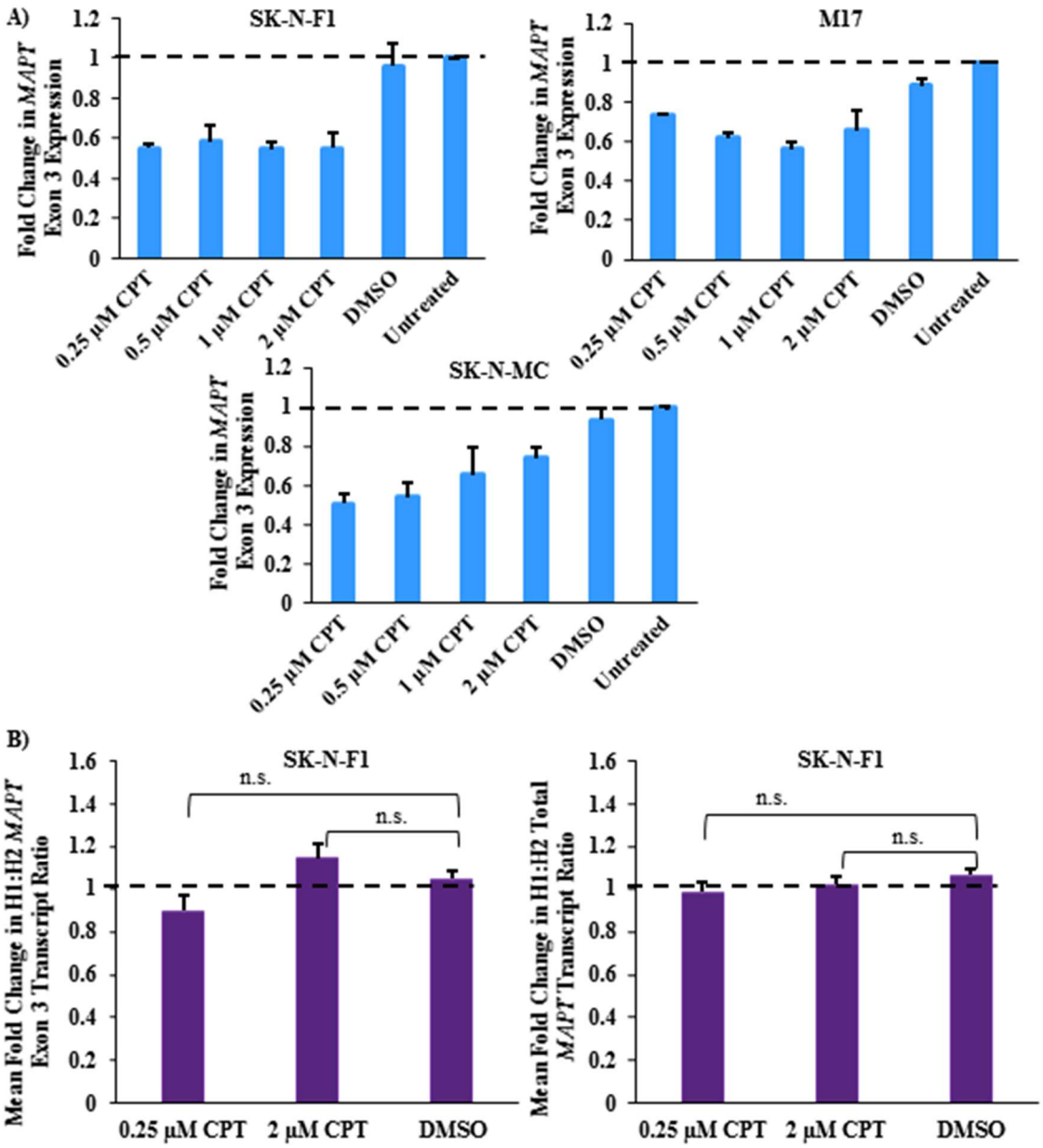


Figure 49. Camptothecin reduced exon 3 inclusion rate.

(A) The effects of camptothecin (CPT) on *MAPT* exon 3 inclusion rate in SK-N-F1, SK-N-MC and BE(2)-M17 cells. CPT led to a reduction in exon 3 inclusion rate. Error bars indicate standard deviation (SK-N-F1 n=2; BE(2)-M17 n=2; SK-N-MC n=3). **(B)** Mean fold change in the normalised H1:H2 total and exon 3-containing *MAPT* transcript ratios in SK-N-F1 cells treated with camptothecin. Error bars indicate standard deviation (n=3).

6.3 Discussion

6.3.1 hnRNP Q and hnRNP F Regulate the Haplotype-Specific Inclusion of Exon 3

The aim of this final results chapter of the thesis is to study the allele-specific regulation of *MAPT* exon 3 inclusion by splice factors in SK-N-F1 neuroblastoma cells. To study allele-specific exon 3 expression, an allele-specific qPCR assay that allows us to detect and quantify the ratio of H1 vs H2 *MAPT* transcripts in our samples was first developed. Gene knockdown by siRNA of the nine candidate splice factors (Table 9) were performed in SK-N-F1 neuroblastoma cells. The reduction in the expression levels of hnRNP Q and hnRNP F resulted in an increase in the H1:H2 exon 3 transcript ratio, indicating they are allele-specific splicing regulators of *MAPT* exon 3. Confirmation of hnRNP Q and hnRNP F interactions with rs17651213 and rs1800547 SNP sequences, respectively, was carried out using RNA-EMSA supershift assays. In this study, haplotype-specific differences were demonstrated in *MAPT* exon 3 splicing by hnRNP Q and hnRNP F using a novel allele-specific Taqman assay and the splice factors are likely to exert their regulatory effects through intronic SNPs rs17651213 and rs1800547.

6.3.2 Detecting Allele-Specific Alternative Splicing Events

Previous microarray based experiments have shown that allelic differences in gene expression levels associated with SNPs are often high in magnitude, with an average difference of 2.6-fold detected (Ge et al. 2009). The high magnitude of differences in gene expression levels is easily detectable using Taqman assays as quantified by the differences in Taqman probe fluorescence (Arai et al. 2011; X. Chen et al. 2008b; Harmon et al. 2010; H. S. Lo et al. 2003; Snow et al. 2015). However, the authors have highlighted limited assay sensitivity as observed in the inaccurate differences in Taqman probes fluorescence measured at different allelic template ratios, which was also observed in our assay standard curve (X. Chen et al. 2008b; Harmon et al. 2010). Subtle changes in allelic gene expression and splicing differences that could be of physiological importance may therefore be more difficult to be detected using the Taqman approach. Despite the potentially reduced sensitivity in the detection of allele-specific differences, allele-specific effects on exon 3

inclusion by hnRNP Q and hnRNP F were successfully identified with reproducible results, demonstrating that Taqman assays can be used to measure allele-specific splicing events. For more subtle allele-specific changes, the quantification of alleles by mass spectrometry methods could be used as they have high sensitivity and accuracy (Braun et al. 1997; Jurinke et al. 2001). More global methodologies using allele-specific hybridisation microarrays and/or RNA sequencing (RNA-seq) can also provide accurate measurement of allelic imbalance with high sensitivity and accuracy (Crowley et al. 2015; DeVeale et al. 2012; Ge et al. 2009). These methods however, would not be cost effective for studying a small number genes or transcript isoforms.

6.3.3 Investigating RNA Polymerase II Elongation Rate on the Allelic Expression of *MAPT* Exon 3

Preliminary data here suggests slowing down RNAPII elongation rate appears to have similar effects on exon 3 inclusion from both the H1 and H2 alleles. However, more experiments will be required to obtain more conclusive evidence. The Taqman assay can be applied for studying further if there is any allele-specific effect by RNAPII elongation speed on exon 3 inclusion. Time point experiments that measure how quickly RNAPII reaches different positions on a gene could be used to determine whether if RNAPII intrinsically processes the H1 and H2 transcripts at different speeds. For example, a reversible inhibitor of RNAPII, 5.6-dichlorobenzimidazole 1-β-D-ribofuranoside (DRB), can be applied to cells for halting transcription and its removal will lead to synchronised transcription and RNA processing (Singh and Padgett 2009). The synthesis of *MAPT* exon 3 pre-mRNA transcripts can then be measured at different time points by our allele-specific assay. If RNAPII does process the H1 and H2 alleles at different rate, a change in H1:H2 ratios of exon 3 transcripts maybe observed at different time points, indicating that RNAPII assembles or reaches exon 3 with different speeds. Other compounds that alter RNAPII elongation speed, such as trichostatin A and camptothecin, could be further employed for together with our allele-specific Taqman assay to study how RNAPII speed influence exon 3 production from the H1 and H2 alleles (Dujardin et al. 2014; Singh and Padgett 2009).

As mentioned previously the Taqman based approach may lack sensitivity. Since the pausing of RNAPII along pre-mRNA transcripts has been demonstrated to directly correlate with transcriptional output as measured by qPCR (Singh and Padgett 2009), ChIP-seq type methods will be incredibly useful in studies of this kind where the preservation of *in vivo* protein-DNA interaction state is preferred. The HaploChIP method could be employed to study how RNAPII distribute on the *MAPT* gene of H1 and H2 haplotypes (Knight et al. 2003). The higher or lower RNAPII density at regulatory sites between the H1 and H2 alleles, such as rs17651213 or rs1800547, may provide mechanistic insights into how different allelic sequences could lead to differential occupation of RNAPII along functional regions on two alleles of the *MAPT* gene. Unlike allele-specific ChIP however, there is currently no method developed yet for quantifying the allele-specific protein loading on RNA molecules *in vivo*. Therefore studying how hnRNP Q and hnRNP F assemble on rs17651213 and rs1800547 sequences on *MAPT* pre-mRNA transcripts will still require indirect *in vitro* approaches such as RNA-EMSA or RNA-affinity pull-down methods, or cell based methods such as overexpressing or knocking down the splice factor of interest.

6.3.4 Regulation of Exon 3 Inclusion by hnRNP Q and hnRNP F

Exon 3 is expressed at a higher level from the H2 allele specifically in the grey matter of the brain but not in the white matter or the peripheral neurons (Trabzuni et al. 2012). Since both hnRNP F and hnRNP Q are ubiquitously expressed in the human body (Honore et al. 1995; Lau et al. 2001), it would be possible that other neuron specific or enriched/depleted proteins are involved with the interactions with hnRNP F and hnRNP Q to regulate their neuronal allele-specific exon 3 inclusion activities. For example, the neuron specific splice factors PTB and Nova-1 interact with each other to mediate the expression of brain specific exons (Polydorides et al. 2000). It has also been shown that PTB protein alone has the ability to differentiate between neuronal and non-neuronal sequences (L. Zhang et al. 1999). hnRNP F has been also shown to directly regulate a neural-specific splicing event where it is coupled to a neuron-specific protein complex (Min et al. 1995).

The actual mechanisms by which hnRNP Q and hnRNP F regulate neuronal allele-specific *MAPT* exon 3 inclusion is likely to be much more complicated. hnRNP F has previously been demonstrated to associate with transcriptional activities (Garneau et al. 2005; C. S. Lo et al. 2012; C. S. Lo et al. 2015) and to bind cap-binding protein complex (Gamberi et al. 1997). Since transcription is usually coupled with pre-mRNA splicing (de la Mata et al. 2003; Dujardin et al. 2014), and the capping of mRNA interacts with splicing and polyadenylation (Proudfoot et al. 2002), the fact that hnRNP F interacts with the above processes would imply that hnRNP F mediated splicing is likely to be regulated at many levels. It is interesting to note that apart from the splicing activity of hnRNP Q (H. H. Chen et al. 2008a; Mourelatos et al. 2001), it is also implicated in neuronal mRNA metabolism processes (Bannai et al. 2004; H. H. Chen et al. 2012), suggesting further potential neuron-specific mechanisms at work.

In summary, a trans-acting splicing regulator hnRNP Q interacting with SNP rs17651213 was discovered by RNA-protein pull-down experiments where such interaction was not previously predicted by computational sequence analysis based on consensus motifs. DNA/RNA-affinity approaches thus provide an informative method to screen for and to identify new RNA-protein interacting partners for further functional studies. hnRNP F was identified to interact with SNP rs17651213 in our RNA-protein pull-down experiment but not in our RNA-EMSA supershift assay. The RNA-EMSA RNA-protein binding reaction conditions were different to those used in the pull-down experiment, which could explain why hnRNP F-rs17651213 interaction is identified using one method but not the other. Our data highlight the significance to complement computational predictions with biological data for identifying true RNA/DNA-protein interaction events.

CHAPTER 7

Concluding Remarks

7.1 A Model for the Differential Regulation of *MAPT* Exon 3 Splicing

7.1.1 A Model for the Haplotype-Specific Regulation of H1 and H2 *MAPT* Exon 3 Inclusion

Based on the results from this thesis, I propose the following model mechanism to explain the differential regulation of *MAPT* exon 3 inclusion by hnRNP Q and hnRNP F via rs17651213 and rs1800547 (Figure 50). hnRNP Q would predominantly act to promote the exclusion of exon 3 via the H1 rs17651213 allele (Figure 50) since hnRNP Q showed higher interaction affinity towards the H1 rs17651213 sequence consistently in replicate RNA-protein pull-down experiments. hnRNP F interacts with SNP rs1800547 and its silencing in SK-N-F1 cells led to a decrease in exon 3: total *MAPT* transcript ratio. hnRNP F may therefore act predominantly via the H2 rs1800547 allele to promote exon 3 inclusion (Figure 50). This model also summarises the findings from this study.

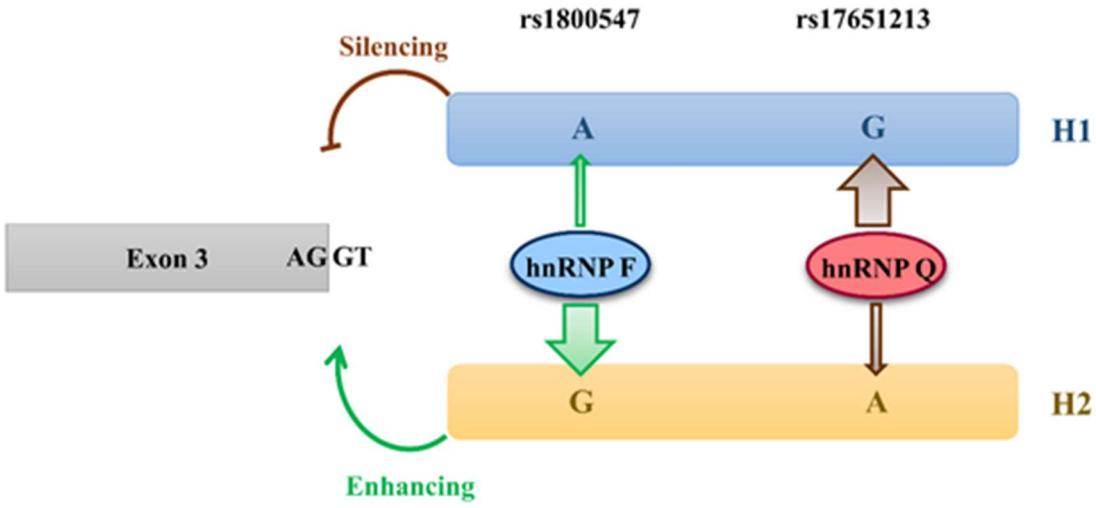


Figure 50. Model for the haplotype-specific regulation of H1 and H2 *MAPT* exon 3 inclusion. H1 rs1800547 and rs17651213 sequences promote the silencing of exon 3 inclusion while the H2 sequences have enhancing effects. hnRNP Q and hnRNP F act predominantly through H1 rs17651213 and H2 rs1800547 sequences, to promote the skipping and inclusion of exon 3, respectively.

Under normal conditions, the H1 rs1800547 and rs17651213 risk SNPs work together to have an overall effect of silencing exon 3 inclusion when compared to the H2 rs1800547 and rs17651213 sequences, which have an enhancing effect (Figure 50). The interaction between rs1800547 and rs17651213 and their individual effect on exon 3 inclusion are likely to be complex and highly context specific since exon 3 has an intrinsically suboptimal branch point at the 3' splice site (Andreadis et al. 1995). Nevertheless, the data from this study highlighted the haplotype-specific differences between H1 and H2 SNP alleles and their combinatorial effects on regulating exon 3 inclusion.

Previous studies have demonstrated that the regulation of inclusion of exons by splicing factors is highly context specific. The same cis-regulatory sequence element could have both enhancer and silencer effects depending on the surrounding sequences. For example, deletion experiments of hnRNP F binding motifs (G-run elements) in the fibrinogen gamma-chain gene pseudoexon showed that the deletion of a silencer G-run element could have enhancing effects on the pseudoexon if neighbouring G-run elements are not present (Rimoldi et al. 2013). Likewise, the same splice factor could promote both exon inclusion and skipping, depending on the sequence context. For example, recent genome wide analyses on alternative splicing events showed that the depletion of hnRNP F proteins led to both activation and repression of alternative exons, strongly indicating that hnRNP F normally regulates both the enhancing and silencing of alternative exons (Huelga et al. 2012; E. Wang et al. 2012).

It is important to note that the H1/H2 haplotype-specific regulation of exon 3 inclusion by rs17651213 is highly dependent on the identity of its upstream functional haplotype-specific SNP rs1800547, demonstrating the complex interactions between the functional SNP and its surrounding haplotypic sequence context.

7.1.2 Allele-Specific Splicing and Disease Susceptibility

The strong association of rs17651213 and rs1800547 with the PD GWAS risk SNP rs17649553 suggests the functional effects of rs17651213 and rsr1800547 on exon 3 inclusion could be contributing to the risk or protection conferred by the H1 and H2 haplotypes respectively. Exon 3 encodes the N-terminal acidic projection domain that mediates the interaction of tau protein with various cellular components such as the plasma membrane, dynactin, actin cytoskeleton, phospholipase C-γ and tyrosine kinase fyn signalling pathways and axonal transport processes (Brandt et al. 1995; Cunningham et al. 1997; M. DiTella et al. 1994; Dixit et al. 2008; Hwang et al. 1996; Jenkins and Johnson 1998; Kempf et al. 1996; G. Lee et al. 1998; Magnani et al. 2007; Shahani and Brandt 2002), many of which are implicated in PD pathogenesis (Panicker et al. 2015; Saha et al. 2004; Sandebring et al. 2009).

In this study, the genetic mechanisms which regulate the inclusion of exon 3 under haplotype-specific control have been investigated. Understanding how different exon 3 tau isoform levels mediate processes implicated in PD could provide insights into the mechanisms of H1/H2 polymorphisms confer risk/protection in neurodegeneration.

7.2 The Post GWAS-Era

7.2.1 Challenges in Understanding Common Traits and Diseases

Common complex neurodegenerative disorders such as Alzheimer's and Parkinson's disease, are often sporadic in nature with no clear familial patterns identifiable, making it extremely difficult to narrow down the mechanisms that cause diseases. Studies in rare Mendelian genes that cause familial forms of the diseases have provided important insights into the molecular pathways in disease pathogenesis. Molecular biology and biochemistry methods have therefore been central to the understanding of biological processes. Experiments in model systems, such as cell and animal models, have relied heavily on the Mendelian mutations which generally have aggressive effects on disease phenotypes. While the pathways unravelled have provided significant understanding of disease mechanisms, it begs the question of whether if the same mechanisms underlie the majority of the idiopathic forms of the diseases.

Between 2002 and 2012, 413 Alzheimer's disease clinical drug trials have been conducted (Cummings et al. 2014). Of the 413 trials, 36.5% have been conducted for symptomatic treatments for improving cognition and 53.5% were aimed at disease-modifying therapies (Cummings et al. 2014). During the decade, 83 of the 413 trials have reached phase 3 study. However, only 1 compound, memantine, has advanced from phase 3 trial to review and approval by the United States food and drug administration agency, placing the attrition rate at 99.6% (Cummings et al. 2014). The high attrition rate in clinical trials has several implications in research into neurodegeneration. First, the trials themselves may be poorly designed, although it is unlikely that the few hundred trials are all poorly designed. Second, successful drug efficacies in experimental models may not reflect their effects in humans. Third, mechanisms identified in rare Mendelian disease models may be poor representations of common non-Mendelian diseases. Fourth, the drug targets themselves could be poor targets for modifying disease progression, reflecting the general poor understanding in both disease mechanisms and in normal biological pathways. Fifth, the process of neurodegeneration occurs over decades, administering drugs in patients already suffering from the disease may be too late. Lastly, biological pathways rarely work in isolation, thus attempting to create a huge impact on the biological networks using just one compound may be extremely inefficient.

7.2.2 Genome-Wide Approaches to Understand Common Variation and Genetic Predisposition

The past decade has seen huge amount of data on the human genome being generated from the Human Genome Project and the International HapMap Project. These have enabled the complex genomic architecture to be much better understood. For example, the genome wide analyses of genetic variations showed SNPs are often associated with differences observed in gene expression and splicing, highlighting their allele-specific functional effects (Degner et al. 2012; Kwan et al. 2008). At the same time, GWAS have associated thousands of low penetrance susceptibility loci to thousands of common traits and diseases.

The systematic analysis and interpretation of GWAS data, however, have not been performed only until very recently beginning with several key publications in 2012 studying the genome wide enrichment of GWAS risk and associated SNPs in regulatory elements (Cowper-Sal lari et al. 2012; Maurano et al. 2012; Schaub et al. 2012). This has been made possible through the development of novel biostatistical approaches. These studies identified SNPs that are in strong linkage disequilibrium with lead GWAS risk SNPs, are often enriched in non-coding regulatory elements (Cowper-Sal lari et al. 2012; Maurano et al. 2012; Schaub et al. 2012), providing the functional basis for how genetic variations contribute towards disease susceptibility. The above studies illustrate the importance of understanding the allele-specific mechanisms of non-coding sequence variations. More importantly the traditional view that a risk SNP simply regulates gene expression of an "interesting" gene in the locus, which is often already known to the field, should be discarded. Instead, understanding the potential combinatorial effects of risk associated SNPs across the genome beyond affecting gene expression should be prioritised. For example, GWAS SNP sets could be enriched in non-coding regulatory elements such as transcription factor binding sites, epigenetic marks, DNase I hypersensitive sites, or splicing motifs. Identifying the enrichment of GWAS SNP sets in regulatory elements related to age, cell type specificity, cellular processes or master regulators, may have the potential to provide important insights into several key unanswered questions in neurodegeneration.

Techniques such as ChIP-seq or DNase-seq, will play a crucial role to provide biological data necessary for piecing together the bigger picture of the mechanisms in allele-specific functional regulation by non-coding variations. While the type of post-GWAS analysis described above was not employed in this thesis, the approaches used here would be hugely valuable for functional the follow-up studies on the regulatory elements of interest to study in detail the allele-specific mechanisms. For example, genomic DNA vectors used in this study that are able to capture physiologically relevant non-coding haplotypic sequence diversity, would provide a novel and powerful tool to study the differential regulation of gene expression and alternative splicing by

non-coding SNPs, both in normal physiology and disease conditions. Ultimately, studies that combine genome-wide bioinformatics approaches and biological experiments will be crucial to the understanding of how genetic variation contributes to disease mechanisms.

7.3 Machine Learning in Understanding the Human Genome

7.3.1 Human Genetics in the Age of Artificial Intelligence

The function of the human genome is highly complex, involving multiple layers and networks of regulation, orchestrated in a precise and timely fashion. To build up knowledge of the functional elements in the genome by experimental approaches, molecular biology is an indispensable tool. However, it will be highly unlikely for humans to understand the multiple layers of higher functions of the genome and to uncover the complex patterns of interactions disguised in a vast number of pathway networks using simply molecular biology and animal models. Machine learning methods may prove to be essential in understanding the breadth and complexity of the human genome that are often abstract and not obvious to the human eye, and are challenging to even state-of-the-art statistical approaches (Libbrecht and Noble 2015; Park and Kellis 2015; Yip et al. 2013).

7.3.2 Application of Machine Learning Algorithms in Genetics and Genomics

Machine learning is a general terminology regarding computational methods which use computer algorithms that improve with experience (Libbrecht and Noble 2015). Methods in machine learning have been applied extensively in a wide range of areas in genomics and genetics (Libbrecht and Noble 2015). For example, algorithms can be designed to “learn” how to recognise patterns in DNA sequences such as transcription start sites, introns, exons and untranslated regions (Picardi and Pesole 2010). Algorithms can use data from genomic assays such as ChIP-seq, DNase-seq or RNA-seq. For example, chromatin data can be used to visualise the genome-wide maps of chromatin-state by algorithms which learn chromatin states, characterise their biological functions and find correlations with large-scale functional datasets (Ernst and Kellis 2012). Another example is the DeepBind algorithm, a deep learning approach to predict the affinity of a protein to DNA or RNA sequences, to predict RNA splicing and editing, to discover regulatory motifs and to interpret

genetic variation by combining multiple types of biological datasets (Alipanahi et al. 2015). Most biological systems do not function as independent entities. It is also possible to use machine learning algorithms for integrating different entities into a meaningful network, such as the network of protein interactions and regulatory interactions between genes (Bacardit and Llorà 2013; Friedman et al. 2000). An example of such is the modelling of how genes regulate each other by using Bayesian networks (Friedman et al. 2000).

7.3.3 Cross-Disciplinary Research in Understanding Common Diseases

To understand the genetic mechanisms in common traits and diseases, cross-disciplinary studies will be crucial to provide a full picture on the functions of genetic variation from the sequence level to higher order network levels. I envisage that genetic variation functions at 5 dimensions, where a range of different techniques from wet to dry lab, will be best suited to be used at different levels (Figure 51).

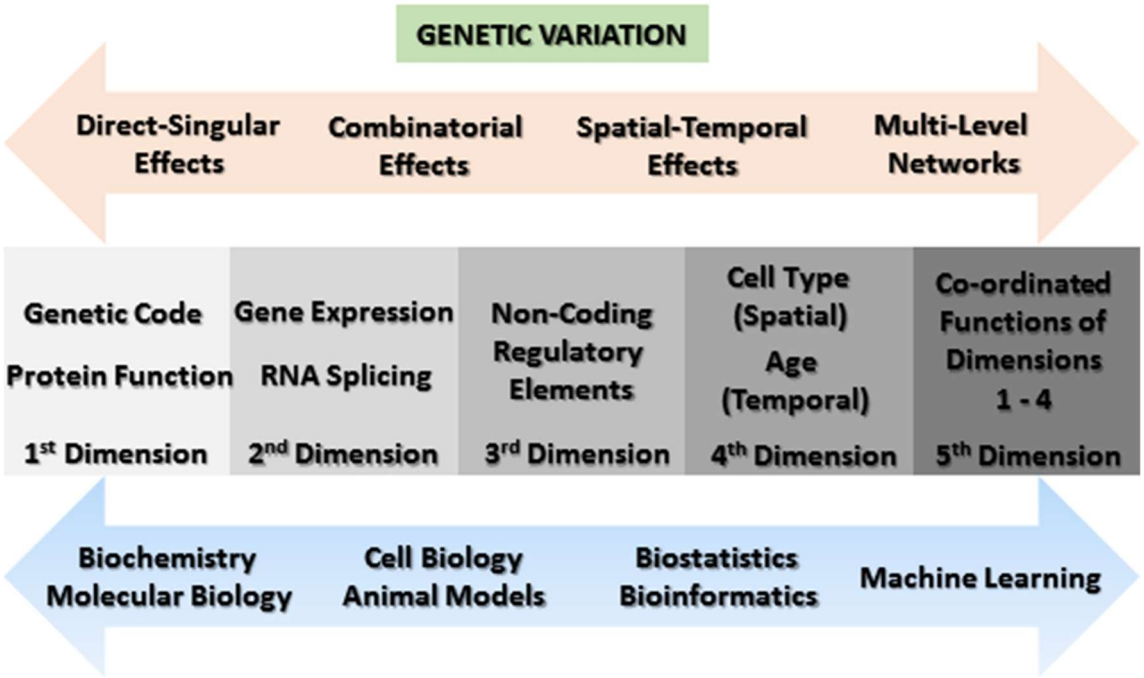


Figure 51. The 5 dimensions of genetic variation.

The functional effects of genetic variation could be grouped into 5 dimensional levels. The combinatorial and network effects increase with each dimensional layer. Different techniques, from biochemistry to machine learning algorithms, are advantageous for understanding the effects of genetic variation at different levels.

For example, biochemical characterisation will be best suited for studying the direct effect of a change in coding sequence by a mutation (1st dimension) (Figure 51). Statistical methods will be useful in understanding how, say GWAS risk SNP sets, enrich in various non-coding regulatory elements (3rd dimension) (Figure 51). Machine learning algorithms will be useful in discerning patterns, to make predictions, to aid interpretation of functions, and to study different biological entities at the network level (5th dimension) (Figure 51).

To illustrate the importance of understanding genetic variation at multiple levels, a study on the chromatin states across human populations would be an excellent exemplar (Kasowski et al. 2013). The authors studied the chromatin states by performing RNA-seq and ChIP-seq to map a diverse class of regulatory elements in individuals from diverse ancestry backgrounds (Kasowski et al. 2013). The chromatin states are segmented and categorised based on the patterns of the elements identified in RNA-seq and ChIP-seq experiments using a learning algorithm (Ernst and Kellis 2012; Kasowski et al. 2013). The authors identified regions of chromatin that are variable in different populations, and further performed statistical analysis to identify the enrichment of SNPs in these variable regions. Lastly, the authors identified a strong correlation of allele-specific signals between daughters and parents in certain chromatin marks, and enhancer and promoter marks (Kasowski et al. 2013). The results from this study indicate that certain patterns of chromatin marks and transcription factor binding are heritable, possibly by disrupting transcription factor binding through chromatin variation caused by genetic variation (Kasowski et al. 2013).

We have recently witnessed the defeat of a world leading champion in the game of Go by the programme AlphaGo, a state-of-the-art deep neural networks approach combined with Monte Carlo simulation with policy networks that minimises overfitting in the model (Silver et al. 2016). The defeat of a world Go champion by artificial intelligence was previously thought to be impossible for at least another decade, and the success of AlphaGo is seen as a huge milestone in the artificial intelligence community. Perhaps machine learning methods are developing faster than ever before, combining with already existing experimental and computational approaches in biology,

cross-disciplinary research projects in the future will likely be leading the way to open new avenues for understanding complex diseases by novel approaches that may ultimately lead to the development of a cure for common complex diseases.

REFERENCES

1000 Genomes Project Consortium, Abecasis, G. R., et al. (2012), 'An integrated map of genetic variation from 1,092 human genomes', *Nature,* 491 (7422), 56-65.

Abecasis, G. R., et al. (2010), 'A map of human genome variation from population-scale sequencing', *Nature,* 467 (7319), 1061-73.

Aguzzi, A. and Lakkaraju, A. K. (2016), 'Cell Biology of Prions and Prionoids: A Status Report', *Trends Cell Biol,* 26 (1), 40-51.

Akerman, M., et al. (2009), 'A computational approach for genome-wide mapping of splicing factor binding sites', *Genome Biol,* 10 (3), R30.

Al-Bassam, J., et al. (2002), 'MAP2 and tau bind longitudinally along the outer ridges of microtubule protofilaments', *J Cell Biol,* 157 (7), 1187-96.

Alipanahi, B., et al. (2015), 'Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning', *Nat Biotechnol,* 33 (8), 831-8.

Allen, C. and Borisy, G. G. (1974), 'Structural polarity and directional growth of microtubules of Chlamydomonas flagella', *J Mol Biol,* 90 (2), 381-402.

Almen, M. S., et al. (2012), 'Genome wide analysis reveals association of a FTO gene variant with epigenetic changes', *Genomics,* 99 (3), 132-7.

Alon, T., et al. (2009), 'Transgenic mice expressing green fluorescent protein under the control of the corticotropin-releasing hormone promoter', *Endocrinology,* 150 (12), 5626-32.

Alonso Adel, C., et al. (2004), 'Promotion of hyperphosphorylation by frontotemporal dementia tau mutations', *J Biol Chem,* 279 (33), 34873-81.

Ameur, A., et al. (2011), 'Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain', *Nat Struct Mol Biol,* 18 (12), 1435-40.

Amos, L. and Klug, A. (1974), 'Arrangement of subunits in flagellar microtubules', *J Cell Sci,* 14 (3), 523-49.

Andorfer, C., et al. (2005), 'Cell-cycle reentry and cell death in transgenic mice expressing nonmutant human tau isoforms', *J Neurosci,* 25 (22), 5446-54.

Andreadis, A., Brown, W. M., and Kosik, K. S. (1992), 'Structure and novel exons of the human tau gene', *Biochemistry,* 31 (43), 10626-33.

Andreadis, A., Broderick, J. A., and Kosik, K. S. (1995), 'Relative exon affinities and suboptimal splice site signals lead to non-equivalence of two cassette exons', *Nucleic Acids Res,* 23 (17), 3585-93.

Andreadis, A., et al. (1993), 'The exon trapping assay partly discriminates against alternatively spliced exons', *Nucleic Acids Res,* 21 (9), 2217-21.

Anko, M. L., et al. (2010), 'Global analysis reveals SRp20- and SRp75-specific mRNPs in cycling and neural cells', *Nat Struct Mol Biol,* 17 (8), 962-70.

--- (2012), 'The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes', *Genome Biol,* 13 (3), R17.

Antoch, M. P., et al. (1997), 'Functional identification of the mouse circadian Clock gene by transgenic BAC rescue', *Cell,* 89 (4), 655-67.

Arai, T., et al. (2001), 'Distinct isoforms of tau aggregated in neurons and glial cells in brains of patients with Pick's disease, corticobasal degeneration and progressive supranuclear palsy', *Acta Neuropathol,* 101 (2), 167-73.

--- (2011), 'Increased expression of NKX2.3 mRNA transcribed from the risk haplotype for ulcerative colitis in the involved colonic mucosa', *Hum Immunol,* 72 (7), 587-91.

Arikan, M. C., et al. (2002), 'Modulation of the membrane-binding projection domain of tau protein: splicing regulation of exon 3', *Brain Res Mol Brain Res,* 101 (1-2), 109-21.

Aronov, S., et al. (2001), 'Axonal tau mRNA localization coincides with tau protein in living neuronal cells and depends on axonal targeting signal', *J Neurosci,* 21 (17), 6577-87.

--- (2002), 'Visualization of translated tau protein in the axons of neuronal P19 cells and characterization of tau RNP granules', *J Cell Sci,* 115 (Pt 19), 3817-27.

Arrasate, M., Perez, M., and Avila, J. (2000), 'Tau dephosphorylation at tau-1 site correlates with its association to cell membrane', *Neurochem Res,* 25 (1), 43-50.

Aubert, M. and Blaho, J. A. (1999), 'The Herpes Simplex Virus Type 1 Regulatory Protein ICP27 Is Required for the Prevention of Apoptosis in Infected Human Cells', *J Virol* (73), 2803-13.

Bacardit, J. and Llorà, X. (2013), 'Large-scale data mining using genetics-based machine learning', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 3 (1), 37-61.

Baker, M., et al. (1999), 'Association of an extended haplotype in the tau gene with progressive supranuclear palsy', *Hum Mol Genet,* 8 (4), 711-5.

Ballas, N., et al. (2005), 'REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis', *Cell,* 121 (4), 645-57.

Ballatore, C., Lee, V. M., and Trojanowski, J. Q. (2007), 'Tau-mediated neurodegeneration in Alzheimer's disease and related disorders', *Nat Rev Neurosci,* 8 (9), 663-72.

Baloyannis, S. J. (2006), 'Mitochondrial alterations in Alzheimer's disease', *J Alzheimers Dis,* 9 (2), 119-26.

Bannai, H., et al. (2004), 'An RNA-interacting protein, SYNCRIP (heterogeneous nuclear ribonuclear protein Q1/NSAP1) is a component of mRNA granule transported with inositol 1,4,5-trisphosphate receptor type 1 mRNA in neuronal dendrites', *J Biol Chem,* 279 (51), 53427-34.

Barrett, J. C., et al. (2005), 'Haploview: analysis and visualization of LD and haplotype maps', *Bioinformatics,* 21 (2), 263-5.

Baumann, K., et al. (1993), 'Abnormal Alzheimer-like phosphorylation of tau-protein by cyclin-dependent kinases cdk2 and cdk5', *FEBS Lett,* 336 (3), 417-24.

Beggs, H. E., Soriano, P., and Maness, P. F. (1994), 'NCAM-dependent neurite outgrowth is inhibited in neurons from Fyn-minus mice', *J Cell Biol,* 127 (3), 825-33.

Bell, C. G., et al. (2010), 'Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus', *PLoS One,* 5 (11), e14040.

Berget, S. M. (1995), 'Exon recognition in vertebrate splicing', *J Biol Chem,* 270 (6), 2411-4.

Bertolotti, A., et al. (2000), 'Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response', *Nat Cell Biol,* 2 (6), 326-32.

Biernat, J., et al. (1993), 'Phosphorylation of Ser262 strongly reduces binding of tau to microtubules: distinction between PHF-like immunoreactivity and microtubule binding', *Neuron,* 11 (1), 153-63.

Binder, L. I., Frankfurter, A., and Rebhun, L. I. (1985), 'The distribution of tau in the mammalian central nervous system', *J Cell Biol,* 101 (4), 1371-8.

Black, D. L. (1991), 'Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells?', *Genes Dev,* 5 (3), 389-402.

--- (1992), 'Activation of c-src neuron-specific splicing by an unusual RNA element in vivo and in vitro', *Cell,* 69 (5), 795-807.

--- (1995), 'Finding splice sites within a wilderness of RNA', *Rna,* 1 (8), 763-71.

--- (2003), 'Mechanisms of alternative pre-messenger RNA splicing', *Annu Rev Biochem,* 72, 291-336.

Black, D. L., et al. (2002), 'The electrophoretic mobility shift assay for RNA binding proteins', in Christopher W.J. Smith (ed.), *RNA:Protein Interactions A Practical Approach* (The Practical Approach Seires; United States: Oxford University Press Inc., New York), 109-36.

Black, J. and Vos, J. M. (2002), 'Establishment of an oriP/EBNA1-based episomal vector transcribing human genomic beta-globin in cultured murine fibroblasts', *Gene Ther,* 9 (21), 1447-54.

Blennow, K., de Leon, M. J., and Zetterberg, H. (2006), 'Alzheimer's disease', *Lancet,* 368 (9533), 387-403.

Bliss, T. V. and Collingridge, G. L. (1993), 'A synaptic model of memory: long-term potentiation in the hippocampus', *Nature,* 361 (6407), 31-9.

Bloom, G. S. (2014), 'Amyloid-beta and tau: the trigger and bullet in Alzheimer disease pathogenesis', *JAMA Neurol,* 71 (4), 505-8.

Boehm, J. (2013), 'A 'danse macabre': tau and Fyn in STEP with amyloid beta to facilitate induction of synaptic depression and excitotoxicity', *Eur J Neurosci,* 37 (12), 1925-30.

Boettger, L. M., et al. (2012), 'Structural haplotypes and recent evolution of the human 17q21.31 region', *Nat Genet,* 44 (8), 881-5.

Bolton, D. C., McKinley, M. P., and Prusiner, S. B. (1982), 'Identification of a protein that purifies with the scrapie prion', *Science,* 218 (4579), 1309-11.

Borst, E. M. and Messerle, M. (2003), 'Construction of a cytomegalovirus-based amplicon: a vector with a unique transfer capacity', *Hum Gene Ther,* 14 (10), 959-70.

Bose, S. K., et al. (2006), 'Identification of Ebp1 as a component of cytoplasmic bcl-2 mRNP (messenger ribonucleoprotein particle) complexes', *Biochem J,* 396 (1), 99-107.

Botstein, D., et al. (1980), 'Construction of a genetic linkage map in man using restriction fragment length polymorphisms', *Am J Hum Genet,* 32 (3), 314-31.

Boutajangout, A., et al. (2004), 'Expression of tau mRNA and soluble tau isoforms in affected and non-affected brain areas in Alzheimer's disease', *FEBS Lett,* 576 (1-2), 183-9.

Braak, E., Braak, H., and Mandelkow, E. M. (1994), 'A sequence of cytoskeleton changes related to the formation of neurofibrillary tangles and neuropil threads', *Acta Neuropathol,* 87 (6), 554-67.

Braak, H. and Braak, E. (1991), 'Neuropathological stageing of Alzheimer-related changes', *Acta Neuropathol,* 82 (4), 239-59.

Braak, H. and Del Tredici, K. (2011), 'The pathological process underlying Alzheimer's disease in individuals under thirty', *Acta Neuropathol,* 121 (2), 171-81.

Braak, H., et al. (2003), 'Staging of brain pathology related to sporadic Parkinson's disease', *Neurobiol Aging,* 24 (2), 197-211.

Bradke, F. and Dotti, C. G. (1997), 'Neuronal polarity: vectorial cytoplasmic flow precedes axon formation', *Neuron,* 19 (6), 1175-86.

--- (1999), 'The role of local actin instability in axon formation', *Science,* 283 (5409), 1931-4.

Bramblett, G. T., et al. (1993), 'Abnormal tau phosphorylation at Ser396 in Alzheimer's disease recapitulates development and contributes to reduced microtubule binding', *Neuron,* 10 (6), 1089-99.

Brandt, R. and Lee, G. (1993), 'Functional organization of microtubule-associated protein tau. Identification of regions which affect microtubule growth, nucleation, and bundle formation in vitro', *J Biol Chem,* 268 (5), 3414-9.

Brandt, R., Léger, J., and Lee, G. (1995), 'Interaction of tau with the neural plasma membrane mediated by tau's amino-terminal projection domain', *J Cell Biol,* 131 (5), 1327-40.

Braun, A., Little, D. P., and Koster, H. (1997), 'Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry', *Clin Chem,* 43 (7), 1151-8.

Brest, P., et al. (2011), 'A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease', *Nat Genet,* 43 (3), 242-5.

Brook, J. D., et al. (1992), 'Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member', *Cell,* 68 (4), 799-808.

Brunden, K. R., Trojanowski, J. Q., and Lee, V. M. Y. (2009), 'Advances in Tau-focused drug discovery for Alzheimer's disease and related tauopathies', *Nat Rev Drug Discov,* 8 (10), 783-93.

Buee, L., et al. (2000), 'Tau protein isoforms, phosphorylation and role in neurodegenerative disorders', *Brain Res Brain Res Rev,* 33 (1), 95-130.

Bullmann, T., et al. (2010), 'Expression of the embryonal isoform (0N/3R) of the microtubule-associated protein tau in the adult rat central nervous system', *J Comp Neurol,* 518 (13), 2538-53.

Buratti, E., et al. (2004), 'RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon', *Mol Cell Biol,* 24 (3), 1387-400.

Burd, C. G. and Dreyfuss, G. (1994), 'RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing', *Embo j,* 13 (5), 1197-204.

Burdett, T (EBI)., et al. 'The NHGRI-EBI Catalog of published genome-wide association studies. Available at: www.ebi.ac.uk/gwas. Accessed 20-03-2016, version v1.0.'.

Burre, J., et al. (2010), 'Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro', *Science,* 329 (5999), 1663-7.

Busch, A. and Hertel, K. J. (2012), 'Evolution of SR protein and hnRNP splicing regulatory factors', *Wiley Interdiscip Rev RNA,* 3 (1), 1-12.

Bush, W. S. and Moore, J. H. (2012), 'Chapter 11: Genome-wide association studies', *PLoS Comput Biol,* 8 (12), e1002822.

Butter, F., et al. (2009), 'Unbiased RNA-protein interaction screen by quantitative proteomics', *Proc Natl Acad Sci U S A,* 106 (26), 10626-31.

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', *Neurobiol Aging,* 29 (12), 1923-9.

Caffrey, T. M., et al. (2006), 'Haplotype-specific expression of exon 10 at the human MAPT locus', *Hum Mol Genet,* 15 (24), 3529-37.

Callow, M. J., et al. (1994), 'Expression of human apolipoprotein B and assembly of lipoprotein(a) in transgenic mice', *Proc Natl Acad Sci U S A,* 91 (6), 2130-4.

Capranico, G., et al. (2007), 'The effects of camptothecin on RNA polymerase II transcription: roles of DNA topoisomerase I', *Biochimie,* 89 (4), 482-9.

Cartegni, L., Chew, S. L., and Krainer, A. R. (2002), 'Listening to silence and understanding nonsense: exonic mutations that affect splicing', *Nat Rev Genet,* 3 (4), 285-98.

Cavaloc, Y., et al. (1999), 'The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers', *Rna,* 5 (3), 468-83.

Cavelier, L. B., et al. (2001), 'Regulation and activity of the human ABCA1 gene in transgenic mice', *J Biol Chem,* 276 (21), 18046-51.

Cente, M., et al. (2009), 'Expression of a truncated human tau protein induces aqueous-phase free radicals in a rat model of tauopathy: implications for targeted antioxidative therapy', *J Alzheimers Dis,* 17 (4), 913-20.

Chambers, C. B., et al. (1999), 'Overexpression of four-repeat tau mRNA isoforms in progressive supranuclear palsy but not in Alzheimer's disease', *Ann Neurol,* 46 (3), 325-32.

Chambers, M. C., et al. (2012), 'A cross-platform toolkit for mass spectrometry and proteomics', *Nat Biotechnol,* 30 (10), 918-20.

Chandler, K. J., et al. (2007), 'Relevance of BAC transgene copy number in mice: transgene copy number variation across multiple transgenic lines and correlations with transgene integrity and expression', *Mamm Genome,* 18 (10), 693-708.

Chen, H. H., et al. (2008a), 'The RNA binding protein hnRNP Q modulates the utilization of exon 7 in the survival motor neuron 2 (SMN2) gene', *Mol Cell Biol,* 28 (22), 6929-38.

--- (2012), 'hnRNP Q regulates Cdc42-mediated neuronal morphogenesis', *Mol Cell Biol,* 32 (12), 2224-38.

Chen, J., et al. (1992), 'Projection domains of MAP2 and tau determine spacings between microtubules in dendrites and axons', *Nature,* 360 (6405), 674-7.

Chen, S., et al. (2010), 'MAPT isoforms: differential transcriptional profiles related to 3R and 4R splice variants', *J Alzheimers Dis,* 22 (4), 1313-29.

Chen, X., et al. (2008b), 'Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk', *Hum Mol Genet,* 17 (9), 1336-48.

Chin, J., et al. (2004), 'Fyn kinase modulates synaptotoxicity, but not aberrant sprouting, in human amyloid precursor protein transgenic mice', *J Neurosci,* 24 (19), 4692-7.

Cho, J. H. and Johnson, G. V. (2003), 'Glycogen synthase kinase 3beta phosphorylates tau at both primed and unprimed sites. Differential impact on microtubule binding', *J Biol Chem,* 278 (1), 187-93.

Choi, S. H., et al. (2014), 'A three-dimensional human neural cell culture model of Alzheimer's disease', *Nature,* 515 (7526), 274-8.

Clark, L. N., et al. (2006), 'Frequency of LRRK2 mutations in early- and late-onset Parkinson disease', *Neurology,* 67 (10), 1786-91.

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', *J Mol Biol,* 116 (2), 227-47.

Cohen, P. and Cohen, P. T. (1989), 'Protein phosphatases come of age', *J Biol Chem,* 264 (36), 21435-8.

Collins, F. S., Guyer, M. S., and Charkravarti, A. (1997), 'Variations on a theme: cataloging human DNA sequence variation', *Science,* 278 (5343), 1580-1.

Conde, C. and Caceres, A. (2009), 'Microtubule assembly, organization and dynamics in axons and dendrites', *Nat Rev Neurosci,* 10 (5), 319-32.

Conrad, C., et al. (1997), 'Genetic evidence for the involvement of tau in progressive supranuclear palsy', *Ann Neurol,* 41 (2), 277-81.

Cooper, T. A. (2005), 'Use of minigene systems to dissect alternative splicing elements', *Methods,* 37 (4), 331-40.

Coren, J. S. and Sternberg, N. (2001), 'Construction of a PAC vector system for the propagation of genomic DNA in bacterial and mammalian cells and subsequent generation of nested deletions in individual library members', *Gene,* 264 (1), 11-8.

Corvelo, A., et al. (2010), 'Genome-wide association between branch point properties and alternative splicing', *PLoS Comput Biol,* 6 (11), e1001016.

Cowper-Sal lari, R., et al. (2012), 'Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression', *Nat Genet,* 44 (11), 1191-8.

Crowley, J. J., et al. (2015), 'Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance', *Nat Genet,* 47 (4), 353-60.

Cruts, M., et al. (2005), 'Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region', *Hum Mol Genet,* 14 (13), 1753-62.

Cummings, J. L., Morstorf, T., and Zhong, K. (2014), 'Alzheimer's disease drug-development pipeline: few candidates, frequent failures', *Alzheimers Res Ther* (6), 37.

Cunningham, C. C., et al. (1997), 'Microtubule-associated protein 2c reorganizes both microtubules and microfilaments into distinct cytological structures in an actin-binding protein-280-deficient melanoma cell line', *J Cell Biol,* 136 (4), 845-57.

Darzacq, X., et al. (2007), 'In vivo dynamics of RNA polymerase II transcription', *Nat Struct Mol Biol,* 14 (9), 796-806.

David, D. C., et al. (2005), 'Proteomic and functional analyses reveal a mitochondrial dysfunction in P301L tau transgenic mice', *J Biol Chem,* 280 (25), 23802-14.

Davies, C. A., et al. (1987), 'A quantitative morphometric analysis of the neuronal and synaptic content of the frontal and temporal cortex in patients with Alzheimer's disease', *J Neurol Sci,* 78 (2), 151-64.

Dawson, H. N., et al. (2001), 'Inhibition of neuronal maturation in primary hippocampal neurons from tau deficient mice', *J Cell Sci,* 114 (Pt 6), 1179-87.

Dayanandan, R., et al. (1999), 'Mutations in tau reduce its microtubule binding properties in intact cells and affect its phosphorylation', *FEBS Lett,* 446 (2-3), 228-32.

de la Mata, M., et al. (2003), 'A slow RNA polymerase II affects alternative splicing in vivo', *Mol Cell,* 12 (2), 525-32.

de Silva, R., et al. (2003), 'Pathological inclusion bodies in tauopathies contain distinct complements of tau with three or four microtubule-binding repeat domains as demonstrated by new specific monoclonal antibodies', *Neuropathol Appl Neurobiol,* 29 (3), 288-302.

--- (2006), 'An immunohistochemical study of cases of sporadic and inherited frontotemporal lobar degeneration using 3R- and 4R-specific tau monoclonal antibodies', *Acta Neuropathol,* 111 (4), 329-40.

Deal, K. K., et al. (2006), 'Distant regulatory elements in a Sox10-beta GEO BAC transgene are required for expression of Sox10 in the enteric nervous system and other neural crest-derived tissues', *Dev Dyn,* 235 (5), 1413-32.

Degner, J. F., et al. (2012), 'DNase I sensitivity QTLs are a major determinant of human expression variation', *Nature,* 482 (7385), 390-4.

Desai, A. and Mitchison, T. J. (1997), 'Microtubule polymerization dynamics', *Annu Rev Cell Dev Biol,* 13, 83-117.

DeVeale, B., van der Kooy, D., and Babak, T. (2012), 'Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective', *PLoS Genet,* 8 (3), e1002600.

Di Maria, E., et al. (2000), 'Corticobasal degeneration shares a common genetic background with progressive supranuclear palsy', *Ann Neurol,* 47 (3), 374-7.

Dickson, D. W. (2012), 'Parkinson's disease and parkinsonism: neuropathology', *Cold Spring Harb Perspect Med,* 2 (8).

Dickson, D. W., et al. (2011), 'Neuropathology of frontotemporal lobar degeneration-tau (FTLD-tau)', *J Mol Neurosci,* 45 (3), 384-9.

Dimas, A. S., et al. (2009), 'Common regulatory variation impacts gene expression in a cell type-dependent manner', *Science,* 325 (5945), 1246-50.

Dimauro, I., et al. (2012), 'A simple protocol for the subcellular fractionation of skeletal muscle cells and tissue', *BMC Res Notes,* 5, 513.

Ding, Y., et al. (2014), 'In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features', *Nature,* 505 (7485), 696-700.

Distefano, J. K. and Taverna, D. M. (2011), 'Technological issues and experimental design of gene association studies', *Methods Mol Biol,* 700, 3-16.

DiTella, M., et al. (1994), 'Microfilament-associated growth cone component depends upon Tau for its intracellular localization', *Cell Motil Cytoskeleton,* 29 (2), 117-30.

DiTella, M. C., et al. (1996), 'MAP-1B/TAU functional redundancy during laminin-enhanced axonal growth', *J Cell Sci,* 109 (Pt 2), 467-77.

Dixit, R., et al. (2008), 'Differential Regulation of Dynein and Kinesin Motor Proteins by Tau', *Science,* 319 (5866), 1086-9.

Do, C. B., et al. (2011), 'Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease', *PLoS Genet,* 7 (6), e1002141.

Donnelly, M. P., et al. (2010), 'The distribution and most recent common ancestor of the 17q21 inversion in humans', *Am J Hum Genet,* 86 (2), 161-71.

Dormann, D., et al. (2010), 'ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import', *Embo j,* 29 (16), 2841-57.

Dotti, C. G., Sullivan, C. A., and Banker, G. A. (1988), 'The establishment of polarity by hippocampal neurons in culture', *J Neurosci,* 8 (4), 1454-68.

Drechsel, D. N., et al. (1992), 'Modulation of the dynamic instability of tubulin assembly by the microtubule-associated protein tau', *Mol Biol Cell,* 3 (10), 1141-54.

Drewes, G., et al. (1997), 'MARK, a novel family of protein kinases that phosphorylate microtubule-associated proteins and trigger microtubule disruption', *Cell,* 89 (2), 297-308.

--- (1992), 'Mitogen activated protein (MAP) kinase transforms tau protein into an Alzheimer-like state', *Embo j,* 11 (6), 2131-8.

--- (1995), 'Microtubule-associated protein/microtubule affinity-regulating kinase (p110mark). A novel protein kinase that regulates tau-microtubule interactions and dynamic instability by phosphorylation at the Alzheimer-specific site serine 262', *J Biol Chem,* 270 (13), 7679-88.

Dreyfuss, G., et al. (1993), 'hnRNP proteins and the biogenesis of mRNA', *Annu Rev Biochem,* 62, 289-321.

Drubin, D. G. and Kirschner, M. W. (1986), 'Tau protein function in living cells', *J Cell Biol,* 103 (6 Pt 2), 2739-46.

DuBoff, B., Gotz, J., and Feany, M. B. (2012), 'Tau promotes neurodegeneration via DRP1 mislocalization in vivo', *Neuron,* 75 (4), 618-32.

Dudbridge, F. and Gusnanto, A. (2008), 'Estimation of significance thresholds for genomewide association scans', *Genet Epidemiol,* 32 (3), 227-34.

Dudek, S. M. and Johnson, G. V. (1995), 'Postnatal changes in serine/threonine protein phosphatases and their association with the microtubules', *Brain Res Dev Brain Res,* 90 (1-2), 54-61.

Dujardin, G., et al. (2014), 'How slow RNA polymerase II elongation favors alternative exon skipping', *Mol Cell,* 54 (4), 683-90.

Edwards, T. L., et al. (2010), 'Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease', *Ann Hum Genet,* 74 (2), 97-109.

Emilsson, V., et al. (2008), 'Genetics of gene expression and its effect on disease', *Nature,* 452 (7186), 423-8.

Ernst, J. and Kellis, M. (2012), 'ChromHMM: automating chromatin-state discovery and characterization', *Nat Methods* (9; United States), 215-6.

Falush, D., Stephens, M., and Pritchard, J. K. (2003), 'Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies', *Genetics,* 164 (4), 1567-87.

Farah, C. A., et al. (2006), 'Tau interacts with Golgi membranes and mediates their association with microtubules', *Cell Motil Cytoskeleton,* 63 (11), 710-24.

--- (2005), 'Interaction of microtubule-associated protein-2 and p63: a new link between microtubules and rough endoplasmic reticulum membranes in neurons', *J Biol Chem,* 280 (10), 9439-49.

Fardaei, M., et al. (2001), 'In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts', *Nucleic Acids Res,* 29 (13), 2766-71.

Feany, M. B., et al. (1995), 'Epitope expression and hyperphosphorylation of tau protein in corticobasal degeneration: differentiation from progressive supranuclear palsy', *Acta Neuropathol,* 90 (1), 37-43.

Fica, S. M., et al. (2013), 'RNA catalyses nuclear pre-mRNA splicing', *Nature,* 503 (7475), 229-34.

Fraefel, C., et al. (1996), 'Helper virus-free transfer of herpes simplex virus type 1 plasmid vectors into neural cells', *J Virol,* 70 (10), 7190-7.

Frandemiche, M. L., et al. (2014), 'Activity-dependent tau protein translocation to excitatory synapse is disrupted by exposure to amyloid-beta oligomers', *J Neurosci,* 34 (17), 6084-97.

Frappier, T. F., et al. (1994), 'tau Regulation of microtubule-microtubule spacing and bundling', *J Neurochem,* 63 (6), 2288-94.

Frazer, K. A., et al. (2009), 'Human genetic variation and its contribution to complex traits', *Nat Rev Genet,* 10 (4), 241-51.

Freedman, M. L., et al. (2011), 'Principles for the post-GWAS functional characterization of cancer risk loci', *Nat Genet,* 43 (6), 513-8.

Fried, M. and Crothers, D. M. (1981), 'Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis', *Nucleic Acids Res,* 9 (23), 6505-25.

Friedman, N., et al. (2000), 'Using Bayesian networks to analyze expression data', *J Comput Biol,* 7 (3-4), 601-20.

Frost, B., Gotz, J., and Feany, M. B. (2015), 'Connecting the dots between tau dysfunction and neurodegeneration', *Trends Cell Biol,* 25 (1), 46-53.

Fu, M. M. and Holzbaur, E. L. (2014), 'Integrated regulation of motor-driven organelle transport by scaffolding proteins', *Trends Cell Biol,* 24 (10), 564-74.

Fu, X. D. and Ares, M., Jr. (2014), 'Context-dependent control of alternative splicing by RNA-binding proteins', *Nat Rev Genet,* 15 (10), 689-701.

Fulga, T. A., et al. (2007), 'Abnormal bundling and accumulation of F-actin mediates tau-induced neuronal degeneration in vivo', *Nat Cell Biol,* 9 (2), 139-48.

Fung, H. C., et al. (2006), 'Association of tau haplotype-tagging polymorphisms with Parkinson's disease in diverse ethnic Parkinson's disease cohorts', *Neurodegener Dis,* 3 (6), 327-33.

Gabizon, R., et al. (1993), 'Mutation and polymorphism of the prion protein gene in Libyan Jews with Creutzfeldt-Jakob disease (CJD)', *Am J Hum Genet,* 53 (4), 828-35.

Galloway, P. G., Perry, G., and Gambetti, P. (1987), 'Hirano body filaments contain actin and actin-associated proteins', *J Neuropathol Exp Neurol,* 46 (2), 185-99.

Galvan, M., et al. (2001), 'Sequence of neurofibrillary changes in aging and Alzheimer's disease: A confocal study with phospho-tau antibody, AD2', *J Alzheimers Dis,* 3 (4), 417-25.

Gamberi, C., et al. (1997), 'Interaction between the human nuclear cap-binding protein complex and hnRNP F', *Mol Cell Biol,* 17 (5), 2587-97.

Garneau, D., et al. (2005), 'Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x', *J Biol Chem,* 280 (24), 22641-50.

Garner, M. M. and Revzin, A. (1981), 'A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system', *Nucleic Acids Res,* 9 (13), 3047-60.

Ge, B., et al. (2009), 'Global patterns of cis variation in human cells revealed by high-density allelic expression analysis', *Nat Genet,* 41 (11), 1216-22.

Gething, M. J. (1999), 'Role and regulation of the ER chaperone BiP', *Semin Cell Dev Biol,* 10 (5), 465-72.

Gibb, G. M., et al. (2004), 'Differential involvement and heterogeneous phosphorylation of tau isoforms in progressive supranuclear palsy', *Brain Res Mol Brain Res,* 121 (1-2), 95-101.

Gibson, G. (2011), 'Rare and common variants: twenty arguments', *Nat Rev Genet,* 13 (2), 135-45.

Gilsbach, B. K. and Kortholt, A. (2014), 'Structural biology of the LRRK2 GTPase and kinase domains: implications for regulation', *Front Mol Neurosci,* 7, 32.

Giulietti, M., et al. (2013), 'SpliceAid-F: a database of human splicing factors and their RNA-binding sites', *Nucleic Acids Res,* 41 (Database issue), D125-31.

Goedert, M. (2015), 'NEURODEGENERATION. Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled Abeta, tau, and alpha-synuclein', *Science,* 349 (6248), 1255555.

Goedert, M. and Jakes, R. (1990), 'Expression of separate isoforms of human tau protein: correlation with the tau pattern in brain and effects on tubulin polymerization', *Embo j,* 9 (13), 4225-30.

--- (2005), 'Mutations causing neurodegenerative tauopathies', *Biochim Biophys Acta,* 1739 (2-3), 240-50.

Goedert, M., Ghetti, B., and Spillantini, M. G. (2012a), 'Frontotemporal Dementia: Implications for Understanding Alzheimer Disease', *Cold Spring Harb Perspect Med* (2).

--- (2012b), 'Frontotemporal dementia: implications for understanding Alzheimer disease.', *Cold Spring Harb Perspect Med,* 2 (2), a006254.

Goedert, M., et al. (1989a), 'Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain', *EMBO J,* 8 (2), 393-9.

--- (1989b), 'Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease', *Neuron,* 3 (4), 519-26.

--- (1997), 'Phosphorylation of microtubule-associated protein tau by stress-activated protein kinases', *FEBS Lett,* 409 (1), 57-62.

Golde, T. E., et al. (1990), 'Expression of beta amyloid protein precursor mRNAs: recognition of a novel alternatively spliced form and quantitation in Alzheimer's disease using PCR', *Neuron,* 4 (2), 253-67.

Gomez-Isla, T., et al. (1997), 'Neuronal loss correlates with but exceeds neurofibrillary tangles in Alzheimer's disease', *Ann Neurol,* 41 (1), 17-24.

Goraczniak, R. and Gunderson, S. I. (2008), 'The regulatory element in the 3'-untranslated region of human papillomavirus 16 inhibits expression by binding CUG-binding protein 1', *J Biol Chem,* 283 (4), 2286-96.

Goren, A., et al. (2006), 'Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers', *Mol Cell,* 22 (6), 769-81.

Grant, S. F. and Hakonarson, H. (2008), 'Microarray technology and applications in the arena of genome-wide association', *Clin Chem,* 54 (7), 1116-24.

Greenwood, J. A., et al. (1994), 'Casein kinase II preferentially phosphorylates human tau isoforms containing an amino-terminal insert. Identification of threonine 39 as the primary phosphate acceptor', *J Biol Chem,* 269 (6), 4373-80.

Habek, M., Brinar, V. V., and Borovecki, F. (2010), 'Genes associated with multiple sclerosis: 15 and counting', *Expert Rev Mol Diagn,* 10 (7), 857-61.

Halliday, M. and Mallucci, G. R. (2015), 'Review: Modulating the unfolded protein response to prevent neurodegeneration and enhance memory', *Neuropathol Appl Neurobiol,* 41 (4), 414-27.

Hallier, M., et al. (1998), 'The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS', *J Biol Chem,* 273 (9), 4838-42.

Hamre, K. M., et al. (1989), 'Alz-50 immunoreactivity in the neonatal rat: changes in development and co-distribution with MAP-2 immunoreactivity', *Neurosci Lett,* 98 (3), 264-71.

Hamza, T. H., et al. (2010), 'Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease', *Nat Genet,* 42 (9), 781-5.

Hanger, D. P., Anderton, B. H., and Noble, W. (2009), 'Tau phosphorylation: the therapeutic challenge for neurodegenerative disease', *Trends Mol Med,* 15 (3), 112-9.

Hanger, D. P., et al. (1992), 'Glycogen synthase kinase-3 induces Alzheimer's disease-like phosphorylation of tau: generation of paired helical filament epitopes and neuronal localisation of the kinase', *Neurosci Lett,* 147 (1), 58-62.

--- (2007), 'Novel phosphorylation sites in tau from Alzheimer brain support a role for casein kinase 1 in disease pathogenesis', *J Biol Chem,* 282 (32), 23645-54.

Harding, H. P., et al. (2000), 'Regulated translation initiation controls stress-induced gene expression in mammalian cells', *Mol Cell,* 6 (5), 1099-108.

Hardy, J., et al. (2005), 'Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens', *Biochem Soc Trans,* 33 (Pt 4), 582-5.

Harmon, B. T., et al. (2010), 'Functional characterization of a haplotype in the AKT1 gene associated with glucose homeostasis and metabolic syndrome', *Hum Genet,* 128 (6), 635-45.

Harris, D., et al. (2006), 'Identification of cellular factors associated with the 3'-nontranslated region of the hepatitis C virus genome', *Mol Cell Proteomics,* 5 (6), 1006-18.

Hecht, M., Bromberg, Y., and Rost, B. (2015), 'Better prediction of functional effects for sequence variants', *BMC Genomics,* 16 Suppl 8, S1.

Hellman, L. M. and Fried, M. G. (2007), 'Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions', *Nat Protoc,* 2 (8), 1849-61.

Helmke, S. and Pfenninger, K. H. (1995), 'Growth cone enrichment and cytoskeletal association of non-receptor tyrosine kinases', *Cell Motil Cytoskeleton,* 30 (3), 194-207.

Henriquez, J. P., et al. (1995), 'Subpopulations of tau interact with microtubules and actin filaments in various cell types', *Cell Biochem Funct,* 13 (4), 239-50.

Hentati, F., et al. (2014), 'LRRK2 parkinsonism in Tunisia and Norway: A comparative analysis of disease penetrance', *Neurology* (83), 568-9.

Hibbitt, O. C., et al. (2010), 'Long-term physiologically regulated expression of the low-density lipoprotein receptor in vivo using genomic DNA mini-gene constructs', *Mol Ther,* 18 (2), 317-26.

--- (2007), 'Delivery and long-term expression of a 135 kb LDLR genomic DNA locus in vivo by hydrodynamic tail vein injection', *J Gene Med,* 9 (6), 488-97.

Higuchi, R., et al. (1993), 'Kinetic PCR analysis: real-time monitoring of DNA amplification reactions', *Biotechnology (N Y),* 11 (9), 1026-30.

Hill-Burns, E. M., et al. (2014), 'Identification of a novel Parkinson's disease locus via stratified genome-wide association study', *BMC Genomics,* 15, 118.

Himmler, A., et al. (1989), 'Tau consists of a set of proteins with repeated C-terminal microtubule-binding domains and variable N-terminal domains', *Mol Cell Biol,* 9 (4), 1381-8.

Hirokawa, N., Shiomura, Y., and Okabe, S. (1988), 'Tau proteins: the molecular structure and mode of binding on microtubules', *J Cell Biol,* 107 (4), 1449-59.

Hirokawa, N., et al. (1996), 'Selective stabilization of tau in axons and microtubule-associated protein 2C in cell bodies and dendrites contributes to polarized localization of cytoskeletal proteins in mature neurons', *J Cell Biol,* 132 (4), 667-79.

Ho, T. H., et al. (2004), 'Muscleblind proteins regulate alternative splicing', *Embo j,* 23 (15), 3103-12.

Hochberg, Y. and Benjamini, Y. (1990), 'More powerful procedures for multiple significance testing', *Stat Med,* 9 (7), 811-8.

Hoglinger, G. U., et al. (2011), 'Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy', *Nat Genet,* 43 (7), 699-705.

Holmes, B. B. and Diamond, M. I. (2012), 'Cellular mechanisms of protein aggregate propagation', *Curr Opin Neurol,* 25 (6), 721-6.

Hong, M., et al. (1998), 'Mutation-specific functional impairments in distinct tau isoforms of hereditary FTDP-17', *Science,* 282 (5395), 1914-7.

Honore, B., et al. (1995), 'Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes', *J Biol Chem,* 270 (48), 28780-9.

Hook, B., et al. (2005), 'RNA–protein interactions in the yeast three-hybrid system: Affinity, sensitivity, and enhanced library screening', *RNA* (11), 227-33.

Hoozemans, J. J., et al. (2009), 'The unfolded protein response is activated in pretangle neurons in Alzheimer's disease hippocampus', *Am J Pathol,* 174 (4), 1241-51.

--- (2005), 'The unfolded protein response is activated in Alzheimer's disease', *Acta Neuropathol,* 110 (2), 165-72.

Houlden, H., et al. (2001), 'Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype', *Neurology,* 56 (12), 1702-6.

Hsiao, K., et al. (1989), 'Linkage of a prion protein missense variant to Gerstmann-Straussler syndrome', *Nature,* 338 (6213), 342-5.

Huelga, S. C., et al. (2012), 'Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins', *Cell Rep,* 1 (2), 167-78.

Hull, J., et al. (2007), 'Identification of common genetic variation that modulates alternative splicing', *PLoS Genet,* 3 (6), e99.

Hutton, M., et al. (1998), 'Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17', *Nature,* 393 (6686), 702-5.

Hwang, S. C., et al. (1996), 'Activation of phospholipase C-gamma by the concerted action of tau proteins and arachidonic acid', *J Biol Chem,* 271 (31), 18342-9.

Ignelzi, M. A., Jr., et al. (1994), 'Impaired neurite outgrowth of src-minus cerebellar neurons on the cell adhesion molecule L1', *Neuron,* 12 (4), 873-84.

Iioka, H., et al. (2011), 'Efficient detection of RNA-protein interactions using tethered RNAs', *Nucleic Acids Res,* 39 (8), e53.

Im, S. Y., Kim, Y. E., and Kim, Y. J. (2015), 'Genetics of Progressive Supranuclear Palsy', *J Mov Disord,* 8 (3), 122-9.

Inoue, R., et al. (2004), 'Infectious delivery of the 132 kb CDKN2A/CDKN2B genomic DNA region results in correctly spliced gene expression and growth suppression in glioma cells', *Gene Ther,* 11 (15), 1195-204.

Inoue, S. and Salmon, E. D. (1995), 'Force generation by microtubule assembly/disassembly in mitosis and related movements', *Mol Biol Cell,* 6 (12), 1619-40.

International Human Genome Sequencing Consortium (2004), 'Finishing the euchromatic sequence of the human genome', *Nature,* 431 (7011), 931-45.

Ioannou, P. A., et al. (1994), 'A new bacteriophage P1-derived vector for the propagation of large human DNA fragments', *Nat Genet,* 6 (1), 84-9.

Ishii, N., et al. (2006), 'Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction', *J Hum Genet,* 51 (12), 1087-99.

Ittner, L. M. and Gotz, J. (2011), 'Amyloid-beta and tau--a toxic pas de deux in Alzheimer's disease', *Nat Rev Neurosci,* 12 (2), 65-72.

Ittner, L. M., Ke, Y. D., and Gotz, J. (2009), 'Phosphorylated Tau interacts with c-Jun N-terminal kinase-interacting protein 1 (JIP1) in Alzheimer disease', *J Biol Chem,* 284 (31), 20909-16.

Ittner, L. M., et al. (2010), 'Dendritic function of tau mediates amyloid-beta toxicity in Alzheimer's disease mouse models', *Cell,* 142 (3), 387-97.

Jaenisch, R. and Bird, A. (2003), 'Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals', *Nat Genet,* 33 Suppl, 245-54.

Jenkins, S. M. and Johnson, G. V. (1998), 'Tau complexes with phospholipase C-gamma in situ', *Neuroreport,* 9 (1), 67-71.

Jenkins, S. M., et al. (2000), 'Modulation of tau phosphorylation and intracellular localization by cellular stress', *Biochem J,* 345 Pt 2, 263-70.

Jiang, D., et al. (2008), 'Purification and identification of positive regulators binding to a novel element in the c-Jun promoter', *Biochemistry,* 47 (35), 9318-34.

Jiang, H., et al. (2004), 'Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons', *Hum Mol Genet,* 13 (24), 3079-88.

Jicha, G. A., et al. (1999), 'Hierarchical phosphorylation of recombinant tau by the paired-helical filament-associated protein kinase is dependent on cyclic AMP-dependent protein kinase', *J Neurochem,* 72 (1), 214-24.

Johnson, G. V. (1992), 'Differential phosphorylation of tau by cyclic AMP-dependent protein kinase and Ca2+/calmodulin-dependent protein kinase II: metabolic and functional consequences', *J Neurochem,* 59 (6), 2056-62.

Josephs, K. A., et al. (2011), 'Neuropathological background of phenotypical variability in frontotemporal dementia', *Acta Neuropathol,* 122 (2), 137-53.

Jurinke, C., et al. (2001), 'Automated genotyping using the DNA MassArray technology', *Methods Mol Biol,* 170, 103-16.

Kanaan, N. M., et al. (2011), 'Pathogenic forms of tau inhibit kinesin-dependent axonal transport through a mechanism involving activation of axonal phosphotransferases', *J Neurosci,* 31 (27), 9858-68.

Kanemaru, K., et al. (1992), 'Fetal-type phosphorylation of the tau in paired helical filaments', *J Neurochem,* 58 (5), 1667-75.

Kasowski, M., et al. (2013), 'Extensive variation in chromatin states across humans', *Science,* 342 (6159), 750-2.

Kempf, M., et al. (1996), 'Tau binds to the distal axon early in development of polarity in a microtubule- and microfilament-dependent manner', *J Neurosci,* 16 (18), 5583-92.

Kerem, B., et al. (1989), 'Identification of the cystic fibrosis gene: genetic analysis', *Science,* 245 (4922), 1073-80.

Keryer-Bibens, C., Barreau, C., and Osborne, H. B. (2008), 'Tethering of proteins to RNAs by bacteriophage proteins', *Biol Cell,* 100 (2), 125-38.

Kidd, M. (1963), 'Paired helical filaments in electron microscopy of Alzheimer's disease', *Nature,* 197, 192-3.

--- (1964), 'ALZHEIMER'S DISEASE--AN ELECTRON MICROSCOPICAL STUDY', *Brain,* 87, 307-20.

King, M. E., et al. (2006), 'Tau-dependent microtubule disassembly initiated by prefibrillar beta-amyloid', *J Cell Biol,* 175 (4), 541-6.

Kirschner, M. and Mitchison, T. (1986), 'Beyond self-assembly: from microtubules to morphogenesis', *Cell,* 45 (3), 329-42.

Klein, C. and Westenberger, A. (2012), 'Genetics of Parkinson's Disease', *Cold Spring Harb Perspect Med* (2).

Klein, R. J., et al. (2005), 'Complement Factor H Polymorphism in Age-Related Macular Degeneration', *Science,* 308 (5720), 385-9.

Knight, J. C. (2004), 'Allele-specific gene expression uncovered', *Trends Genet,* 20 (3), 113-6.

Knight, J. C., et al. (2003), 'In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading', *Nat Genet,* 33 (4), 469-75.

Konzack, S., et al. (2007), 'Swimming against the tide: mobility of the microtubule-associated protein tau in neurons', *J Neurosci,* 27 (37), 9916-27.

Koressaar, T. and Remm, M. (2007), 'Enhancements and modifications of primer design program Primer3', *Bioinformatics,* 23 (10), 1289-91.

Kosik, K. S., et al. (1989a), 'Developmentally regulated expression of specific tau sequences', *Neuron,* 2 (4), 1389-97.

--- (1989b), 'Tau in situ hybridization in normal and Alzheimer brain: localization in the somatodendritic compartment', *Ann Neurol,* 26 (3), 352-61.

Kouri, N., et al. (2011), 'Corticobasal degeneration: a pathologically distinct 4R tauopathy', *Nat Rev Neurol,* 7 (5), 263-72.

--- (2015), 'Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy', *Nat Commun,* 6, 7247.

Ksiezak-Reding, H., et al. (1994), 'Ultrastructure and biochemical composition of paired helical filaments in corticobasal degeneration', *Am J Pathol,* 145 (6), 1496-508.

Kunda, P., et al. (2001), 'Evidence for the involvement of Tiam1 in axon formation', *J Neurosci,* 21 (7), 2361-72.

Kuo, H. C., Nasim, F. H., and Grabowski, P. J. (1991), 'Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle', *Science,* 251 (4997), 1045-50.

Kwak, H., et al. (2013), 'Precise maps of RNA polymerase reveal how promoters direct initiation and pausing', *Science,* 339 (6122), 950-3.

Kwan, T., et al. (2008), 'Genome-wide analysis of transcript isoform variation in humans', *Nat Genet,* 40 (2), 225-31.

Kwiatkowski, T. J., Jr., et al. (2009), 'Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis', *Science,* 323 (5918), 1205-8.

LaFerla, F. M. and Green, K. N. (2012), 'Animal models of Alzheimer disease', *Cold Spring Harb Perspect Med,* 2 (11).

Lambert, M. P., et al. (1998), 'Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins', *Proc Natl Acad Sci U S A,* 95 (11), 6448-53.

Lander, E. S. (1996), 'The new genomics: global views of biology', *Science,* 274 (5287), 536-9.

Lane, D., Prentki, P., and Chandler, M. (1992), 'Use of gel retardation to analyze protein-nucleic acid interactions', *Microbiol Rev,* 56 (4), 509-28.

Lappalainen, T., et al. (2013), 'Transcriptome and genome sequencing uncovers functional variation in humans', *Nature,* 501 (7468), 506-11.

Lau, P. P., Chang, B. H., and Chan, L. (2001), 'Two-hybrid cloning identifies an RNA-binding protein, GRY-RBP, as a component of apobec-1 editosome', *Biochem Biophys Res Commun,* 282 (4), 977-83.

Laws, S. M., et al. (2007), 'Fine mapping of the MAPT locus using quantitative trait analysis identifies possible causal variants in Alzheimer's disease', *Mol Psychiatry,* 12 (5), 510-7.

Lee, G. (2005), 'Tau and src family tyrosine kinases', *Biochim Biophys Acta,* 1739 (2-3), 323-30.

Lee, G., Neve, R. L., and Kosik, K. S. (1989), 'The microtubule binding domain of tau protein', *Neuron,* 2 (6), 1615-24.

Lee, G., et al. (1998), 'Tau interacts with src-family non-receptor tyrosine kinases', *J Cell Sci,* 111 (Pt 21), 3167-77.

--- (2004), 'Phosphorylation of tau by fyn: implications for Alzheimer's disease', *J Neurosci,* 24 (9), 2304-12.

Lee, V. M., et al. (2011), 'Developing therapeutic approaches to tau, selected kinases, and related neuronal protein targets', *Cold Spring Harb Perspect Med,* 1 (1), a006437.

Lee, Y. and Rio, D. C. (2015), 'Mechanisms and Regulation of Alternative Pre-mRNA Splicing', *Annu Rev Biochem,* 84, 291-323.

Lees, A. J., Hardy, J., and Revesz, T. (2009), 'Parkinson's disease', *Lancet,* 373 (9680), 2055-66.

Leroy, K., et al. (2012), 'Lack of tau proteins rescues neuronal cell death and decreases amyloidogenic processing of APP in APP/PS1 mice', *Am J Pathol,* 181 (6), 1928-40.

Leroy, O., et al. (2006), 'ETR-3 represses Tau exons 2/3 inclusion, a splicing event abnormally enhanced in myotonic dystrophy type I', *J Neurosci Res,* 84 (4), 852-9.

Lewis, C. M. (2002), 'Genetic association studies: design, analysis and interpretation', *Brief Bioinform,* 3 (2), 146-53.

Lewis, J., et al. (2001), 'Enhanced neurofibrillary degeneration in transgenic mice expressing mutant tau and APP', *Science,* 293 (5534), 1487-91.

Li, X., et al. (2011), 'Novel diffusion barrier for axonal retention of Tau in neurons and its failure in neurodegeneration', *Embo j,* 30 (23), 4825-37.

Liao, G. and Gundersen, G. G. (1998), 'Kinesin is a candidate for cross-bridging microtubules and intermediate filaments. Selective binding of kinesin to detyrosinated tubulin and vimentin', *J Biol Chem,* 273 (16), 9797-803.

Liao, H., et al. (1998), 'Protein phosphatase 1 is targeted to microtubules by the microtubule-associated protein Tau', *J Biol Chem,* 273 (34), 21901-8.

Libbrecht, M. W. and Noble, W. S. (2015), 'Machine learning applications in genetics and genomics', *Nat Rev Genet,* 16 (6), 321-32.

Lill, C. M., et al. (2012), 'Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database', *PLoS Genet,* 8 (3), e1002548.

Lim, F., Spingola, M., and Peabody, D. S. (1994), 'Altering the RNA binding specificity of a translational repressor', *J Biol Chem,* 269 (12), 9006-10.

Lim, L. P. and Burge, C. B. (2001), 'A computational analysis of sequence features involved in recognition of short introns', *Proc Natl Acad Sci U S A,* 98 (20), 11193-8.

Lin, M. K. and Farrer, M. J. (2014), 'Genetics and genomics of Parkinson's disease', *Genome Med,* 6 (6), 48.

Lindwall, G. and Cole, R. D. (1984), 'Phosphorylation affects the ability of tau protein to promote microtubule assembly', *J Biol Chem,* 259 (8), 5301-5.

Listerman, I., Sapra, A. K., and Neugebauer, K. M. (2006), 'Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells', *Nat Struct Mol Biol,* 13 (9), 815-22.

Litersky, J. M. and Johnson, G. V. (1992), 'Phosphorylation by cAMP-dependent protein kinase inhibits the degradation of tau by calpain', *J Biol Chem,* 267 (3), 1563-8.

Litman, P., et al. (1993), 'Subcellular localization of tau mRNA in differentiating neuronal cell culture: implications for neuronal polarity', *Neuron,* 10 (4), 627-38.

Liu, C. and Götz, J. (2013), 'Profiling Murine Tau with 0N, 1N and 2N Isoform-Specific Antibodies in Brain and Peripheral Organs Reveals Distinct Subcellular Localization, with the 1N Isoform Being Enriched in the Nucleus', *PLoS One* (8).

Liu, F. and Gong, C. X. (2008), 'Tau exon 10 alternative splicing and tauopathies', *Mol Neurodegener,* 3, 8.

Liu, H. X., et al. (2000), 'Exonic splicing enhancer motif recognized by human SC35 under splicing conditions', *Mol Cell Biol,* 20 (3), 1063-71.

Liu, W. K., et al. (1995), 'Detection of a Cdc2-related kinase associated with Alzheimer paired helical filaments', *Am J Pathol,* 146 (1), 228-38.

--- (2001), 'Relationship of the extended tau haplotype to tau biochemistry and neuropathology in progressive supranuclear palsy', *Ann Neurol,* 50 (4), 494-502.

Livak, K. J. and Schmittgen, T. D. (2001), 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods,* 25 (4), 402-8.

Livak, KJ (1997 & 2001), 'ABI Prism 7700 Sequence detection System User Bulletin #2 Relative quantification of gene expression', (ABI company publication).

Lo, C. S., et al. (2012), 'Heterogeneous nuclear ribonucleoprotein F suppresses angiotensinogen gene expression and attenuates hypertension and kidney injury in diabetic mice', *Diabetes,* 61 (10), 2597-608.

--- (2015), 'Overexpression of heterogeneous nuclear ribonucleoprotein F stimulates renal Ace-2 gene expression and prevents TGF-beta1-induced kidney injury in a mouse model of diabetes', *Diabetologia,* 58 (10), 2443-54.

Lo, H. S., et al. (2003), 'Allelic variation in gene expression is common in the human genome', *Genome Res,* 13 (8), 1855-62.

Loomis, P. A., et al. (1990), 'Identification of nuclear tau isoforms in human neuroblastoma cells', *Proc Natl Acad Sci U S A,* 87 (21), 8422-6.

Lovci, M. T., et al. (2013), 'Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges', *Nat Struct Mol Biol,* 20 (12), 1434-42.

Lowary, P. T. and Uhlenbeck, O. C. (1987), 'An RNA mutation that increases the affinity of an RNA-protein interaction', *Nucleic Acids Res,* 15 (24), 10483-93.

Lu, T., et al. (2014), 'REST and stress resistance in ageing and Alzheimer's disease', *Nature,* 507 (7493), 448-54.

Luco, R. F., et al. (2011), 'Epigenetics in alternative pre-mRNA splicing', *Cell,* 144 (1), 16-26.

Ludolph, A. C., et al. (2009), 'Tauopathies with parkinsonism: clinical spectrum, neuropathologic basis, biological markers, and treatment options', *Eur J Neurol,* 16 (3), 297-309.

Lufino, M. M., Manservigi, R., and Wade-Martins, R. (2007), 'An S/MAR-based infectious episomal genomic DNA expression vector provides long-term regulated functional complementation of LDLR deficiency', *Nucleic Acids Res,* 35 (15), e98.

Lufino, M. M., Edser, P. A., and Wade-Martins, R. (2008), 'Advances in high-capacity extrachromosomal vector technology: episomal maintenance, vector delivery, and transgene expression', *Mol Ther,* 16 (9), 1525-38.

Lufino, M. M., et al. (2011), 'Episomal transgene expression in pluripotent stem cells', *Methods Mol Biol,* 767, 369-87.

--- (2013), 'A GAA repeat expansion reporter model of Friedreich's ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds', *Hum Mol Genet,* 22 (25), 5173-87.

Maas, T., Eidenmuller, J., and Brandt, R. (2000), 'Interaction of tau with the neural membrane cortex is regulated by phosphorylation at sites that are modified in paired helical filaments', *J Biol Chem,* 275 (21), 15733-40.

Mackenzie, I. R., Rademakers, R., and Neumann, M. (2010), 'TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia', *Lancet Neurol,* 9 (10), 995-1007.

Macnab, S. and Whitehouse, A. (2009), 'Progress and prospects: human artificial chromosomes', *Gene Ther,* 16 (10), 1180-8.

Maeda, S., et al. (2006), 'Increased levels of granular tau oligomers: an early sign of brain aging and Alzheimer's disease', *Neurosci Res,* 54 (3), 197-201.

--- (2007), 'Granular tau oligomers as intermediates of tau filaments', *Biochemistry,* 46 (12), 3856-61.

Magnani, E., et al. (2007), 'Interaction of tau protein with the dynactin complex', *Embo j,* 26 (21), 4546-54.

Mailliot, C., et al. (1998), 'Phosphorylation of specific sets of tau isoforms reflects different neurofibrillary degeneration processes', *FEBS Lett,* 433 (3), 201-4.

Mandelkow, E. M. and Mandelkow, E. (2012), 'Biochemistry and cell biology of tau protein in neurofibrillary degeneration', *Cold Spring Harb Perspect Med,* 2 (7), a006247.

Maness, P. F., et al. (1988), 'c-src gene product in developing rat brain is enriched in nerve growth cone membranes', *Proc Natl Acad Sci U S A,* 85 (14), 5001-5.

Maniatis, T. and Reed, R. (2002), 'An extensive network of coupling among gene expression machines', *Nature,* 416 (6880), 499-506.

Marin, I., van Egmond, W. N., and van Haastert, P. J. (2008), 'The Roco protein family: a functional perspective', *Faseb j,* 22 (9), 3103-10.

Marquis, J., et al. (2006), 'CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding', *Biochem J,* 400 (2), 291-301.

Martinez-Canto, A., et al. (2012), 'TGFBR1 intralocus epistatic interaction as a risk factor for colorectal cancer', *PLoS One,* 7 (1), e30812.

Masliah, E., et al. (2001), 'Altered expression of synaptic proteins occurs early during progression of Alzheimer's disease', *Neurology,* 56 (1), 127-9.

Matera, A. G. and Wang, Z. (2014), 'A day in the life of the spliceosome', *Nat Rev Mol Cell Biol,* 15 (2), 108-21.

Matlin, A. J., Clark, F., and Smith, C. W. (2005), 'Understanding alternative splicing: towards a cellular code', *Nat Rev Mol Cell Biol,* 6 (5), 386-98.

Maurano, M. T., et al. (2012), 'Systematic localization of common disease-associated variation in regulatory DNA', *Science,* 337 (6099), 1190-5.

Maynard, N. D., et al. (2008), 'Genome-wide mapping of allele-specific protein-DNA interactions in human cells', *Nat Methods,* 5 (4), 307-9.

Mazanetz, M. P. and Fischer, P. M. (2007), 'Untangling tau hyperphosphorylation in drug design for neurodegenerative diseases', *Nat Rev Drug Discov,* 6 (6), 464-79.

Mazzulli, J. R., et al. (2011), 'Gaucher disease glucocerebrosidase and alpha-synuclein form a bidirectional pathogenic loop in synucleinopathies', *Cell,* 146 (1), 37-52.

McDaniell, R., et al. (2010), 'Heritable individual-specific and allele-specific chromatin signatures in humans', *Science,* 328 (5975), 235-9.

McHugh, C. A., Russell, P., and Guttman, M. (2014), 'Methods for comprehensive experimental identification of RNA-protein interactions', *Genome Biol,* 15 (1), 203.

Mehta, A. and Driscoll, D. M. (1998), 'A sequence-specific RNA-binding protein complements apobec-1 To edit apolipoprotein B mRNA', *Mol Cell Biol,* 18 (8), 4426-32.

Mhatre, S. D., et al. (2013), 'Invertebrate models of Alzheimer's disease', *J Alzheimers Dis,* 33 (1), 3-16.

Mili, S. and Steitz, J. A. (2004), 'Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses', *RNA* (10), 1692-4.

Min, H., Chan, R. C., and Black, D. L. (1995), 'The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event', *Genes Dev,* 9 (21), 2659-71.

Miniard, A. C., et al. (2010), 'Nucleolin binds to a subset of selenoprotein mRNAs and regulates their expression', *Nucleic Acids Res,* 38 (14), 4807-20.

Mittal, N., et al. (2009), 'Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks', *Proc Natl Acad Sci U S A* (106), 20300-5.

Montes, M., et al. (2012), 'Functional coupling of transcription and splicing', *Gene,* 501 (2), 104-17.

Moraga, D. M., et al. (1993), 'A tau fragment containing a repetitive sequence induces bundling of actin filaments', *J Neurochem,* 61 (3), 979-86.

Moreno, J. A., et al. (2013), 'Oral treatment targeting the unfolded protein response prevents neurodegeneration and clinical disease in prion-infected mice', *Sci Transl Med,* 5 (206), 206ra138.

--- (2012), 'Sustained translational repression by eIF2alpha-P mediates prion neurodegeneration', *Nature,* 485 (7399), 507-11.

Morita, T. and Sobue, K. (2009), 'Specification of neuronal polarity regulated by local translation of CRMP2 and Tau via the mTOR-p70S6K pathway', *J Biol Chem,* 284 (40), 27734-45.

Morris, H. R., et al. (1999), 'The tau gene A0 polymorphism in progressive supranuclear palsy and related neurodegenerative diseases', *J Neurol Neurosurg Psychiatry,* 66 (5), 665-7.

Morris, M., et al. (2011), 'The many faces of tau', *Neuron,* 70 (3), 410-26.

Morsch, R., Simon, W., and Coleman, P. D. (1999), 'Neurons may live for decades with neurofibrillary tangles', *J Neuropathol Exp Neurol,* 58 (2), 188-97.

Mourelatos, Z., et al. (2001), 'SMN interacts with a novel family of hnRNP and spliceosomal proteins', *Embo j,* 20 (19), 5443-52.

Myers, A. J., et al. (2007), 'The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts', *Neurobiol Dis,* 25 (3), 561-70.

Nacharaju, P., et al. (1999), 'Accelerated filament formation from tau protein with specific FTDP-17 missense mutations', *FEBS Lett,* 447 (2-3), 195-9.

Nahnsen, S., et al. (2013), 'Tools for label-free peptide quantification', *Mol Cell Proteomics,* 12 (3), 549-56.

Nakata, T. and Hirokawa, N. (2003), 'Microtubules provide directional cues for polarized axonal transport through interaction with kinesin motor head', *J Cell Biol,* 162 (6), 1045-55.

Nakaya, T., et al. (2013), 'FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns', *Rna,* 19 (4), 498-509.

Nalls, M. A., et al. (2011), 'Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies', *Lancet,* 377 (9766), 641-9.

--- (2014), 'Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease', *Nat Genet,* 46 (9), 989-93.

NCI-NHGRI Working Group on Replication in Association Studies, et al. (2007), 'Replicating genotype-phenotype associations', *Nature,* 447 (7145), 655-60.

Nelson, P. T., et al. (2012), 'Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature', *J Neuropathol Exp Neurol,* 71 (5), 362-81.

Nembaware, V., et al. (2004), 'Allele-specific transcript isoforms in human', *FEBS Lett,* 577 (1-2), 233-8.

Nijholt, D. A., et al. (2012), 'The unfolded protein response is associated with early tau pathology in the hippocampus of tauopathies', *J Pathol,* 226 (5), 693-702.

Nogales, E., et al. (1999), 'High-resolution model of the microtubule', *Cell,* 96 (1), 79-88.

Nojima, T., et al. (2015), 'Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing', *Cell* (161), 526-40.

Obeso, J. A., et al. (2014), 'The expanding universe of disorders of the basal ganglia', *Lancet,* 384 (9942), 523-31.

Olanow, C. W. and Prusiner, S. B. (2009), 'Is Parkinson's disease a prion disorder?', *Proc Natl Acad Sci U S A,* 106 (31), 12571-2.

Oliva, R., et al. (1998), 'Significant changes in the tau A0 and A3 alleles in progressive supranuclear palsy and improved genotyping by silver detection', *Arch Neurol,* 55 (8), 1122-4.

Orozco, D., et al. (2012), 'Loss of fused in sarcoma (FUS) promotes pathological Tau splicing', *EMBO Rep,* 13 (8), 759-64.

Owen, F., et al. (1989), 'Insertion in prion protein gene in familial Creutzfeldt-Jakob disease', *Lancet,* 1 (8628), 51-2.

Ozelius, L. J., et al. (2006), 'LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews', *N Engl J Med* (354; United States), 424-5.

Palmer, M. S., et al. (1991), 'Homozygous prion protein genotype predisposes to sporadic Creutzfeldt-Jakob disease', *Nature,* 352 (6333), 340-2.

Palop, J. J. and Mucke, L. (2010), 'Amyloid-beta-induced neuronal dysfunction in Alzheimer's disease: from synapses toward neural networks', *Nat Neurosci,* 13 (7), 812-8.

Pan, K. M., et al. (1993), 'Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins', *Proc Natl Acad Sci U S A,* 90 (23), 10962-6.

Pandit, S., et al. (2013), 'Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing', *Mol Cell,* 50 (2), 223-35.

Panicker, N., et al. (2015), 'Fyn Kinase Regulates Microglial Neuroinflammatory Responses in Cell Culture and Animal Models of Parkinson's Disease', *J Neurosci,* 35 (27), 10058-77.

Papasozomenos, S. C. and Binder, L. I. (1987), 'Phosphorylation determines two distinct species of Tau in the central nervous system', *Cell Motil Cytoskeleton,* 8 (3), 210-26.

Park, Y. and Kellis, M. (2015), 'Deep learning for regulatory genomics', *Nat Biotechnol,* 33 (8), 825-6.

Pastor, P., et al. (2004), 'Novel haplotypes in 17q21 are associated with progressive supranuclear palsy', *Ann Neurol,* 56 (2), 249-58.

Patel, G. P., Ma, S., and Bag, J. (2005), 'The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex', *Nucleic Acids Res,* 33 (22), 7074-89.

Paz, I., et al. (2010), 'SFmap: a web server for motif analysis and prediction of splicing factor binding sites', *Nucleic Acids Res,* 38 (Web Server issue), W281-5.

Peddigari, S., et al. (2013), 'hnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition', *Nucleic Acids Res,* 41 (1), 575-85.

Peltz, S. W. and Ross, J. (1987), 'Autogenous regulation of histone mRNA decay by histone proteins in a cell-free system', *Mol Cell Biol,* 7 (12), 4345-56.

Peruzzi, P. P., et al. (2009), 'Physiological transgene regulation and functional complementation of a neurological disease gene deficiency in neurons', *Mol Ther,* 17 (9), 1517-26.

Picardi, E. and Pesole, G. (2010), 'Computational methods for ab initio and comparative gene finding', *Methods Mol Biol,* 609, 269-84.

Pickrell, J. K., et al. (2010), 'Understanding mechanisms underlying human gene expression variation with RNA sequencing', *Nature,* 464 (7289), 768-72.

Pittman, A. M., et al. (2004), 'The structure of the tau haplotype in controls and in progressive supranuclear palsy', *Hum Mol Genet,* 13 (12), 1267-74.

--- (2005), 'Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration', *J Med Genet,* 42 (11), 837-46.

Piva, F., et al. (2012), 'SpliceAid 2: a database of human splicing factors expression data and RNA target motifs', *Hum Mutat,* 33 (1), 81-5.

Plath, K., et al. (2002), 'Xist RNA and the mechanism of X chromosome inactivation', *Annu Rev Genet,* 36, 233-78.

Polman, C. H., et al. (2005), 'Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria"', *Ann Neurol,* 58 (6), 840-6.

Polydorides, A. D., et al. (2000), 'A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing', *Proc Natl Acad Sci U S A,* 97 (12), 6350-5.

Pomerantz, M. M., et al. (2009), 'The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer', *Nat Genet,* 41 (8), 882-4.

Ponthier, J. L., et al. (2006), 'Fox-2 splicing factor binds to a conserved intron motif to promote inclusion of protein 4.1R alternative exon 16', *J Biol Chem,* 281 (18), 12468-74.

Poorkaj, P., et al. (1998), 'Tau is a candidate gene for chromosome 17 frontotemporal dementia', *Ann Neurol,* 43 (6), 815-25.

Pope, W., et al. (1993), 'Phosphorylated tau epitope of Alzheimer's disease is coupled to axon development in the avian central nervous system', *Exp Neurol,* 120 (1), 106-13.

Prendergast, J. G., et al. (2012), 'A genome-wide screen in human embryonic stem cells reveals novel sites of allele-specific histone modification associated with known disease loci', *Epigenetics Chromatin,* 5 (1), 6.

Price, A. L., et al. (2006), 'Principal components analysis corrects for stratification in genome-wide association studies', *Nat Genet,* 38 (8), 904-9.

Pritchard, J. K. (2001), 'Are rare variants responsible for susceptibility to complex diseases?', *Am J Hum Genet,* 69 (1), 124-37.

Proudfoot, N. J., Furger, A., and Dye, M. J. (2002), 'Integrating mRNA processing with transcription', *Cell,* 108 (4), 501-12.

Prusiner, S. B. (1982), 'Novel proteinaceous infectious particles cause scrapie', *Science,* 216 (4542), 136-44.

--- (1998), 'Prions', *Proc Natl Acad Sci U S A,* 95 (23), 13363-83.

Purcell, S., et al. (2007), 'PLINK: a tool set for whole-genome association and population-based linkage analyses', *Am J Hum Genet,* 81 (3), 559-75.

Qiang, L., et al. (2006), 'Tau protects microtubules in the axon from severing by katanin', *J Neurosci,* 26 (12), 3120-9.

Quinn, N., et al. (2009), 'Movement Disorders', in C.Howard Clarke, R., M. Rossor, and S. Shorvon (eds.), *Neurology: A Queen Square Textbook* (Wiley-Blackwell, A John Wiley & Sons, Ltd., Publication), 155-87.

Quintanilla, R. A., et al. (2012), 'Truncated tau and Aβ cooperatively impair mitochondria in primary neurons', *Neurobiol Aging,* 33 (3), 619 e25-35.

Rademakers, R., et al. (2005), 'High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy', *Hum Mol Genet,* 14 (21), 3281-92.

Ranish, J. A., et al. (2003), 'The study of macromolecular complexes by quantitative proteomics', *Nat Genet,* 33 (3), 349-55.

Rapoport, M., et al. (2002), 'Tau is essential to beta -amyloid-induced neurotoxicity', *Proc Natl Acad Sci U S A,* 99 (9), 6364-9.

Rattenbacher, B. and Bohjanen, P. R. (2012), 'Evaluating posttranscriptional regulation of cytokine genes', *Methods Mol Biol,* 820, 71-89.

Ray, D., et al. (2013), 'A compendium of RNA-binding motifs for decoding gene regulation', *Nature,* 499 (7457), 172-7.

Redon, R., et al. (2006), 'Global variation in copy number in the human genome', *Nature,* 444 (7118), 444-54.

Reed, N. A., et al. (2006), 'Microtubule acetylation promotes kinesin-1 binding and transport', *Curr Biol,* 16 (21), 2166-72.

Refenes, N., et al. (2009), 'Role of the H1 haplotype of microtubule-associated protein tau (MAPT) gene in Greek patients with Parkinson's disease', *BMC Neurol,* 9, 26.

Ren, Y. R., et al. (2012), 'Unbiased discovery of interactions at a control locus driving expression of the cancer-specific therapeutic and diagnostic target, mesothelin', *J Proteome Res,* 11 (11), 5301-10.

Reynolds, C. H., et al. (1997), 'Stress-activated protein kinase/c-jun N-terminal kinase phosphorylates tau protein', *J Neurochem,* 68 (4), 1736-44.

--- (2008), 'Phosphorylation regulates tau interactions with Src homology 3 domains of phosphatidylinositol 3-kinase, phospholipase Cgamma1, Grb2, and Src family kinases', *J Biol Chem,* 283 (26), 18177-86.

Rhein, V., et al. (2009), 'Amyloid-beta and tau synergistically impair the oxidative phosphorylation system in triple transgenic Alzheimer's disease mice', *Proc Natl Acad Sci U S A,* 106 (47), 20057-62.

Riaz, M., et al. (2012), 'Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival, and mRNA expression of the nearest genes', *Breast Cancer Res Treat,* 133 (3), 843-51.

Rimoldi, V., et al. (2013), 'Dual role of G-runs and hnRNP F in the regulation of a mutation-activated pseudoexon in the fibrinogen gamma-chain transcript', *PLoS One,* 8 (3), e59333.

Risch, N. and Merikangas, K. (1996), 'The future of genetic studies of complex human diseases', *Science,* 273 (5281), 1516-7.

Risch, N. J. (2000), 'Searching for genetic determinants in the new millennium', *Nature,* 405 (6788), 847-56.

Robberson, B. L., Cote, G. J., and Berget, S. M. (1990), 'Exon definition may facilitate splice site selection in RNAs with multiple exons', *Mol Cell Biol,* 10 (1), 84-94.

Roberson, E. D., et al. (2007), 'Reducing endogenous tau ameliorates amyloid beta-induced deficits in an Alzheimer's disease mouse model', *Science,* 316 (5825), 750-4.

Ron, D. and Walter, P. (2007), 'Signal integration in the endoplasmic reticulum unfolded protein response', *Nat Rev Mol Cell Biol,* 8 (7), 519-29.

Rosner, H., et al. (1995), 'Developmental expression of tau proteins in the chicken and rat brain: rapid down-regulation of a paired helical filament epitope in the rat cerebral cortex coincides with the transition from immature to adult tau isoforms', *Int J Dev Neurosci,* 13 (6), 607-17.

Ross, A. F., et al. (1997), 'Characterization of a beta-actin mRNA zipcode-binding protein', *Mol Cell Biol,* 17 (4), 2158-65.

Ross, O. A., et al. (2008), 'Genomic investigation of alpha-synuclein multiplication and parkinsonism', *Ann Neurol,* 63 (6), 743-50.

Rossor, M., et al. (2009), 'Cognitive Impairment and Dementia', in C.Howard Clarke, R., M. Rossor, and S. Shorvon (eds.), *Neurology: A Queen Square Textbook* (Wiley-Blackwell, A John Wiley & Sons, Ltd., Publication), 245-88.

Rouskin, S., et al. (2014), 'Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo', *Nature,* 505 (7485), 701-5.

Roy, S., et al. (2005), 'Axonal transport defects: a common theme in neurodegenerative diseases', *Acta Neuropathol,* 109 (1), 5-13.

Rutledge, R. G. and Cote, C. (2003), 'Mathematics of quantitative kinetic PCR and the application of standard curves', *Nucleic Acids Res,* 31 (16), e93.

Saeki, Y., Breakefield, X. O., and Chiocca, E. A. (2003), 'Improved HSV-1 amplicon packaging system using ICP27-deleted, oversized HSV-1 BAC DNA', *Methods Mol Med,* 76, 51-60.

Saeki, Y., et al. (2001), 'Improved helper virus-free packaging system for HSV amplicon vectors using an ICP27-deleted, oversized HSV-1 DNA in a bacterial artificial chromosome', *Mol Ther,* 3 (4), 591-601.

Saha, A. R., et al. (2004), 'Parkinson's disease alpha-synuclein mutations exhibit defective axonal transport in cultured neurons', *J Cell Sci,* 117 (Pt 7), 1017-24.

Sandebring, A., et al. (2009), 'Parkin deficiency disrupts calcium homeostasis by modulating phospholipase C signalling', *Febs j,* 276 (18), 5041-52.

Sander, J. D. and Joung, J. K. (2014), 'CRISPR-Cas systems for editing, regulating and targeting genomes', *Nat Biotechnol,* 32 (4), 347-55.

Sanders, D. W., et al. (2016), 'Prions and Protein Assemblies that Convey Biological Information in Health and Disease', *Neuron,* 89 (3), 433-48.

Sandoval, P. C., et al. (2013), 'Proteome-wide measurement of protein half-lives and translation rates in vasopressin-sensitive collecting duct cells', *J Am Soc Nephrol,* 24 (11), 1793-805.

Santarella, R. A., et al. (2004), 'Surface-decoration of microtubules by human tau', *J Mol Biol,* 339 (3), 539-53.

Schaal, T. D. and Maniatis, T. (1999), 'Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences', *Mol Cell Biol,* 19 (3), 1705-19.

Schaap, I. A., et al. (2007), 'Tau protein binding forms a 1 nm thick layer along protofilaments without affecting the radial elasticity of microtubules', *J Struct Biol,* 158 (3), 282-92.

Schagger, H. and Ohm, T. G. (1995), 'Human diseases with defects in oxidative phosphorylation. 2. F1F0 ATP-synthase defects in Alzheimer disease revealed by blue native polyacrylamide gel electrophoresis', *Eur J Biochem,* 227 (3), 916-21.

Schaub, M. A., et al. (2012), 'Linking disease associations with regulatory information in the human genome', *Genome Res,* 22 (9), 1748-59.

Scheff, S. W., et al. (2006), 'Hippocampal synaptic loss in early Alzheimer's disease and mild cognitive impairment', *Neurobiol Aging,* 27 (10), 1372-84.

--- (2007), 'Synaptic alterations in CA1 in mild Alzheimer disease and mild cognitive impairment', *Neurology,* 68 (18), 1501-8.

Schmitt, K., et al. (2012), 'Insights into mitochondrial dysfunction: aging, amyloid-beta, and tau-A deleterious trio', *Antioxid Redox Signal,* 16 (12), 1456-66.

Schmittgen, T. D. and Livak, K. J. (2008), 'Analyzing real-time PCR data by the comparative C(T) method', *Nat Protoc,* 3 (6), 1101-8.

Schork, N. J., et al. (2009), 'Common vs. Rare Allele Hypotheses for Complex Diseases', *Curr Opin Genet Dev,* 19 (3), 212-9.

Schreiber, E., et al. (1989), 'Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells', *Nucleic Acids Res,* 17 (15), 6419.

Schulz, K. L., et al. (2012), 'A new link to mitochondrial impairment in tauopathies', *Mol Neurobiol,* 46 (1), 205-16.

Schwanhäusser, Björn, et al. (2011), 'Global quantification of mammalian gene expression control', *Nature,* 473, 337-42.

Seelaar, H., et al. (2011), 'Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review', *J Neurol Neurosurg Psychiatry,* 82 (5), 476-86.

Sena-Esteves, M., et al. (2000), 'HSV-1 amplicon vectors--simplicity and versatility', *Mol Ther,* 2 (1), 9-15.

Senear, D. F. and Brenowitz, M. (1991), 'Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay', *J Biol Chem,* 266 (21), 13661-71.

Sennvik, K., et al. (2007), 'Tau-4R suppresses proliferation and promotes neuronal differentiation in the hippocampus of tau knockin/knockout mice', *Faseb j,* 21 (9), 2149-61.

Sergeant, N., Wattez, A., and Delacourte, A. (1999), 'Neurofibrillary degeneration in progressive supranuclear palsy and corticobasal degeneration: tau pathologies with exclusively "exon 10" isoforms', *J Neurochem,* 72 (3), 1243-9.

Sergeant, N., et al. (2001), 'Dysregulation of human brain microtubule-associated tau mRNA maturation in myotonic dystrophy type 1', *Hum Mol Genet,* 10 (19), 2143-55.

Seubert, P., et al. (1995), 'Detection of phosphorylated Ser262 in fetal tau, adult tau, and paired helical filament tau', *J Biol Chem,* 270 (32), 18917-22.

Shahani, N. and Brandt, R. (2002), 'Functions and malfunctions of the tau proteins', *Cell Mol Life Sci,* 59 (10), 1668-80.

Shankarling, G., et al. (2014), 'Transcriptome-wide RNA interaction profiling reveals physical and functional targets of hnRNP L in human T cells', *Mol Cell Biol,* 34 (1), 71-83.

Sharp, A. J., et al. (2005), 'Segmental duplications and copy-number variation in the human genome', *Am J Hum Genet,* 77 (1), 78-88.

Shaw-Smith, C., et al. (2006), 'Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability', *Nat Genet,* 38 (9), 1032-7.

Shen, L. X., Basilion, J. P., and Stanton, V. P., Jr. (1999), 'Single-nucleotide polymorphisms can cause different structural folds of mRNA', *Proc Natl Acad Sci U S A,* 96 (14), 7871-6.

Shipton, O. A., et al. (2011), 'Tau protein is required for amyloid {beta}-induced impairment of hippocampal long-term potentiation', *J Neurosci,* 31 (5), 1688-92.

Shizuya, H., et al. (1992), 'Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector', *Proc Natl Acad Sci U S A,* 89 (18), 8794-7.

Sidransky, E. and Lopez, G. (2012), 'The link between the GBA gene and parkinsonism', *Lancet Neurol,* 11 (11), 986-98.

Silver, D., et al. (2016), 'Mastering the game of Go with deep neural networks and tree search', *Nature,* 529 (7587), 484-9.

Simon-Sanchez, J., et al. (2009), 'Genome-wide association study reveals genetic risk underlying Parkinson's disease', *Nat Genet,* 41 (12), 1308-12.

Singh, J. and Padgett, R. A. (2009), 'Rates of in situ transcription and splicing in large human genes', *Nat Struct Mol Biol,* 16 (11), 1128-33.

Singleton, A. and Hardy, J. (2011), 'A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci', *Hum Mol Genet,* 20 (R2), R158-62.

Sjoberg, M. K., et al. (2006), 'Tau protein binds to pericentromeric DNA: a putative role for nuclear tau in nucleolar organization', *J Cell Sci,* 119 (Pt 10), 2025-34.

Skipper, L., et al. (2004), 'Linkage disequilibrium and association of MAPT H1 in Parkinson disease', *Am J Hum Genet,* 75 (4), 669-77.

Smith, A. J., et al. (2012), 'Use of allele-specific FAIRE to determine functional regulatory polymorphism using large-scale genotyping arrays', *PLoS Genet,* 8 (8), e1002908.

Smith, C. W. and Valcarcel, J. (2000), 'Alternative pre-mRNA splicing: the logic of combinatorial control', *Trends Biochem Sci,* 25 (8), 381-8.

Smith, P. J., et al. (2006), 'An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers', *Hum Mol Genet,* 15 (16), 2490-508.

Snow, A. K., et al. (2015), 'APC promoter 1B deletion in seven American families with familial adenomatous polyposis', *Clin Genet,* 88 (4), 360-5.

Soeno, Y., et al. (2010), 'Identification of novel ribonucleo-protein complexes from the brain-specific snoRNA MBII-52', *Rna* (16; United States), 1293-300.

Sontag, E., et al. (1999), 'Molecular interactions among protein phosphatase 2A, tau, and microtubules. Implications for the regulation of tau phosphorylation and the development of tauopathies', *J Biol Chem,* 274 (36), 25490-8.

Spencer, C. C., et al. (2011), 'Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21', *Hum Mol Genet,* 20 (2), 345-53.

Spillantini, M. G., et al. (1998a), 'Tau pathology in two Dutch families with mutations in the microtubule-binding region of tau', *Am J Pathol,* 153 (5), 1359-63.

--- (1998b), 'Mutation in the tau gene in familial multiple system tauopathy with presenile dementia', *Proc Natl Acad Sci U S A,* 95 (13), 7737-41.

Spires-Jones, T. L., et al. (2008), 'In vivo imaging reveals dissociation between caspase activation and acute neuronal death in tangle-bearing neurons', *J Neurosci,* 28 (4), 862-7.

Spitale, R. C., et al. (2013), 'RNA SHAPE analysis in living cells', *Nat Chem Biol,* 9 (1), 18-20.

Spriggs, K. A., Bushell, M., and Willis, A. E. (2010), 'Translational regulation of gene expression during conditions of cell stress', *Mol Cell,* 40 (2), 228-37.

Stamer, K., et al. (2002), 'Tau blocks traffic of organelles, neurofilaments, and APP vesicles in neurons and enhances oxidative stress', *J Cell Biol,* 156 (6), 1051-63.

Stefansson, H., et al. (2005), 'A common inversion under selection in Europeans', *Nat Genet,* 37 (2), 129-37.

Stranger, B. E., et al. (2007), 'Relative impact of nucleotide and copy number variation on gene expression phenotypes', *Science,* 315 (5813), 848-53.

Sultan, A., et al. (2011), 'Nuclear tau, a key player in neuronal DNA protection', *J Biol Chem,* 286 (6), 4566-75.

Sun, H. and Chasin, L. A. (2000), 'Multiple splicing defects in an intronic false exon', *Mol Cell Biol,* 20 (17), 6414-25.

Suter, D. M. and Forscher, P. (2001), 'Transmission of growth cone traction force through apCAM-cytoskeletal linkages is regulated by Src family tyrosine kinase activity', *J Cell Biol,* 155 (3), 427-38.

Tacheny, A., et al. (2013), 'Mass spectrometry-based identification of proteins interacting with nucleic acids', *J Proteomics,* 94, 89-109.

Tacke, R. and Manley, J. L. (1995), 'The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities', *Embo j,* 14 (14), 3540-51.

Takahashi, M., et al. (1995), 'Involvement of tau protein kinase I in paired helical filament-like phosphorylation of the juvenile tau in rat brain', *J Neurochem,* 64 (4), 1759-68.

Takanashi, M., et al. (2002), 'Expression patterns of tau mRNA isoforms correlate with susceptible lesions in progressive supranuclear palsy and corticobasal degeneration', *Brain Res Mol Brain Res,* 104 (2), 210-9.

Takei, Y., et al. (2000), 'Defects in axonal elongation and neuronal migration in mice with disrupted tau and map1b genes', *J Cell Biol,* 150 (5), 989-1000.

Takuma, H., Arawaka, S., and Mori, H. (2003), 'Isoforms changes of tau protein during development in various species', *Brain Res Dev Brain Res,* 142 (2), 121-7.

Talkish, J., et al. (2014), 'Mod-seq: high-throughput sequencing for chemical probing of RNA structure', *RNA* (20), 713-20.

Talukdar, I., et al. (2011), 'hnRNP A1 and hnRNP F modulate the alternative splicing of exon 11 of the insulin receptor gene', *PLoS One,* 6 (11), e27869.

Tanaka, E., Ho, T., and Kirschner, M. W. (1995), 'The role of microtubule dynamics in growth cone motility and axonal growth', *J Cell Biol,* 128 (1-2), 139-55.

Tanaka, E. M. and Kirschner, M. W. (1991), 'Microtubule behavior in the growth cones of living neurons during axon elongation', *J Cell Biol,* 115 (2), 345-63.

Tashiro, K., et al. (1997), 'Somatodendritic localization of phosphorylated tau in neonatal and adult rat cerebral cortex', *Neuroreport,* 8 (12), 2797-801.

Tellez-Nagel, I. and Wiśniewski, H. M. (1973), 'Ultrastructure of neurofibrillary tangles in Steele-Richardson-Olszewski syndrome', *Arch Neurol,* 29 (5), 324-7.

Terada, S., et al. (2005), 'Exon 3 insert of tau protein in neurodegenerative diseases', *Acta Neuropathol,* 110 (1), 12-8.

Teslovich, T. M., et al. (2010), 'Biological, clinical and population relevance of 95 loci for blood lipids', *Nature,* 466 (7307), 707-13.

The International HapMap Consortium (2003), 'The International HapMap Project', *Nature,* 426 (6968), 789-96.

The International HapMap Consortium, et al. (2007), 'A second generation human haplotype map of over 3.1 million SNPs', *Nature,* 449 (7164), 851-61.

Tobin, J. E., et al. (2008), 'Haplotypes and gene expression implicate the MAPT region for Parkinson disease: the GenePD Study', *Neurology,* 71 (1), 28-34.

Tomidokoro, Y., et al. (2001), 'Abeta amyloidosis induces the initial stage of tau accumulation in APP(Sw) mice', *Neurosci Lett,* 299 (3), 169-72.

Tomonaga, M. (1977), 'Ultrastructure of neurofibrillary tangles in progressive supranuclear palsy', *Acta Neuropathol,* 37 (2), 177-81.

Trabzuni, D., et al. (2012), 'MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies', *Hum Mol Genet,* 21 (18), 4094-103.

Trinczek, B., et al. (1995), 'Domains of tau protein, differential phosphorylation, and dynamic instability of microtubules', *Mol Biol Cell,* 6 (12), 1887-902.

Trojanowski, J. Q., et al. (1989), 'Distribution of tau proteins in the normal human central and peripheral nervous system', *J Histochem Cytochem,* 37 (2), 209-15.

Tsai, B. P., et al. (2011), 'Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach', *Mol Cell Proteomics,* 10 (4), M110.007385.

Tuerk, C. and Gold, L. (1990), 'Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase', *Science,* 249 (4968), 505-10.

Ubersax, J. A. and Ferrell, J. E., Jr. (2007), 'Mechanisms of specificity in protein phosphorylation', *Nat Rev Mol Cell Biol,* 8 (7), 530-41.

Untergasser, A., et al. (2012), 'Primer3--new capabilities and interfaces', *Nucleic Acids Res,* 40 (15), e115.

Utton, M. A., et al. (2005), 'Molecular motors implicated in the axonal transport of tau and alpha-synuclein', *J Cell Sci,* 118 (Pt 20), 4645-54.

Vacic, V., et al. (2014), 'Genome-wide mapping of IBD segments in an Ashkenazi PD cohort identifies associated haplotypes', *Hum Mol Genet,* 23 (17), 4693-702.

van den Boogaard, M., et al. (2014), 'A common genetic variant within SCN10A modulates cardiac SCN5A expression', *J Clin Invest,* 124 (4), 1844-52.

van den Oord, E. J. (2008), 'Controlling false discoveries in genetic studies', *Am J Med Genet B Neuropsychiatr Genet,* 147b (5), 637-44.

van der Brug, M. P., et al. (2015), 'Parkinson's disease: From human genetics to clinical trials', *Sci Transl Med,* 7 (305), 205ps20.

Vance, C., et al. (2009), 'Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6', *Science,* 323 (5918), 1208-11.

Vandesompele, J., et al. (2002), 'Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes', *Genome Biol,* 3 (7), RESEARCH0034.

Vandrovcova, J., et al. (2009), 'Association of MAPT haplotype-tagging SNPs with sporadic Parkinson's disease', *Neurobiol Aging,* 30 (9), 1477-82.

Varela, M. C., et al. (2006), 'A 17q21.31 microdeletion encompassing the MAPT gene in a mentally impaired patient', *Cytogenet Genome Res,* 114 (1), 89-92.

Veloso, A., et al. (2014), 'Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications', *Genome Res,* 24 (6), 896-905.

Verlaan, D. J., et al. (2009), 'Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease', *Am J Hum Genet,* 85 (3), 377-93.

Vermersch, P., et al. (1996), 'Specific tau variants in the brains of patients with myotonic dystrophy', *Neurology,* 47 (3), 711-7.

Vershinin, M., et al. (2007), 'Multiple-motor based transport and its regulation by Tau', *Proc Natl Acad Sci U S A,* 104 (1), 87-92.

Violet, M., et al. (2014), 'A major role for Tau in neuronal DNA and RNA protection in vivo under physiological and hyperthermic conditions', *Front Cell Neurosci,* 8, 84.

Visscher, P., et al. (2012), 'Five Years of GWAS Discovery', *Am J Hum Genet* (90), 7-24.

Visser, M., Kayser, M., and Palstra, R. J. (2012), 'HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter', *Genome Res,* 22 (3), 446-55.

Vulliet, R., et al. (1992), 'Proline-directed phosphorylation of human Tau protein', *J Biol Chem,* 267 (31), 22570-4.

Wade-Martins, R., Frampton, J., and James, M. R. (1999), 'Long-term stability of large insert genomic DNA episomal shuttle vectors in human cells', *Nucleic Acids Res,* 27 (7), 1674-82.

Wade-Martins, R., Saeki, Y., and Chiocca, E. A. (2003), 'Infectious delivery of a 135-kb LDLR genomic locus leads to regulated complementation of low-density lipoprotein receptor deficiency in human cells', *Mol Ther,* 7 (5 Pt 1), 604-12.

Wade-Martins, R., et al. (2000), 'Stable correction of a genetic deficiency in human cells by an episome carrying a 115 kb genomic transgene', *Nat Biotechnol,* 18 (12), 1311-4.

--- (2001), 'An infectious transfer and expression system for genomic DNA loci in human and mouse cells', *Nat Biotechnol* (19; United States), 1067-70.

Wadsworth, J. D., et al. (2003), 'Molecular and clinical classification of human prion disease', *Br Med Bull,* 66, 241-54.

Wahl, M. C., Will, C. L., and Luhrmann, R. (2009), 'The spliceosome: design principles of a dynamic RNP machine', *Cell,* 136 (4), 701-18.

Walker, S. C., et al. (2008), 'RNA affinity tags for the rapid purification and investigation of RNAs and RNA-protein complexes', *Methods Mol Biol,* 488, 23-40.

Wang, E., et al. (2012), 'Global profiling of alternative splicing events and gene expression regulated by hnRNPH/F', *PLoS One,* 7 (12), e51266.

Wang, J. and Bell, L. R. (1994), 'The Sex-lethal amino terminus mediates cooperative interactions in RNA binding and is essential for splicing regulation', *Genes Dev,* 8 (17), 2072-85.

Wang, J., et al. (2006), 'An improved recombineering approach by adding RecA to lambda Red recombination', *Mol Biotechnol,* 32 (1), 43-53.

Wang, X., et al. (2009), 'Impaired Balance of Mitochondria Fission and Fusion in Alzheimer Disease', *J Neurosci,* 29 (28), 9090-103.

Wang, Y. and Mandelkow, E. (2016), 'Tau in physiology and pathology', *Nat Rev Neurosci,* 17 (1), 22-35.

Wang, Z. and Burge, C. B. (2008), 'Splicing regulation: from a parts list of regulatory elements to an integrated splicing code', *Rna,* 14 (5), 802-13.

Ward, A. M., et al. (2011), 'Quantitative mass spectrometry of DENV-2 RNA-interacting proteins reveals that the DEAD-box RNA helicase DDX6 binds the DB1 and DB2 3' UTR structures', *RNA Biol,* 8 (6), 1173-86.

Webb, A., et al. (2008), 'Role of the tau gene region chromosome inversion in progressive supranuclear palsy, corticobasal degeneration, and related disorders', *Arch Neurol,* 65 (11), 1473-8.

Weingarten, M. D., et al. (1975), 'A protein factor essential for microtubule assembly', *Proc Natl Acad Sci U S A,* 72 (5), 1858-62.

Wellcome Trust Case Control Consortium (2007), 'Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls', *Nature,* 447 (7145), 661-78.

Welter, D., et al. (2014), 'The NHGRI GWAS Catalog, a curated resource of SNP-trait associations', *Nucleic Acids Res,* 42 (Database issue), D1001-6.

Wider, C., et al. (2010), 'Association of the MAPT locus with Parkinson's disease', *Eur J Neurol,* 17 (3), 483-6.

Williams, D. R., et al. (2007), 'Genetic variation at the tau locus and clinical syndromes associated with progressive supranuclear palsy', *Mov Disord,* 22 (6), 895-7.

Xing, L., et al. (2013), 'Creation and characterization of BAC-transgenic mice with physiological overexpression of epitope-tagged RCAN1 (DSCR1)', *Mamm Genome,* 24 (1-2), 30-43.

Xing, W., et al. (2004), 'HSV-1 amplicon-mediated transfer of 128-kb BMP-2 genomic locus stimulates osteoblast differentiation in vitro', *Biochem Biophys Res Commun,* 319 (3), 781-6.

Yamamoto, H., et al. (1988), 'Dephosphorylation of microtubule proteins by brain protein phosphatases 1 and 2A, and its effect on microtubule assembly', *J Neurochem,* 50 (5), 1614-23.

Ye, J., et al. (1994), 'Characterization of a silencer regulatory element in the human interferon-gamma promoter', *J Biol Chem,* 269 (41), 25728-34.

Yip, K. Y., Cheng, C., and Gerstein, M. (2013), 'Machine learning and genome annotation: a match meant to be?', *Genome Biol,* 14 (5), 205.

Yuan, A., et al. (2008), 'Axonal transport rates in vivo are unaffected by tau deletion or overexpression in mice', *J Neurosci,* 28 (7), 1682-7.

Zabetian, C. P., et al. (2007), 'Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson's disease', *Ann Neurol,* 62 (2), 137-44.

Zeggini, E. and Ioannidis, J. P. (2009), 'Meta-analysis in genome-wide association studies', *Pharmacogenomics,* 10 (2), 191-201.

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 (1), 117-30.

Zhang, X., Bailey, S. D., and Lupien, M. (2014), 'Laying a solid foundation for Manhattan--'setting the functional basis for the post-GWAS era'', *Trends Genet,* 30 (4), 140-9.

Zhang, X. H., Leslie, C. S., and Chasin, L. A. (2005), 'Dichotomous splicing signals in exon flanks', *Genome Res,* 15 (6), 768-79.

Zhang, Y., et al. (1998), 'A new logic for DNA engineering using recombination in Escherichia coli', *Nat Genet,* 20 (2), 123-8.

--- (2000), 'DNA cloning by homologous recombination in Escherichia coli', *Nat Biotechnol,* 18 (12), 1314-7.

Zhong, Q., et al. (2012), 'Tau isoform composition influences rate and extent of filament formation', *J Biol Chem,* 287 (24), 20711-9.

Zhou, Y., Lu, Y., and Tian, W. (2012), 'Epigenetic features are significantly associated with alternative splicing', *BMC Genomics,* 13, 123.

Zody, M. C., et al. (2008), 'Evolutionary toggling of the MAPT 17q21.31 inversion region', *Nat Genet,* 40 (9), 1076-83.