December 2012

The role of genetic variations on gene expression and splicing in control human brain: dissection of the aetiology of complex neurological diseases

---

Daniah M. Trabzuni

Department of Molecular Neuroscience

Institute of Neurology

University College London (UCL)

Thesis submitted in fulfillment of the degree of Doctor of Philosophy of UCL, December 2012

Declaration

I, Daniah M. Trabzuni, confirm that the work contained within this thesis is my own original work. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

“Enjoy your journey, satisfy your curiosity and never give up”

Abstract

Over the past two decades there has been a realization of the importance of understanding the underlying molecular mechanism of complex neurological diseases. GWAS studies confirmed a significant association between SNPs and complex neurologic and psychiatric diseases such as Parkinson's disease and Alzheimer's disease. In this project, the impact of genetic variations on gene expression and alternative splicing in control post-mortem human brain tissues from twelve different regions were assessed. These are disease associated regions and support different functional roles. They are: frontal cortex, temporal cortex, occipital cortex, white matter, hippocampus, thalamus, hypothalamus, putamen, substantia nigra, medulla, cerebellum and spinal cord.

Based on 1231 RNA human exon arrays, genotyped and imputed DNA samples from 137 control human brain, brain transcriptome profiles, gene and exon expression quantitative trait loci (QTL) were identified in multiple brain regions.

Significant region-specific exon and gene expression QTLs were reported. Cerebellum and white matter show more unique expression profiles and expression QTLs in comparison with other brain regions. Furthermore, alternative splicing patterns were in a specific group of regions such as the cortical regions. In addition, two neurodegenerative disease related genes were investigated in detail, namely *LRRK2* and *MAPT*. Significant regional differences in expression at mRNA and protein levels were shown. Moreover, exon QTLs correlated with the expression of specific exons located in functional protein domains of *LRRK2*. Also, an exon QTL has been found that shows a protective effect against Parkinson's disease with an increase in the inclusion of exon 3 in grey matter for *MAPT*.

This study has yielded novel regional specific expression QTLs and novel insights into the expression, regulation and function of specific genes in different regions of control human brain that are related to neurological diseases. This reference dataset is a valuable resource to complement other datasets for research into the complex genetics of neurological diseases.

Dedication

I dedicate this thesis to all of the people and their families that donated their brains for research for the sake of knowledge and enlightenment. Without these generous donations, the research presented in this thesis would not have been possible.

I thank my beloved family for their support and encouragement throughout this long journey of seeking the never ending knowledge.

Acknowledgments

I would like to thank Allah, above all, for giving me the patience and determination to conduct this research, to reach the best places and meet the best people.

I would really thank Dr Mina Ryten, as being the main team member and the main drive to facilitate this big project. We worked twenty four hours a day, seven days a week, and even over holidays to solve challenges and overcome all difficulties in this project.

I thank my supervisor Professor John Hardy for his continuous support and invaluable guidance with my work. His patience, suggestions, and wise comments throughout this work made this thesis possible. I also would like to thank my second supervisor Professor Henry Houlden for his support.

Special thanks with a heart full of gratitude and appreciation to Professor Peter Thomson for his help, comments, valuable input and constant support to make this process more fun and enjoyable on the daily basis through the time in writing this thesis especially the writing of chapters 1, 4 and 7.

Special thanks to Dr Patrick Lewis for his constant help, ideas, and input, especially for the protein aspects of the *LRRK2* work. His help landed both of us the Michael J. Fox foundation for Parkinson's Research (MJFF) grant for "Characterizing region specific splice isoforms of LRRK2." for follow-up investigations of the results presented in Chapter 6 of this thesis.

I also thank June Smalley, Dr Lee Stanyer for the help they offered to facilitate the work on this project.

I would like to show my great appreciation for all my friends for their presence in my life and make all challenges we face as motivation to live our dreams.

Special thanks for all the collaborators who made this dataset available through the thesis and associated publications as a useful source in science in general and specifically in the area of neurodegenerative disease.

I would like to thank AROS Applied Biotechnology AS company laboratories and Affymetrix for their valuable input, especially, Goeff Scopes, Jane Ramsey and Thomas Thykjaer.

In addition, this work was supported by the MRC through the MRC Sudden Death Brain Bank (especially Colin Smith and Robert Walker) and by a Project Grant (to John Hardy). This project was further supported by the Cure PSP, Irene and Abe Pollin Fund for CBD Research (Rohan DeSliva and John Hardy). Selina Wray is funded by an Alzheimer's Research UK research fellowship. The work performed by the North American Brain Expression Consortium was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health.

Last but not least, I would like to thank the King Faisal Specialist Hospital and Research Centre in Saudi Arabia for my PhD scholarship. I especially want to thank Abdulaziz Albahkaly, Dr Sultan Al-Sedairy and Dr Brian Meyer for their support from Saudi Arabia throughout my challenging times.

Collaborations

This thesis has involved collaborations with other research groups; the nature of these collaborations included the supply of tissues, data and the exchange of scientific thoughts and ideas.

Chapter 2

The UKBEC brain tissues collection and dissection were performed by Dr Colin Smith and Dr Robert Walker at the MRC Sudden Death Brain Bank.

Chapter 5

MAPT Array expression validation by using quantitative real time polymerase chain reaction (qRT-PCR) with the TaqMan assay was performed at the Queen Square Brain Bank (QSBB) by Dr Jana Vandrovcova.

MAPT protein expression (tau), extraction and Western blot analysis was performed at the Institute of Neurology by Dr Selina Wray.

MAPT protein isoforms expression (tau) using sandwich ELISA analysis was performed at the QSBB by Dr Connie Luk.

Collection and processing of the human post-mortem tissue of the North American Brain Expression Consortium (NABEC) was performed at the Molecular Genetics Section and Laboratory of Neurogenetics, NIA, NIH, USA.

Chapter 6

LRRK2 Array expression validation using quantitative real time polymerase chain reaction (qRT-PCR) using the TaqMan assay was performed at the Queen Square Brain Bank (QSBB) by Dr Jana Vandrovcova.

LRRK2 protein expression, extraction and Western blot analysis was performed at the Institute of Neurology by Dr Patrick Lewis.

LRRK2 immunohistochemistry (IHC) staining analysis was performed at the Queen Square Brain Bank (QSBB) by Dr Rina Bandopadhyay and Dr. Adamantios Mamais.

Publications

Parts of this thesis have been published in the following manuscripts:

Trabzuni, D., et al. (2011). **Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies.** *Journal of Neurochemistry*, ***119****(2): p. 275-82.*

Trabzuni D., et al. (2012). ***MAPT*** **expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies**. *Human Molecular Genetics*. ***21****(18): 4094-4103.*

Trabzuni D., et al., (2012). **Fine-mapping, gene expression and splicing analysis of the disease associated LRRK2 locus**. *PLoS One*. *Submitted. December 2012.*

Trabzuni D., et al. (2012). **Sex-biased gene expression, splicing and regulation in the adult human brain.** *Nature. Submitted December 2012.*

Hardy J, Trabzuni D, et al. (2009). **Whole genome expression as a quantitative trait.** *Biochemical Society Transactions. **37**(Pt 6):1276-7.*

Ryten M, Trabzuni D, Hardy J, (2009). **Genotypic analysis of gene expression in the dissection of the aetiology of complex neurological and psychiatric diseases.** *Brief Functional Genomic Proteomic.**8**(3):194-8.*

Charlesworth, G., Plagnol, V., Holmström, K. M., Bras, J., Sheerin, U.-M., Preza, E., Rubio-Agusti, I., Ryten, M., Schneider, S.A, Stamelou, M., Trabzuni, D., et al. (2012). **Mutations in Ano3 Cause Dominant Craniocervical Dystonia: Ion Channel Implicated in Pathogenesis.** *The American Journal of Human Genetics. **91**(6):1041-50.*

Kumar A, Gibbs JR, Beilina A, Dillman A, Kumaran R, Trabzuni D. et al. (2012). **Age-associated changes in gene expression in human brain and isolated neurons.** *Neurobiology of Aging*. *http://dx.doi.org/10.1016/j.neurobiolaging.2012.10.021*

Patani, R., Lewis, P.A., Trabzuni, D. et al. (2012). **Investigating the utility of human embryonic stem cell-derived neurons to model ageing and neurodegenerative disease using whole genome gene expression and splicing analysis.** *Journal of Neurochemistry*. ***122****(4):738-51.*

Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, Toro R, Appel K, Bartecek R, Bergmann O, Bernard M, Brown AA, Cannon DM, Chakravarty MM, Christoforou A, Domin M, Grimm O, Hollinshead M, Holmes AJ, Homuth G, Hottenga JJ, Langan C, Lopez LM, Hansell NK, Hwang KS, Kim S, Laje G, Lee PH, Liu X, Loth E, Lourdusamy A, Mattingsdal M, Mohnke S, Maniega SM, Nho K, Nugent AC, O'Brien C, Papmeyer M, Pütz B, Ramasamy A, Rasmussen J, Rijpkema M, Risacher SL, Roddey JC, Rose EJ, Ryten M, Shen L, Sprooten E, Strengman E, Teumer A, Trabzuni D, et al. (2012). **Identification of common variants associated with human hippocampal and intracranial volumes.** *Nature Genetics.****44****(5):552-61.*

Hernandez DG, Nalls MA, Moore M, Chong S, Dillman A, Trabzuni D, et al. (2012). **Integration of QWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain.** *Neurobiology of Disease. **47**(1):20-8.*

International Parkinson's Disease Genomics Consortium (IPDGC) 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.*

International Parkinson's Disease Genomics Consortium (IPDGC); Wellcome Trust Case Control Consortium 2 (WTCCC2), (2011). **A two-stage meta-analysis identifies several new loci for Parkinson's disease.** *PLoS Genetics. **7**(6):e1002142.*

Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW, Hernandez DG, Johnson JO, Mok K, Ryten M, Trabzuni D, et al. (2011). **A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD**. *Neuron. **72**(2):257-68*

Devine MJ, Kaganovich A, Ryten M, Mamais A, Trabzuni D, et al. (2011). **Pathogenic LRRK2 mutations do not alter gene expression in cell model systems or human brain tissue.** *PLoS One. **6**(7):e22489.*

Table of Contents

List of tables and figures

Chapter 4

Chapter 5

Chapter 6

Abbreviations

µg	Microgram(s)
µl	Microliter(s)
3`UTR	3` un translated region
3R	3 repeats tau isoform
4R	4 repeats tau isoform
5`UTR	5` un translated region
AD	Alzheimer's disease
ANOVA	Analysis of variance
bDNA	Branched DNA
BLs	Blockers
bp	Base pairs
CBD	Corticobasal degeneration
cDNA	Complementary deoxyribonucleic acid
CEs	Capture Extenders
CNS	Central nervous system
CNV	Copy number variation
CRBL	Cerebellum
cRNA	Complementary ribonucleic acid
Ct	Cycle threshold
DABG	Detection above background
DAVID	The Database for Annotation, Visualization and Integrated Discovery
DNA	Deoxyribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
ENCODE	Encyclopaedia of DNA Elements
FC	Fold change
FCTX	Frontal Cortex
FDR	False discovery rate
FTLD	Frontotemporal lobar degeneration
GO	Gene ontology
GSEA	Gene set enrichment analysis
GWAS	Genome wide associating studies
HIPP	Hippocampus
HK	House keeping genes
HYPO	Hypothalamus
IgG	Immunoglobulin G antibody
IHC	Immunohistochemistry
IVT	In Vitro Transcription
LD	Linkage disequilibrium
LEs	Label Extenders
MAF	Minor allele frequency
Mb	Mega base(s)
MEDU	Medulla
miRNA	Micro ribonucleic acid

MJFF2	Michael J. Fox foundation antibody 2
MRC	Medical Research Council
mRNA	Messenger ribonucleic acid
NABEC	North American Brain Expression Consortium
NFT	Neurofibrillary tangles
ng	Nano gram
nm	Nano meter (wave length)
OCTX	Occipital Cortex
OD	Optical density
OMIM	Online Mendelian Inheritance in Man
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCTX	Parietal Parasagittal
PD	Parksinson's disease
PMI	Post Mortem Interval
PSP	Progressive supranuclear palsy
PUTM	Putamen
QC	Quality control
QG	QuantiGene
QTL	Quantitative trait loci
RIN	RNA integrity number
RNA	Ribonucleic acid
Rpm	Revolutions per minute
SHRI	Sun Health Research Institute
SNIG	*Substania nigra*
SNP	Single nucleotide polymorphism
SPCO	Spinal Cord
SSPE-T	Sodium chloride + sodium phosphate + EDTA + Triton X-100 buffer
TCTX	Temporal Cortex
THAL	Thalamus
TSS	Transcription start site
UKBEC	UK Brain Expression Consortium
WGCNA	Weighted gene co-expression network analysis
WHMT	White Matter (Intralobular)
WT	Whole Transcript

Gene abbreviations

AGTR1	Angiotensin II receptor, type 1
ANO3	Anoctamin 3
BACE2	Beta-site APP-cleaving enzyme 2
BIN1	Bridging integrator 1
CBLN3	cerebellin 3 precursor
CNTNAP2	Contactin-associated protein-like 2
CR1	Erythrocyte complement component, C3b/C4b receptor 1
CRTM	Class I MHC-restricted T cell-associated molecule
EN1	Engrailed homeobox 1
GABRP	gamma-aminobutyric acid (GABA) A receptor, subunit π
IDE	Insulin degrading enzyme
IGFBP3	Insulin-like growth factor-binding protein 3
IGHMBP2	immunoglobulin helicase μ- binding protein 2
KCND3	Potassium voltage-gated channel, Shal-related subfamily, member 3
LINGO2	Leucine rich repeat and Ig domain containing 2
LRRK2	Leucine-rich repeat kinase 2
MAPT	Microtubule-associated protein tau
NLGN4X	Neuroligin 4, X-linked
NRK	Nik related kinase
NRXN3	Neurexin 3
PDZK1	PDZ domain containing 1
PICALM	Phosphatidylinositol binding clathrin assembly protein
PLP1	Proteolipid protein 1
RPLP0	ribosomal protein, large, P0
SCN8A	Sodium channel, voltage gated, type VIII, alpha subunit
SETX	Senataxin
SLC18A2	solute carrier family 18, member 2
SLC6A3	solute carrier family 6, member 3
TH	Tyrosine hydroxylase
TMPRSS5	Transmembrane protease, serine 5
TPM4	Tropomyosin 4
TUBB	Tubulin beta class I
UBC	Ubiquitin C
VSIG4	V-set and immunoglobulin domain containing 4

1 Introduction

The human brain has a unique and a complex structure. The heterogeneous nature of this organ is due to different types of cells, neurons and the levels that these components are structured at. The knowledge about the human central nervous system (CNS) structure and function is increasing greatly, and this will be used to benefit human health. All the research efforts in the molecular and cellular neuroscience fields are vital to develop our understanding of the human CNS function and to dissect the complexity of neurological disease development. These rapid advancements were mainly determined by the genomic, transcriptomic and diagnostic technological innovations (Geschwind and Konopka 2009). These technologies are applicable at the genome-wide scale, and their application results in more insightful information being produced, and this can lead toward more informed investigations and in turn through to clinical applications. Until now, although these efforts have resulted in achieving some of the required aims, the field still lacks the knowledge and the comprehensive understanding of the complex machinery of the biological connections in the human CNS.

It was anticipated that different morphological and cognitive features between human and other organisms were due to the variability in gene expression profiles (Enard, Khaitovich et al. 2002). Studying the transcriptome in human CNS or even in different organisms is a reasonable approach to dissect the functional complexity and the progress of the diseases. In 2002, this concept was highlighted by the first microarray study that was performed to compare expression patterns between different tissues in different organisms (Enard, Khaitovich et al. 2002). Results demonstrated, in

the human brain, that there was considerable variation in gene expression profiles between different individuals that may contribute to the brain function processes.

In order to create that specific knowledge, research on different cell lines and other model organisms was performed to understand more about the gene expression underlying functional regulation in these models (Ross, Scherf et al. 2000; Jin, Riley et al. 2001; Brem, Yvert et al. 2002; Wade, Kulbokas et al. 2002; Waterston, Lindblad-Toh et al. 2002). This approach was applied for many different reasons, including cost, quality, quantity and the accessibility of tissues and samples, specifically CNS samples and specific cell types such as neurons. Human based research was limited for some of the reasons mentioned above. However, the necessity to dissect the complexity of human neurological diseases required researchers to perform their experiments on human tissues and samples rather than model organisms. At the same time, some transcriptomic studies revealed that the gene expression signatures are distinctive and different between human and chimpanzee brains even though the DNA similarity between them is 98.7% (Enard, Khaitovich et al. 2002). That was not the case for gene expression patterns in liver and lymphoblast cells as they were more similar to each other (Enard, Khaitovich et al. 2002; Gu and Gu 2003; Karaman, Houck et al. 2003). Results from these studies emphasized the importance of performing human based research. It is worth noting that the recent advancement of technologies has made this research approach achievable in a relatively short time line (Prof. John Hardy, personal communication).

Focusing on the genome-wide and regional human CNS expression studies, previous studies demonstrated the usability of post-mortem human tissues (Franz, Ullmann et al. 2005) and the success of investigating and comparing gene expression

patterns in control versus disease human post-mortem brains or CNS samples for some neurological diseases (Mirnics, Middleton et al. 2001; Hodges, Strand et al. 2006). In the study of Hodge *et al*. differential gene expression profiles in control and disease samples were detected between different brain regions, namely caudate nucleus, motor cortex and cerebellum, but not in prefrontal cortex. This result raised the concept that some of the pathological features are not necessarily revealed by the histopathological assessments of the neurological diseases (Hodges, Strand et al. 2006).

Likewise, not only were gene expression profiles investigated in detail in specific tissues, but also the alternative splicing profiles were closely investigated by different groups (Castle, Zhang et al. 2008; Xi, Feber et al. 2008; Qin, Haroutunian et al. 2009; Tollervey, Wang et al. 2011). The fact that 92-94 % of the human genes are alternatively spliced (Wang, Sandberg et al. 2008), indicates that the brain has a high level of alternative splicing events compared with other tissues (Yeo, Holste et al. 2004) and there is evidence that 50% of mutations in human genetic diseases are driven by pre-mRNA splicing (Cartegni, Chew et al. 2002) highlighted the importance of these studies.

The availability of this knowledge in addition to the development of advanced technology and the advanced downstream analysis tools (Clark, Schweitzer et al. 2007; Anton, Aramburu et al. 2010) facilitates some significant and key studies to be performed to discover novel expression and splicing events that are involved in the functional mechanisms and regulation process in diseased as well as normal human tissues including the brain (Haverty, Weng et al. 2002; Roth, Hevezi et al. 2006; Clark, Schweitzer et al. 2007; French, Peeters et al. 2007; Castle, Zhang et al. 2008; Johnson, Kawasawa et al. 2009; Anton, Aramburu et al. 2010; Fugier, Klein et al. 2011; Kang,

Kawasawa et al. 2011; Prescott and Chamberlain 2011; Tollervey, Wang et al. 2011; Hawrylycz, Lein et al. 2012). As a result of these developments, a few tissue specific databases of mRNA expression profiles from normal human tissues were created for the scientific community (Haverty, Weng et al. 2002; Su, Wiltshire et al. 2004; Son, Bilke et al. 2005).

For example, in the study of French *et al.*, novel differentially spliced isoforms and exons were identified between the subgroups of Glial brain tumour cases. These novel isoforms are believed to have a role in the initiation or progression of the glioma (French, Peeters et al. 2007). Another recent study by Tollervey and colleagues identified disease related alternative splicing events in patients with frontotemporal lobar degeneration (FTLD) or Alzheimer's disease (AD) (Tollervey, Wang et al. 2011).

Furthermore, it is important to profile the normal variability of gene expression and splicing patterns. This will provide the complete mapping of these transcriptional regulation patterns in healthy tissues and cells which are not affected with the pathology of the disease. These samples will not have altered composition, molecular mechanisms or gene function, and hence will make a useful reference sample for comparison with diseases samples. There are two main recent studies that were performed on normal brains that this project complements.

Firstly, in the study of Kang and colleagues, the transcriptome of developmental and adult brains were profiled. In total, 57 brains (40 brains representing different developmental stages and 17 representing adults) were included in this study from 16 brain regions, and the expression levels analyzed by using the Affymetrix GeneChip®Human Exon 1.0 ST arrays. 90 % of the genes were regionally differentially expressed at gene and exon usage levels (Kang, Kawasawa et al. 2011).

The second study of note is the anatomical atlas of the adult human brain by Hawrylycz and colleagues. A very precise and comprehensive neuroanatomical brain transcriptome evaluation from two individuals were performed. Histology and MRI scan analyses were also combined with the mRNA transcription data (Hawrylycz, Lein et al. 2012).

These studies showed firstly, the normal and regional variability in the human brain transcriptome profiles; secondly, the correlation between different brain regions and their functional connectivity; thirdly, the ideal approach to maximize the use of the human brain transcriptomic data. This last point was achieved by a) integrating the transcriptomic data with genomic data, expression quantitative trait loci (QTL) analyses, b) advanced functional analyses, using weighted gene co expression network analysis (WGCNA), c) functional annotations analyses, using the Database for Annotation, Visualization and Integrated Discovery (DAVID) and d) use of a bioinformatics database, Encyclopaedia of DNA Elements (ENCODE) Consortium. This provides a way forward to obtain more information about transcriptional regulation in the human CNS in order to understand the effect of genetic disorders targeting specific human CNS regions.

Expression quantitative trait loci (QTL) analysis is the mapping of genetic data that influence the level of the expression patterns. Integrating these two types of data is important to understand to what extent the patterns of gene expression are under genetic control (Schadt, Monks et al. 2003; Morley, Molony et al. 2004; Cookson, Liang et al. 2009) and more importantly, to understand which groups of genes or biological processes are affected by specific genetic variants such as risk variants associated with complex diseases (Monks, Leonardson et al. 2004).

Performing expression QTL analysis in human is feasible approach by considering mRNA transcription profiles as quantitative traits or phenotypes. An initial study using lymphoblastoid cell lines showed significant value of using this analysis in normal tissues on a small cohort (56 individuals). The results showed 29% of differentially expressed genes had a genetic component (Schadt, Monks et al. 2003). Another study was performed by using lymphoblastoid cell lines from 94 normal individuals. This study introduced the concept of *cis, trans* and master genetic regulators. Results showed that there are many genetic variants that co-regulate the expression level of a gene. In addition, master regulators that affect expression levels of many genes were reported and genes that share the same genetic regulator are correlated (Morley, Molony et al. 2004). As well as the results from these studies shedding light on the complexity of the transcription regulation mechanisms by genetic variants, they also provided some insights about the variants' contribution in biological processes in normal human tissues.

Implementing an expression QTL analysis after a genome-wide association (GWA) study of human common complex diseases has become an essential systematic approach (Cookson, Liang et al. 2009). GWA studies have provided a systematic approach to identify common low risk variants that lead to amino acid changes in the protein (McCarthy, Abecasis et al. 2008; Manolio, Collins et al. 2009). However, it was observed that most of the identified disease associated risk variants from the GWA studies were in the noncoding regions of the genome and not mapped to amino acid code in the protein (Hardy and Singleton 2009; Manolio 2010). This highlights the importance of generating genome-wide data to understand the effect of common non coding variants on gene expression in control human CNS tissues. Expression QTL analysis can be applied at genome-wide scale in order to assess the functional

role and the biological basis of how these common variants are involved in the common disease pathology by affecting the gene expression and splicing patterns in the human CNS (Cookson, Liang et al. 2009; Hardy and Singleton 2009).

Previous studies demonstrated the potential and feasibility of this approach in the human CNS and other human tissues, such as liver (Schadt, Molony et al. 2008), adipose (Emilsson, Thorleifsson et al. 2008), blood (Emilsson, Thorleifsson et al. 2008; Heinzen, Ge et al. 2008), human monocytes (Zeller, Wild et al. 2010), lymphoblastoid cell lines (Monks, Leonardson et al. 2004) and human brain (Myers, Gibbs et al. 2007; Heinzen, Ge et al. 2008; Zou, Carrasquillo et al. 2010). In order to perform these studies, the availability of the tissues was the main consideration. For example most of the tissues, except for the human brain, are easy to access and collect. This is one of the reasons that expression QTL studies were mainly conducted using blood or lymphoblastoid cell lines (Emilsson, Thorleifsson et al. 2008) (**Table 1.1**). However, the study by Meyer *et al.* was performed using 193 of normal human brain samples dissected from cortical regions. The results showed significant associations between single variants and transcription levels, as the frequency of the common risk variant will be predictable in a cohort of normal individuals (Myers, Gibbs et al. 2007).

Furthermore, Heinzen and colleagues performed a comprehensive association test not only with the transcript expression level, but also with the individual exon levels to explore exon splicing QTLs in blood and brain (Heinzen, Ge et al. 2008). Moreover, a case control study was performed to combine the analysis of the disease risk SNPs from GWAS and brain gene expression of Late-onset Alzheimer disease (LOAD). This study was conducted on a large cohort (~2300 AD cases – control). The results demonstrated an expression QTL association between a protective risk SNP

(rs7910977) and an increase in the expression of the *IDE* (Insulin degrading enzyme) gene. Further validations were performed on ~193 samples to confirm the association. This study provided strong evidence for a new candidate gene for AD (Zou, Carrasquillo et al. 2010).

Performing expression QTL analyses on different tissues has its own advantages and disadvantages. The table below shows the key factors that affect the performance of this type of study (**Table 1.1**).

Table 1.1: Advantages and disadvantages of using different human tissues for expression QTLs.

Tissue type	Advantages	Disadvantages
Cell lines	High quality RNA	May not be relevant to disease
	Specific cell type	No connectivity between different cell types
	Large sample size	Genomic instability
	Easy to manipulate	artificial
	Availability	
Blood	High quality RNA	May not be relevant to disease
	specific cell type	No connectivity between different cell types
	Large sample size	
	Availability	
Brain	Relevant to disease	Low quality RNA
	Unique to the region	Mixed cell types
		Small sample size
		Availability

This table shows the different limitations and advantages when performing expression QTL analyses in different human tissues.

It is clear from Table 1.1 that each tissue type can provide a useful resource depending on the hypotheses of the study that will be conducted despite its own limitations.

Focusing on the normal human CNS or brain studies, more recent studies were performed using advanced expression and genotyping platforms (Gibbs, van der Brug et al. 2010; Kang, Kawasawa et al. 2011). The Gibbs *et al.* study was performed on ~600 normal human brains dissected from four regions: cerebellum, frontal cortex, pons and temporal cortex. The results showed an abundance of *cis* regulation of mRNA expression and DNA methylation at a genome-wide scale (Gibbs, van der Brug et al. 2010). In addition, this study demonstrated the complexity of the expression QTL signals controlling the gene expression patterns in the human CNS.

Another key study is by Kang *et al.* As mentioned previously, this study was performed on developmental and adult human brains in a relatively small cohort (57 samples in total), dissected from 16 anatomical brain regions. The results showed that the expression patterns differ more during the developmental stages of the human brain, and that there is evidence of sex related difference in the expression and splicing pattern (Kang, Kawasawa et al. 2011).

In summary, all the studies that were performed on the human CNS tissues are valuable recourses in the scientific field to help understand and evaluate critical pathways and genes that are related to the development of neurodegenerative diseases such as Parkinson's disease (PD) and AD and target the molecular and cellular functions and connections in the human CNS. However, all these data sets cover different human CNS human regions, have different coverage across the genome, and are based on different sample sizes.

The importance of expanding the human tissue deposit especially in the human CNS is increasing. Thus, in this project the study was performed with large samples number (*n* = 137 individuals) dissected from twelve brain regions. These regions are: frontal cortex (FCTX) Brodmann areas 9 and 46, temporal cortex (TCTX) Brodmann areas 21,41 and 42, parietal parasaggital (PCTX) Brodmann areas 3,1, and 2, occipital cortex (specifically primary visual cortex, OCTX) Brodmann areas 17, hippocampus (HIPP), thalamus (THAL), hypothalamus (HYPO), putamen (PUTM), *Substania nigra* (SNIG), medulla (specifically inferior olivary nucleus, MEDU), cerebellum (CRBL), spinal cord (SPCO), and intralobular white matter (WHMT) below Brodmann areas 39 and 40. Genome-wide profiling for expression and splicing was performed using the Affymetrix GeneChip®Human Exon 1.0 ST Arrays, which have comprehensive coverage for all exons as well as transcripts. In parallel, all individuals were SNP genotyped and imputed to perform: i) Regional differential expression patterns, ii) regional differential spicing patterns and finally iii) Regional expression QTL analyses, on human CNS samples.

Aims and objectives

This project was performed to fulfil specific aims to benefit research in the neuroscience field. The main aims for this thesis are as follows:

Firstly, the study aimed to provide the neuroscience community with a reliable human reference data set from post-mortem multiple CNS regions. Although this resource is unique, it is an accompaniment to other human genotypic-phenotypic data sets in the field. Consequently, this study aimed to provide substantial human data for researchers to answer long standing questions in various directions in relation to the neurological diseases.

The second aim was to identify the human brain transcriptome in twelve CNS regions in a large cohort. This is achieved by investigating and exploring the CNS differentially gene expression and alternatively gene splicing profiles between different functional regions in control individuals.

Thirdly and mainly, this study aims at expanding our understanding of the functional genomics in the human CNS. Integrating the genomic data with transcriptomic data is the step forward to study the functional effects of the non-coding genetic variants on gene expression and splicing and their contribution to the human neurological diseases.

2 Materials and Methods

2.1 Characteristics of the human post-mortem tissue of the UK Brain Expression Consortium (UKBEC)

The brain tissues collection and dissection were performed by Dr Colin Smith and Dr Robert Walker at the MRC Sudden Death Brain Bank.

2.1.1 Human post-mortem brain tissue collection and dissection

A total of 137 frozen human post-mortem brain and central nervous system (CNS) tissue samples of Caucasian controls with no neurological conditions were obtained from the Medical Research Council (MRC) Sudden Death Brain and Tissue Bank, Edinburgh, UK (Millar, Walker et al. 2007), and the Sun Health Research Institute (SHRI), an affiliate of Sun Health Corporation, USA (Beach, Sue et al. 2008). All samples had fully informed consent for retrieval and were authorized for ethically approved scientific investigation (UCL Research Ethics Committee number 10/H0716/3). Histology was performed on sections prepared from all paraffin embedded (FFPE) brain blocks and the neuropathological diagnosis was determined by a consultant neuropathologist (Trabzuni, Ryten et al. 2011).

Brain tissues originating from 101 individuals from the MRC Sudden Death Brain and Tissue Bank were removed as fresh tissue. Tissue handling and dissection methods were applied exactly in the same way for all cases. The bodies were stored refrigerated and were brought up to the PM suite just prior to the start of the autopsy. Each post-mortem brain dissection was carried out in the same way. The whole brain was removed within 15 minutes and was immediately cut into coronal slices and the

various anatomical regions of interest were immediately sampled from brain coronal slices at autopsy and immediately flash frozen. These regions are: frontal cortex (FCTX) Brodmann areas 9 and 46, temporal cortex (TCTX) Brodmann areas 21,41 and 42, parietal parasaggital (PCTX) Brodmann areas 3,1, and 2, occipital cortex (specifically primary visual cortex, OCTX) Brodmann areas 17, hippocampus (HIPP), thalamus (THAL), hypothalamus (HYPO), putamen (PUTM), *Substania nigra* (SNIG), medulla (specifically inferior olivary nucleus, MEDU), cerebellum (CRBL), spinal cord (SPCO), and intralobular white matter (WHMT) below Brodmann areas 39 and 40. Furthermore, the samples once removed from the coronal slices were placed in sealed containers which in turn were placed on cool blocks chilled to -20°C and stored within an insulated box. The samples were dissected into various size pieces approximately 250 mg – 500 mg and were placed in tubes which were immediately snap-frozen in liquid nitrogen. Samples were processed with every effort to ensure that tissue was maintained chilled, was not subjected to drying prior to preservation and had no extra freeze/ thaw cycles, with subsequent loss of RNA quality.

The MRC Sudden Death Brain and Tissue Bank cohort included 23 females and 78 males, the age at death ranged from 16 to 83 years with a mean of 50.4 years. The brain pH ranged from 5.42 to 6.63 with a mean of 6.31. The post mortem interval (PMI) was calculated from the time of death to the time of removal of the brain from the skull, and ranged from 28 to 114 hours with a mean of 52.2 hours. The total number of samples was 1,844 as some regions were not available from certain individual brains.

The brain tissues collection and dissection were performed at the Sun Health Research Institute.

An additional 36 frozen human post-mortem brain tissues of Caucasian controls were collected by the Sun Health Research Institute (SHRI), an affiliate of Sun Health Corporation (Beach, Sue et al. 2008). Whole brains were removed as fresh tissue at autopsy and anatomical regions of interest were sampled from frozen brain coronal slices on pre-chilled -20°C plastic Petri dishes placed on top of -20°C ice packs at and were snap-frozen in liquid nitrogen with subsequent storage at -80°C. These regions are: FCTX, TCTX, OCTX, HIPP, THAL, PUTM, SNIG, MEDU, CRBL and WHMT.

The time interval from the removal of the brain at the mortuary to the completion of the dissection and placement of samples within the storage freezer ranged from 2.5 to 4 hours.

The SHRI cohort included 12 females and 24 males, the age at death ranged from 53 to 102 years with a mean of 80 years. The PMI ranged from 1 to 5.5 hours with a mean of 2.6 hours. The total number of samples was 476 as some regions were not available from all brains. Note that the PCTX, HYPO and SPCO regions were not available from the Sun Health cohort.

The cohort of the two data sets combined is called the UK Brain Expression Consortium (UKBEC) and included 35 females and 102 males, the age at death ranged from 16 to 102 years with a mean of 59 years. The brain pH ranged from 5.42 to 6.63 with a mean of 6.31. The PMI ranged from 1 to 114 hours with a mean of 42 hours. This is the largest sample resource of this nature described to date, totalling 2320 RNA samples to which RNA quality control has been applied and observed on different

regions from control human post-mortem brains. Detailed phenotypic information is described in **Table 2.1.1** (Trabzuni, Ryten et al. 2011).

Table 2.1.1: Demographics of the samples studied.

		Sex		Age / year		Brain pH		PMI[1] (hours)		RIN[2]	
Brain bank	**individuals**	**Male**	**Female**	**range**	**mean**	**range**	**mean**	**range**	**mean**	**range**	**mean**
MRC-UK	101	78	23	16-83	50.4	5.42-6.31	6.3	28-114	52.2	1-8.5	4
SHRI-USA	36	24	12	53-102	80	NA	NA	1-5.5	2.6	1-8	3.6
UKBEC	137	102	35	16-102	59	5.42-6.31	6.3	1-114	43.7	1-8.5	3.85

[1] PMI = post mortem interval (hours); [2] RIN = RNA integrity number

This table shows the values (range, mean) for variables in the cohort from both MRC-UK and SHRI-USA data set separately and joined as UKBEC. Of note, however, because of the different practices of the two tissue resources, there is no overlap in the PMIs between the MRC-UK (long PMI) and SHRI-USA (short PMI). This table is modified from (Trabzuni, Ryten et al. 2011).

2.1.2 Post-mortem determination of brain pH from MRC Sudden Death Brain and Tissue Bank samples

The brain pH from the MRC Sudden Death Brain and tissue bank was recorded using a Hanna HI8424 hand held pH meter with a glass bodied electrode (Fisher Scientific, Loughborough, UK). Consistent with previous studies, the pH did not vary between different CNS regions in a normal brain (Stan, Ghose et al. 2006; Monoranu, Apfelbacher et al. 2009). A single pH value was used and measured at the lateral ventricle. The electrode was inserted into the lateral ventricle after the brain was coronally sliced behind the mammillary bodies. It will be influenced by any remaining CSF fluid within the lateral ventricle but predominantly by the actual tissue comprising the wall of the lateral ventricle.

Brain pH from SHRI samples was not available for the following reasons: 1) it was not part of the USA brain bank standard operating procedures; 2) the restriction of tissue availability for most USA samples.

2.2 RNA extraction using single-step TRIZOL®Reagent based kit

Total RNA was isolated from human post-mortem brain tissues using the miRNeasy 96 kit (Qiagen, UK). Brain tissues were collected and weighed in RNAse-free 96 -well plates; they were between 50-100 mg in weight. Each brain tissue sample was subjected to a minimum of one extraction; with a number of samples extracted in duplicates and triplicates. All steps were performed on dry ice prior to the addition of the QIAzol®Reagent.

2.2.1 Tissue homogenization

Brain tissues from different regions were cut and weighed in RNase-free 96-well plates. All tissue samples were homogenized completely at 4°C using the TissueLyser II (Retsch, UK) for 4-5 minutes maximum at 30 Hz in 800 µl of QIAzol with the addition of two 3 mm stainless steel beads.

2.2.2 Phase separation

The homogenized tissue samples were incubated at room temperature (15-25°C) for 5 minutes. After incubation, 160 µl of chloroform ($CHCl_3$) (Sigma,UK) was added, the plates were mixed vigorously using the TissueLyser II for 30 seconds at 30 Hz. Then plates were incubated at room temperature (15-25°C) for 8-10 minutes. The plates were centrifuged at 6,000 g for 45 minutes at 4°C.

2.2.3 RNA precipitation, wash and elution

The aqueous phase (upper layer, with approximately 60 % of the total volume after the QIAzol was added) was transferred to a fresh RNase-free 96-well plates. It was very important not to disturb the bottom phase that contains the DNA and protein. RNA was precipitated by adding 800 µl of 100% ethanol to the aqueous phase (1.5 volume of the aqueous phase), followed by mixing the samples. The samples were then transferred into the wells of the RNeasy 96-well plates to allow the RNA in the solution to bind to the membrane and to be filtered through by centrifugation.

The plates were centrifuged at 5600 g for 2 minutes at room temperature, and the supernatant was discarded. Samples were then washed to remove salt traces and impurities by adding 800 µl of washing buffers and centrifuging plates for 2 minutes each wash, for 3 to 4 times. Finally, RNA was eluted in two steps in 65 µl in total

volume of pre-heated RNase-free water to 50°C. RNA quality checks were then immediately performed on the samples, followed by the application of expression profiling by using human microarrays.

2.2.4 RNA quality assessments

2.2.4.1 Concentration and purity

The concentration and purity of each RNA sample were assessed by measuring the optical density (OD) of each RNA sample at a wavelength of 260 nm, using the NanoDrop ND-1000 Spectrophotometer V3.3.0. The concentration of each sample was calculated, together with the ratio of absorbance at 260 nm/280 nm and 260 nm/230 nm.

2.2.4.2 RNA integrity numbers (RIN)

After purification, all RNA samples were applied to a RNA 6000 Nano-LabChip using Agilent 2100 bioanalyzer for electrophoresis applied lab-on-a-chip technology (Agilent Technologies UK Ltd, UK) to obtain RNA Integrity Numbers (RIN) (Schroeder, Mueller et al. 2006). RIN and total RNA electropherograms were calculated automatically by the 2100 Expert Software (Agilent Technologies UK Ltd, UK). Further assessment for the calculated RIN number was performed by checking each electropherogram visually.

2.3 Genome-wide expression profiling using Affymetrix GeneChip®Human Exon 1.0 ST Arrays

Exon array experimental work was performed by AROS Applied Biotechnology AS Company as a service laboratory for the samples that I provided.

The sent RNA samples were prepared and randomized in big batches to reduce the batch effect variability on the expression profiles as much as possible.

2.3.1 Preparation of biotin labelled cDNA

In total, 1266 total RNA samples were examined using expression arrays. For each array, 200 ng of total RNA was used as starting material for the complementary DNA (cDNA) preparation. The procedure started with the first and second strand cDNA synthesis, and then the In Vitro Transcription (IVT) reaction was applied to generate cRNA. This was performed using the Ambion®WT Whole Transcript Expression Kit according to the manufacturer's instructions. Biotin labelling was performed using the Terminal Labelling Kit (Affymetrix, UK Ltd.) according to the manufacturer's instructions. Following the IVT reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen, UK).

2.3.2 Array hybridization and scanning

Prior to hybridization, the fragmented cDNA sample was heated to 95°C for 5 minutes and subsequently to 45°C for 5 minutes before loading onto the Affymetrix Human Exon 1.0 ST array cartridge. The probe array cartridge was then incubated for 16 hours at 45°C at constant rotation (60 rpm). The washing and staining procedure was performed in the Affymetrix Fluidics Station 450. The probe array was exposed to 10 washes in 6X SSPE-T at 25°C followed by 4 washes in 0.5×SSPE-T at 50°C. The

biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, final concentration 2 mg/ml (Affymetrix, UK Ltd.) in 6xSSPE-T for 30 minutes at 25°C followed by 10 washes in 6×SSPE-T at 25°C. This was followed by an antibody amplification step using normal goat IgG as blocking reagent, final concentration 0.1 mg/ml (Affymetrix, UK Ltd.) and biotinylated anti-streptavidin antibody (goat), final concentration 3 mg/ml (Affymetrix, UK Ltd.). Thena staining step took place with a streptavidin-phycoerythrin conjugate, final concentration 2 mg/ml (Affymetrix, UK Ltd.) in 6×SSPE-T for 30 minutes at 25°C and 10 washes in 6×SSPE-T at 25°C. The probe arrays were scanned at 560 nm using a confocal laser-scanning microscope (GeneChip® Scanner 3000 7G) and visually inspected for hybridization artefacts.

2.4 Genome-wide expression profiling using Illumina HumanHT-12 v3 Expression BeadChip

Eighty-eight individuals of the UKBEC dataset from FCTX and CRBL regions were run on Illumina arrays in addition to the exon arrays. These data will be analyzed and combined with the North American Brain Expression Consortium (NABEC) dataset (see Section 2.17 at the end of this chapter) for microtubule-associated protein tau (*MAPT*) gene analysis (see Chapter 5) to maximize the number of samples available for the selection of individuals with H1/H1, H1/H2 and H2/H2 haplotype status.

2.4.1 Preparation and amplification of biotin labelled cRNA

A sample of 230 ng of total RNA was used as starting material for the first and second strand cDNA synthesis by using Illumina® TotalPrep RNA Amplification Kit

(Ambion, USA). First strand cDNA was synthesized by using oligo (dt) primers and incubating the reverse transcription master mix reaction at 42°C for 2 hours. Then the second strand master mix was added and incubated at 16°C for an additional 2 hours. This was followed by a cDNA purification step by using the cDNA filter cartridges. Finally, the cRNA synthesis was performed by incubating the samples with the In Vitro Transcription (IVT) master mix at 37°C overnight (14 hours). Following the IVT reaction, the unincorporated nucleotides, salts and unused NTPs were removed using cRNA filter cartridges (Ambion, USA).

2.4.2 cRNA quality assessments

The concentration and purity of each cRNA sample were assessed by measuring the optical density (OD) of each RNA sample at a wavelength of 260 nm, using the NanoDrop ND-1000 Spectrophotometer V3.3.0. The concentration of each sample was calculated, together with the ratio of absorbance at 260 nm/280 nm and 260 nm/230 nm. cRNA samples with concentrations less than 150 ng/µl were reprocessed for amplification to get a higher cRNA yield. At this stage samples were ready for the hybridization step.

2.4.3 Array hybridization, washing and scanning

Prior to hybridization, the sample was heated to 65°C for 5 minutes and subsequently cooled down to room temperature before loading onto the Illumina HumanHT-12 v3 Expression BeadChip array. After loading 15 µl of the samples, the array cartridge was then incubated in the Hyp chamber for 16-20 hours at 58°C at constant rotation (60 rpm). The washing and staining procedure for the bead chip arrays were performed in the slide rack. The arrays were exposed to a first wash for 5 minutes in pre-heated E1BC washing solution at 55°C, followed by a first room

temperature wash for 5 minutes in E1BC washing solution. Then the ethanol wash in 100% ethanol for 10 minutes at room temperature, followed by a second room temperature wash for 2 minutes. The blocking and detecting steps were performed for 10 minutes each using Block E1 buffer and streptavidin-Cy3 solution. Last room temperature wash was performed for 5 minutes and arrays were dried at room temperature for 10 minutes by centrifugation. The bead chip arrays were scanned using an Illumina® BeadArray Reader to read and extract the image. The expression data were normalized and analyzed using Gene Expression Module 3.2.7 in the Illumina® BeadStudio Application in combination with the NABEC dataset (see Section 2.17).

2.5 Affymetrix GeneChip®Human Exon 1.0 ST arrays quality control

2.5.1 Array quality performance

The Expression Console™ software version 1.1 (EC) (Affymetrix, UK Ltd.) was used to evaluate the performance quality of the arrays including the labelling, hybridization, scanning and background signals by probe set summarization and CHP file generation using the robust multichip analysis (RMA) algorithm.

The quality assessment was performed by generating different parameters for all the probe sets analyzed by EC. Boxplots of the relative log expression for all the probe sets were obtained, as well as the present call (%P) which was the main parameter used for the array quality in this study. All parameters are shown in **Table 2.5.1** below. Samples that failed any one of the quality criteria cut-offs were re-run on arrays from different RNA aliquots.

In addition, cDNA and cRNA Agilent 2100 Bioanalyzer profiles were generated for samples with a wide range of RIN numbers (RINs from 2 to7) to assess the cDNA preparation and cRNA production nucleotide lengths from RNA samples with different levels of degradation.

Table 2.5.1: Arrays quality control parameters in Expression Console™ software.

Parameter name	Definition
Boxplots	The relative log expression for all the probe sets analyzed.
%P ranges from (2 - 76.7).	The percent of probe sets present - For the exon level analysis, this is defined as the DABG (Detection above background) probe level *P*-value is less than or equal to 0.01.
%A ranges from (98.2 - 23.2). In this analysis the cut-off 65%. Note that %A = 100 - %P	The percent of probe sets called present absent - For exon level analysis, the DABG *P*-value is greater than 0.01.
All_probeset_mean ranges from (4.8 - 6.2).	The mean signal value for all the probe sets analyzed.
All_probeset_stdev ranges from (1.7 - 2.2).	The standard deviation for all the probe sets analyzed.
All_probeset_mad_residual_mean ranges from (0.2 - 1).	The mean of the absolute deviation of the residuals from the median for all probe sets analyzed.
All_probeset_mad_residual_stdev ranges from (0.16 - 0.68).	The standard deviation of the absolute deviation of the residuals from the median for all probe sets analyzed.
All_probeset_rle_mean ranges from (0.28 - 1.9).	The mean absolute relative log expression (RLE) for all the probe sets analyzed.
All_probeset_rle_stdev ranges from (0.27 - 1.47).	The standard deviation of the mean absolute relative log expression (RLE) for all the probe sets analyzed.
Pos_control_mean ranges from (5.2 - 8.26).	The mean for the set of putative exon based probe sets from putative housekeeping genes.
Pos_control_rle_mean ranges from (0.19 - 3).	The mean absolute relative log expression (RLE) for the set of putative exon based probe sets from putative housekeeping genes.
Pos_control_rle_stdv ranges from (0.17 – 1.75).	The standard deviation absolute relative log expression (RLE) for the set of putative exon based probe sets from putative housekeeping genes.
Bgrd_mean ranges from (31 – 234).	The mean of the raw intensity for the probes used to calculate background prior to any intensity transformations.
bac_spike_mean ranges from (8 – 13).	The mean for the set of probe sets which hybridize to the pre-labelled bacterial spike controls (BioB, BioC, BioD, and Cre).

pos_vs_neg_auc is the area under the curve (AUC) for a receiver operator curve (ROC) comparing the intron controls to the exon controls by applying a threshold to the probe set summary. The ROC curve is generated by evaluating how well the probe set summary separates the positive controls from the negative controls (e.g., exon from intron). The assumption (which is only valid in part) is that the negative controls are a measure of false positives and the positive controls are a measure of true positives. An AUC of 1 reflects perfect separation whereas as an AUC value of 0.5 would reflect no separation. Note that the AUC of the ROC curve is equivalent to a rank sum statistic used to test for differences in the centre of two distributions.

all_probeset parameters are important to assess the quality of many steps in the microarrays preparation including RNA, hybridization efficiency and the scanning quality.

Pos_control and neg_control parameters are important to assess the true positive results in the pos_vs_neg_auc.

bac_spike parameters will help to assess any possible problem with the hybridization and chip quality in specific.

2.5.2 Statistical analysis for the effect of post-mortem variables on RNA and array quality performance

Previous studies revealed that post-mortem variables have been considered to affect the RNA quality (Weis, Llenos et al. 2007). In this study the effect of post-mortem variables such as pH of the brain tissue, PMI, cause of death, age and gender were investigated further with additional CNS regions in a large combined data set.

These variables were obtained from the sample data sheets from the brain banks. Linear mixed regression and analysis of variance (ANOVA) model analyses were performed to investigate factors affecting RIN and present call (%P). Terms included in the model were age, gender, cause of death, region, PMI, pH, as well as variability within individuals and brain tissue blocks. Separate models were constructed for RIN and %P. Statistical analyses were conducted using Partek® Genomics Suite™ version 6.6 and PASW Statistics version 18 software. The explanatory ability of the model in a forwards stepwise manner was assessed, by examining the increase in variation explained when a new covariate or set of covariates were added to the existing model, together with a *P*-value for that increase and adjusted R^2 measure.

2.6 Affymetrix GeneChip®Human Exon 1.0 ST arrays analysis

2.6.1 Pre-processing, normalization and summarization of the raw expression intensity (CEL) files

Raw signal intensities were pre-processed to generate $\log_2$ scale values by using robust multi-array average (RMA) quantile normalization with probe set summarization with median polish according to the RMA (Irizarry, Hobbs et al. 2003) with GC background correction (GC-RMA) algorithm in Affymetrix Power Tools 1.14.3 software (Wu and Irizarry 2004) (http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx). This was applied after re-mapping the Affymetrix probe sets onto human genome ref Sequence build 19 (GRCh37) as documented in the Netaffx annotation file (HuEx-1_0-st-v2 Probe set Annotations, Release 31) and restricting the analysis to ~ 308,717 probe sets (out of a total of 1.4 million probe sets) that: i) had gene annotation according to RefSeq build hg19 using custom scripts, ii) did not target intronic regions and iii) contained at least three 25-mer probes with a unique hybridization to target sequence (single exon). Each probe set in the Affymetrix exon arrays is prepared to measure a single exon expression level except in rare cases.

Principal component analysis and unsupervised hierarchical clustering were then used to identify outliers and to visualize the clustering of all arrays samples by different functional regions; Partek® Genomics Suite™ software was used for this step. At this point, the array data were ready to be used for different types of analyses such as differential regional expression (see Section 2.6.2) and expression QTL analysis (see Section 2.16) as different probe set filtering methods were applied to different analyses.

2.6.2 Region differential expression and alternative splicing profiles in 12 CNS regions

After removing the outliers, 917 samples from twelve CNS regions were included in this analysis. The regions were as follow: FCTX, n = 96, TCTX, n = 84, OCTX, n = 95, HIPP, n = 93, THAL, n = 83, HYPO, n = 13, PUTM, n = 94, SNIG, n = 68, MEDU, n = 92, CRBL, n = 95, SPCO, n = 13 and WHMT, n = 91.

2.6.2.1 Probe sets filtering using detection above background (DABG) method

In order to reduce the back ground noise in the expression data for each case, the probe sets list (308,717 probe sets) was filtered further by calculating the detection-above-background (DABG, detection metric) using Affymetrix Power Tools (Affymetrix) (Clark, Schweitzer et al. 2007). Thus, the regional expression analysis list was restricted to probe sets that have gene annotation according to RefSeq build hg19, contained at least 3 probes with unique hybridization and had DABG P values ≤ 0.001 in $\geq 50\%$ of samples in one brain region. This list contained 169,827 probe sets which were considered as detected. These probe sets were re-normalized and summarized based on the RMA quantile normalization method in Section 2.6.1. In addition they were summarized to a trancript-level summary for ~ 19,215 transcripts by calculating the 10% trimmed Winsorized mean of the expression values from the corresponding probe sets to each transcript. Gene symbols for the detected transcripts were assigned to the transcript identification numbers using Netaffx annotation file (HuEx-1_0-st-v2 Transcript Annotations, Release 31) to produce a list containing 18,107 detected genes which will be used for the rest of the analyses.

Furthermore, shorter lists that contain detected genes in one or multi CNS regions were constructed from this main list.

2.6.3 Analysis of variance (ANOVA) statistical analysis

The date of array hybridization (batch effects), gender, region and individual were included as covariates to eliminate the possibility of variability that influences the expression and splicing profiles.

ANOVA modules (method of moments) were performed using Partek® Genomics Suite™ to determine differentially expressed and alternatively spliced transcripts amongst 12 CNS regions. All *P* values were corrected for multiple comparisons using the false discovery rate (FDR) step-up method (Benjamini and Hochberg 1995). This analysis was applied to the list of genes that was generated in the previous section 2.6.2.1 by considering only genes that were detected in all 12 CNS regions after applying DABG filter method to reduce the probability of false positive results (12,790 genes).

2.6.3.1 Analysis of differentially expressed genes

Model 1: ANOVA 4-way model to identify differentially expressed transcripts between CNS regions

$$Y_{ijklm} = \mu + Region_i + Gender_j + Scan\ Date_k + Individual\ (Gender)_{jl} + \varepsilon_{ijklm}$$

- Where Y_{ijklm} represents the m^{th} observation on the i^{th} Region, j^{th} Gender, k^{th} Scan Date, and l^{th} Individual.
- μ is the common effect for the whole experiment.
- ε_{ijklm} represents the random error present in the m^{th} observation on the i^{th} Region, j^{th} Gender, k^{th} Scan Date, and l^{th} Individual.

- The errors ε_{ijklm} are assumed to be normally and independently distributed with mean 0 and standard deviation δ for all measurements.
- Scan Date and Individual are random effects.

Then, a list of differentially expressed genes was identified by applying a statistical threshold of minimum fold change (FC) difference of ≥ 1.7 and FDR < 0.05 to the ANOVA result sheet. This was performed using Partek® Genomics Suite™ to compare all possible combinations between the 12 CNS regions using contrast method (Tamhane and Dunlop 2000) and define the differentially expressed genes lists using the list that was generated in the previous section 2.6.2.1 after applying the DABG filter method (12,790 genes).

2.6.3.2 Analysis of differentially spliced genes

Model 2: Alt-splicing ANOVA 4-way model to identify alternatively spliced transcripts between CNS regions.

$Y_{ijklm} = \mu + \text{Region}_i + \text{Gender}_j + \text{Scan Date}_k + \text{Marker ID}_l + \text{Individual (Gender)}_{jm} + \text{Sample ID (Region * Gender * Scan Date * Individual)}_{ijkmn} + \text{Region * Marker ID}_{il} + \varepsilon_{ijklmno}$

- Where $Y_{ijklmno}$ represents the o^{th} observation on the i^{th} Region, j^{th} Gender, k^{th} Scan Date, l^{th} Marker ID, m^{th} Individual, and n^{th} Sample ID.
- μ is the common effect for the whole experiment.
- $\varepsilon_{ijklmno}$ represents the random error present in the o^{th} observation on the i^{th} Region, j^{th} Gender, k^{th} Scan Date, l^{th} Marker ID, m^{th} Individual, and n^{th} Sample ID.
- The errors $\varepsilon_{ijklmno}$ are assumed to be normally and independently distributed with mean 0 and standard deviation δ for all measurements.

- Marker ID_l is an exon-to-exon effect (alt-splicing independent of tissue type).
- Region * Marker ID_{il} represent whether an exon expresses differently in different levels of the specified alternative splice factor(s).
- Sample ID (Individual * Region * Gender * Scan Date)$_{ijkmn}$ is a sample-to-sample effect.
- Scan Date, Individual and Sample ID are random effects.

Then, a list of differentially spliced genes was identified by applying a statistical threshold FDR < 10^{-5} to the ANOVA result sheet. This was performed using Partek® Genomics Suite™ to compare all possible combinations between the 12 CNS regions and cortical regions: FCTX, TCTX and OCTX (Tamhane and Dunlop 2000).

In addition, to confirm of the results of the differentially expressed and spliced genes, analysis of variance (ANOVA) procedures were run independently on the expression data for selected transcripts, using the same models as described in the previous section. The *P* values that were generated from these results were adjusted using a Bonferroni procedure, using the total number of tests undertaken.

2.7 Affymetrix arrays expression validation using a direct RNA quantification with branched DNA: QuantiGene®2.0 Assay (QG)

This experimental work was performed at Affymetrix, USA Ltd. as a service laboratory for the samples that I provided.

CRBL, OCTX, PUTM and WHMT total RNA samples from 12 individuals that were analyzed on exon arrays were also analyzed using the QG platform for validation of exon array results. Three target genes were selected for validation, namely, leucine-rich repeat kinase 2 (*LRRK2*), sodium channel voltage gated type VIII alpha subunit (*SCN8A*), and microtubule-associated protein tau (*MAPT*). ribosomal protein, large, P0 (*RPLP0*) and ubiquitin C (*UBC*) were selected as housekeeping genes to normalize the target genes as they showed relatively low variability in expression levels (i.e. low coefficient of variation) in all regions in the UKBEC dataset. The approach to the selection of reference genes is explained in previous studies (de Jonge, Fehrmann et al. 2007; Coulson, Brockbank et al. 2008). In addition, a recent study confirms the efficiency of using this approach in selecting housekeeping genes to normalize in different tissues (Chervoneva, Li et al. 2010). A summary of the QG probes that were used for analysis of all five genes is provided in **Table 2.7.1.**

QuantiGene 2.0 Reagent System (QG) was used and the protocol in the QuantiGene 2.0 Reagent System User Manual was followed with the exception of the substrate step. Lumigen® Lumi-Phos® Plus and 10% lithium lauryl sulphate and a 30 minutes incubation at 46°C was used in place of the Lumigen® APS-5 substrate. A Biotek ELx 405 select plate washer was used for all of the wash steps in the assay. The QG 2.0 plates were then read on a Molecular Devices LMax luminometer (Turner BioSystems Modulus Microplate Luminometer (P/N 9300-001), Molecular

Devices LMAX) with the plate incubator set to 45°C to maintain the temperature of the Lumigen® Lumi-Phos® Plus substrate. In total, 13 QG 2.0 plates were run to cover all target genes and the house keeping genes. Both housekeeping genes (*RPLP0* and *UBC*) were loaded in duplicates at 12.5 ng/well. In addition, target genes were loaded in duplicates at 75 ng/well.

Table 2.7.1: Quanti-Gene's probe sets designed and used to perform the validation.

Symbol Part	Catalogue number	Specificity
LRRK2	83322 SA-26988	This probe set is designed specifically to hybridize to human *LRRK2*.
LRRK2	83326 SA-50260	This probe set is designed specifically to hybridize to human *LRRK2*, exon 33.
LRRK2	83327 SA-50259	This probe set is designed specifically to hybridize to human *LRRK2*, exon 36.
SCN8A	83324 SA-17320	This probe set is designed specifically to hybridize to human *SCN8A* transcript variants 1 and 2.
SCN8A	83323 SA-50261	This probe set is designed specifically to hybridize to human *SCN8A*, exon 6.
MAPT	81849 SA-15486	This probe set is designed specifically to hybridize to human *MAPT*, all six variants.
MAPT	83321 SA-50258	This probe set is designed specifically to hybridize to human *MAPT*, exon 3
MAPT	83325 SA-50257	This probe set is designed specifically to hybridize to human *MAPT*, exon 4.
UBC	80041 SA-10061	This probe set is designed specifically to hybridize to human *UBC*.
RPLP0	81152 SA-11148	This probe set is designed specifically to hybridize to Human *RPLP0* transcript variants 1 and 2

2.8 Array expression validation using quantitative real time polymerase chain reaction (qRT-PCR) using the TaqMan essay

Aliquots of the total RNA previously extracted from each brain region and analyzed on exon arrays were used for validation by quantitative RT-PCR analysis. The relative expression values were calculated using the Ct value of each target gene which was normalized by the subtraction of the geometric mean of the Ct values from the three housekeeping genes to obtain the ΔCt and ΔΔCt values.

This work was performed at the Queen Square Brain Bank (QSBB) by Dr Jana Vandrovcova for the samples that I provided.

2.8.1 *MAPT* and *LRRK2* qRT-PCR

MAPT transcript-specific assays used were (Hs00902188, variants 3 and 4, exon boundary 2-3, Hs00902978, variants 5 and 7, exon boundary 3-4 and Hs00902314, variants 1, 2, 6 and 8, exon boundary 3-4) (Invitrogen, UK). The *LRRK2* gene expression was quantified using the qRT-PCR technique (Invitrogen, UK) from 30 individuals in 3 brain regions: CRBL, OCTX and WHMT. The two following *LRRK2* specific assays which cover the exon-exon boundary were used: Hs00968193: exon boundary 39-40; Hs00411194: exon boundary 44-45. Fluorescence was measured using the MX3000P system (Agilent, UK). All runs were performed in triplicate and were normalized to a geometric mean of three housekeeping genes (*PPIA, RPL0* and *UBC*). Details of the assays are given in **Table 2.8.1.**

2.8.2 Selected genes for qRT-PCR validation

This experimental work was performed by AROS Applied Biotechnology AS Company as a service laboratory for the samples that I provided.

These experiments were performed using the Fluidigm BioMark system (Fluidigm, Europe) on a subset of samples from THAL (n = 34), WHMT (n = 37), HIPP (n = 29) and HYPO (n = 13) for validations. The expression selected genes/transcripts, including three endogenous controls, using human-specific TaqMan assays (Applied Biosystems UK) were obtained. Details of the assays and genes names are given in **Table 2.8.1**). *RPLP0*, *TUBB* and *UBC* specific assays were used as endogenous controls. Samples were analyzed using Fluidigm 96.96 Dynamic (Fluidigm Europe) arrays with assay triplicates in accordance with the manufacturer‘s protocol. An extract of 100 ng of total RNA was used as input and reverse transcription was performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems UK) in accordance with the manufacturer‘s protocol. After reverse transcription the cDNA, samples were amplified as described in the Fluidigm® Specific Target Amplification Quick Reference Manual (Fluidigm, Europe). The Ct value (cycle number at threshold) was used to calculate the relative amount of mRNA molecules. Gene expression or splicing candidates were considered confirmed if the *P*-value calculated by unpaired ***t***-test was < 0.05.

Table 2.8.1: Details of the Applied Biosystem Taq Man assays for the selected genes.

Gene name	Assay ID	Assay location	Different targeted isoforms
BIN1	Hs01120903_m1	Exon boundary 6-7	Three
BIN1	Hs01120891_m1	Exon boundary 6-7	Seven
CR1	Hs00559348_m1	Exon boundary 33-34 and 41-42	Two
IGFBP3	Hs00181211_m1	Exon boundary 2-3	Two
LRRK2	Hs00968193_m1	Exon boundary 39-40	One
LRRK2	Hs00411194_m1	Exon boundary 44-45	One
MAPT	Hs00902188_m1	Exon boundary 2-3	Two
MAPT	Hs00902978_m1	Exon boundary 3-4	Two
MAPT	Hs00902314_m1	Exon boundary 3-4	Four
NLGN4X	Hs01934144_s1	Exon boundary 6-6	Two
NRXN3	Hs01028181_m1	Exon boundary 8-9	One
NRXN3	Hs01031277_m1	Exon boundary 1-2	Two
PICALM	Hs00999745_m1	Exon boundary 12-13	Two
PICALM	Hs00999727_m1	Exon boundary 12-13	One
RPLP0	Hs99999902_m1	Exon boundary 3-3	Two
TUBB	Hs03929064_g1	Exon boundary 4-4	One
UBC	Hs00824723_m1	Exon boundary 1-2	One
VSIG4	Hs00200695_m1	Exon boundary 1-2	Four

This table shows the selected genes and their corresponding specific Taq Man assays for their different transcripts/isoforms. The last two columns show the location of the specific probe on the exon-exon boundary for specific and different transcripts, in addition, the numbers of isoforms that each assay covers, based on the human genome build 19.

2.9 MAPT protein expression (tau), extraction and Western blot analysis

This work was performed at the Institute of Neurology by Dr Selina Wray for the samples I provided.

From 12 individuals, brain tissue from five brain regions (FCTX, OCTX, CRBL, PUTM and WHMT) were homogenized in 10 mM Tris-HCl (pH 7.4), 0.8 M NaCl, 1 mM EDTA, 10% sucrose and protease inhibitor tablets (Roche, UK). Homogenates were clarified by centrifugation at 10,000 g for 10 minutes at 4°C prior to aliquoting and storage at -80°C. Protein concentrations were measured by the bicinchoninic acid assay (BCA) assay and equal amounts of protein were dephosphorylated using λ protein phosphatase (New England Biolabs, UK) as described previously (Hanger, Gibb et al. 2002). Briefly, proteins were incubated with λ protein phosphatase at a final concentration of 40 U/μl for 3 hours at 30°C. Dephosphorylation reactions were stopped by the addition of LDS buffer (Invitrogen, UK) followed by heating at 100°C for 10 minutes. Samples were centrifuged at 10,000 g prior to separation on 10% Bis-Tris gels alongside recombinant tau protein ladder (Sigma, UK). Protein was transferred to nitrocellulose membrane and probed with rabbit polyclonal antibody to total tau (Dako, UK) and mouse monoclonal to actin (Sigma, UK). Blots were visualized and quantified using an Odyssey Infrared imaging system (LI-COR Biosciences, UK). The levels of total tau in each brain region were normalized to actin protein, and the levels of individual tau isoforms were calculated as a percentage of total tau for the same sample. The *t-test* was performed to calculate the significance of the protein levels differences between different regions is significant.

2.10 MAPT protein isoforms expression (tau) using sandwich ELISA analysis

This work was performed at the QSBB by Dr Connie Luk for the samples that I provided.

From the homogenized samples in Section 2.12 in the five brain regions (CRBL, FCTX, OCTX, PUTM and WHMT), sandwich ELISAs for 3R- and 4R- tau were carried out as previously described by Luk *et al.* group (Luk, Giovannoni et al. 2009). Briefly, microlitre plates were coated with 150 µl of 10 µg/ml of capture antibody (RD3 or RD4) (de Silva, Lashley et al. 2003) in coating buffer (150 µl, sodium tetraborate buffer, pH 9.4). The plate was washed and 150 µl of diluent buffer was added into the plate after which 25 µl of diluted supernatants obtained from the brain homogenates (1 in 1,000 and 1 in 150 for 3R- and 4R- tau ELISAs respectively) were added in duplicates to the blocked plates previously coated with RD3 or RD4 antibodies. The plates were incubated at room temperature for 2 hours, washed and 150 µl affinity-purified sheep anti-tau–HRP conjugates, diluted to 1 in 1,000 and 1 in 3,500 for 3R- and 4R-tau assays, was added, respectively, and incubated for 1 hour on a shaker. After washing, the plates were developed with tetramethylbenzidine (TMB) substrate as previously described. The isoform composition in each sampled region is expressed as ng/µg of total brain protein.

2.11 *LRRK2* array splicing validation using Reverse Transcriptase RT-PCR

2.11.1 Splicing index analysis for array data

To test for differential exon specific expression in the array data, the ratio of exon probe set value by overall gene/trancript expression (Winsorized mean across exons) was computed for each individual and brain region. Then the two linear models with or without the inclusion of the regions as covariates were compared and the *P*-value from this *F* test reported was reported.

2.11.2 Semi-quantitative Reverse-Transcriptase RT- PCR

The validation for the array splicing events was performed by using QIAGEN LongRange 2stepS RT-PCR kit (Qiagen,UK). All primers for this analysis were designed using the Primer3 software (http://fokker.wi.mit.edu/primer3/input.htm), and then they were BLAST-searched against UCSC human In-silico PCR tools. The sequences of the PCR primers used in this study are listed in **table 2.11.1** below.

An aliquot of total RNA from 48 samples, originating from four brain regions (OCTX, SNIG, MEDU and CRBL) and 12 individuals were used as a subset for further validation. cDNA was synthesis from 1-2 µg of total RNA using gene specific designed primers under the following conditions: incubation at 42ºC for 90 minutes, followed by enzyme activation at 85°C for 5 minutes. 2.5 µl of cDNA was used to perform the semi-quantitative RT-PCR for the targeted exons under the following conditions: Initial activation at 93ºC for 3 minutes, followed by denaturation at 93°C for 15 seconds, annealing 59°C for 50 seconds, elongation at 68ºC for 50 seconds for 30-40 cycles. PCR products were run on a 2% agarose gel (Invitrogen,UK) and

photographed using UV illumination to visualize GelRed staining. Images were inverted in Adobe Photoshop.

Table 2.11.1: ***LRRK2*** **Reverse Transcriptase PCR primers.**

Exon	Primer name	Sequence (5′ >>> 3′)	Location
32-33	*LRRK2*-Ex32-33 Forward	ACCATCATAAACGAG AGCCTTAATTTC	In exon 31
32-33	*LRRK2*-Ex32-33 Reverse	AGCAATCTGGAATTT TTCTAGGAGCTT	In exon 34
33	*LRRK2*-Exon-33 Forward	CAGCTGCAGTTAGAT GAAAATGAGC	In exon 32
33	*LRRK2*-Exon-33 Reverse	TGCCCTTAGGGTGTT TTGGACAACCTT	In exon 34
42-43-44	*LRRK2*-Exon43 Forward	AGTTTGATGAATTAG AAATACAAGGAAAA	In exon 42
42-43-44	*LRRK2*-Exon43 Reverse	TTTAGGTAATAAAAT GCGTCTCGTCAG	In junction of exon 44-45

This table shows the sequences of the primers that were used for *LRRK2* exons splicing validation on a subset of the human brain RNA samples.

2.12 LRRK2 protein expression, extraction and Western blot analysis

This work was performed at the Institute of Neurology by Dr Patrick Lewis with samples that I provided.

From five individuals, human brain tissues were homogenized from five brain regions: FCTX, OCTX, CRBL, PUTM and WHMT in phosphate buffered saline (Life Technologies, UK) 10% W/V using a Tissue rupture (Qiagen, UK) on ice for 30 seconds, aliquoted and frozen prior to further use. Samples were denatured by the addition of LDS loading buffer with beta mercaptoethanol. Samples were denatured at 100°C for 10 minutes and clarified by centrifugation at 12,000 g for 1 minute prior to loading onto 4-12% polyacrylamide gels (Life Sciences, UK), with 20 µl loaded per sample. Gels were run at 140 V, 4°C for 3 hours and transferred onto Polyvinalidene Fluoride PVDF membrane (Millipore, UK) at 25v for 24 hours. Following transfer, membranes were blocked in PBST plus 5% milk powder for 1 hour and probed with primary antibodies. For LRRK2, a rabbit monoclonal antibody (MJFF2, Epitomics/Michael J. Fox Foundation, USA) was used at a 1:1000 dilution in phosphate-buffered saline Tween 20 (PBST)/5% milk. For β Actin, a mouse monoclonal antibody (Sigma, UK) was used at a 1:5000 dilution. Membranes were probed at 4°C for 16 hours, washed three times with PBST and then probed with secondary horseradish peroxidase (HRP) conjugated antibodies (anti-Rabbit, 1:2000 dilution in PBST milk, and anti-mouse 1:10000 dilution for LRRK2 and β Actin respectively) for 1 hour at room temperature. Membranes were washed three times in PBST and then immunoreactivity visualized by exposure of membranes to enhanced chemiluminescence (ECL) substrate (Pierce) and then photographic film (Super-RX, Fujifilm, UK). Films were developed using a Konica developer.

2.13 LRRK2 immunohistochemistry (IHC) staining analysis

This work was performed at the Queen Square Brain Bank (QSBB) by Dr Rina Bandopadhyay.

Flash-frozen sections of 12 µm thickness were cut using a cryostat (Leica Microsystems, UK) from three control brains (no diagnosed neurodegenerative disorders) which were obtained from the QSBB archive from two brain regions; CRBL and FCTX. Sections were air-dried for 30 minutes and fixed in 4% freshly made paraformaldehyde for 30 minutes followed by three 5 minute washes in 1XPBS. Endogenous peroxidase activity was blocked by incubating sections in methanol/0.3% H2O2, followed by three 5 minute washes in 1XPBS. Non-specific protein binding was blocked in 10% non-fat milk in 1X PBS at room temperature for 30 minutes. A rabbit monoclonal primary antibody to LRRK2 (MJFF2, Epitomics/Michael J. Fox Foundation, USA) diluted at 1:50 in 1X PBS, incubated initially for one hour at room temperature and then at 4°C for 16 hours. This was followed by incubations with biotinylated secondary antibody (anti-rabbit, Dako, UK) 1:200, and then with avidin-biotin reagent (Vector laboratories, UK) for 30 minutes each. Bound antibody was visualized using the chromogen diaminobenzidine, and sections were counterstained with Mayer's haematoxylin. Finally, sections were dehydrated in graded alcohols (70%-100%), cleared with xylene and mounted with coverslips with DPX mounting medium (VWR, UK). To validate specific staining, one slide without primary antibody was always included.

To assess LRRK2 specific staining further, a recombinant protein block was included. The LRRK2 antibody in 1X PBS was incubated with 100 µg of LRRK2 recombinant protein (Life Technologies, UK) initially at 37°C for 4 hours and then

overnight at 4°C. The solution was then centrifuged at 12,000 g for 30 minutes at 4°C. The supernatant was carefully decanted and used on a cerebellar section.

2.14 DNA extraction using Gentra Puregene Tissue Kit (4 g)

In parallel, genomic DNA was extracted from sub-dissected samples (100–200 mg) of human post-mortem brain tissue using Gentra Puregene Tissue Kit (4 g) (Qiagen, UK). A single DNA extraction was performed on each brain tissue sample. All steps were performed on dry ice prior to the addition of the Cell Lysis Solution.

2.14.1 Tissue homogenization and lysing

Brain tissues from CRBL or OCTX regions were collected and weighed in sterilized 15 ml conical centrifuge tubes (VWR, UK); they were between 50-100 mg in weight. All tissue samples were homogenized on ice using the Tissue rupture (Qiagen, UK) for 30 seconds at maximum speed in 4 ml of Cell Lysis Solution. In this step, the tissue was homogenized until no big pieces from the tissue remained or visualized. The homogenized tissue samples were incubated at 55°C with 20 µl Proteinase K (Qiagen, UK) overnight for maximum DNA yields.

2.14.2 RNA digestion

After incubation, 20 µl of RNase A solution (Qiagen, UK) was added, the tubes were mixed gently by inverting 25 times. Then the tubes were incubated at 37°C for 60 minutes. After 3 minutes, tubes were placed on ice to quickly cool samples and stop the enzymatic reaction.

2.14.3 Protein and DNA precipitation, wash and elution

Following the incubation, 1.3 ml of Protein precipitation solution was added, and vortexed vigorously for 40-50 seconds at maximum speed, centrifuged for 15 minutes at 2000 g. It was important to ensure that a tight white protein pellet was seen.

The supernatant was transferred to a fresh 15 ml conical centrifuge tube. DNA was precipitated by adding 4 ml of 100% propanol and mixed by inverting gently 50 times. Tubes were centrifuged at 2000 g for 45 minutes at 4°C. The supernatant was discarded carefully and the DNA pellet was washed by adding 70% ethanol and inverted several times. After centrifugation for 1 minute at 2000 g, the supernatant was discarded and the pellet was air-dried for 10-15 minutes at room temperature. Finally, the DNA pellet was dissolved by adding 400 µl pre-heated DNA hydration solution (55°C) and vortexed for 10 seconds. DNA samples were incubated at room temperature overnight to ensure that all the DNA was dissolved. The next day, DNA samples were ready for quality checks followed by the application of Illumina 1M Quad BeadChip genotyping.

2.14.4 DNA quality assessments

2.14.4.1 Concentration and purity

The concentration and purity of each DNA sample was assessed by measuring the optical density (OD) of each DNA sample at a wavelength of 260 nm, using the NanoDrop ND-1000 Spectrophotometer V3.3.0. The concentration of each sample was calculated, together with the ratio of absorbance at 260 nm/280 nm and 260 nm /230 nm. Samples with 260 nm/280 nm ratio lower than 1.7 were re-extracted from a fresh tissue block.

2.15 Genome-wide DNA genotyping using Illumina Human Omni1-Quad BeadChips and genotyping data quality control

2.15.1 Illumina Human Omni1-Quad BeadChips and custom ImmunoChip

This experimental work was performed by AROS Applied Biotechnology AS Company as a service laboratory for the samples that I provided.

For this procedure, 4 µl of 50 ng/µl genomic DNA samples were marked on the Illumina Infinium Human Omni1-Quad BeadChip (Illumina, USA) according to the manufacturer's instructions. The BeadChips were scanned using an iScan (Illumina, USA) with an AutoLoader (Illumina, USA). GenomeStudio v.1.8.X (Illumina Corp.) was used for analyzing the data and generating SNP calls. SNP calls were created using the HumanOmni1-Quad_v1-0_B (C) cluster file provided by Illumina as a reference.

The following experimental work was performed by UCL Genomics as a service laboratory for the samples that I provided.

In addition, all individuals were also genotyped using ImmunoChip, a custom genotyping array was designed for the fine-mapping of auto-immune disorders. The ImmunoChip also contains 2,000 SNPs selected for replication of a Parkinson's disease genome-wide association study performed by the International Parkinson's Disease Genomics Consortium (International Parkinson's Disease Genomics Consortium and Wellcome Trust Case Control Consortium 2 2011; Nalls, Plagnol et al. 2011). This provided the resolution to fine map the *LRRK2* gene (see Chapter 6).

2.15.2 Genotyping data quality control

This analysis work was performed at the Department of Medical & Molecular Genetics, King's College London by Dr Michael E. Weale and Dr Adaikalavan Ramasamy for the data that I provided.

Standard quality control checks carried out on both datasets include: the removal of the genotyped markers that are labelled as monomorphic, or have less than 95% genotyping call rate across samples, *P* value for deviation from Hardy-Weinberg equilibrium ($P < 0.0001$), had no genomic position information and had less than two heterozygotes or redundant with an existing SNP. In addition, the reported sex status was checked and three Individuals were removed as they were suspected of being of non- Caucasian European ancestry, based on principal components projection of population stratification generated using HapMap3 data. This resulted in a final dataset of 134 individuals.

After quality control, 815,859 SNPs from the Illumina Omi1-Quad chip and 137,456 from the ImmunoChip were included and obtained a total of 905,943 after merging the two datasets and resolving for overlapping SNPs. All manipulations of genotypic data were performed in PLINK version 1.07 (Purcell, Neale et al. 2007).

Next, un-typed, missing SNPs and indels were imputed from the autosomes and chromosome X using the MACH (Li, Willer et al. 2009; Li, Willer et al. 2010) and minimac (http://genome.sph.umich.edu/wiki/Minimac) softwares using the European panel of the 1000 Genomes Project (March 2012: Integrated Phase I haplotype release version 3, based on the 2010-11 data freeze and 2012-03-14 haplotypes) which is based on 381 individuals of European descent. The genotyped and imputed data were

restricted to ~5.88 million SNPs that had a minor allele frequency of at least 5% and acceptable imputation quality ($r^2 > 0.50$).

For *MAPT* analysis in Chapter 5, the selection of individuals with H1/H1 haplotype status was based on genotyped SNPs rs1800547 and rs1052553 in both datasets.

2.16 Expression Quantitative trait loci (QTL) analysis

This work was performed at the Department of Medical & Molecular Genetics, King's College London by Dr Michael E. Weale and Dr Adaikalavan Ramasamy for the data I provided.

2.16.1 Identification of polymorphism-containing probes using reference datasets

Before the expression QTL analysis was performed, probes containing common polymorphisms using different public datasets of genetic variation were identified. The main datasets that were used initially were the genotyping data from Illumina HumanHap550 array from the NABEC dataset, and the Illumina Omni-1M Quad array from the UKBEC dataset. Then the CEU panel of the final release of HapMap (release #28, merged Phase I+II+III data) was used. Next, the SNP and indel data of the European panel (n = 381) of the latest version of the 1000 Genomes Project (March 2012: Integrated Phase I haplotype release version 3, based on the 2010-11 data freeze and 2012-03-14 haplotypes) was also considered. Finally, the SNPs (average read depth ≥ 10) from the Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA (http://evs.gs.washington.edu/EVS/) [accessed 11 May 2012) which is taken from 3,510 European Americans from multiple ESP cohorts were

used. SNPs that were identified with at least 1% minor allele frequency in European descent samples in all the datasets were included in the analysis.

2.16.2 Masking probes that contain polymorphisms for the whole genome Affymetrix Exon 1.0 ST expression probe sets

After polymorphism identification, this information was used to mask or discard the probes that contain the polymorphism in the expression arrays pre-processing and normalization steps.

The design of the Affymetrix expression exon array is made to have four probes which are grouped into one probe set to measure in most cases the expression level of a single given exon. If one or two of the four probes contain a polymorphism and the remainder do not, the probe set and estimated the exon signal from the remaining probes were included in the analysis. If less than two probes remained (because one or more probes had a polymorphism or because of other quality control filters dropped-out of probes) then the remaining information was considered insufficient and the probe set was discarded. Probe masking has the advantage that both false positive and false negative expression QTL signals can be recovered. The masking, array CEL files processing, normalization and summarization for 291,704 probe sets were performed as in Section 2.6.1 in Affymetrix Power Tools to generate the whole genome expression profile intensities as $\log_2$ transformed values for the ten brain regions for the expression QTL analysis after correcting for the polymorphisms in probes. The ten brain regions that were included for the expression QTL association analysis are as follow: FCTX, $n = 127$; TCTX, $n = 119$; OCTX, $n = 129$; HIPP, $n = 122$; THAL, $n = 124$; CRBL, $n = 130$; SNIG, $n = 101$; PUTM, $n = 129$; MEDU, $n = 119$; WHMT, $n = 131$, originating from 134. It should be noted that SPCO and HYPO

regions were excluded from the expression QTL analysis, as the sample size for these two regions was too small (only 13 samples each).

2.16.3 Expression QTL analysis and linkage disequilibrium (LD) signal identification

Once the data were converted into a suitable format, the association between each SNP (genetic marker) and each expression profile (1231 arrays) was performed. The expression profile is defined as a probe set signal as covering an exon (in total 291,704 probe sets) or a transcript set which is a group of probe sets covering the whole transcript (in total 26,493 transcripts), assuming an additive genetic model for SNPs. The computation was done using the MatrixEQTL software (Shabalin 2012) (http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL/) and Revolution R (http://www.revolutionanalytics.com/) on a high performance Linux-based computer cluster.

The process of imputation and linkage disequilibrium across the genome creates a problem in the expression QTLs from SNP-rich high LD regions, which was represented several times by LD proxy. Therefore, multiple associations for a given probe set were treated as a single signal if the associated SNPs were in LD ($r^2 > 0.5$) with each other. For a given transcript ID/ cluster pairing, SNP with the smallest *P*-value was classified as the 'LD-resolved' sentinel marker expression QTL. An expression QTL signal was defined as *cis*-acting if the target SNP is located within 1 Mb of the transcription start site of the associated transcript and *trans*-acting if the SNP is located outside the *cis* region on the genome.

2.16.4 False discovery rate (FDR) calculation

The results were classified by the marker type: SNP or indel, expression type: exon or gene/transcript-level and the distance of SNP to the transcription start site: *cis* or *trans*. Then, the FDR was calculated based on these classifications. For each class, the number of tests was conducted and converted the nominal *P* values into a false discovery rates (FDR). Only the associations with FDR < 1% were considered for the subsequent analyses.

2.16.5 SNP and functional annotations

The haplotype variant call format (VCF) file for SNPs and indels was submitted separately to the Seattle Sequence Annotation hg19 server (http://snp.gs.washington.edu/SeattleSeqAnnotation134/index.jsp). The Genome variation server functional annotation (functionGVS) and transcription factor columns were retained (SeattleSNPs. 2012).

2.16.6 Integrating expression QTLs with GWAS catalogue

The NHGRI GWAs catalogue (Hindorff, Sethupathy et al. 2009; Hindorff, MacArthur et al. 2012) was downloaded from the link http://www.genome.gov/gwastudies/. This catalogue contained 8.922 records; 337 records that reported haplotypes or indels (instead of SNPs) were ignored. After removing duplicate entries, there were 7,042 SNPs of which 6,387 (~91%) were identified in the imputed genotype data of UKBEC. The GWAS hits were classified as either brain or non-brain related according to the reported trait/diseases. For each sentinel SNP or sentinel indel, the LD with these GWAS SNPs (within a 10 Mb or 1000 SNP moving window) was calculated using the vcftools utility (version 0.1.9;

http://vcftools.sourceforge.net/) and haplotype data of the European panel (n = 379) from the 1000 Genomes Project (March 2012). If an LD between SNPs of $r^2 \geq 0.50$ was observed, the corresponding GWAS SNPs along with the LD measure was recorded.

2.17 Characteristics of the human post-mortem tissue of the North American Brain Expression Consortium (NABEC)

This work was performed at the Molecular Genetics Section and Laboratory of Neurogenetics, NIA, NIH, USA, resulting in two publications (Gibbs, van der Brug et al. 2010; Hernandez, Nalls et al. 2012).

In this study, in addition to the expression profiles that were generated from 88 individuals from the UKBEC dataset on Illumina HT-12 arrays (see Section 2.5), the expression profiles that were generated from 305 individuals from the NABEC dataset, who were expression profiled on the Illumina HumanHT-12 v3 Expression BeadChip arrays, were downloaded and used to increase the number of samples for *MAPT* analysis (see Chapter 5).

2.17.1 Human post-mortem brain tissue collection and dissection

CRBL and FCTX samples originating from 305 control individuals were collected as previously described (Gibbs, van der Brug et al. 2010; Nalls, Plagnol et al. 2011; Hernandez, Nalls et al. 2012). A small number of individuals were expression profiled by both consortia but these were removed in this study.

2.18 RNA extraction using single-step TRIZOL®Reagent based protocol

Total RNA was extracted from 100–200 mg of human post-mortem brain tissue using a glass-Teflon homogenizer and 1 ml TRIzol (Invitrogen, Carlsbad, CA) as described previously (Gibbs, van der Brug et al. 2010; Hernandez, Nalls et al. 2012).

2.19 Genome-wide expression profiling using Illumina HumanHT-12 v3 Expression BeadChip

RNA was biotinylated and amplified using the Illumina® TotalPrep-96 RNA Amplification Kit and directly hybridized onto HumanHT-12 v3 Expression BeadChips (Illumina Inc., USA) in accordance with the manufacturer's instructions. In this section the same protocol in Section 2.5 was applied.

2.20 Illumina HumanHT-12 v3 Expression BeadChip arrays analysis

Expression data were analyzed using the Gene Expression Module 3.2.7 within Illumina® BeadStudio. The raw intensity values for each probe were transformed using the cubic spline normalization method (Workman, Jensen et al. 2002) (http://www.illumina.com/documents/products/technotes/technote_beadstudio_normalization.pdf) and then $\log_2$ transformed for mRNA analysis. The annotations for probes according to ReMOAT (Barbosa-Morais, Dunning et al. 2010) were re-mapped on the human genome build 19 and then restricted the analysis to genes that were reliable, uniquely hybridized and were associated with gene descriptions. In the analysis for MAPT eQTL, the analyzed data that was generated by the following two probes: ILMN_1710903 and ILMN_2310814 was used. Prior to use in expression QTL analyses, data were corrected for gender, batch effects, age, post-mortem interval and brain bank as described previously (Gibbs, van der Brug et al. 2010).

2.21 Genome-wide DNA genotyping using Illumina® Infinium HumanHap550-v3 BeadChips

The DNA extraction and imputation protocol was similar to that employed in UKBEC. The individuals were genotyped on Illumina® Infinium HumanHap550 v3 Genotyping BeadChips (Illumina, USA) and approximately 5.3 million SNPs were available after imputation and quality control (Gibbs, van der Brug et al. 2010). Finally, the selection of individuals with H1/H1 haplotype status for the *MAPT* gene was based on genotyped SNPs rs1800547 and rs1052553 in both datasets.

3 Quality control assessment on a large dataset of regionally-dissected human control brains for genome-wide expression studies.

3.1 Summary

Human brain tissue is a challenging source to perform genome-wide expression and splicing profiling by using advanced, high throughput techniques such as: expression microarrays and RNA sequencing. Several studies have previously shown that pre- and post-mortem variables can alter total RNA integrity in a complex manner. As a result, researchers need to be cautious when interpreting results that originated from post-mortem tissues.

In order to investigate the viability of the post-mortem tissue data and assess the effect of post-mortem interval (PMI), agonal state and age on array performance and gene expression profiles, and the utility of pH and RNA integrity number (RIN) as predictors of RNA and expression quality; these variables were measured and assessed on 1,266 human Affymetrix Exon Arrays originating form 137 control individuals from thirteen CNS regions. In addition, the accuracy of the array results using QuantiGene (QG), as an independent non-PCR based validation method.

The results revealed that post-mortem delay, agonal state and age have little impact on array performance quality, the quality of array results are robust and not necessarily proportionate to variable RNA integrity, and brain pH has only a small effect on array performance. QuantiGene, as a validation yielded very similar expression profiles as array results did.

This analysis is the first step in this project to build a reliable human, regional brain expression open-access database that is designed to allow the genome-wide assessment of genetic variability of expression. These quality control parameters will allow dataset users to assess data accuracy from control human post-mortem tissues.

3.2 Introduction

Microarray analysis and RNA sequencing of the transciptome in human post-mortem brain tissue is a vital tool in investigating the underlying complex genetic mechanisms involved in neurodegenerative and psychiatric disorders (Gilbert, Brown et al. 1981; Glanzer, Haydon et al. 2004; Myers, Gibbs et al. 2007). The restriction of brain tissue availability, the fact that brain tissues can only be obtained after PMI of unpredictable duration, varying agonal states, medical history, cause of death, different procedures in tissue handling and tissue dissection accuracy are making these types of projects extremely challenging. However, these variables which potentially influence the RNA integrity in post-mortem human brain tissues need to be accounted for in order for such data to be highly reliable (Sajdel-Sulkowska, Majocha et al. 1988; Burke, O'Malley et al. 1991; Leonard, Logel et al. 1993; Glasel 1995; Imbeaud, Graudens et al. 2005; Schroeder, Mueller et al. 2006; Durrenberger, Fernando et al. 2010; Birdsill, Walker et al. 2011). A previous study using 89 human post-mortem brains showed that pre- and post-mortem factors have different levels of effect on different locations in the mRNA transcript. For example, the 3' UTR end is more sensitive to post-mortem factors than other locations in the transcript (Johnston, Cervenak et al. 1997).

In this project, the selection of the array type was made based on the advantages that the commercially available arrays were offering at the time. Affymetrix GeneChip®Human Exon 1.0 ST Arrays offered more comprehensive coverage of the genome, compared with different versions of previous expression arrays, as it was designed to contain 1.4 million probe sets covering 90% of the coding exons. This had been annotated from different genomic sources over the whole transciptome (median number of probes is 30 to 40 per transcript) (Gardina, Clark et

al. 2006). This helped to eliminate RNA degradation that can occur with 3'UTR designed arrays. Degraded RNA will be inefficiently targeted and amplified to cDNA using the olig-DT priming protocol in 3'UTR arrays, and as a result, this decreases the ability to detect degraded and low expressed transcripts. In contrast, exon array probe sets target all coding exons along the targeted transcript to give higher resolution and more sensitivity for the exon and gene expression level analysis (**Figure 3.2.1**). Moreover, the exon array technology is designed to uncover splicing events in the transciptome (Clark, Schweitzer et al. 2007), such as skipping or inclusion of exons using ANOVA based method that is based on the Splicing Index concept (Srinivasan, Shiue et al. 2005). However, every technique has its own limitations, it is noteworthy that the exon arrays have some restrictions with regard to their ability to detect certain splicing events (false negative). For example, small exons with low inclusion/expression rates may prove difficult to detect and thus may have a weak signal to detect and thus may be filtered out. The alternative spliced mRNA exon/ transcript half-life is too short and is degraded before detection. In addition, the probe design of this array does not include the exon-exon junctions (Clark, Schweitzer et al. 2007). More detailed information is available about the exon array design on the Affymetrix website (http://www.affymetrix.com/estore/) by searching for 'GeneChip Human Exon ST Array'.

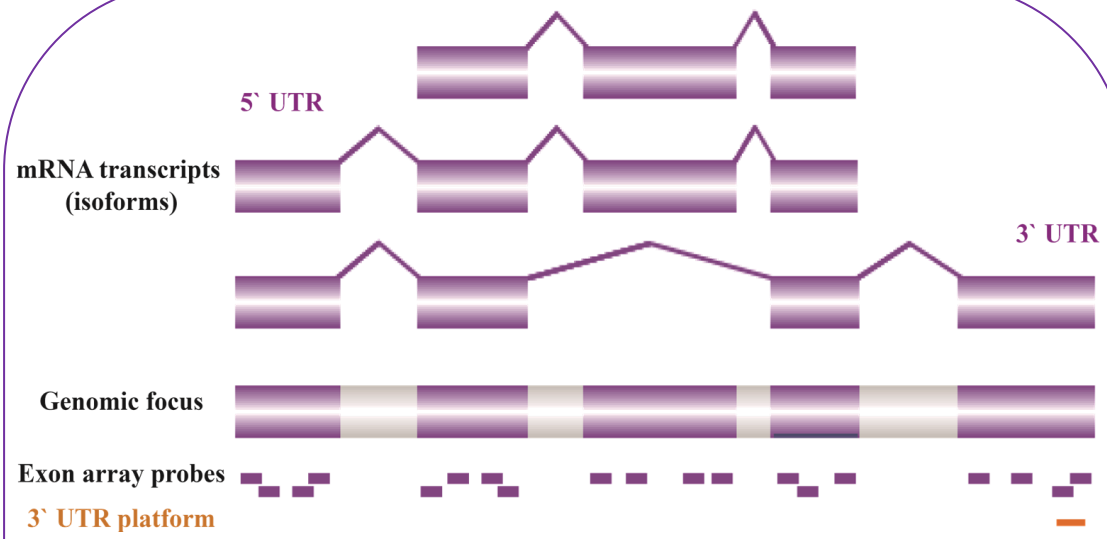

Figure 3.2.1:

Diagram of the exon array probes / probe sets design.

This figure shows the coverage of the probes along the whole length of the transcript on the arrays. Purple boxes are exons from different isoforms. Purple dashes are individual probes targeting the genomic locus, which is termed a probe selection region (PSR). 3-4 probes covering the same exon form a probe set. The orange dash shows the probe location on the 3'UTR platform. Modified from the Affymetrix website, 2011.

Moreover, it is important to have a reliable and stable method to assess the quality of RNA samples that are generated from precious heterogeneous tissues, especially from small anatomical regions, such as the *substantia nigra* and hypothalamus. In order to prevent unnecessary costs, such assessment is important before samples were run through the whole experimental process. The most widespread measure for estimating the integrity of RNA samples at present is the RIN as calculated by the Agilent 2100 Bioanalyzer for electrophoresis (Agilent Technologies UK Ltd, UK) (Becker, Hammerle-Fickinger et al. 2010). The RIN ranges from 1 (undetectable) to 10, with 1 being completely degraded and 10 being the most intact RNA.

The algorithm for the RIN calculation is largely based on the ratio of ribosomal RNA 28S/18S signal heights, areas under the signal peaks, position and separation times (**Figure 3.2.2**). In addition, the algorithm counts for the level of the signal in the fast, inter-regions and post regions to look for RNA degradation products. It assigns relative weight to each of these factors and calculates the final RIN from the combination of these factors (Schroeder, Mueller et al. 2006).

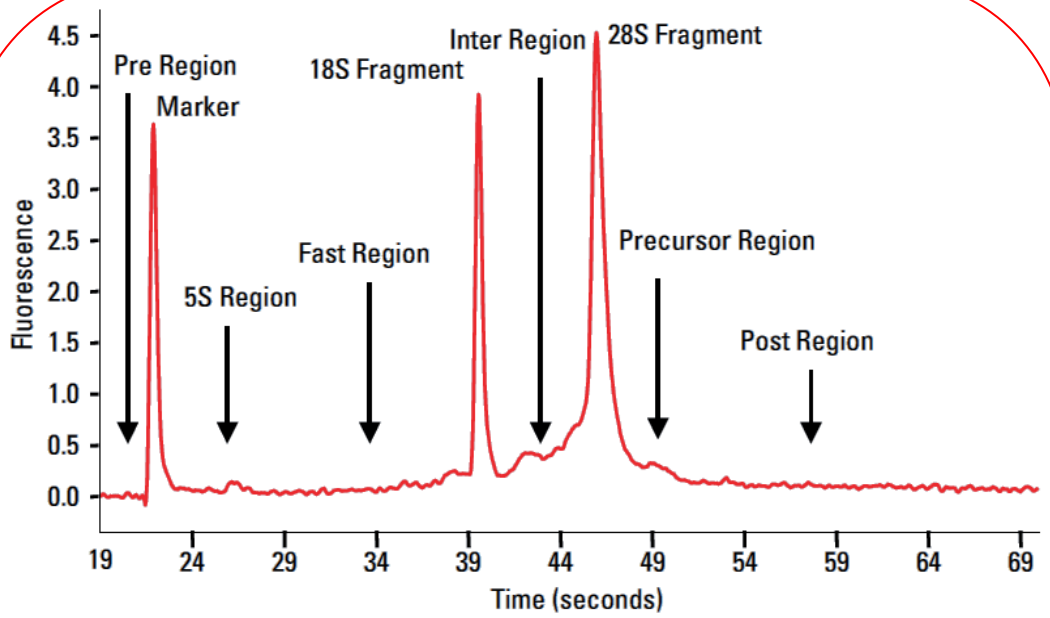

Figure 3.2.2:

Electrophoresis results from the bioanalyzer software clarifying the RIN calculation parameters.

This figure shows all the parameters that the RIN calculation is based on from total RNA samples. This sample shows a high integrity RNA sample with RIN = 10 (Agilent Technologies, 2004).

This provides an indication that it is an indirect measure of the mRNA quality; and further, it has been shown to be inconsistent even for the same RNA sample that was measured at different times (Imbeaud, Graudens et al. 2005; Schroeder, Mueller et al. 2006; Sherwood, Head et al. 2011).

This project of regional expression of control human brain expression was constructed to allow the assessment of the genetic variability in gene expression QTLs and splicing QTLs as well as detailed genome-wide and regional expression analysis in the adult human brain (Hardy, Trabzuni et al. 2009). A large series of 137 control human brain tissues in which were dissected from thirteen different central nervous system (CNS) regions: FCTX, n = 124, TCTX, n = 118, PCTX, n = 19, OCTX, n = 132, HIPP, n = 126, THAL, n = 121, HYPO, n = 14, PUTM, n = 130, SNIG, n = 104, MEDU, n = 116, CRBL, n = 137, SPCO, n = 14 and WHMT, n = 129. From each individual brain region, the RNA was isolated for whole transcriptome exon array analysis. This resulted in a total of 2,318 extracted RNA samples, but only 1,266 RNA samples analysed on Affymetrix Exon arrays and represent by far the largest single CNS expression dataset to date. For this quality control analysis, the factors that affected the reliability of the RNA samples quality were assessed.

The assessment of the following aspects was performed; i) the effects of brain bank (BB), age, gender, cause of death, region of the brain, post-mortem delay and brain pH on RIN-based RNA quality, and, ii) the effects of RNA quality on the performance quality of the array experiment, which was measured by a reliable and widely used parameter, present call (%P) (Weis, Llenos et al. 2007). %P is the percentage of probe sets with signal detection above background noise. The effects of RNA quality on the cDNA preparation and cRNA production were examined as part of the quality control of the array experiment, and finally the reproducibility of array results using QuantiGene (QG), a novel, PCR-independent platform was confirmed (Canales, Luo et al. 2006; Arikawa, Sun et al. 2008; Hall, Leong et al. 2011).

3.3 Results

In total, 1,302 tissue blocks were available for the analysis. They were generated from 137 individuals from 13 CNS regions, as some regions were not available from all individuals. In 870 tissue blocks more than one extraction was performed, and the RIN was measured to assess the RNA integrity. One RNA sample from each tissue was selected to be run on the Affymetrix Exon arrays giving a total of 1,266 exon array expression profiles (**Figure 3.3.1**).

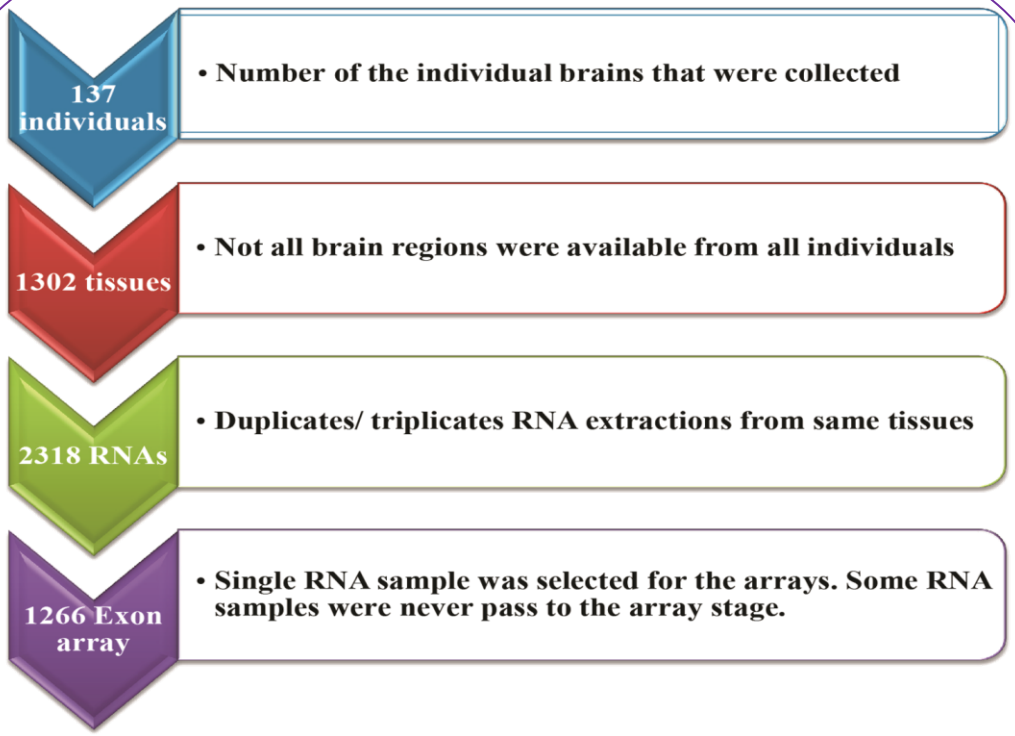

Figure 3.3.1:

Diagram of the samples and data used for the quality control analysis.

This figure shows the full process for the tissues and RNA samples through the extraction, quality control and array experiments (Trabzuni, Ryten et al. 2011).

3.3.1 Factors affecting RIN-based RNA quality

The dependence of RIN values on the other covariates assessed in this study was evaluated. The focus was on the explanatory power of the available covariates on the RIN for each RNA extraction. In this study 2,318 RNA samples were assessed. The calculated RIN ranged from 1 to 8.5 with a mean of 3.85. 43% of the samples had RIN values < 3. 33% of the variation in RIN was explained by differences among tissue blocks (adjusted R^2 measure), which set an upper limit for the explanatory power of the covariates, which all act at a between-tissue-sample level. 16% of the variation in RIN was explained by individual-level differences.

Among all these factors, the most important covariates were brain region (explaining 9.2% of the variation in RIN, $P = 1.7\times10^{-42}$), age (explaining an additional 1.1%, $P = 4.1\times10^{-05}$) and cause of death (explaining an additional 1.9%, $P = 0.022$). Brain bank and gender together explained only an additional 0.4 % of the variation.

pH and PMI were investigated separately, as pH was not measured in the SHRI-USA dataset and because the range of PMIs from the two brain banks did not overlap.

In this part of the analysis 1842 RNA samples were assessed. pH explained 2.1% of the variation in the RIN from the MRC-UK dataset ($P = 1.0\times10^{-4}$) (**Figure 3.3.2-A**). Six individuals had very low pH (< 5.90) and, when excluded, the correlation was no longer significant ($P = 0.82$) (**Figure 3.3.2-B**).

The effect of PMI on RIN differed between the MRC-UK and SHRI-USA datasets. There was no significant correlation between PMI and RIN in the MRC-UK dataset (1842 samples), which had PMIs ranging from 28 to 114 hours. In contrast, PMI explained 8.4% of the variation in RIN in the SHRI-USA dataset (476 samples) (*P*-value = 0.0053), which had PMIs ranging from 1 – 5.5 hours. This result was consistent with results from Birdsill et. al. and Kang et. al studies on control and diseased human post-mortem brain tissues (Birdsill, Walker et al. 2011; Kang, Kawasawa et al. 2011). Longer PMI was associated with a higher RIN in this data set, a counterintuitive result which might represent confounding with some unmeasured variables in this analysis.

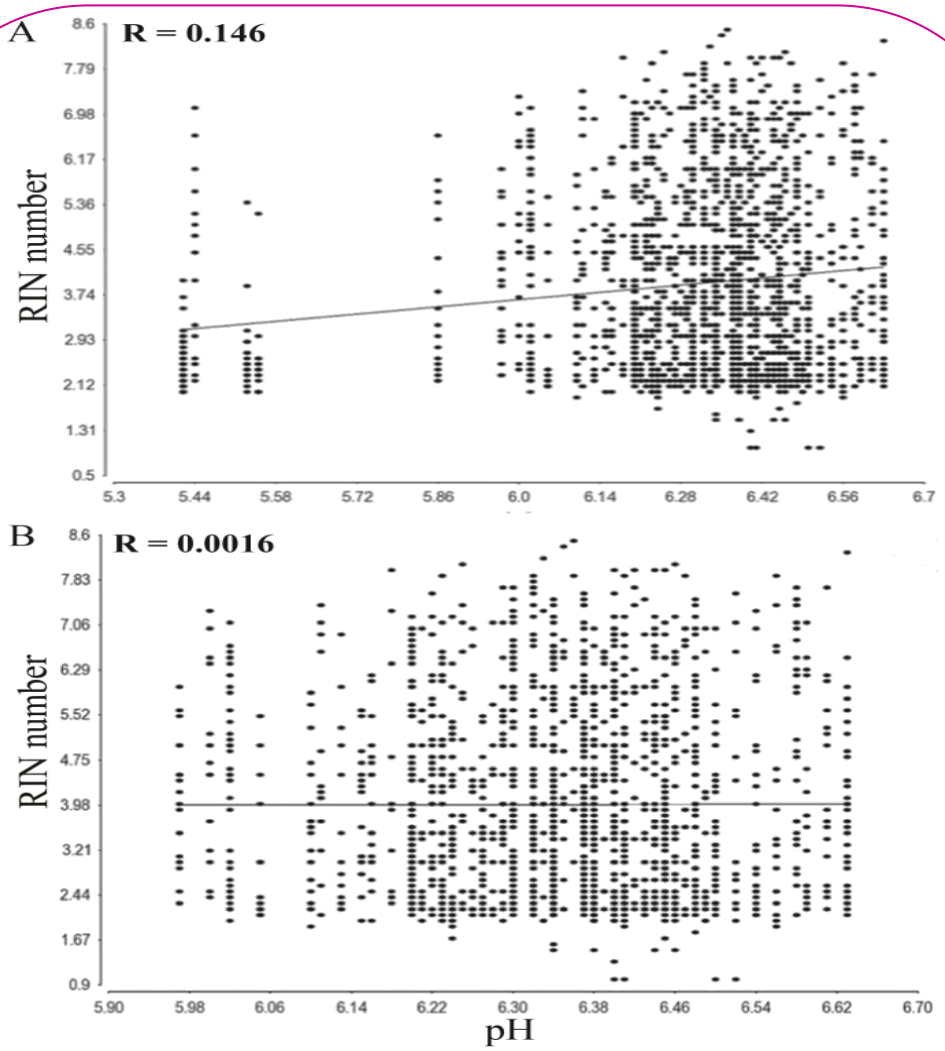

Figure 3.3.2:

The relationship between RIN and pH.

(A) Scatter plot for total RNA samples from 13 region of control brain tissue with linear regression line of RIN numbers for pH where samples with pH < 5.9 were included. A significant correlation between PIN and pH was obtained ($P = 1.0\times10^{-4}$, $r = 0.146$. **(B)** Scatter plot as in (A) but where samples with pH < 5.9 were excluded for total RNA samples with linear regression line of RIN numbers for pH where samples with pH < 5.9 were excluded. No significant correlation was obtained ($r = 0.0016$ and $P = 0.82$) (Trabzuni, Ryten et al. 2011).

3.3.2 Factors affecting present call (%P), cDNA and cRNA profile

The dependence of RNA array performance quality on RIN-based RNA quality and the other covariates available for the analysis were assessed. A systematic quality control check of the arrays was performed using Expression Console™ software. This software produces a number of array quality measures. The most reliable and widely used parameter is present call (%P) (Tomita, Vawter et al. 2004; Weis, Llenos et al. 2007). The present call is the percentage of probe sets with signal detection above background (DABG) probe level *P*-value ≤ 0.01. The range of %P in this study was 1.7% to 77.3% (mean of 61.4%). 36% of the variation in %P was explained by individual-level differences (adjusted R^2 measure).

For 1,266 RNA samples, among all the factors, the most important covariates were brain region (explained 12.4% of the variation in %P, $P = 3.6\times10^{-53}$), followed by brain bank (explained an additional 4.7%, $P = 9.6\times10^{-6}$), and then RIN (explained an additional 2.7%, $P = 9.1\times10^{-5}$). Age, gender, and cause of death together explained only an additional 2.2% of the variation. The effect of brain region on %P was most obvious when comparing CRBL and WHMT. CRBL showed the highest %P (mean = 68%) while WHMT showed the lowest %P (mean = 57.5%) (**Figure 3.3.3**).

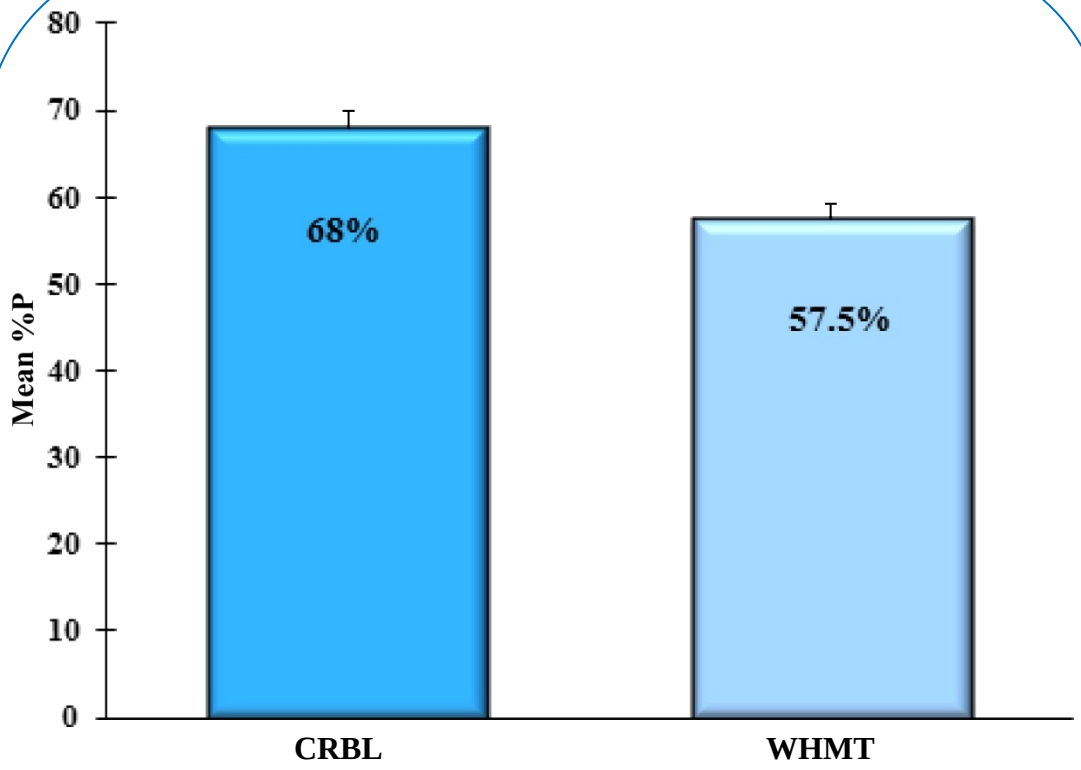

Figure 3.3.3:

The variation in %P by brain region. CRBL = cerebellum, WHMT = white matter.

This graph shows the different performance of samples from specific brain regions on the array. The results from two-sample *t-test* are highly significant $P = 1.3\times10^{-24}$. The heights of the bars represent the mean. The error bars represent the S.E.M. (Trabzuni, Ryten et al. 2011).

For the same reasons described previously pH and PMI were analyzed separately. pH as a factor explained 12.0% of the variation in %P from the MRC-UK dataset (984 samples) ($P = 2.3\times10^{-9}$) (**Figure 3.3.4**). However as with RIN, this correlation was highly dependent on the six individuals with low pH (< 5.90). When these samples were excluded from the analysis no significant correlation was obtained.

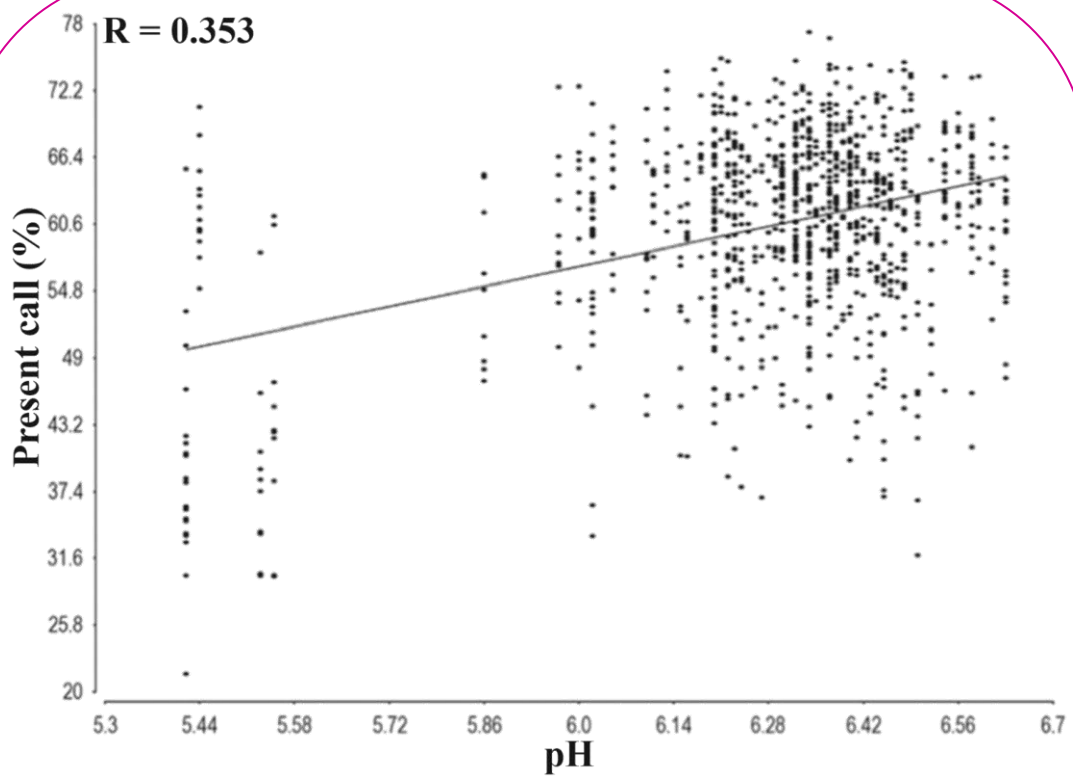

Figure 3.3.4:

The linear regression of Present call (%P) for pH.

Scatter plot shows that pH explains 12.0% of the variation in %P ($P = 2.3\times10^{-9}$, $r = 0.353$), including samples with low pH values. Low pH values are driving the regression analysis (Trabzuni, Ryten et al. 2011).

Finally, no significant correlation was found between PMI and %P, either for the MRC-UK dataset (984 samples) or the SHRI-USA dataset (282 samples). Moreover, no difference was observed in the expected nucleotide lengths in cDNA (~ 200 - 400 nt) and cRNA (ranges from 200 to 2000 nt) production in the Agilent profiles when different RNA samples with different RIN values (from 1- 8) were used.

3.3.3 Validation of array results using QuantiGene, a PCR-independent platform

QuantiGene is a technique for mRNA expression quantification using branched DNA (bDNA) technology. This eliminates the need to purify or amplify RNA, so it is a non-PCR based technique, and is therefore not subject to the systematic biases that PCR or reverse transcriptase could create when applied to degraded RNA samples (Hall, Leong et al. 2011). In addition, it is based on the cooperative hybridization of specifically designed probe sets for the targeted mRNA. Lastly, the amplification and the detection are for the signal that represents the target quantity, and not the quantification of the target itself (**Figure 3.3.5**, modified from the Panomics website, 2011). This technique can therefore be used in place of qRT-PCR (Canales, Luo et al. 2006; Hall, Leong et al. 2011).

Each oligonucleotide probe set contains three types of synthetic probes, Capture Extenders (CEs), Label Extenders (LEs), and Blockers (BLs) that hybridize to specific sequences of the target mRNA. CEs are ~ 40 nucleotides long, around half of the CEs' sequence is complementary to the target mRNA, and the other half is complementary to the capture probes that are at the bottom of the Capture Plates. The LEs are also ~ 40 nucleotides long and half of the LEs' sequence is complementary to the target mRNA, and the other half is complementary to the branched DNA Amplifiers. Finally, the BLs are used to hybridize across the target mRNA sequence

that are not covered by CEs and LEs. These three probes hybridize contiguously to a region of 300 to 500 nucleotides, creating a region of a double-stranded RNA/DNA hybrid molecule. This double stranded molecule is stable and does not create the secondary structures that single stranded mRNAs can form.

The exon array results in a subset of 12 individuals were validated by comparing the normalized mRNA expression level for three genes: *LRRK2* (Leucine-rich repeat kinase 2), *SCN8A* (Sodium channel, voltage gated, type VIII, alpha subunit) and *MAPT* (microtubule-associated protein tau), in four brain regions (CRBL, OCTX, PUTM, and WHMT). Similar regional patterns of expression were observed across the two platforms for both high (*SCN8A*, *MAPT*) (**Figure 3.3.6**) and low expression transcripts in brain (*LRRK2*) (**Figure 3.3.7**).

Furthermore, there was a very high correlation in fold change ($r = 0.92$) at both the gene and exon expression levels and signal intensity values ($r = 0.95$) in the four brain regions that were studied by both techniques.

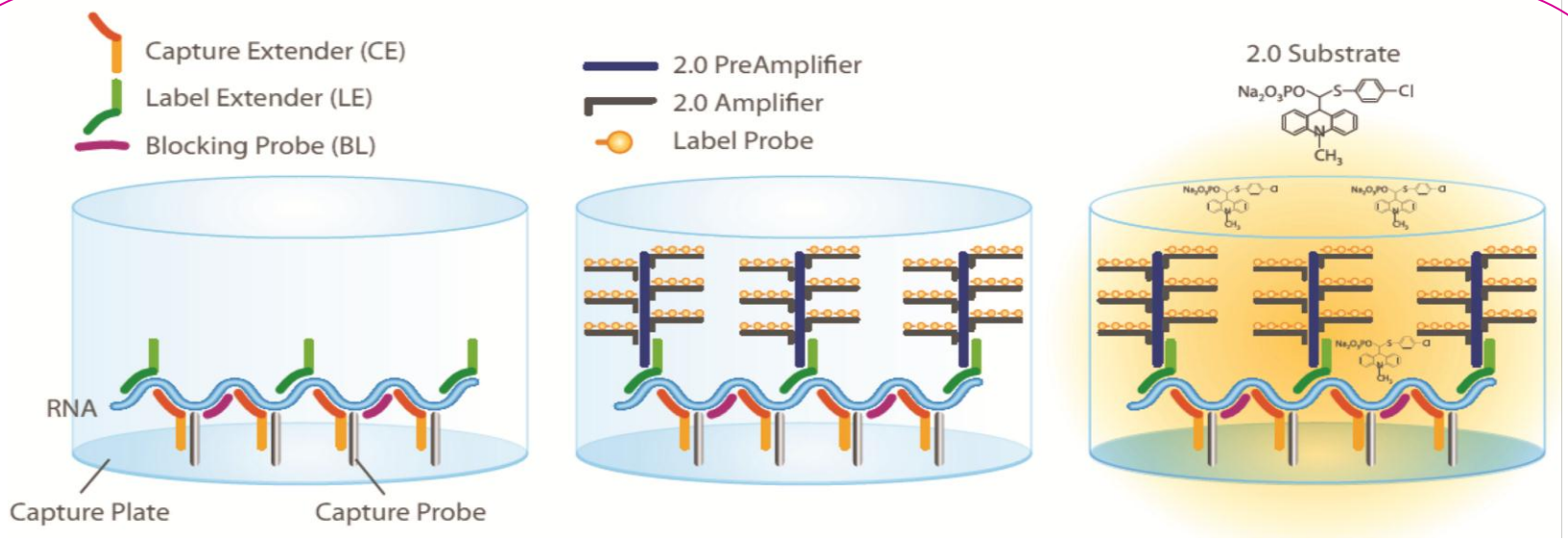

Figure 3.3.5:

An overview of the QuantiGene Technology.

This technology based on the hybridization of the target RNA and specific probe set, then the signal amplification with a luminescent readout. **(A)** Probe Set design to capture the target RNA. Probe Set oligonucleotides (CEs, LEs, BLs) bind a contiguous region of the target RNA and the CEs, selectively capture target RNA during an overnight incubation by cooperative hybridization**. (B)** Signal amplification is performed via sequential hybridization of 2.0 Pre-Amplifier to LEs, 2.0 Amplifier to the 2.0 Pre-Amplifier and Label Probe to the 2.0 Amplifier. The number of LEs determines assay sensitivity. **(C)** Addition of a chemilumigenic 2.0 Substrate* generates a luminescent signal that is proportional to the amount of target mRNA present. The figure was modified from Affymetrix website, Panomics Quantitative Assays.

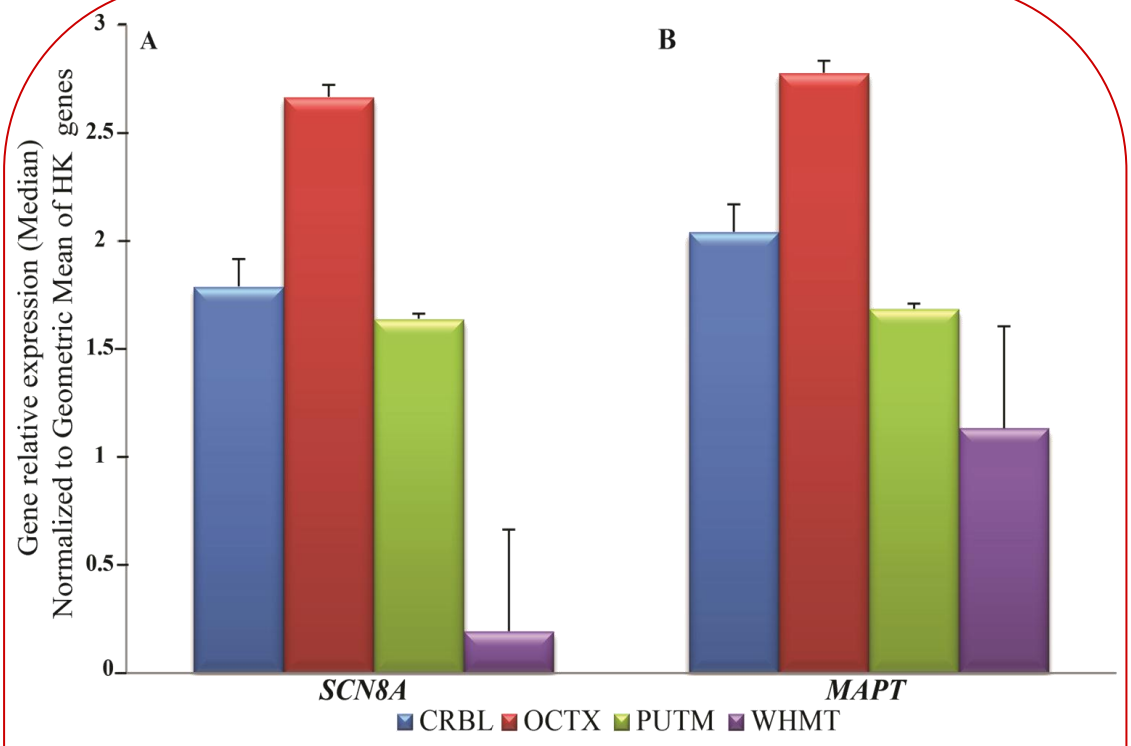

Figure 3.3.6:

QuantiGene validation of microarray expression results for *SCN8A and MAPT* mRNA expression level in four brain regions. CRBL = cerebellum, OCTX = occipital cortex, PUTM = putamen, WHMT = white matter.

(A) *SCN8A* mRNA expression level between different brain regions. The graph shows high expression in OCTX compared with other regions. It is clear that this gene in mostly not expressed in the WHMT region (expression level close to zero) and also has a large standard error. **(B)** *MAPT*, showing higher expression in OCTX compared with other brain regions. These results confirm the array results with significant *P*-values <0.01. The expression level is presented as the mean and the error bars is S.E.M. (Trabzuni, Ryten et al. 2011)

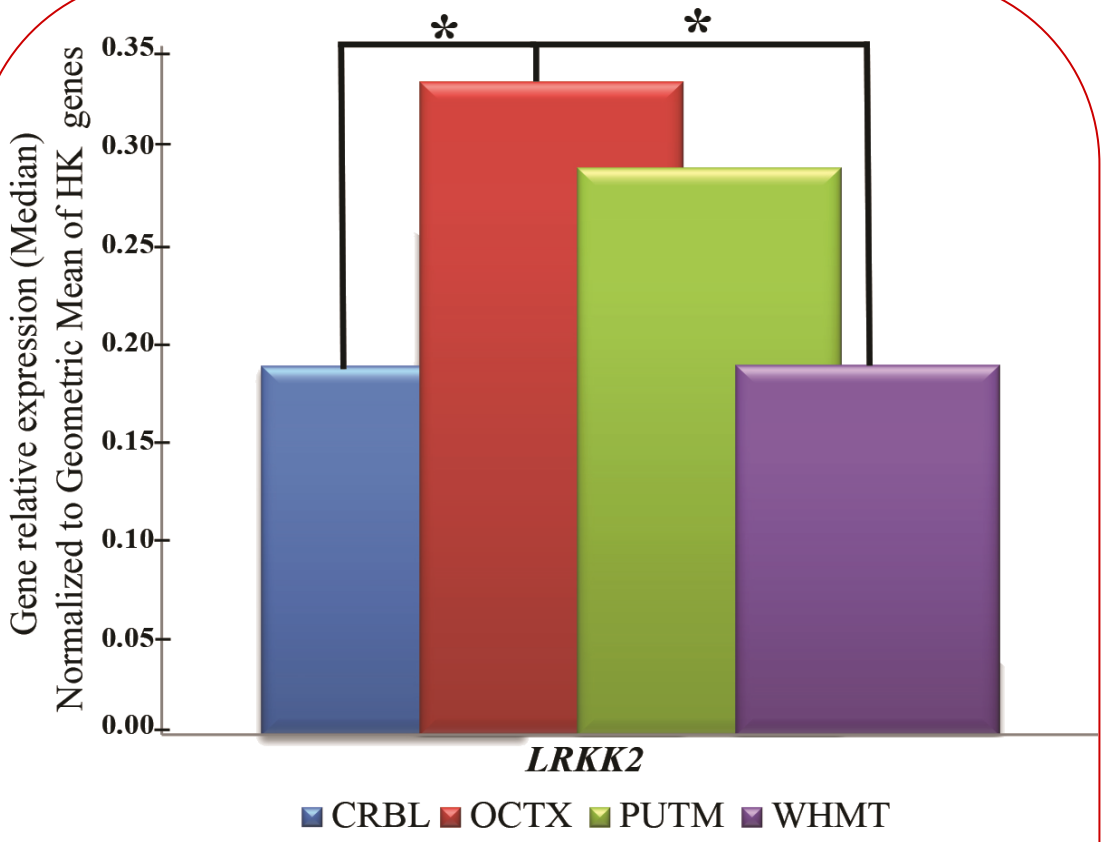

Figure 3.3.7:

QuantiGene validation of microarray expression results for *LRRK2* mRNA expression level in four brain regions. CRBL = cerebellum, OCTX = occipital cortex, PUTM = putamen, WHMT = white matter.

This graph shows the expression of this gene at high level in OCTX compared with other brain regions. Wilcoxon signed rank test was performed and these results confirm the difference of the expression between brain regions in array data is significant. The expression level is presented as the median and the stars indicating the significant difference in expression with *P*-values < 0.01. In this case, as *LRRK2* expression level was very low, it was unreliable to use S.E.M. and Wilcoxon test was performed using the median values (Trabzuni, Ryten et al. 2011).

3.4 Discussion

This study reports quality control data on the largest human control brain exon microarray data set generated to date. This data set was based on the analysis of tissue samples from 13 different CNS regions originating from 137 individuals and containing 2,318 processed RNA samples. It is noteworthy that there are a number of previous studies of gene expression in control human post-mortem brains, however these used a different platform that generated only gene-level signals, a maximum of four brain regions were dissected and a smaller sample size was tested (Mistry and Pavlidis 2010).

The results showed considerable variation in RIN values among RNA samples. 67% of the variation resides in differences among extractions from the same tissue blocks, and most of the remaining variation is unexplained by the available covariate information. Brain pH is the most important post-mortem factor influencing RIN-based RNA integrity, a result consistent with previous studies (Hardy, Wester et al. 1985; Mexal, Berger et al. 2006; Chevyreva, Faull et al. 2008; Monoranu, Apfelbacher et al. 2009; Durrenberger, Fernando et al. 2010). Samples with very low pH values (ranging from 5.42 – 5.90) were responsible for the positive correlation seen between pH and %P (and also RIN). However, when these low pH samples were removed from the analysis no significant correlation was observed. This may in part explain contradictory observations regarding the effect of pH on RNA integrity and sample performance on arrays (Hardy, Wester et al. 1985; Monoranu, Apfelbacher et al. 2009; Birdsill, Walker et al. 2011) and confirms the findings of a recent smaller study on control brain tissues (Sherwood, Head et al. 2011).

The array-based expression data were validated by the QuantiGene PCR-independent method, when tested on two highly expressed genes (*MAPT* & *SCN8A*) and one low expressed gene (*LRRK2*). However, the performance quality of the array, as defined by %P, was not profoundly affected by age, gender, region, PMI, RIN or cause of death. This confirms findings from previous studies on much smaller sample sizes (Tomita, Vawter et al. 2004; Durrenberger, Fernando et al. 2010; Birdsill, Walker et al. 2011). Only 2.7% of the variation in %P was explained by RIN. Indeed, 80 RNA samples with undetectable RINs performed well on the arrays with %P values ranging from 45% to 76%. Thus RIN was found to be a poor predictor of array quality performance even at the low end of the RIN scale. Furthermore, the latter was confirmed since the cDNA and cRNA nucleotide length synthesis was not affected by the wide range of RIN-based RNA quality values (from 2 to 7) in the current study. Furthermore, it is important to note that the values for the RIN that were reported in this study are much lower than those reported for comparable samples in other laboratories. This is most likely due to differences in the RNA isolation kits that were used. Most RNA kits, including the RNAeasy kit, encourage the loss of small or degraded RNA species. For this project the miRNAeasy kit was used, which specifically allows for the conservation of short microRNA species. When short RNA is present, the RIN algorithm generates lower RIN values. As a result RIN values generated here are not directly comparable with those reported in other studies where a different RNA isolation kit was used. An example of such a study was carried out on the Sun Health brain bank (Beach, Sue et al. 2008).

The robust performance of the Affymetrix Exon arrays using degraded RNA samples was as a result of the recent modifications to the RNA amplification Ambion® WT Expression kit. This kit uses both non-polyA and polyA-based mRNA

priming for first strand cDNA synthesis. This meant that RNA amplification did not need an intact polyA tail. In addition, increasing the quantity of the starting material of RNA from 500 ng to 750 ng during the troubleshooting of the array experiments for this project the array's performance was improved.

A number of limiting factors may have impacted on the outcomes of this study. For example the cause of death is an imperfect reflection of the true medical and drug treatment history of the individual, and that access to that history, was it available, would reveal other factors of greater relevance. In this study the cause of death only explained 1.9% of variation in RIN and there was no significant relationship with %P. Likewise, in the range of 28-114 hours there was no effect of PMI on either RIN or %P, nor a loss of RIN-based RNA quality over the 1-5 hour range was observed. It remains possible that there may be selective loss of RNA within an hour since the half-life of some mRNA species has been reported to be as short as 15 minutes, while others may be as long as 22 days depending on the tissue type and storage conditions (Ross 1995; Barrachina, Castano et al. 2006; Bahar, Monahan et al. 2007; Beach, Sue et al. 2008; Vennemann and Koppelkamm 2010). This issue has been studied in detail for a range of PMIs by Harrison et al. (1995) and confirmed by a more recent study by Tomita et al. (2010). The same conclusion was made that a limited effect of PMI was observed on mRNA (Harrison, Heath et al. 1995; Tomita, Vawter et al. 2004; Kang, Kawasawa et al. 2011).

Furthermore, there is a possibility that samples with the shortest PMIs (1-2.5 hours) within the SHRI-USA sample set may originate from those individuals who suffered longer agonal states prior to death than individuals with longest PMIs (3-5.5 hours) and agonal stress has been shown to affect gene expression differently in

different brain regions (Li, Meng et al. 2007). Other factors may contribute to this such as intermittent edge effects, especially on small samples during dissecting and tissue handling procedures.

Finally, the brain samples were derived from two sources: one was a rapid death brain bank with long post-mortem intervals and the other specialises in obtaining very short post-mortem intervals. Both brain banks had separately optimised their collection and dissection protocols to facilitate gene expression studies and this may limit the generalisability of these conclusions.

These results are important for several reasons. First, they confirm the practical viability of using post-mortem control brain tissue to study the transcriptome of the human brain by array technology and RNA sequencing (Franz, Ullmann et al. 2005). Secondly, they show that microarrays can give reliable results over a wide range of RIN (1 - 8.5) and pH measurements with a drop off in array validity only being observed below brain pH 5.9. Thirdly, they show that the results from Affymetrix exon arrays are reproducible by other technologies, making it possible for database users to analyse the data generated with confidence.

Furthermore, this study is the first step of an on-going multi-regional control human brain expression project that has been established to build an open-access database of identified genome-wide genetic variability in relation with gene expression and splicing QTLs as well as for detailed expression analysis (Hardy, Trabzuni et al. 2009). This will move the field forward in our understanding of the underlying molecular mechanisms of complex neurologic diseases, and will support the neuroscience community with a resource which will bring functional insights.

The results of the work described in this chapter were published in the *Journal of Neurochemistry:*

Trabzuni, D., Ryten, M., Walker, R., Smith, C., Imran, S., Ramasamy, A., Weale, M.E. and Hardy, J. (2011) Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *Journal of Neurochemistry*, **119**, 275-282.

4 Regional specific transcriptome and expression quantitative trait loci (QTL) analyses of UKBEC dataset

4.1 Summary

Different levels of complexity in the structure of the human brain are demonstrated and driven by the variability of the structure of different cell types, distinct gene expression profiles, various transcripts/exon splicing patterns and complex functional connections in different anatomical brain regions. Despite major advances in genomics and transcriptomics, it is still challenging to achieve a comprehensive understanding of the genetic regulation, functional mechanisms and transcriptome mapping of the adult control and diseased human brain.

In order to expand our understanding of the human brain transcriptome and its regulation, association analyses of the QTLs with transcript and exon expression profiles were performed on 1231 exon arrays that were generated from ten brain regions in 134 individuals. In parallel, the genome-wide analyses of regional expression profiles of transcripts and exons levels were performed on 917 exon arrays. These arrays were generated from a subset of samples, which represented 101 individuals from twelve brain regions including cortical areas, and these were obtained from the UKBEC.

The results showed that 75% of the genes were expressed, and 80% of these genes were differentially regulated at the whole-transcript or exon level across CNS regions. 9% of expressed genes are specifically enriched in the CRBL region. An alternative splicing pattern was the dominant pattern in the cortical regions. Regional isoforms were identified to have a functional effect. Brain regional expression QTL signals were reported genome-wide.

This study provides comprehensive information and insights into the transcriptional architecture of the adult human CNS in different anatomical regions, facilitating genomic/transcriptomic correlation investigations. It provides a valuable and reliable resource for future focused cell biology and functional studies.

4.2 Introduction

The difficulty and complexity of studying the brain transcriptome architecture arises from the nature of the human brain as it is a heterogeneous structure containing different cell types at different ratios in different anatomical regions (Azevedo, Carvalho et al. 2009). In addition, the vulnerability of different brain anatomical regions and severity of pathology in different diseases contributes an additional level of complication to understanding the aetiology of these diseases.

Profiling the expression and splicing patterns will provide comprehensive information for different CNS regions with different cell types in relation to neurological diseases. Previously it was reported that the variability in the transcription profiles of the human CNS can lead to different functional features (Enard, Khaitovich et al. 2002). In addition, different mRNA isoforms have different structures and opposing functions. These alternative splicing patterns can promote the progress of human diseases (Wang, Sandberg et al. 2008). This demonstrates the importance of having complete coverage of the human CNS.

The transcriptome patterns in the different regions and cells of the CNS has been attempted by other groups to expand the understanding of the transcriptomic architecture (Roth, Hevezi et al. 2006; Johnson, Kawasawa et al. 2009; Kang, Kawasawa et al. 2011; Hawrylycz, Lein et al. 2012). The Roth *et al.* study showed differential expression patterns between 20 CNS regions (Roth, Hevezi et al. 2006). Region specific changes were also apparent in a study by Hawrylycz *et al.* in addition to revealing differences at the cellular level (Hawrylycz, Lein et al. 2012). Although these studies provide insight into connectivity and functional regulation in the human CNS, they are limited by the sample size and alternative splicing detection technology

that was employed. Furthermore, integrating the transcriptomic data with the genomic data was utilized to map expression quantitative trait loci in the control human brain (Myers, Gibbs et al. 2007; Heinzen, Ge et al. 2008; Gibbs, van der Brug et al. 2010; Kang, Kawasawa et al. 2011). Expression QTL signals associated with whole transcript and exon levels were reported and tissue specific signals revealed the unique expression and splicing patterns in different regions (Heinzen, Ge et al. 2008). The mapping of expression QTLs was shown to be a successful approach for exploring the effect of genetic control on transcriptional regulation in the human CNS, in order to understand the molecular mechanism of this effect on gene expression and splicing patterns.

Investigating and understanding the regional expression and splicing patterns in normal human CNS tissues are of necessity in order to make progress in this field. To date, all mRNA expression databases are lacking in two important ways. Firstly, they are not based on a large scale cohort due to brain tissue availability, and secondly, there are no comprehensive investigations of different regions representing the exon level as well as the gene level. In order to address this challenge, data on 917 samples from twelve CNS regions were included in this analysis. The regions and the sample sizes were as follows: FCTX, n = 96, TCTX, n = 84, OCTX, n = 95, HIPP, n = 93, THAL, n = 83, HYPO, n = 13, PUTM, n = 94, SNIG, n = 68, MEDU, n = 92, CRBL, n = 95, SPCO, n = 13 and WHMT, n = 91. This part of the data set was used for regional differential expression and splicing analysis in 12 regions. In addition, the expression QTL analysis was performed genome-wide using the entire UKBEC data set of 134 individuals in ten regions. The number of samples in each brain region were FCTX, n = 127, TCTX, n = 119, OCTX, n = 129, HIPP, n = 122, THAL, n = 124, PUTM, n = 129, SNIG, n = 101, MEDU, n = 119, CRBL, n = 130 and WHMT, n =

131 (1,231 Affymetrix exon arrays) (Trabzuni, Ryten et al. 2011). This analysis was performed to translate the GWAS results to a functional level and to identify more functional loci that are involved and related to neurological diseases.

4.3 Results

4.3.1 Detected transcripts in different CNS regions using detection above background (DABG) method

According to the Affyemtrix Exon arrays comprehensive design, there are a total of around 24,164 estimated genes on the array (~26,493 transcripts) that were detected and included in the analysis after considering the Affymetrix probe sets re-mapping as described in detail in Chapter 2, Section 2.6.1. After applying the DABG filtering (P values ≤ 0.001) 19,215 transcripts (~73% of ~26,493 transcripts) were defined as detected transcripts in at least one CNS region in at least 50% of the samples as explained in detail in Chapter 2, Section 2.6.2 (Clark, Schweitzer et al. 2007). This list contained 18,107 detected genes (~75% of 24,164 genes) and was generated to for use in the further analyses.

Results showed that across all CNS regions out of 24,164 genes the percentage of detected genes ranged from 57% (13,834 genes) in WHMT to 71% (17,259 genes) in CRBL (**Table 4.3.1**). Furthermore, out of the 18,107 detected genes, 71% (12,790) were detected in all twelve CNS regions and 9% (1,666) were detected only in one region, mainly CRBL (**Table 4.3.2**). More precisely, out of the 1,666 genes, 87% (1,456 genes) were specifically detected only in CRBL while in other CNS regions the number of enriched genes was lower (**Table 4.3.3**). These results were in agreement with previous studies that were performed using microarray analyses on human post-mortem brains on a smaller sample size, different platforms and different CNS regions. These studies showed that gene expression patterns in human cerebellum were distinctive compared with other CNS regions such as cortical regions (Khaitovich,

Muetzel et al. 2004; Roth, Hevezi et al. 2006; Johnson, Kawasawa et al. 2009; Gibbs, van der Brug et al. 2010).

Table 4.3.1: The detected probes sets, transcripts and genes in each CNS region.

CNS regions	Total	FCTX	TCTX	OCTX	HIPP	THAL	HYPO	PUTM	SNIG	MEDU	CRBL	SPCO	WHMT
Number of samples	917	96	84	95	93	83	13	94	68	92	95	13	91
Number of detected probe sets	NA	134,002	139,521	132,396	126,446	126,727	126,392	128,862	115,029	119,000	149,762	112,501	110,751
Number of detected transcripts	NA	16,034	16,379	16,184	15,437	15,259	15,475	15,711	14,614	15,085	18,152	14,878	14,262
Number of detected genes	NA	15,294	15,615	15,496	14,817	14,652	14,882	15,057	14,095	14,544	17,259	14,400	13,834
Percentage of detected genes of 24,164 genes	75%	63%	65%	64%	61%	61%	62%	62%	58%	60%	**71%**	60%	**57%**
Percentage of detected genes of 18,107 genes	NA	84%	86%	86%	82%	81%	82%	83%	78%	80%	**95%**	80%	**76%**

This table shows the estimated numbers and percentages for the detected transcripts and genes that were used in the regional expression analysis across twelve CNS regions. CRBL represents the region that has the highest number of detected genes, while WHMT shows the lowest numbers of detected genes compared to the other regions.

Table 4.3.2: The detected genes in one or more CNS regions.

Number of CNS regions	Number of detected genes	Percentage of detected genes
1	**1,666**	**9%**
2	453	2.5%
3	365	2.0%
4	332	1.8%
5	302	1.7%
6	268	1.5%
7	269	1.5%
8	279	1.5%
9	296	1.6%
10	384	2.1%
11	703	3.9%
12	**12,790**	**71%**
Total	**18107**	**100%**

This table shows the estimated numbers and percentages for the detected genes that were specifically detected in one CNS region up to twelve CNS regions.

Table 4.3.3: Specifically detected genes in a named CNS region.

CNS regions	Number of detected genes	Percentage of detected genes
FCTX	5	0.30%
MEDU	7	0.42%
HIPP	8	0.48%
OCTX	10	0.60%
SNIG	11	0.66%
THAL	18	1.0%
HYPO	20	1.2%
TCTX	21	1.2%
SPCO	28	1.68%
WHMT	38	2.28%
PUTM	44	2.66%
CRBL	**1,456**	**87%**
Total	1,666	100%

This table shows the estimated numbers and percentages for the detected genes that were detected in a specific CNS region.

In the study by Roth and colleagues, tissue specific genes were named and validated. For example they reported that *CRTM* (Class I MHC-restricted T cell-associated molecule), *GABRP* (gamma-aminobutyric acid (GABA) A receptor, subunit pi), *NRK* (Nik related kinase), *PDZK1* (PDZ domain containing 1) and *TPM4* (Tropomyosin 4) as being specifically expressed in CRBL. In addition, they reported *AGTR1* (Engiotensin II receptor, type 1), *EN1* (Engrailed homeobox 1) and *TH* (Tyrosine hydroxylase) to be specific for the SNIG region (Simon, Saueressig et al. 2001; Roth, Hevezi et al. 2006; Johnson, Kawasawa et al. 2009). These results were confirmed in this study. However, there are some differences in the gene lists provided by these studies and this may be due to different sample sizes, dissection protocols, expression techniques, platforms, quality control parameters and analytic approaches. In the study by Simon *et al.* they used RNA in situ hybridization to detect the mRNA expression levels of targeted genes in the mouse brain tissue (Simon, Saueressig et al. 2001). In contrast, the Roth *et al.* study identified the transcriptional profiling of 20 CNS regions in ten human post mortem tissues, using the GeneChip Human Genome U133 Plus 2.0 Array that targets the 3' end of transcripts (Roth, Hevezi et al. 2006). The study of Johnson *et al.* was performed on four human brains (from fatal injuries) dissected from 13 different regions. They used the same Human Exon 1.0 ST Array that was used in this study that is designed to cover the coding sequences genome-wide (Johnson, Kawasawa et al. 2009). These studies, in addition to this study complement each other by providing a complete and accurate human transcriptome map.

4.3.2 Examples of regional specific expression genes

One of the specifically expressed genes in CRBL is *IGHMBP2* the (immunoglobulin helicase μ- binding protein 2) gene. The expression results showed high expression in CRBL and very low expression in the SPCO (close to background signals) (**Figure 4.3.1**). Mutations in this gene are reported to be linked to spinal muscular atrophy with respiratory distress type 1 (SMARD1) (Grohmann, Schuelke et al. 2001). Although histological examination of the spinal cord was normal in two cases in the study of Pitt *et al.* (Pitt, Houlden et al. 2003), spinal muscular atrophy (SMA) is the most severe clinical manifestation. There is a possibility that this gene was not detected in SPCO in this study, because the RNA samples were extracted from the cell bodies and not be from the nerve axons or synaptic regions. However, having this gene highly expressed in CRBL might suggest that clinicians should perform brain scan in SMARD1 patients and look for CRBL pathology.

One other gene that was detected only in CRBL is *CBLN3* (cerebellin 3 precursor) (**Figure 4.3.2**). This gene is involved in synaptic functions in the human CNS and it is a CRBL specific gene/protein (Mizuno, Takahashi et al. 1995).

In addition, two genes were detected specifically in SNIG. The first is *SLC18A2* (solute carrier family 18 (vesicular monoamine), member 2) which is the dopamine vesicular transporter (Roghani, Welch et al. 1996). The second gene is *SLC6A3* (solute carrier family 6 (neurotransmitter transporter, dopamine), member 3) which is a member of the family of sodium- and chloride- dependent neurotransmitter transporters in the CNS (Schultz 1998). These two genes were expected to be specifically expressed in SNIG as this reflects their main functions in the human CNS.

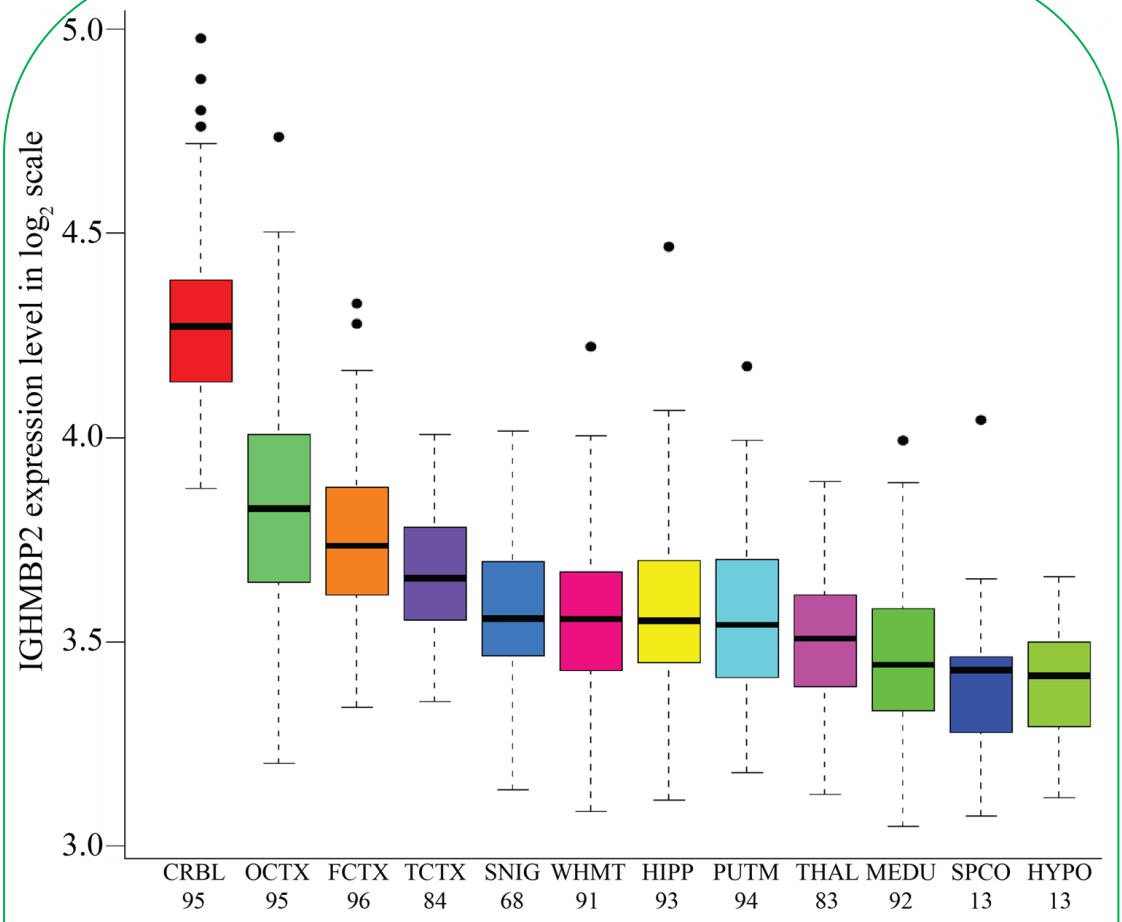

Figure 4.3.1:

Regional Distribution of *IGHMBP2* mRNA expression. CRBL = cerebellum, OCTX = temporal cortex, FCTX = frontal cortex, TCTX = temporal cortex, SNIG = *substantia nigra*, WHMT = white matter, HIPP = hippocampus, PUTM = putamen, THAL = thalamus, MEDU = medulla, SPCO = spinal cord, HYPO = hypothalamus.

Boxplot of mRNA expression levels for *IGHMBP2* in 12 brain regions, from microarray experiments on a log2 scale (*y*-axis). This plot shows the enrichment of *IGHMBP2* transcript expression in CRBL. Whiskers extend from the box to 1.5 times the inter-quartile range.

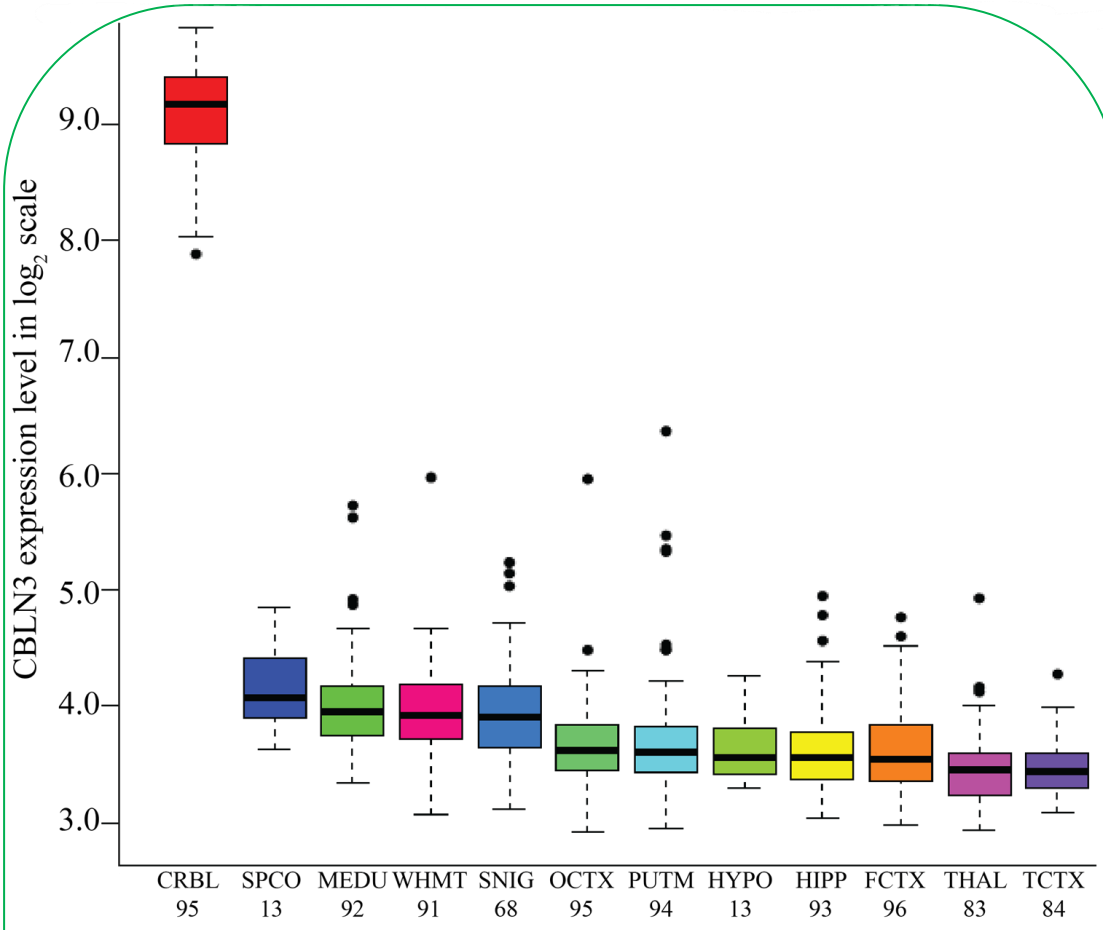

Figure 4.3.2:

Enrichment of *CBLN3* mRNA expression in CRBL. CRBL = cerebellum, SPCO = spinal cord, MEDU = medulla, WHMT = white matter, SNIG = *substantia nigra*, OCTX = temporal cortex, PUTM = putamen, HYPO = hypothalamus, HIPP = hippocampus, FCTX = frontal cortex, THAL = thalamus, TCTX = temporal cortex.

Boxplot of mRNA expression levels for *CBLN3* in 12 CNS regions, from microarray experiments on a log2 scale (*y*-axis). This plot shows the detection of *CBLN3* transcript expression only in CRBL. Whiskers extend from the box to 1.5 times the inter-quartile range.

4.3.3 Human CNS regional differential gene expression

In order to expand the available data and knowledge about the different human CNS expression profiles, the following analysis was conducted. From the genes that were detected in all 12 CNS regions, differentially expressed genes were identified by applying the statistical threshold of a minimum fold change (FC) of ≥ 1.7 and FDR < 0.05 to the genes list that contained 12,790 genes as explained in detail in Chapter 2, Section 2.6.3.

The results showed that 96% of genes (12,376 of 12,790 genes) were differentially expressed across the 12 CNS regions at FDR < 0.05.

In order to provide an overview of the differences in expression patterns between the 12 regions, principal component analysis (PCA) was performed by using all expressions profiles at transcript and exon levels in all regions (**Figure 4.3.3**). This analysis demonstrates that CRBL as a region has a distinctive expression pattern on the PC2 axis compared to other regions; this was expected, as this confirms the results from previous and smaller studies on human and mouse brain tissues (Haverty, Weng et al. 2002; Khaitovich, Muetzel et al. 2004; Roth, Hevezi et al. 2006; Gibbs, van der Brug et al. 2010). In addition, WHMT was the second region after CRBL to show a distinct expression pattern from the other 11 CNS regions (**Figure 4.3.3**, dark purple spheres). Other CNS regions displayed a greater overlap in expression patterns, this was especially true of the cortical regions including FCTX, TCTX and OCTX. These regions will be discussed in greater detail later in this section.

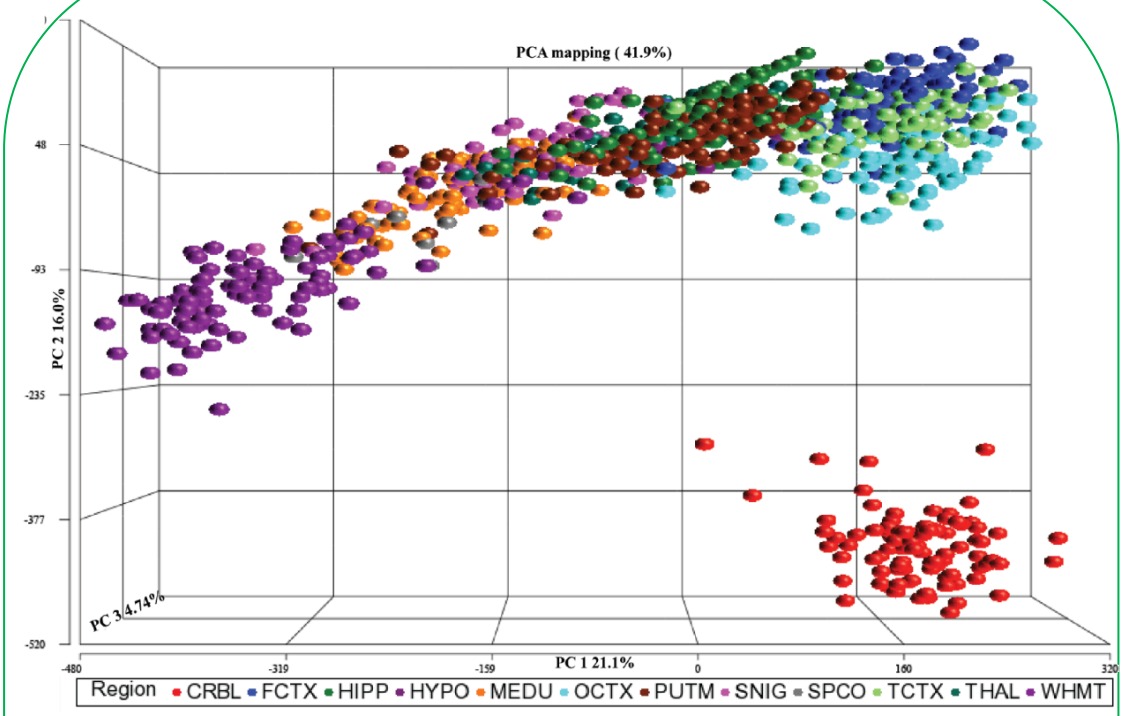

Figure 4.3.3:

Principal component analysis (PCA) of 12 CNS regions. CRBL = cerebellum, FCTX = frontal cortex, HIPP = hippocampus, HYPO = hypothalamus, MEDU = medulla, OCTX = occipital cortex, PUTM = putamen, SNIG = *substantia nigra*, SPCO = spinal cord, TCTX = temporal cortex, THAL = thalamus and WHMT = white matter.

This figure shows a three-dimensional scatterplot of the first three principal component scores obtained from a PCA on expression data of all CNS-region samples. The first three components explain 41.9% of the variability in expression values across all transcripts. Each point represents a single CNS sample from a specific region. Regions have been colour coded, to demonstrate regional differences in expression values. CRBL region (red spheres) cluster separately from the other 11 regions indicating that the expression pattern in this region is unique. In addition, WHMT region (purple spheres) also shows a separate cluster demonstrating different expression patterns than other regions.

Furthermore, focusing on the cortical regions; PCA plots showed a separate clustering for the OCTX region from FCTX and TCTX regions (**Figure 4.3.4**) supporting the results in Table 4.3.4 in the following section.

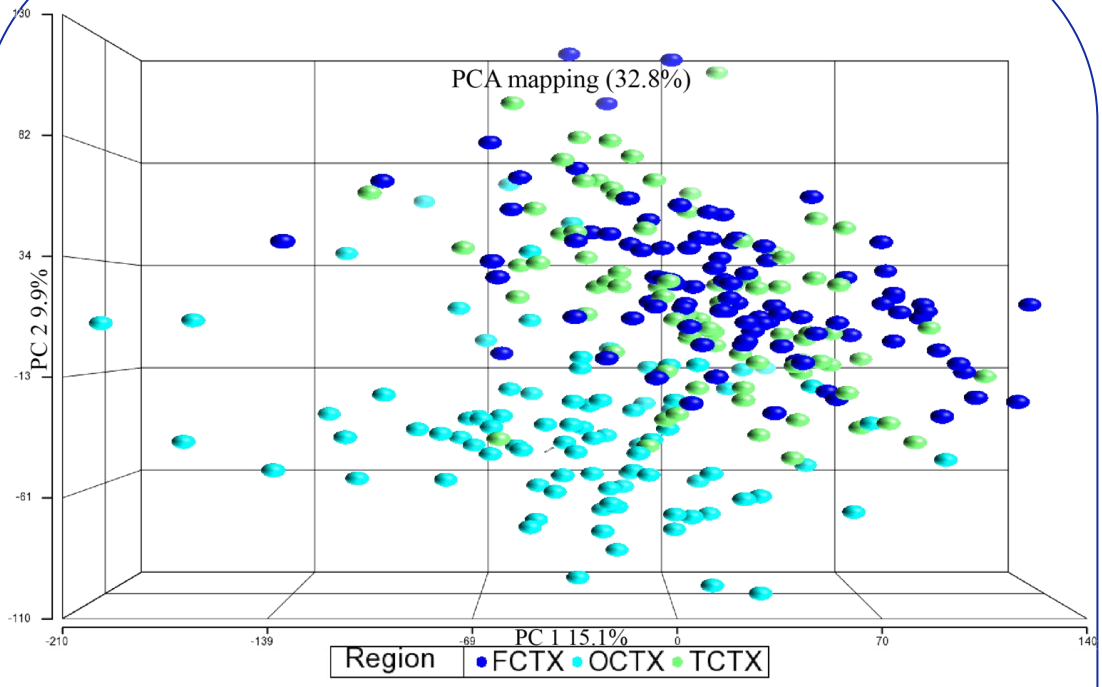

Figure 4.3.4:

Principal component analysis (PCA) of cortical regions. FCTX = frontal cortex, OCTX = occipital cortex and TCTX = temporal cortex.

This figure shows a three-dimensional scatterplot of the first three principal component scores obtained from a PCA on expression data of FCTX, TCTX and OCTX region samples. The first three components explain 32.8% of the variability in expression values across all transcripts. Each point represents a single sample from a specific region. Regions have been colour coded, to demonstrate regional differences in expression values. OCTX region is clustering separately from FCTX and TCTX regions, indicating that the expression pattern is more distinct in this region.

To identify and investigate the expression profiles in more detail, a comparison between each possible pair of regions across the 12 CNS regions was performed and Table 4.3.4 was generated (**Table 4.3.4**). This table facilitates the identification of the expression patterns in term of high and low expressed genes between regions based on the criteria of fold change (FC) difference of ≥ 1.7 or ≤ -1.7. Furthermore, Table 4.3.5 was generated from calculating the differences between the number of high and low expressed genes in each region to give more detailed information about the dominant patterns in each region (**Table 4.3.5**).

The CRBL has the highest number of differentially expressed genes followed by WHMT compared with each of the other 11 regions, specifically the cortical areas, suggesting that CRBL and WHMT have more distinctive functions and the other regions might have many more functions in common. In CRBL, there were 604 genes having higher expression levels (**Table 4.3.4**, red column) and 638 genes with lower expression levels in comparison with the other 11 CNS regions (**Table 4.3.4**, red row and **Figure 4.3.5**, red circles). In addition, there were 216 genes having higher expression (**Table 4.3.4**, purple column and **Figure 4.3.5**, purple/pink circles) and 271 genes having lower expression in WHMT in comparison to the other 11 regions (**Table 4.3.4**, purple row). Looking at the table for the comparison between CRBL and every other region in detail, it is clear that there are more genes that have lower expression levels in CRBL than those genes with higher expression levels (**Figure 4.3.5**, red circles). However, the reverse is the case for WHMT as there are more high expressed genes than low expressed genes as compared with the other CNS regions (**Figure 4.3.5**, purple/pink circles) with the exception of HYPO, SNIG, MEDU and

SPCO where the dominant pattern was that the number of low expressed genes is more than the number of high expressed genes (**Table 4.3.5**).

Table 4.3.4: Differentially expressed genes across twelve CNS regions.

CNS regions	FCTX	TCTX	OCTX	HIPP	THAL	HYPO	PUTM	SNIG	MEDU	CRBL	SPCO	WHMT	All regions
FCTX	NA	**1**	**39**	427	725	989	710	1,482	1,679	1,663	1,821	2,477	1
TCTX	**2**	NA	**25**	468	806	1,025	772	1,541	1,724	1,555	1,815	2,448	0
OCTX	**83**	**62**	NA	743	961	1,213	992	1,677	1,822	1,395	1,889	2,466	0
HIPP	349	329	462	NA	480	238	374	573	841	1,770	933	1,688	0
THAL	628	623	625	517	NA	204	569	379	671	1,800	833	1,683	0
HYPO	812	775	819	492	351	NA	526	185	414	1,671	382	1,230	0
PUTM	715	675	727	440	600	392	NA	721	984	1,706	1,024	1,642	9
SNIG	1,032	1,034	1,051	691	507	214	720	NA	266	1,905	292	918	0
MEDU	1,153	1,123	1,138	800	623	321	817	224	NA	1,830	81	478	1
CRBL	1,927	1,805	1,620	2,283	2,423	2,376	2,258	2,697	2,741	NA	2,667	2,832	**638**
SPCO	1,340	1,297	1,262	887	844	329	871	369	139	1,855	NA	513	0
WHMT	1,932	1,866	1,852	1,469	1,537	1,246	1,461	1,137	899	2,028	773	NA	**271**
All regions	1	9	5	20	53	15	82	31	9	**604**	10	**216**	NA

The table shows the number of genes for which the region shown in the column has higher expression level compared with the region shown in the row. By implication, the values also show the number of genes for which the region shown in the row has lower expression level compared with the region shown in the column. So the total number of genes which show differential expression between any two regions is the sum of the (column, row) and (row, column) values. The row labelled 'All regions' contains the number of genes for which the region shown in the column has higher expression compared with all 11 other regions. Similarly, the column labelled 'All regions' contains the number of genes for which the region shown in the row has lower expression compared with all 11 other regions. Genes were considered high expressed if the fold change ratio ≥ 1.7 and FDR < 0.05 (see Section 2.6.3).

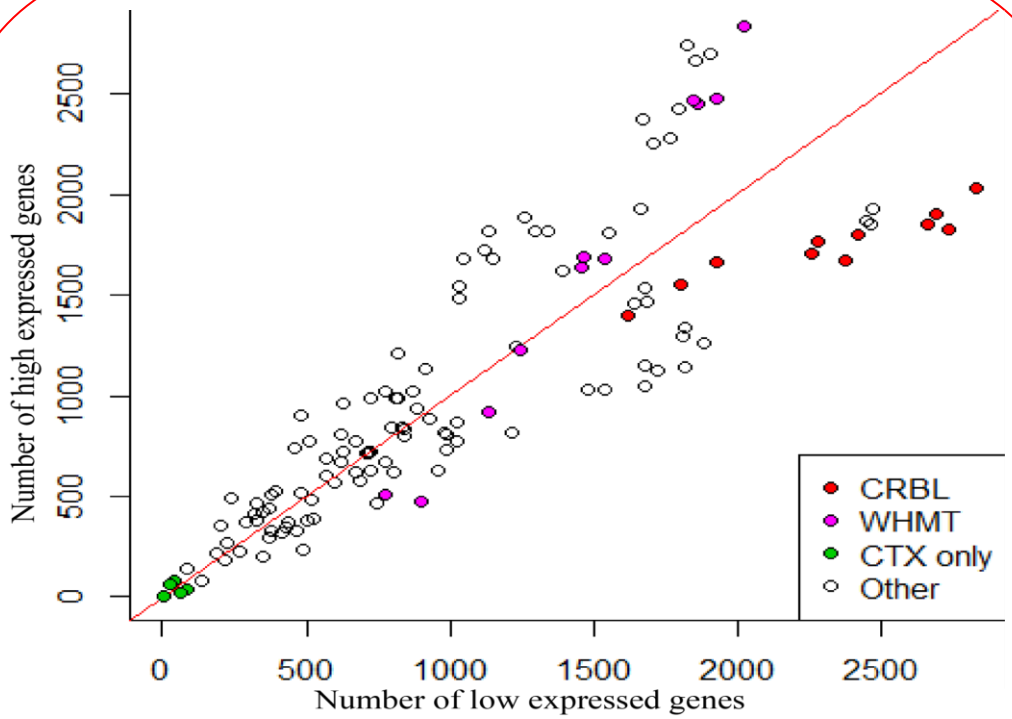

Figure 4.3.5:

Scatter plot of the number of differentially expressed genes of each pair of the 12 CNS regions (Table 4.3.4). CRBL = cerebellum, WHMT = white matter and CTX = cortical regions including frontal, temporal and occipital. Other = the other 7 CNS regions, these regions were not coloured as the focus is on the CRBL, WHMT and CTX regions.

This graph shows the number of high expressed genes (*y*-axis) versus low expressed genes (*x*-axis) for each possible pair comparisons between the 12 CNS regions. The data in this figure were derived from Table 4.3.4. The results are in symmetry in the scatterplot around the 45 degree line (of equal numbers of high and low expressed genes). In CRBL (red circles), there were more low expressed genes than high expressed in comparison with the other 11 CNS regions, while in WHMT (purple/pink circles), there were more high expressed genes than low expressed genes. For Cortical regions (green circles), it shows that the number of the differentially expressed genes between these regions were very small.

Table 4.3.5: The difference between the differentially expressed genes across twelve CNS regions.

CNS regions	FCTX	TCTX	OCTX	HIPP	THAL	HYPO	PUTM	SNIG	MEDU	CRBL	SPCO	WHMT	All regions
FCTX	**NA**	**-1**	**-44**	78	97	177	-5	450	526	-264	481	545	0
TCTX	**1**	**NA**	**-37**	139	183	250	97	507	601	-250	518	582	-9
OCTX	**44**	**37**	**NA**	281	336	394	265	626	684	-225	627	614	-5
HIPP	-78	-139	-281	NA	-37	-254	-66	-118	41	-513	46	219	-20
THAL	-97	-183	-336	37	NA	-147	-31	-128	48	-623	-11	146	-53
HYPO	-177	-250	-394	254	147	NA	134	-29	93	-705	53	-16	-15
PUTM	5	-97	-265	66	31	-134	NA	1	167	-552	153	181	-73
SNIG	-450	-507	-626	118	128	29	-1	NA	42	-792	-77	-219	-31
MEDU	-526	-601	-684	-41	-48	-93	-167	-42	NA	-911	-58	-421	-8
CRBL	264	250	225	513	623	705	552	792	911	NA	812	804	**34**
SPCO	-481	-518	-627	-46	11	-53	-153	77	58	-812	NA	-260	-10
WHMT	-545	-582	-614	-219	-146	16	-181	219	421	-804	260	NA	**55**
All regions	0	9	5	20	53	15	73	31	8	**-34**	10	**-55**	NA

The table shows the differences in the numbers of the high expressed and low expressed genes between the twelve CNS regions. Positive values showing that the number of the high expressed genes for which the region in the column is more than the low expressed genes compared with the region shown in the row. Negative values showing that the number of the low expressed genes for which the region in the column is more than the high expressed genes compared with the region shown in the row.

Moreover, looking in detail at the cortical regions, namely FCTX, TCTX and OCTX, these data showed that the lowest number of differentially expressed genes was between these regions (**Table 4.3.4**, green box and **Figure 4.3.5**, green circles). Only three genes were detected as differentially expressed between FCTX and TCTX, while the OCTX region showed that there are 122 genes that were differentially expressed in comparison to FCTX and 87 genes in comparison to TCTX. This highlights that, although cortical regions have different functions, they share a common similar expression profile except for the OCTX, as this region has a more distinctive pattern compared to the other two cortical regions as the PCA plot showed (**Figure 4.3.4**). It worth noting that the number of differentially expressed genes between the cortical regions in comparison to SNIG, MEDU, CRBL, SPCO and WHMT was large compared with HIPP, THAL, HYPO and PUTM, implying that the cortical regions have more common functions compared with the latter (**Table 4.3.5**).

In summary, 72% (13,187 genes of 18,107 genes) were detected specifically in these three cortical regions. Out of this 72%, 1% (138 genes), 0.9% (123 genes) and 0.7% (89 genes) were differentially expressed in OCTX, FCTX and TCTX respectively in comparison with the other two regions (**Figure 4.3.6**). Although the cortical regions are considered to have functions in common with each other, OCTX showed a small but significantly different expression pattern from FCTX and TCTX regions. In the study by Roth *et al,* although the expression patterns were compared using a different platform in CNS regions, it was mentioned briefly in relation to the PCA plot that OCTX clustered separately from other cortical regions (Roth, Hevezi et al. 2006).

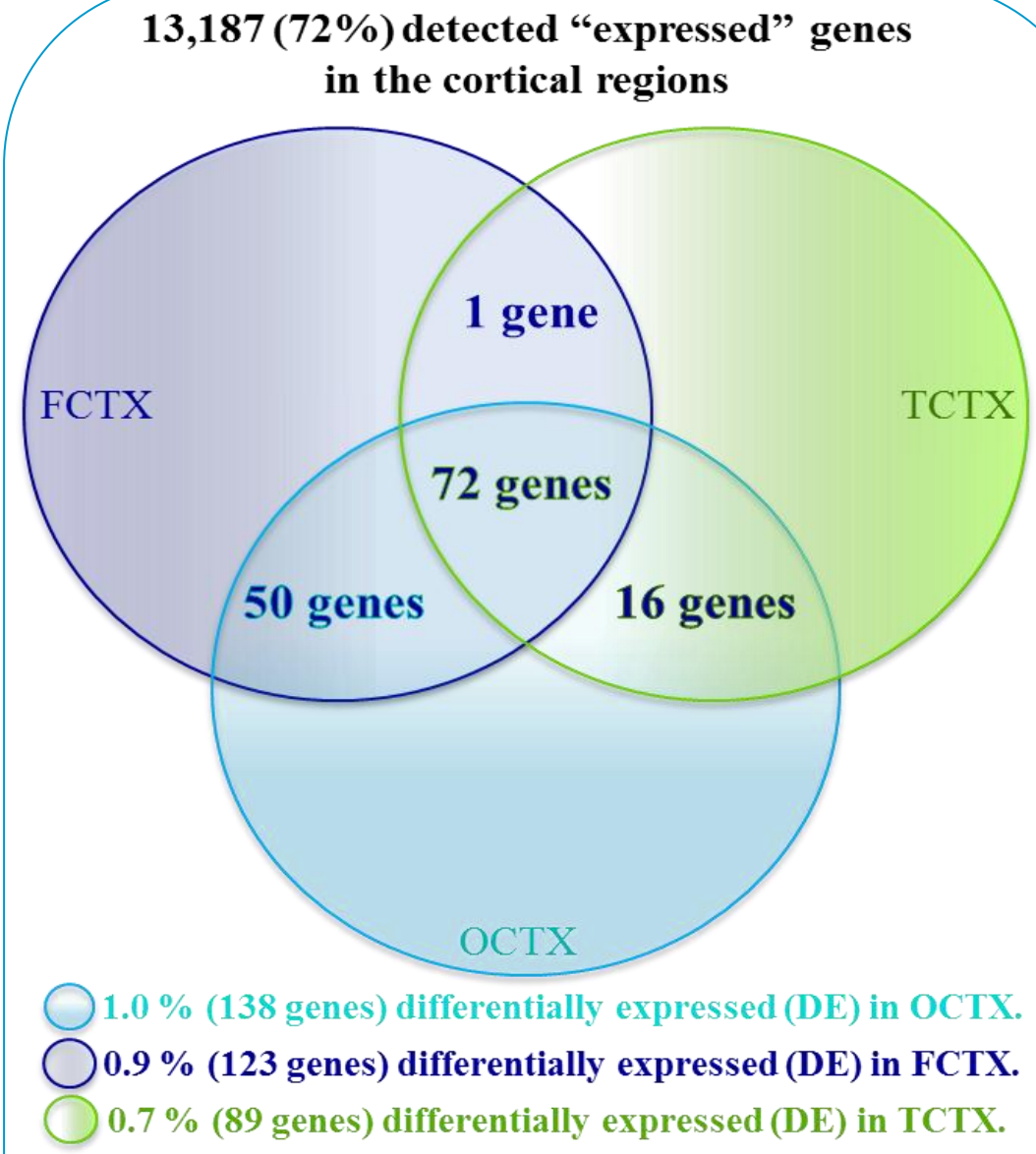

Figure 4.3.6:

Venn diagram representing the differentially expressed (DE) genes within the cortical regions. OCTX = occipital cortex, FCTX = frontal cortex and TCTX = temporal cortex.

This figure shows the numbers and percentages of the DE genes across the cortical regions. The number of the DE genes between OCTX and the other two regions is higher in comparison with FCTX and TCTX.

4.3.4 Examples of human CNS regional differential expression genes

In this section selected genes will be presented as examples of genes with significant regional differences in expression. Selection of these genes was based on the fact that they were reported to be genetically associated with neurological diseases from GWAS results or they were reported in biological studies to have specific functions in the human CNS.

Because of the importance of *LRRK2* (Leucine-rich repeat kinase 2) and *MAPT* (microtubule-associated protein tau) genes in relation to neurological diseases comprehensive investigations were performed on these two specific genes in this thesis (see Chapter 5 for *MAPT* and Chapter 6 for *LRRK2*). Differential expression and splicing patterns for these genes were further investigated and validated using two different platforms. Regional differences in gene expression were validated using Quantigene for *LRRK2*, *MAPT* and *SCN8A* (sodium channel, voltage gated, type VIII, alpha subunit) genes (see Chapter 3, Section 3.3.3), and qRT-PCRs were performed for *LRRK2* and *MAPT* (see Chapter 5, Section 5.3.1 and Chapter 6, Section 6.3.1).

ANO3 (anoctamin 3) gene expression was expressed significantly higher in PUTM by 4 FC ($P < 1.0 \times 10^{-45}$, at FDR 5%) (all Bonferroni adjusted $P < 4.8 \times 10^{-52}$) as compared with the cortical regions (FCTX, TCTX and OCTX) and 12 FC compared with CRBL (Bonferroni adjusted $P = 2.0 \times 10^{-240}$), the region with the lowest expression. Similar expression levels were observed amongst the other CNS regions (**Figure 4.3.7**). The *ANO3* gene encodes for a transmemebrane protein that activates calcium and chloride channels. *ANO3* expression is common in neuronal tissues (Schreiber, Uliyakina et al. 2010). This gene was reported to be associated with late-onset Alzheimer's disease (LOAD) (Milenkovic, Brockmann et al. 2010). In addition, a

recent study reported that mutations in this gene cause Dominant Craniocervical Dystonia (Charlesworth, Plagnol et al. 2012).

BIN1 (Bridging integrator 1) gene was significantly differentially expressed by 3 FC ($P < 1.0 \times 10^{-42}$, at FDR 5%) (Bonferroni adjusted $P = 2.7 \times 10^{-196}$) between WHMT, the region with the highest expression, and CRBL the region with the lowest expression. The region with the second highest expression was HIPP. There was a 1.6 fold difference compared with WHMT. Furthermore, *BIN1* was not only differentially expressed by CNS regions, but also was alternatively spliced by region. The differentially splicing results for this gene will be discussed in detail later in this chapter (see Section 4.3.6). *BIN1* gene encodes for several isoforms of a nucleocytoplasmic adaptor protein involved in regulating synaptic vesicle endocytosis (Wigge and McMahon 1998). This gene locus was reported to be associated with Alzheimer's disease (AD) (Kamboh, Demirci et al. 2012; Karch, Jeng et al. 2012).

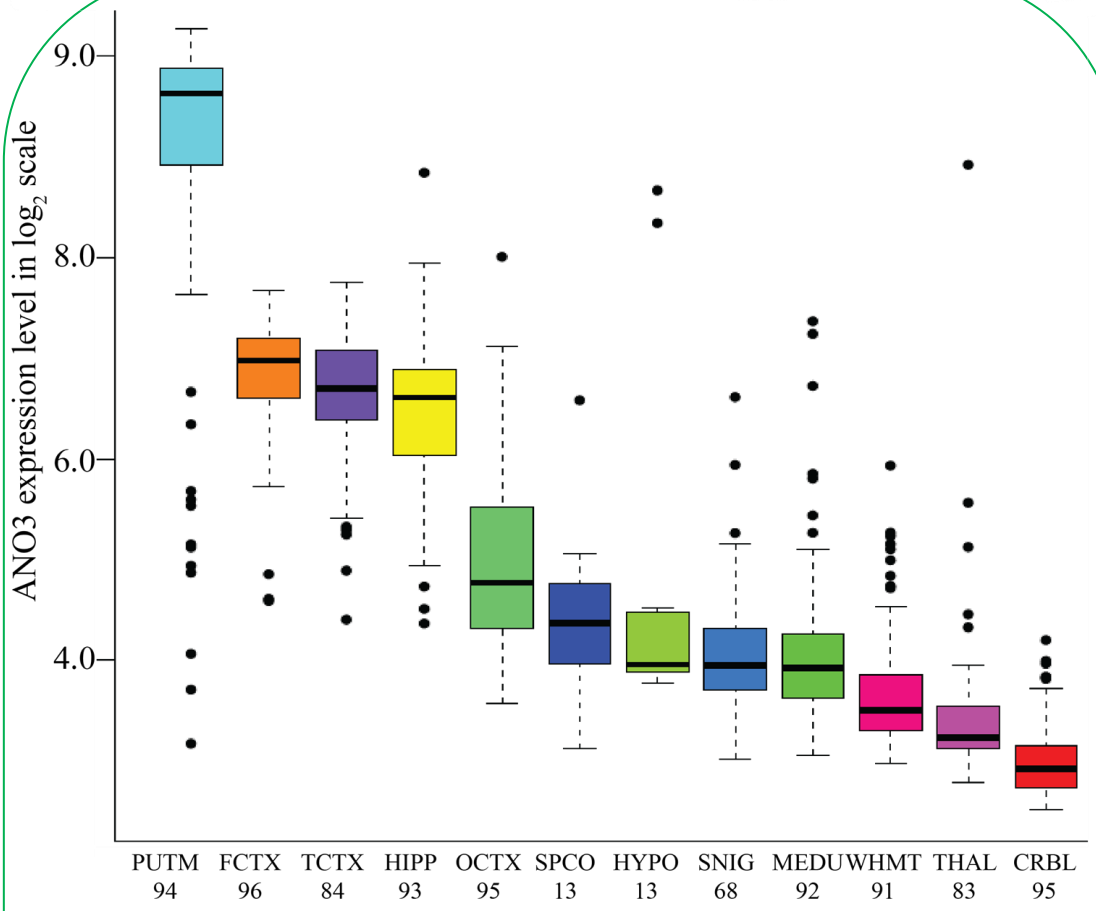

Figure 4.3.7:

Enrichment of *ANO3* mRNA expression in PUTM. PUTM = putamen, FCTX = frontal cortex, TCTX = temporal cortex, HIPP = hippocampus, OCTX = temporal cortex, SPCO = spinal cord, HYPO = hypothalamus, SNIG = *substantia nigra*, MEDU = medulla, WHMT = white matter, THAL = thalamus, CRBL = cerebellum.

Boxplot of mRNA expression levels for *ANO3* in 12 CNS regions, from microarray experiments on a log2 scale (*y*-axis). This plot shows *ANO3* expression in PUTM is higher by 12 FC compared with CRBL. Whiskers extend from the box to 1.5 times the inter-quartile range.

CNTNAP2 (Contactin-associated protein-like 2) gene was significantly differentially expressed by 5.4 - 7 FC higher ($P < 1.0 \times 10^{-45}$, at FDR 5%) in the cortical regions and THAL compared with CRBL (Bonferroni adjusted $P = 5.9 \times 10^{-126}$) and WHMT (Bonferroni adjusted $P = 1.1 \times 10^{-152}$) which showed the lowest expression levels (**Figure 4.3.8**). This result was consistent with a previous study (Alarcon, Abrahams et al. 2008). *CNTNAP2* encodes for a protein that is a member of the neurexin family. This protein is in the nervous system and acts as cell adhesion molecules and receptors. This gene was reported to be associated with different diseases such as autism, epilepsy, schizophrenia, and mental retardation (Alarcon, Abrahams et al. 2008; Friedman, Vrijenhoek et al. 2008).

PLP1 (proteolipid protein 1) gene was significantly differentially expressed in WHMT and SPCO by 6.0 FC higher ($P < 1.0 \times 10^{-45}$, at FDR 5%) compared with other CNS regions, especially CRBL (Bonferroni adjusted $P = 1.1 \times 10^{-170}$). *PLP1* gene encodes for a transmembrane proteolipid protein that is the main myelin protein in the CNS. It is involved in the compaction, stabilization, and maintenance of myelin sheaths. In addition, it is involved in oligodendrocyte development and axonal survival (Diehl, Schaich et al. 1986). Detecting this gene as being highly expressed in WHMT in control individuals is expected and confirms the reliability of the array data. In addition, the expression of *CR1* (Erythrocyte complement component, C3b/C4b receptor *1*) (**Figure 4.3.9, A** and **B**), *IGFBP3* (Insulin-like growth factor-binding protein 3) (**Figure 4.3.9, C** and **D**), *NLGN4X* (Neuroligin 4, X-linked) (**Figure 4.3.10, A** and **B**) and *VSIG4* (V-set and immunoglobulin domain containing 4) (**Figure 4.3.10, C** and **D**) genes were validated on a subset of samples from selected regions (**Table 4.3.6**). Results from the qRT-PCR are consistent with the array expression patterns. These genes were differentially expressed by ~ 2.0 FCs between the selected regions.

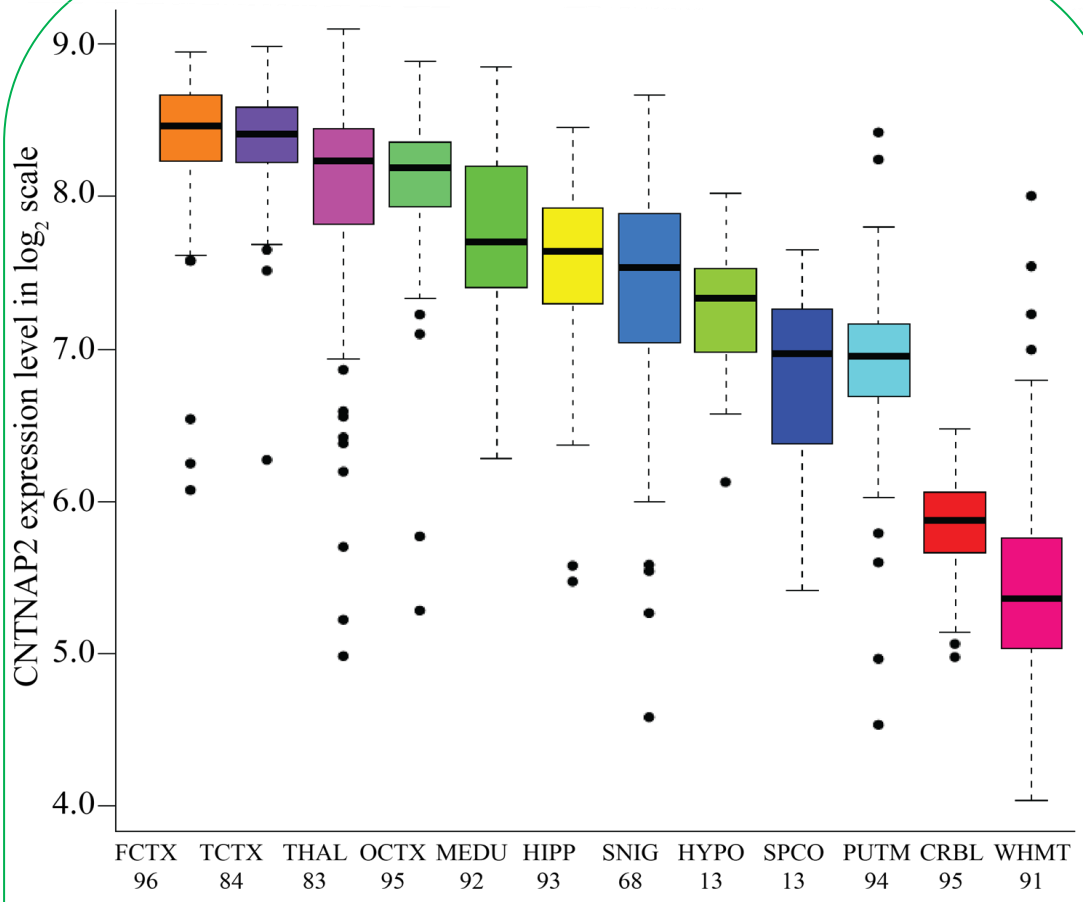

Figure 4.3.8:

Enrichment of *CNTNAP2* mRNA expression in PUTM. FCTX = frontal cortex, TCTX = temporal cortex, THAL = thalamus, OCTX = temporal cortex, MEDU = medulla, HIPP = hippocampus, SNIG = *substantia nigra*, HYPO = hypothalamus, SPCO = spinal cord, PUTM = putamen, CRBL = cerebellum, WHMT = white matter.

Boxplot of mRNA expression levels for *CNTNAP2* in 12 CNS regions, from microarray experiments on a log2 scale (*y*-axis). This plot shows *CNTNAP2* expression in FCTX is higher by ~ 5.4 FC compared with CRBL. Whiskers extend from the box to 1.5 times the inter-quartile range.

Table 4.3.6: Differential expression patterns validation using qRT-PCR.

			Affymetrix exon array results			qRT-PCR results	
Gene Symbol	**Assay ID**	**Regions compared**	***P* values for regional-bias**		**Fold change**	***P* values for regional expression**	**Fold change**
			Expression	**splicing**			
CR1	Hs00559348_m1	HIPP < WHMT	7.7×10^{-38}	Not validated	-1.82	3.14×10^{-02}	-2.01
IGFBP3	Hs00181211_m1	HIPP > WHMT	4.2×10^{-39}	Not validated	2.10	4.05×10^{-02}	1.43
NLGN4X	Hs01934144_s1	THAL > WHMT	1.9×10^{-45}	Not validated	2.68	1.41×10^{-02}	1.41
VSIG4	Hs00200695_m1	HIPP > HYPO	2.6×10^{-06}	Not validated	2.18	4.7×10^{-02}	2.45

This table shows the performed qRT-PCR validations for the expression arrays results of selected genes. The validation for the expression patterns was performed for *CR1, IGFBP3*, *NLGN4X* and *VSIG4* genes in four CNS regions. Significant fold changes were observed for both platforms for the selected differentially expressed genes. (This table was produced by Dr Mina Ryten). The array results *P* values were adjusted using FDR procedure.

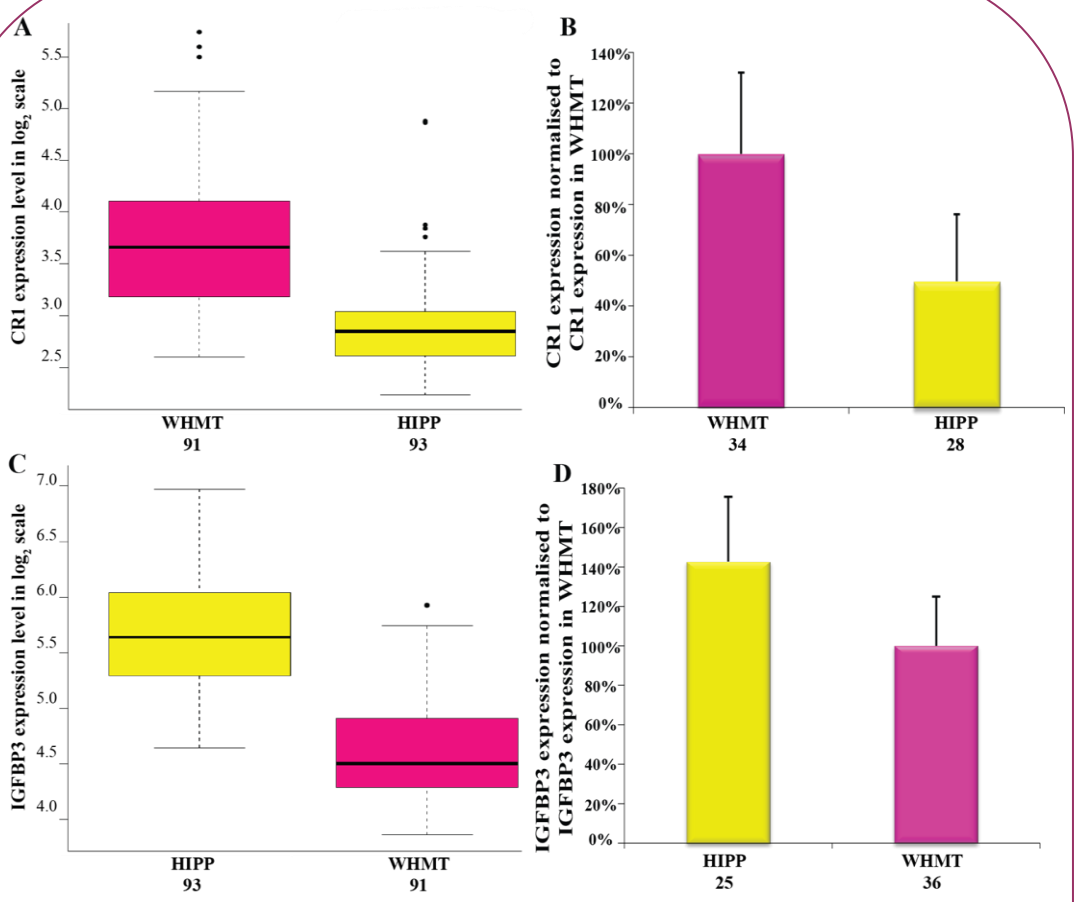

Figure 4.3.9:

Array and qRT-PCR mRNA expression of *CR1* and *IGFBP3*. WHMT = white matter, HIPP = hippocampus.

This figure shows regional variation in: **(A)** Boxplot of *CR1* expression level between WHMT and HIPP as measured using arrays on a log2 scale (*y*-axis)**, (B)** Bar chart of *CR1* expression level in same regions as measured using qRT-PCR validation. **(C)** Boxplot of *IGFBP3* mRNA expression level between HIPP and WHMT as measured using arrays on a log2 scale (*y*-axis)**, (D)** Bar chart of *IGFBP3* level in the same regions as measured using qRT-PCR validation. In both cases expression levels were normalized to the expression of the target gene in WHMT (100%). Box plot whiskers extend from the box to 1.5 times the inter-quartile range. The error bars represent the S.E.M.

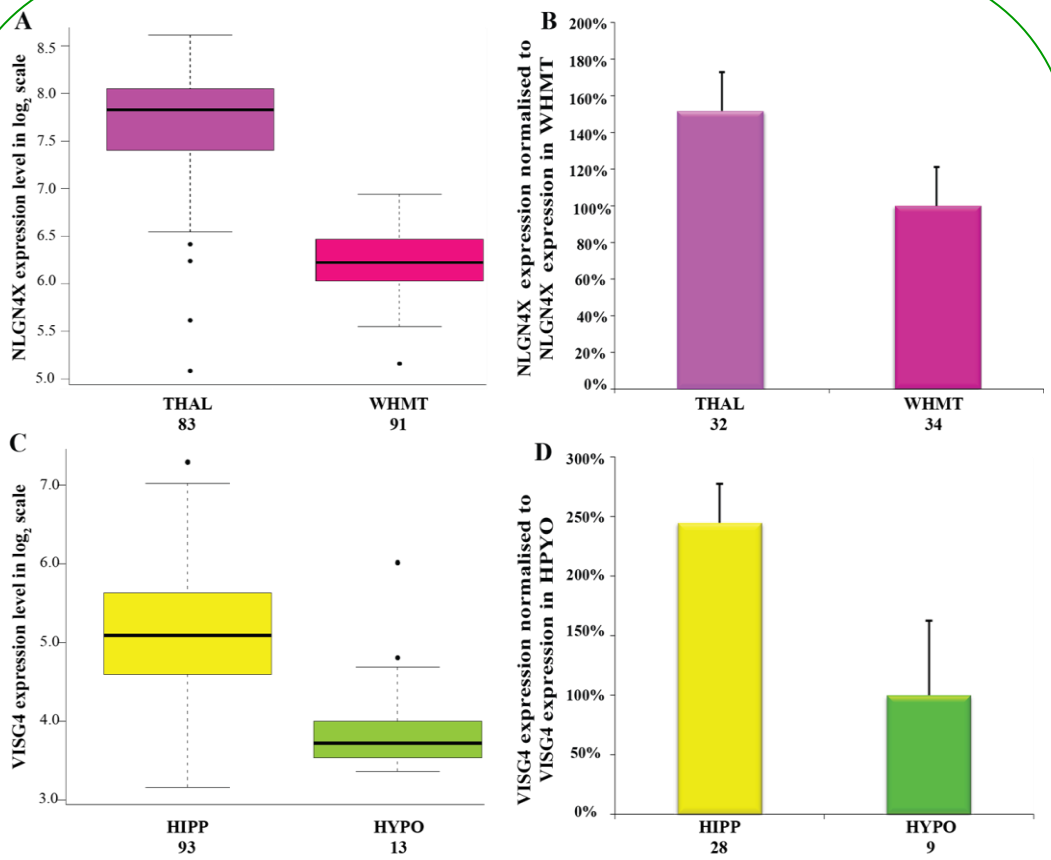

Figure 4.3.10:

Array and qRT-PCR mRNA expression of *NLGN4X* and *VSIG4*. WHMT = white matter, HIPP = hippocampus, THAL = thalamus, HYPO = hypothalamus.

This figure shows regional variation in: **(A)** Boxplot of *NLGN4X* expression level between THAL and WHMT as measured using arrays on a log2 scale (*y*-axis)**, (B)** Bar chart of *NLGN4X* expression level in the same regions as measured using qRT-PCR validation. Expression levels were normalized to the expression of the target gene in WHMT (100%). **(C)** Boxplot of the *VSIG4* mRNA expression level between HIPP and WHMT as measured using arrays on a log2 scale (*y*-axis)**, (D)** Bar chart of *VSIG4* level in the same regions as measured using qRT-PCR. Expression levels were normalized to the expression of the target gene in HYPO (100%). Box plot whiskers extend from the box to 1.5 times the inter-quartile range. The error bars represent the S.E.M.

4.3.5 Human CNS regional differential splicing (exon) expression

Over the last decade, a number of studies have highlighted the importance of alternative splicing events in the human genome (Lander, Linton et al. 2001) and nervous system in particular (Grabowski and Black 2001; Lee and Irizarry 2003) and their role in human disease (Krawczak, Reiss et al. 1992; Black 2003). These studies, and many others, motivated the experimental design stage of this project to conduct an analysis of the CNS samples on a human exon array platform, allowing the identification of differential exon usage and by implication alternative splicing events (Clark, Schweitzer et al. 2007) (see Chapter 3, Section 3.2). In order to define differentially (alternative) spliced transcripts/genes across the 12 CNS regions, the statistical threshold of FDR < 10^{-5} was applied to the alternative splicing ANOVA result sheet that contained 12,790 genes as explained in detail in Chapter 2, Section 2.6.3.1.

The results of this filtering showed that 91% (11,655 genes) of the 12,790 detected (expressed) genes displayed differential exon usage in different regions, strongly suggestive of alternative splicing patterns. Out of these 11,655 genes, 96% (11,241 genes) were differentially expressed as well as differentially spliced. In addition, 3.5% (414 genes) of those genes were differentially spliced but not differentially expressed and 9.2% (1,135 of 12,376 genes) were differentially expressed but not differentially spliced across the 12 CNS regions (**Figure 4.3.11**). These results are consistent with data from a previous study, which revealed that the adult human brain has a very high and specific alternative splicing profile compared with other human tissues (Yeo, Holste et al. 2004).

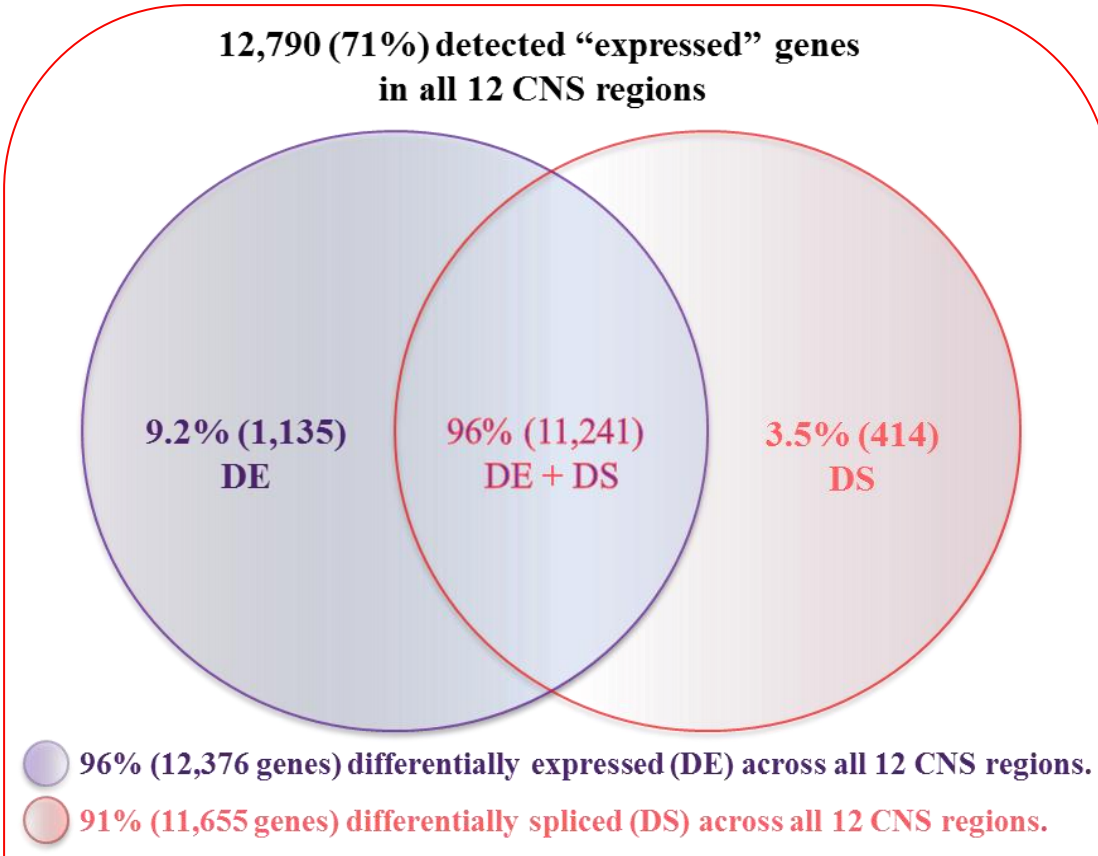

Figure 4.3.11:

Venn diagram representing the differentially expressed (DE) and spliced (DS) genes across the 12 CNS regions.

This figure shows the numbers and percentages of the DE and DS genes across the 12 CNS regions, the percentage of DE and DS genes is high in the adult human CNS compared with genes that are only DE (9.2%). With regard to the expression profiles, 11,241 genes (91%) are both differentially expressed as well as spliced.

The data resulting from analysing the alternative slicing profiles from the different cortical regions in this study was of particular interest. The analysis was performed on the three regions combined together, with the results showing that 72% (13,187 of 18,107 genes) were detected in the three cortical regions: FCTX, TCTX and OCTX. Although there was a relatively small number of differentially expressed genes between these regions (**Table 4.3.4** and **Figure 4.3.5**), the alternative splicing ANOVA detected a high number of differentially spliced genes (**Figure 4.3.12**).

The three possible comparisons within the cortical regions, namely OCTX versus TCTX, FCTX versus OCTX, and FCTX versus TCTX, were considered in relation to the detection and identification of the alternative splicing patterns (**Figure 4.3.13, A-C**). These results showed a similar pattern to the results when the three regions were combined (**Figure 4.3.12**).

Out of the 13,187 detected genes, 74% of genes (9794 genes) in the comparison between OCTX and TCTX (**Figure 4.3.13, A**), 52% of genes (6890 genes) in the comparison between FCTX and OCTX regions (**Figure 4.3.13, B**) and 40% (5280 genes) of genes in the comparison between FCTX and TCTX (**Figure 4.3.13, C**) were identified as differentially spliced. The OCTX, displayed a different alternative splicing pattern compared with the other cortical regions, similar to the results for deferential expression pattern (**Figures 4.3.4** and **4.3.6**), but even more pronounced. FTCX and TCTX had a lower percentage (40%) of differentially spliced genes when compared to each other.

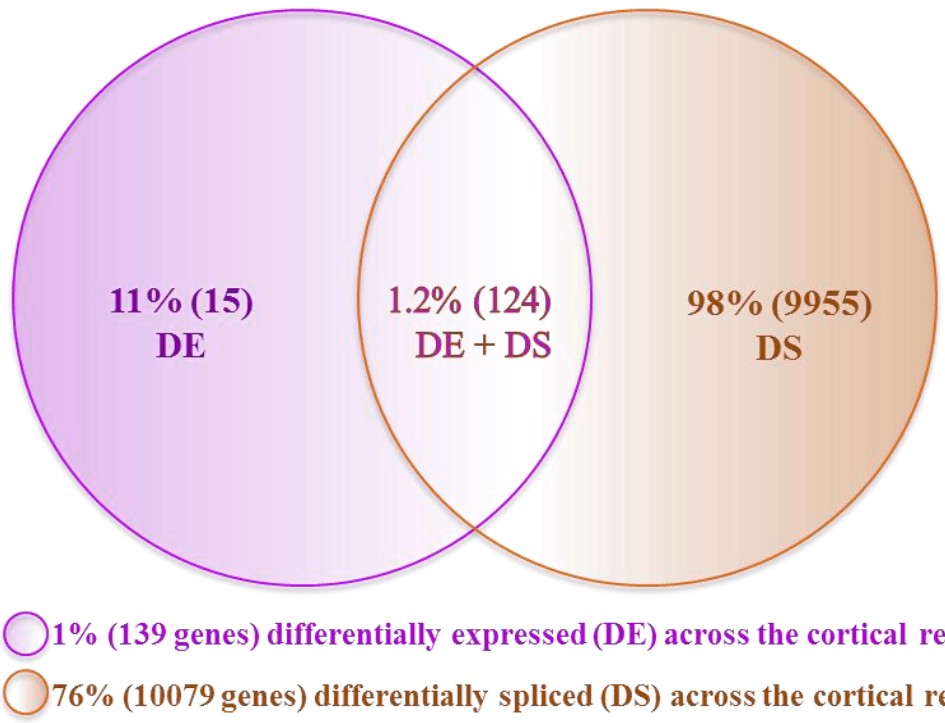

Figure 4.3.12:

Venn diagram representing the differentially expressed (DE) and spliced (DS) genes across the cortical regions.

This figure shows the numbers and percentages of the DE and DS genes across the cortical regions. It shows that the number of genes that are DS is much higher (98%) compared with the number of genes that are DE.

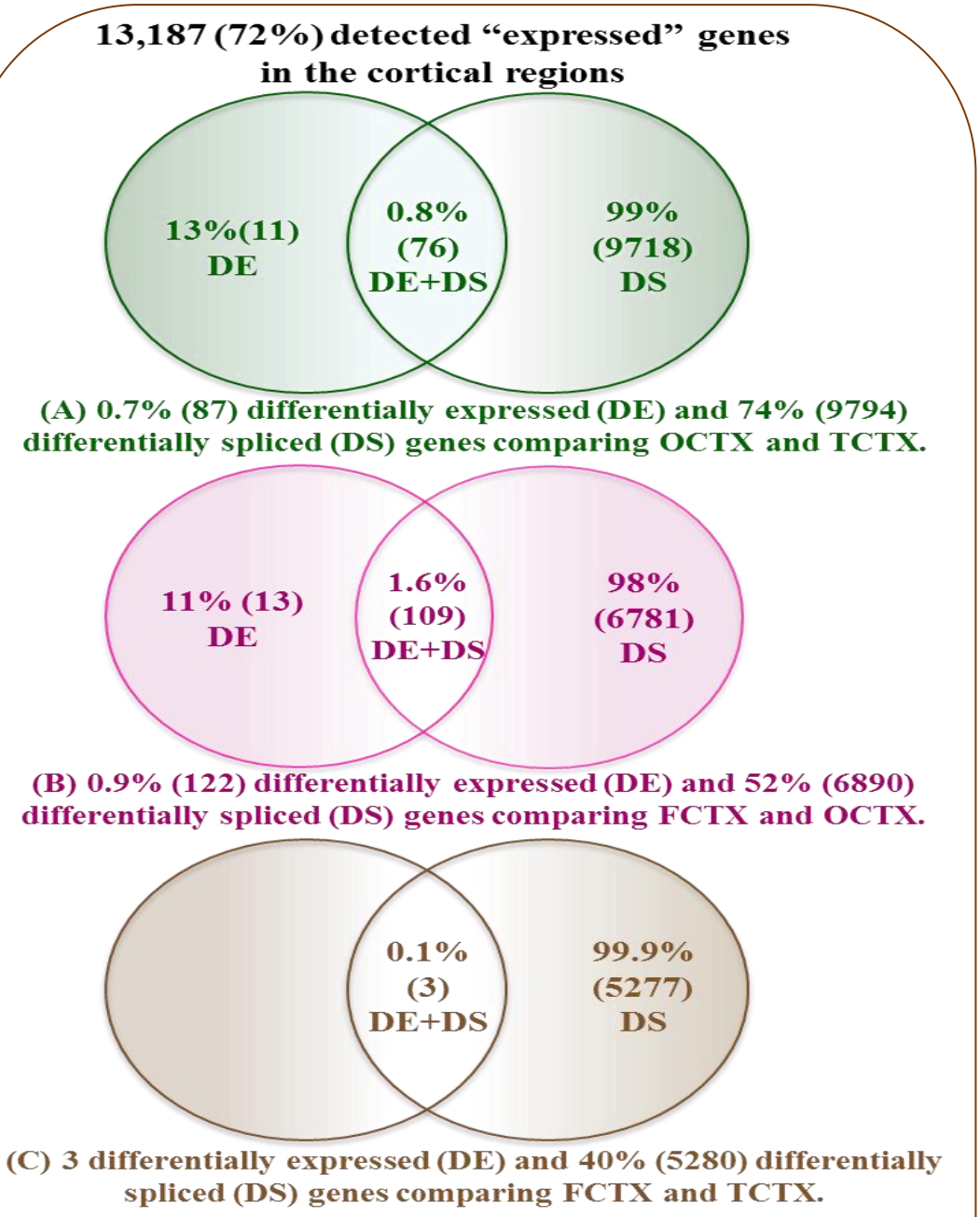

Figure 4.3.13:

Venn diagram representing the differentially expressed (DE) and spliced (DS) genes across: (A) OCTX versus TCTX, (B) FCTX versus OCTX and (C) FCTX versus TCTX.

This figure shows the numbers and percentages of the DE and DS genes across the cortical regions. In the three comparisons, the alternative splicing patterns are the dominant events in these regions compared with the differential expression patterns.

4.3.6 Examples of human CNS regional differential splicing (exon) expression genes

In this section examples of genes with regional differential splicing will be presented. The selection of these genes was based on the fact that they were reported to be genetically associated with neurological diseases from the GWAS results or they were reported in biological studies to have specific functions in the human CNS. The additional examples of *MAPT* and *LRRK2* genes which showed regional alternative splicing patterns will be discussed in detail later in this thesis (see Chapter 5 for *MAPT* and Chapter 6 for *LRRK2*),

KCND3 (potassium voltage-gated channel, Shal-related subfamily, member 3) gene was not only expressed at significantly higher level in SNIG by 7 FC higher (ANOVA $P < 1.0 \times 10^{-45}$, at FDR 5%) (Bonferroni adjusted $P = 9.9 \times 10^{-47}$) as compared with CRBL, but also was differentially spliced by region. *KCND3* has two isoforms, a long (L) isoform (Exon 6 included, probe set 2428126) and a short (S) isoform (Exon 6 spliced out). The arrays detected the long isoform to be expressed specifically in SNIG compared with other CNS regions (ANOVA $P < 1.0 \times 10^{-45}$, at FDR $< 10^{-05}$) (**Figure 4.3.14**). This result is consistent with a previous study that showed by using single-cell RT-PCR that the long isoform of this gene is only expressed in dopaminergic (DA) substantia nigra (SN) neurons (Liss, Franz et al. 2001). *KCND3* gene encodes for the K+ channel that underlies the transient outward potassium current. It is expressed in abundance in the adult brain and its function is to regulate neurotransmitter release and neuronal excitability (Serodio, Vega-Saenz de Miera et al. 1996; Dilks, Ling et al. 1999).

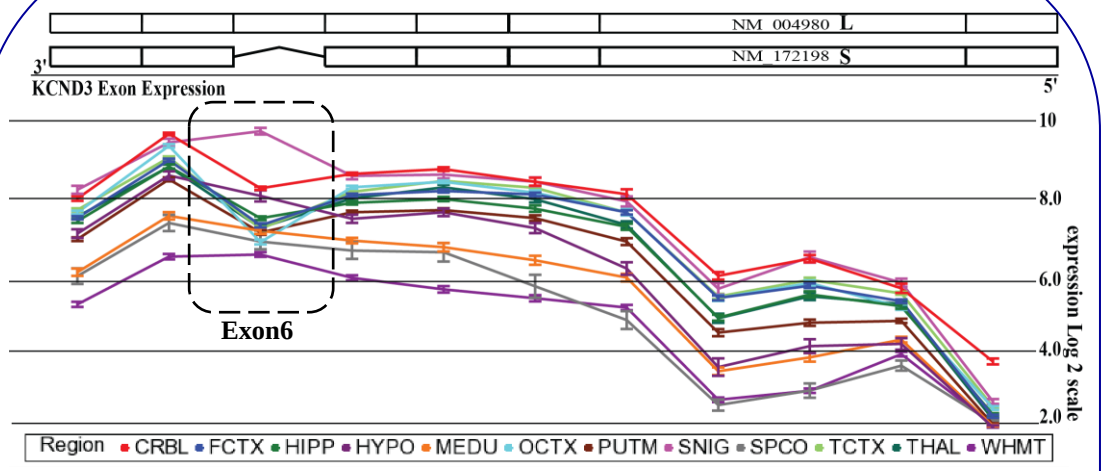

Figure 4.3.14:

Exon array expression data at the probe set level suggesting alternative splicing by CNS region of the *KCND3* transcript. CRBL = cerebellum, FCTX = frontal cortex, HIPP = hippocampus, HYPO = hypothalamus, MEDU = medulla, OCTX = occipital cortex, PUTM = putamen, SNIG = *substantia nigra*, SPCO = spinal cord, TCTX = temporal cortex, THAL = thalamus and WHMT = white matter.

This figure shows the alternative splicing pattern between SNIG and other CNS regions for exon 6 (in the dashed black box). Long and short forms of *KCND3* mRNA as measured using arrays are plotted using the mean expression levels (*y*-axis) for each probe set (*x*-axis) covering the *KCND3* gene. Non-parallel probe set expression levels, in the dashed black box, suggest differences in the expression and splicing of the corresponding exons. Alternative splicing Bonferroni adjusted $P = 9.9 \times 10^{-47}$ for SNIG versus CRBL. This plot was generated using Partek Genomics Suite software.

NRXN3 (Neurexin) gene was significantly differentially expressed by 4 FC higher ($P < 1.0 \times 10^{-45}$, at FDR 5%) in the cortical regions (OCTX, FCTX and TCTX) as compared with WHMT the region with the lowest expression. The region with the second highest expression region was CRBL. Furthermore, *NRXN3* was alternatively spliced by region. From the results of the arrays, two major isoforms of this gene were detected, alpha-*NRXN3* (long isoform) and the beta-*NRXN3* (short isoforms). Alpha-*NRXN3* is the dominant (expressed higher) isoform in all CNS regions except for the CRBL where beta-*NRXN3* is the dominant isoform (**Figure 4.3.15**). A previous study showed a similar pattern of expression of the two isoforms (Hishimoto, Liu et al. 2007)**.** Further validations for these two isoforms were performed on THAL and WHMT regions by using the qRT-PCR platform. A similar significant difference in the expression pattern was observed confirming the array results (**Table 4.3.7** and **Figure 4.3.16**). *NRXN3* gene encodes for cell surface proteins that are expressed in neurons (Tabuchi and Sudhof 2002). This protein has an essential role in synaptic development and functions, acting as a synaptic cell-adhesion molecule that connects presynaptic and postsynaptic neurons at synapses (Craig and Kang 2007; Sudhof 2008). It was reported that this locus may be involved in the development of Autism (Sudhof 2008; Vaags, Lionel et al. 2012).

Regional differences in the expression of *BIN1* were discussed previously in Section 4.3.4, and the results showed that this gene was differentially expressed between regions. Furthermore, the results from the arrays for *BIN1* strongly suggested differential isoform expression amongst CNS regions (ANOVA $P < 1.0 \times 10^{-45}$, at FDR $< 10^{-05}$). The results showed that the isoforms that have exon 7 spliced out (probe set 2574686) are expressed higher by 4 FC in WHMT as compared with CRBL (Bonferroni adjusted $P = 3.5 \times 10^{-35}$) (**Figure 4.3.17**).

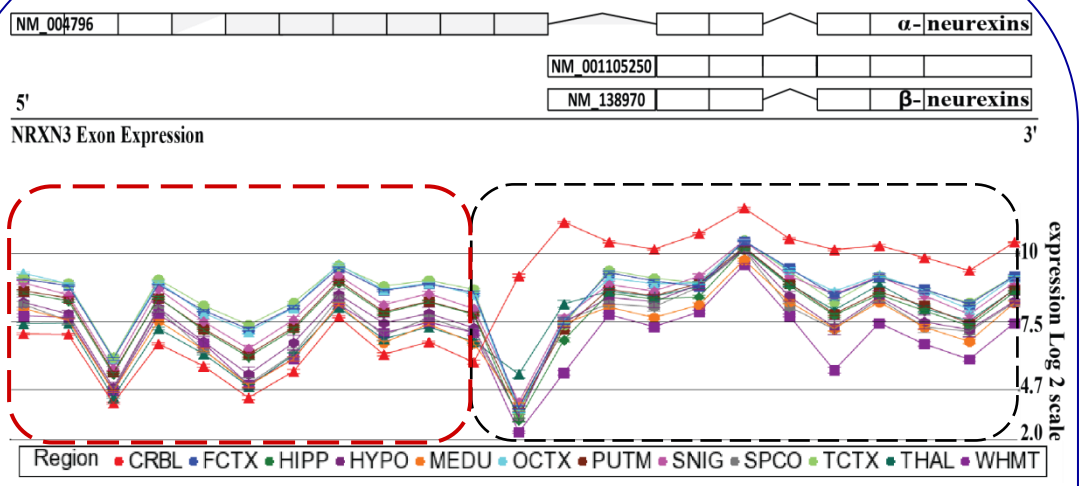

Figure 4.3.15:

Exon array expression data at the probe set level showing different isoforms of the *NRXN3* transcript. CRBL = cerebellum, FCTX = frontal cortex, HIPP = hippocampus, HYPO = hypothalamus, MEDU = medulla, OCTX = occipital cortex, PUTM = putamen, SNIG = *substantia nigra*, SPCO = spinal cord, TCTX = temporal cortex, THAL = thalamus and WHMT = white matter.

This figure shows the expression level of the alpha-*NRXN3* and beta-*NRXN3* isoforms in all 12 CNS regions. beta-*NRXN3* mRNA expression level is high in CRBL (~6 FC) as compared with the other regions (in the dashed black box) (Bonferroni adjusted $P = 4.3 \times 10^{-178}$). In contrast, alpha-*NRXN3* mRNA expression is low in CRBL compared with other regions (in the dashed red box). Mean expression levels (*y*-axis) plotted for each probe set (*x*-axis) covering the *NRXN3* gene. Non-parallel probe set expression levels, in the dashed black box, suggest regional differences in the expression and splicing of the corresponding exons. This plot was generated using Partek Genomics Suite software.

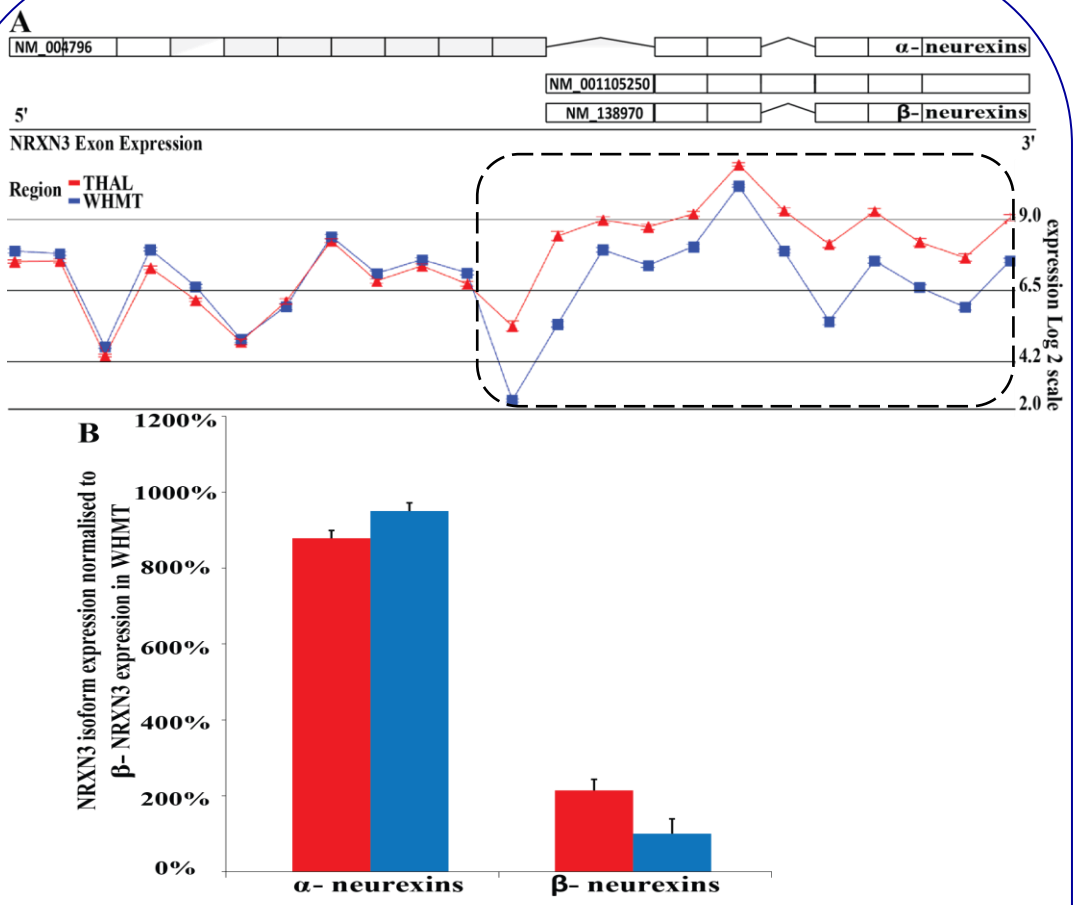

Figure 4.3.16:

Exon array expression data at the probe set level suggesting alternative splicing by CNS region of the *NRXN3* transcript. qRT-PCR validation of the results in WHMT and THAL. WHMT = white matter and THAL = thalamus.

This figure shows the alternative splicing pattern between WHMT and THAL for: **(A)** Alpha and beta *NRXN3* mRNA expression levels between WHMT and THAL (~2 FC) as measured using arrays**,** mean expression levels (*y*-axis) plotted for each probe set (*x*-axis) covering the *NRXN3* gene. Non-parallel probe set expression levels, in the dashed black box, suggest regional differences in the expression and splicing of the corresponding exons. All the differentially spliced exons correspond to exons present in β-neurexins (Bonferroni adjusted $P = 9.5 \times 10^{-99}$). This plot was generated using Partek Genomics Suite software. **(B)** Quantitative qRT-PCR validation of *NRXN3* alpha and beta expression levels in the same regions confirms the array results. The expression level is presented as the mean and the error bars are S.E.M.

In addition, all the *BIN1* isoforms that are expressed in the human CNS have exon 12 spliced out. Further validation for the isoforms that have exon 7 spliced out were performed on THAL and WHMT regions by using the qRT-PCR platform. A similar and significant expression pattern was observed confirming the array results (**Table 4.3.7** and **Figure 4.3.18**).

The last example for this category is the *SETX* (senataxin) gene. The results from the arrays strongly suggested an alternative splicing event for exon 22 (probe set 3228025) between CRBL and SPCO. In CRBL exon 22 expression was higher by 1.8 FC ($P = 2.5 \times 10^{-15}$, at FDR < 10^{-05}) (Bonferroni adjusted $P = 1.5 \times 10^{-15}$) as compared with SPCO region (**Figure 4.3.19**). This suggests that different isoforms are expressed in different CNS regions. Mutations in the *SETX* genes were associated with ataxia-ocular apraxia-2 (AOA2) (Moreira, Klur et al. 2004) and an autosomal dominant form of juvenile amyotrophic lateral sclerosis (ALS4) (Arning, Epplen et al. 2012). It is involved in the regulation of transcription and RNA processing such as RNA polymerase II (Suraweera, Lim et al. 2009). Further validation of this splicing event using a different technique is required.

It is noteworthy the number of extremely small *P* values in this study. This may be the result of (1) sufficiently large sample size, but more importantly (2) small residual background variability, contributing to small standard errors of expression means. Pairwise comparisons of region means were assessed and Bonferroni adjustments were then applied to the *P* values generated from this manual analysis, as indicated in the results above. Despite the fact that the Bonferroni adjustment is more conservative than the FDR adjustment that was used in the Partek analysis, the *P* values were still highly significant.

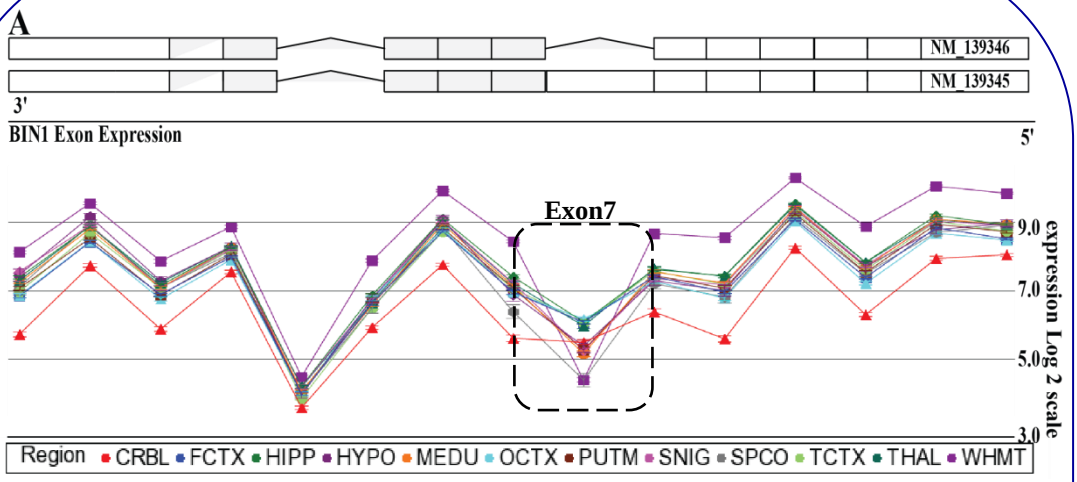

Figure 4.3.17:

Exon array expression data at the probe set level of different isoforms of the *BIN1* transcript. CRBL = cerebellum, FCTX = frontal cortex, HIPP = hippocampus, HYPO = hypothalamus, MEDU = medulla, OCTX = occipital cortex, PUTM = putamen, SNIG = *substantia nigra*, SPCO = spinal cord, TCTX = temporal cortex, THAL = thalamus and WHMT = white matter.

This figure shows the expression level of the different isoforms in all 12 CNS regions. mRNA expression levels of *BIN1* isoforms that have exon 7 spliced out are higher in WHMT (~4 FC) as compared with other regions such as CRBL (In the dashed black box). All the expressed isoforms in the human CNS have exon 12 spliced out. Mean expression levels (*y*-axis) plotted for each probe set (*x*-axis) covering the *BIN1* gene. Non-parallel probe set expression levels, in the dashed black box, suggest regional difference in the expression and splicing of the corresponding exons. Alternative splicing Bonferroni adjusted $P = 3.5 \times 10^{-35}$. This plot was generated using Partek Genomics Suite software.

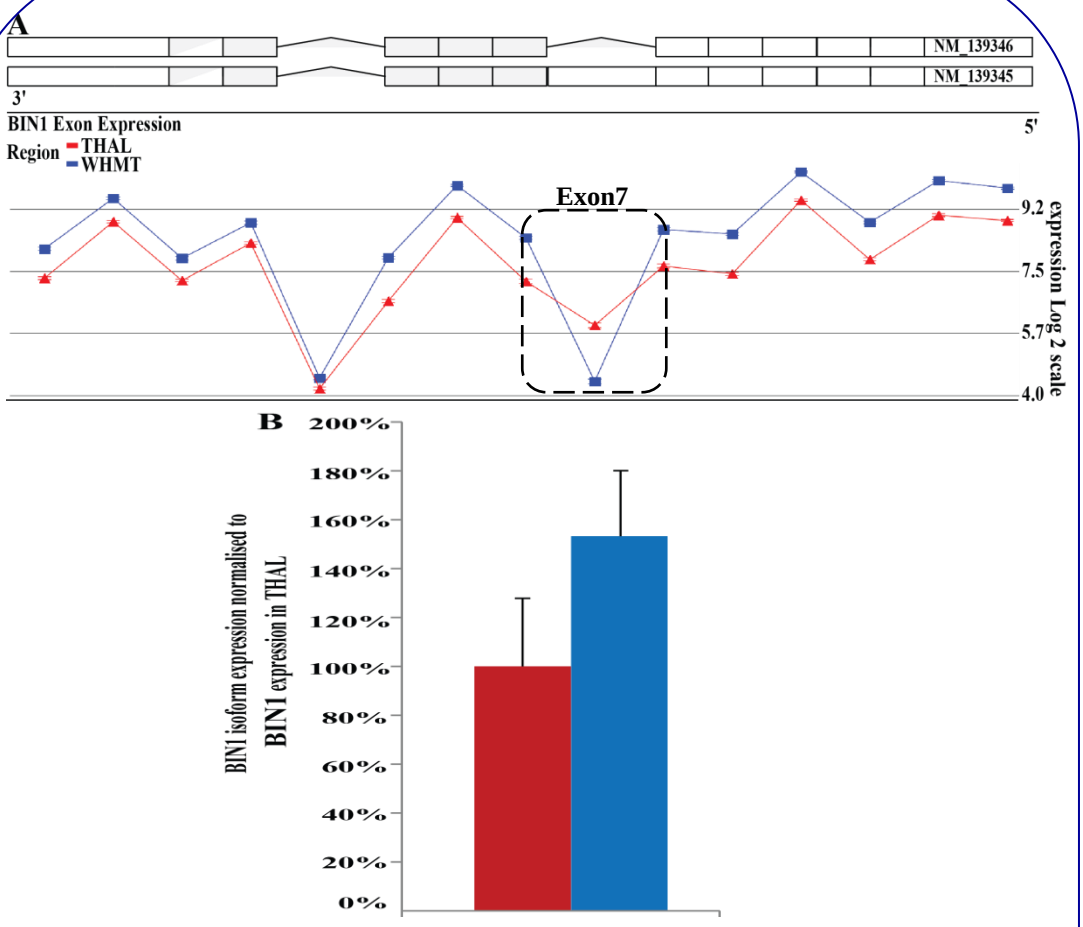

Figure 4.3.18:

Exon array expression data at the probe set level suggesting alternative splicing by CNS region of the *BIN1* transcript. qRT-PCR validation of the results in WHMT and THAL. WHMT = white matter and THAL = thalamus.

This figure shows the alternative splicing pattern between WHMT and THAL for: **(A)** Different isoforms of *BIN1* mRNA expression levels between WHMT and THAL (~2 FC) as measured using arrays**,** mean expression levels (*y*-axis) plotted for each probe set (*x*-axis) covering the *BIN1* gene. Non-parallel probe set expression levels, in the dashed black box, suggest regional difference in the expression and splicing of the corresponding exons. The alternative splicing Bonferroni adjusted $P = 3.2 \times 10^{-59}$. This plot was generated using Partek Genomics Suite software. **(B)** Quantitative qRT-PCR validation of *BIN1* mRNA expression level in the same regions confirms the array results. The expression level is presented as the mean and the error bar are S.E.M.

Table 4.3.7: Differential splicing patterns validation using qRT-PCR.

			Affymetrix exon array results			qRT-PCR results	
Gene Symbol	**Assay ID**	**Array prediction**	***P* values for regional-bias**		**Fold change**	***P* values for regional expression**	**Fold change**
			Expression	**splicing**			
BIN1	Hs01120891_m1	THAL < WHMT	NA	$< 1.0 \times 10^{-45}$	NA	$*4.06 \times 10^{-02}$	1.53
	Hs01120903_m1	THAL > WHMT				9.90×10^{-02}	1.38
NRXN3	Hs01028181_m1 (alpha)	THAL = WHMT	NA	$< 1.0 \times 10^{-45}$	NA	1.53×10^{-01}	1.15
	Hs01031277_m1 (beta)	THAL > WHMT				$*1.98 \times 10^{-02}$	2.14
PICALM	Hs00999727_m1	THAL = WHMT	NA	$< 1.0 \times 10^{-45}$	NA	$*3.90 \times 10^{-02}$	1.33
	Hs00999745_m1	THAL < WHMT				$*3.00 \times 10^{-03}$	1.40

This table shows the qRT-PCR validations performed for the splicing array results for *BIN1*, *NRXN3* and *PICALM* genes. Significant **P* values for both platforms for the alternative splicing were observed from the arrays and the validation experiments. (This table was produced by Dr Mina Ryten). The array results *P* values were adjusted using FDR procedure. The high level of significance of the array results in comparison to the moderate to non-significance of the qRT-PCR results can be due to the smaller sample size in the qRT-PCR study (THAL, $n = 34$, WHMT, $n = 37$) as compared with the array expression study (THAL, $n = 83$, WHMT, $n = 91$) and the larger background variation in the qRT-PCR compared with the array platform. It is noteworthy that the direction of FC values is mostly consistent across the expression and qRT-PCR platforms (fold changes are based on THAL compared with WHMT).

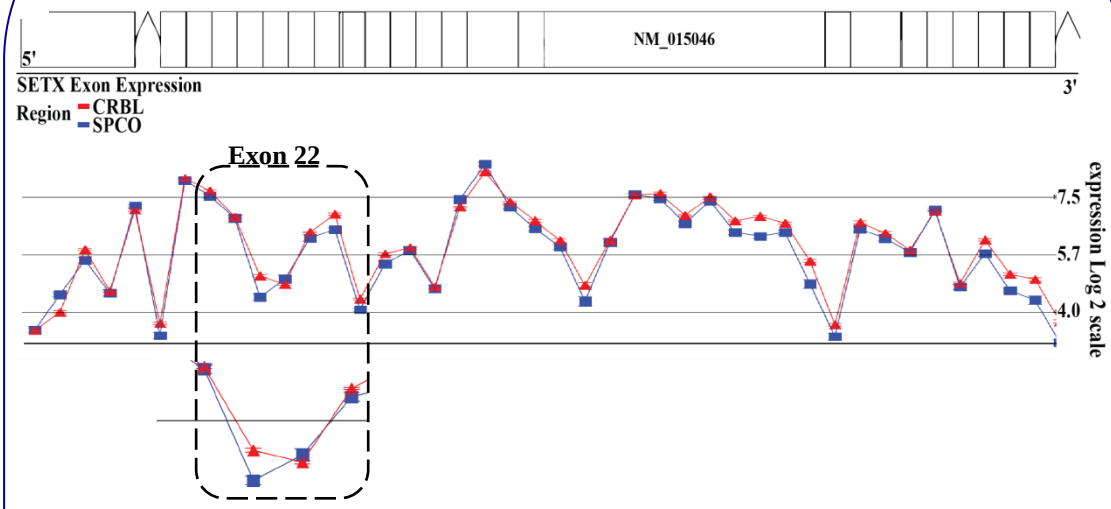

Figure 4.3.19:

Exon array expression data at the probe set level suggesting alternative splicing by CNS region of the *SETX* transcript. SPCO = spinal cord and CRBL = cerebellum.

This figure shows the alternative splicing patterns for exon 22 between CRBL and SPCO. Exon 22 expression level was 1.5 FC higher in CRBL compared with SPCO as measured using arrays**,** mean expression levels (*y*-axis) plotted for each probe set (*x*-axis) covering the *SETX* gene. Non-parallel probe set expression levels, in the dashed black box, suggest regional difference in the expression and splicing of the corresponding exons. Alternative splicing Bonferroni adjusted $P = 1.5 \times 10^{-15}$. This plot was generated using Partek Genomics Suite software.

4.3.7 Initial brain expression quantitative trait loci (QTL) analysis: toward overall integration of genomic information

In order to generate the brain expression QTL results for the ten CNS regions, the association tests between each probe set (exon level) and each SNP/Indel were performed. As a result, 19,648 unique expression QTL signals reached significance based on an FDR < 1%. The QTL was categorized as *cis* or *trans* based on the distance of the SNP to the transcription start site (TSS). In this analysis the *cis* window was within 1 Mb from the TSS.

Results showed that 84% (16,421) of the significant expression QTL signals were specific to only one CNS region (**Table 4.3.8**). Out of these 16,421 QTL signals, 25% (4,083), 18% (2,921) and 12% (1,991) were mapped only in CRBL, TCTX and WHMT respectively (**Table 4.3.9**). These three regions showed the highest number of region specific QTL signals.

In total, there was a significant difference between the *cis* and *trans* acting signals. The *cis*-acting QTL signals accounted for 69.2% while *trans*-acting QTL signals accounted for 30.8% of all expression QTL signals (**Table 4.3.8**). The enrichment pattern of *cis*-acting versus *trans*-acting signals was similar when looking at all region specific QTL signals (**Table 4.3.9**). In order to check if there was a constant pattern between expression and QTL signals in each pair of regions, a table similar to Table 4.3.5 was generated (**Table 4.3.10**). The expression QTL signals pattern was not clearly constant with the expression pattern. There was no distinctive pattern and difficult to interpret.

Table 4.3.8: The significant expression QTL signals in one or more CNS regions.

Number of CNS regions	Number of *cis*-acting QTLs	Percentage of *cis*-acting QTLs	Number of *trans*-acting QTLs	Percentage of *trans*-acting QTLs	Number of total QTLs	Percentage of total QTLs
1	**10,490**	**63.9%**	**5,931**	**36.1%**	**16,421**	**84%**
2	990	95.2%	50	4.8%	1,040	5%
3	594	97.5%	15	2.5%	609	3%
4	360	97.0%	11	3.0%	371	2%
5	273	94.5%	16	5.5%	289	1%
6	195	96.5%	7	3.5%	202	1%
7	160	97.0%	5	3.0%	165	1%
8	165	94.3%	10	5.7%	175	1%
9	143	94.1%	9	5.9%	152	1%
10	224	100.0%	0	0.0%	224	1%
Total	**13,594**	**69.2%**	**6,054**	**30.8%**	**19,648**	**100%**

This table shows the numbers and percentages for the significant expression QTL signals in one CNS region up to ten CNS regions.

Table 4.3.9: Significant expression QTL signals that were specifically mapped in a named CNS region.

CNS regions	Number of *cis*-acting QTLs	Percentage of *cis*-acting QTLs	Number of *trans*-acting QTLs	Percentage of *trans*-acting QTLs	Total number of QTLs	Percentage of total QTLs
CRBL	**2,919**	**71.5%**	**1,164**	**28.5%**	**4,083**	**25%**
WHMT	1,735	87.1%	256	12.9%	1,991	12%
TCTX	1,301	44.5%	1,620	55.5%	2,921	18%
FCTX	994	59.7%	670	40.3%	1,664	10%
HIPP	784	46.0%	921	54.0%	1,705	10%
OCTX	774	69.5%	340	30.5%	1,114	7%
THAL	630	71.4%	252	28.6%	882	5%
MEDU	620	79.5%	160	20.5%	780	5%
PUTM	426	57.1%	320	42.9%	746	5%
SNIG	307	57.4%	228	42.6%	535	3%
Total	**10,490**	**63.9%**	**5,931**	**36.1%**	**16,421**	**100%**

This table shows the numbers and percentages for the expression QTL signals that were mapped in a specific CNS region.

Table 4.3.10: Significant expression QTLs that were specifically mapped in two CNS regions.

CNS regions	CRBL									
CRBL	**4083**	**WHMT**								
WHMT	91	**1991**	**TCTX**							
TCTX	71	42	**2921**	**FCTX**						
FCTX	62	23	96	**1664**	**HIPP**					
HIPP	41	47	35	23	**1705**	**OCTX**				
OCTX	57	6	33	35	4	**1114**	**THAL**			
THAL	13	23	25	19	10	8	**882**	**MEDU**		
MEDU	13	98	16	11	11	10	12	**780**	**PUTM**	
PUTM	6	8	7	11	9	9	8	1	**746**	**SNIG**
SNIG	7	15	4	2	7	2	1	7	1	**535**

The table shows the number of significant expression QTL signals exclusively in a pair of regions. The number of mapped QTL for that pair of regions is specified by the row and column position in the table. The yellow diagonal cells are presenting the number of QTL signals that were mapped exclusively in a named region.

4.3.8 Examples of brain expression QTLs

In this section three examples were selected to be presented. The QTLs were selected to demonstrate an example of a region specific expression QTL and a follow up of GWAS associations reported for the particular gene. Expression QTLs for *MAPT* and *LRRK2* genes will be investigated in detail in the next chapters.

The first example is a region specific expression QTL signal mapped between an intergenic SNP (rs4938050) at genomic position chr11:113590031 (human genome build 19) and *TMPRSS5* (Transmembrane protease, serine 5) gene (Affymetrix transcript 3391724). The SNP is located 12,424 kb downstream of the 3' end of the *TMPRSS5* gene and the minor homozygous allele of this SNP is associated with higher expression of the gene mainly in WHMT ($P = 7.9 \times 10^{-17}$, at FDR 1%) (**Figure 4.3.20**). *TMPRSS5* encodes a serine protease family protein (spinesin), which is expressed specifically in the human CNS and mainly in neurons and axons. In addition, this protein is located in synaptic regions (Yamaguchi, Okui et al. 2002).

The second example is a region specific expression QTL signal that was detected between an intronic SNP (rs6661489) at genomic position chr1:207698044 (human genome build 19) and *CR1* (complement component (3b/4b) receptor 1) gene (Affymetrix transcript 2377332). The SNP is within intron 6-5 of the *CR1* gene and the minor homozygous allele of this SNP is associated with higher expression of the gene mainly in WHMT ($P = 1.5 \times 10^{-03}$, at FDR 1%) followed by the HIPP region (**Figure 4.3.21)**. From the expression results, *CR1* was enriched in WHMT compared with other CNS regions. This QTL is in strong linkage disequilibrium (LD) ($R^2 = 0.87$) with two intronic SNPs rs3818361, $P = 4.0 \times 10^{-14}$ and rs6701713, $P = 4.6 \times 10^{-10}$ that were

reported to be associated with late onset Alzaheimer's disease (LOAD) in GWAS (Hollingworth, Harold et al. 2011; Naj, Jun et al. 2011).

The last example is an expression QTL signal that was between an intergenic SNP (rs10968921) at genomic position chr9:29071728 (human genome build 19) and *LINGO2* (Leucine rich repeat and Ig domain containing 2) gene (Affymetrix transcript 3202528). The SNP is located 14,049 kb upstream of the 5` end of the *LINGO2* gene and the minor homozygous allele of this SNP is associated with higher expression of the gene only in CRBL ($P = 1.6 \times 10^{-33}$, at FDR 1%) (**Figure 4.3.22)**. This gene locus was reported to be associated with Parkinson's disease from a GWAS on an Asian population (Wu, Prakash et al. 2011).

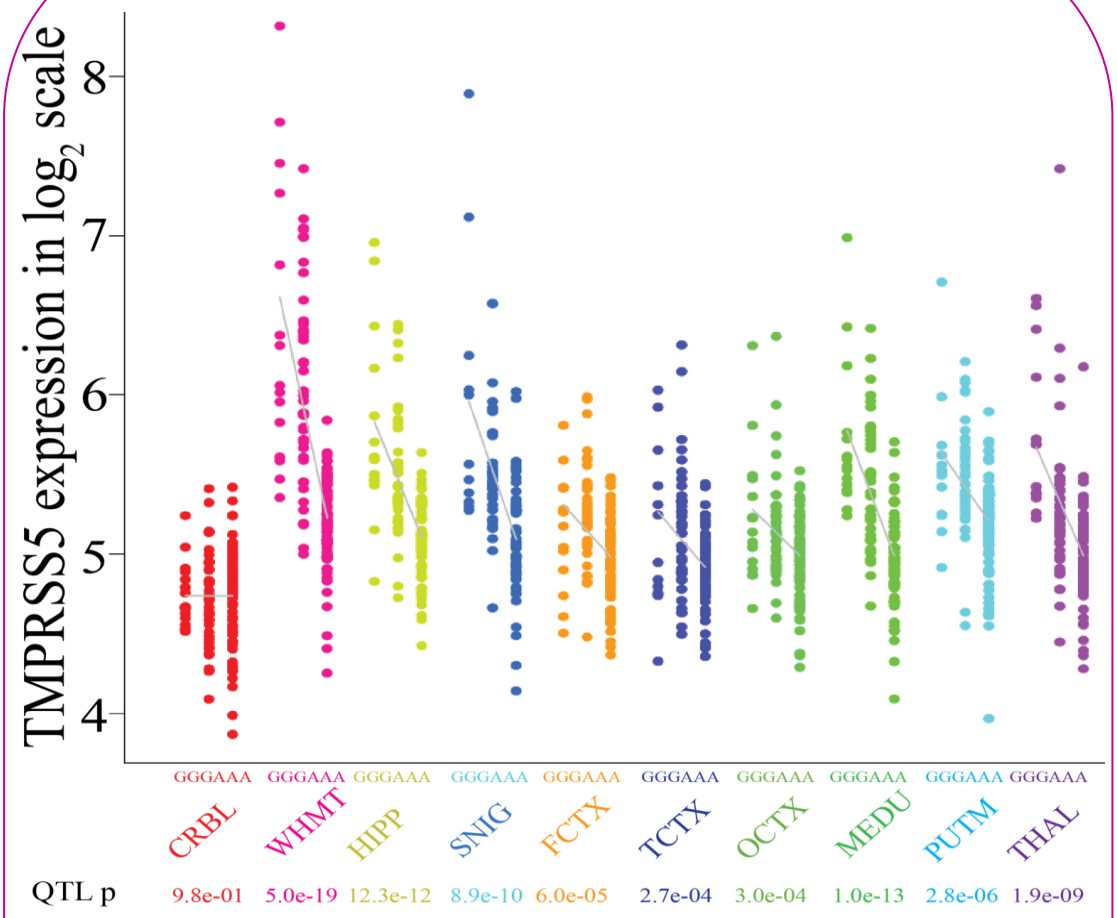

Figure 4.3.20:

The effect of rs4938050 on the expression of *TMPRSS5* transcript. CRBL = cerebellum, WHMT = white matter, HIPP = hippocampus, SNIG = *substantia nigra*, FCTX = frontal cortex, TCTX = temporal cortex, OCTX = occipital cortex, MEDU = medulla, PUTM = putamen and THAL = thalamus.

TMPRSS5 expression (Affymetrix transcript 3391724) stratified by rs4938050 for ten brain regions in 134 brain samples. Increased expression was associated with the homozygous minor allele (GG) in WHMT. A similar pattern was observed in all brain regions, but not as significantly.

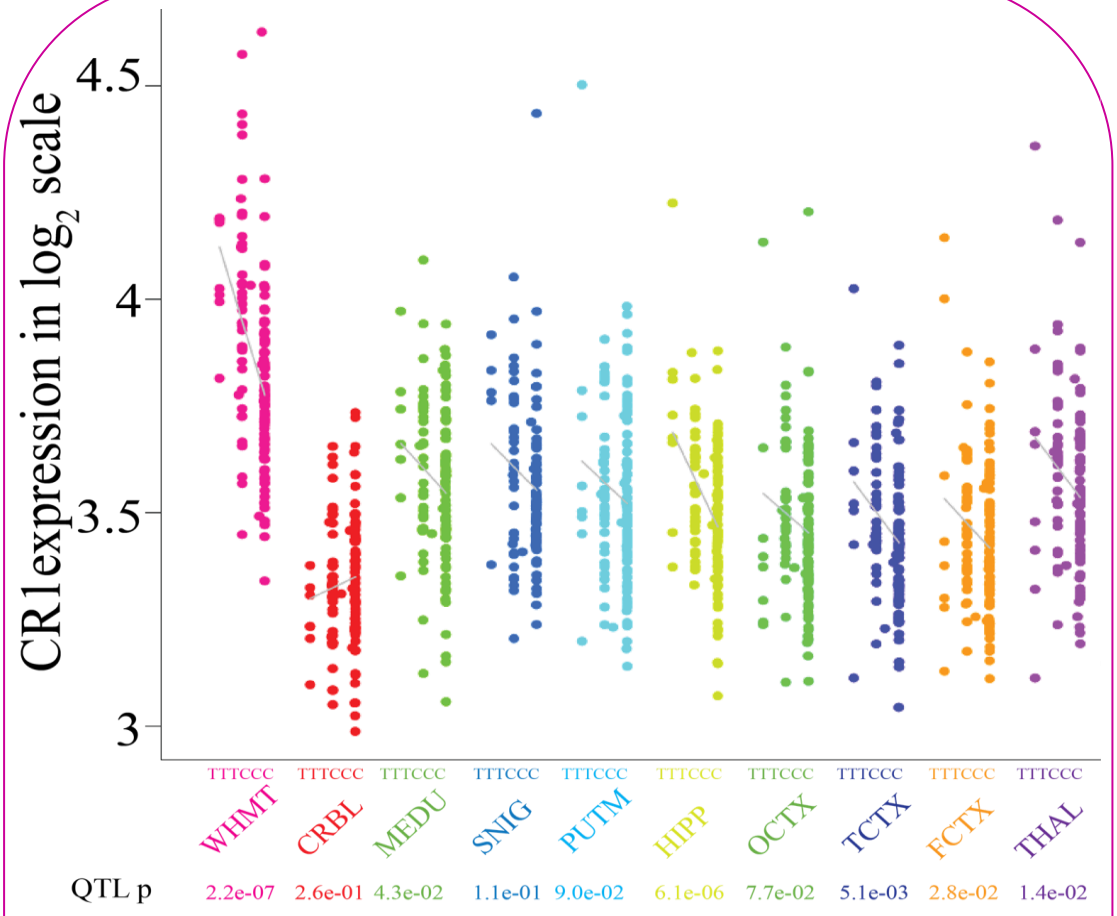

Figure 4.3.21:
The effect of rs6661489 on the expression of *CR1* transcript. WHMT = white matter, CRBL = cerebellum, MEDU = medulla, SNIG = *substantia nigra*, PUTM = putamen, HIPP = hippocampus, OCTX = occipital cortex, TCTX = temporal cortex, FCTX = frontal cortex and THAL = thalamus.

CR1 expression (Affymetrix transcript 2377332) stratified by rs6661489 for ten brain regions in 134 brain samples. Increased expression was associated with the homozygous minor allele (TT) in WHMT. A similar pattern was observed in all brain regions, but not as significantly with the exception of CRBL.

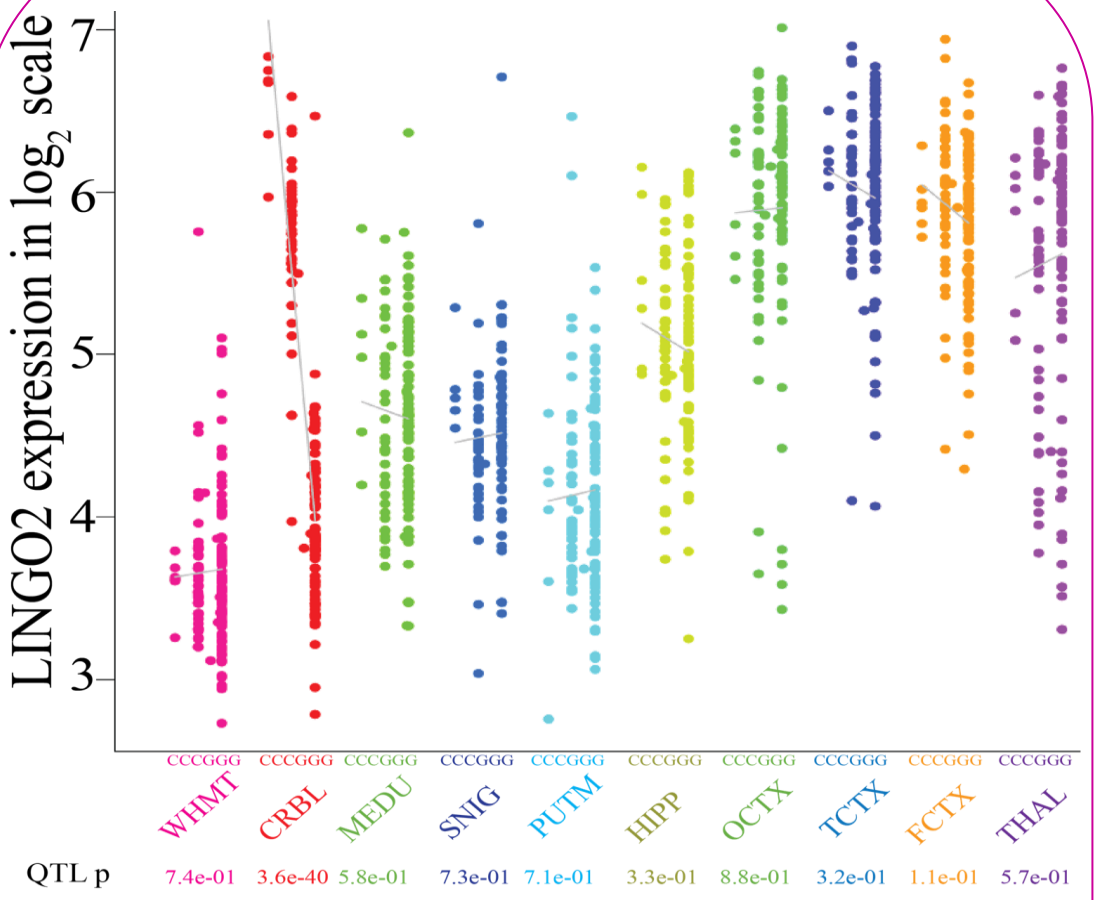

Figure 4.3.22:

The effect of rs10968921 on the expression of *LINGO2* transcript. WHMT = white matter, CRBL = cerebellum, MEDU = medulla, SNIG = *substantia nigra*, PUTM = putamen, HIPP = hippocampus, OCTX = occipital cortex, TCTX = temporal cortex, FCTX = frontal cortex and THAL = thalamus.

LINGO2 expression (Affymetrix transcript 3202528) stratified by rs10968921 for ten brain regions in 134 brain samples. Increased expression was associated with the homozygous minor allele (CC) in CRBL. A similar pattern, not significant was observed in all other brain regions, but not in THAL, PUTM and SNIG.

4.4 Discussion

In this study the transcriptome expression and splicing pattern of the human CNS was investigated in 12 CNS regions. This analysis was based on 101 individuals (917 processed Affymetrix exon arrays) which were a subset of the UKBEC. Furthermore, genome-wide QTL analyses were performed for 10 brain regions of 137 individuals of the UKBEC. This data set is a unique cohort of post mortem human brain samples to complement the other human CNS studies in the field (Myers, Gibbs et al. 2007; Gibbs, van der Brug et al. 2010; Kang, Kawasawa et al. 2011; Hawrylycz, Lein et al. 2012). It is essential to realize that precise mapping of the human transcriptome in different CNS regions will greatly assist in the understanding of region specific pathology in neurological diseases.

The results revealed that using the DABG method to detect region specific genes not only helped to reduce the background noise in the expression signals, but also reduced the false positive alternative splicing signals that might be over detected by the alternative splicing ANOVA module that was used in the downstream analysis. In this study although the DABG filtering threshold used is different from previous studies, there was an overlap in the region specific genes that were reported (Johnson, Kawasawa et al. 2009; Kang, Kawasawa et al. 2011; Hawrylycz, Lein et al. 2012). It is important to be aware that applying different statistical thresholds in these types of analyses will change the detected transcript list that is generated for further analysis, in addition to the transcript alternative splicing detection proportions (Whistler, Chiang et al. 2010). However, some of the examples that were discussed in the result sections confirmed the validity of the array results. For example, *CBLN3* was specific for CRBL, *SLC18A2* and *SLC6A3* for SNIG and *PLP1* for WHMT and SPCO.

CRBL was the region with the most distinctive gene specific and differential expression patterns compared with other CNS regions (Roth, Hevezi et al. 2006; Johnson, Kawasawa et al. 2009; Gibbs, van der Brug et al. 2010). 8% of the detected genes were CRBL specific. These results might be expected, considering the unique anatomical microscopic structure that reflects the special biochemical and cellular functions of CRBL (Llinas RR 2004; Azevedo, Carvalho et al. 2009). Additionally, WHMT is the second region after CRBL that showed a differential expression pattern, despite the fact that this region showed the lowest number of detected genes between the other 11 regions (57%, **Table 4.3.1**). This may be due to the nature of this region as the axonal mRNA expression pattern is different as compared with the neuronal body expression pattern (Giuditta, Chun et al. 2008).

Region specific expression results suggest insights for certain genes in relation to some diseases. For example, the *IGHMBP2* gene is linked with SMARD1 and the existing diagnostic protocol, involves a peripheral nerve biopsy with no requirement to scan regions other than SPCO. This gene was not detected in SPCO in this study, possibly because the RNA samples were extracted from the cell bodies and not from the nerve axons or synaptic regions. However, having this gene enriched in CRBL might open the way for clinicians to look at CRBL pathology and scan the patient's brain for this peripheral nerve disease. A previous study highlighted the concept that the molecular pathology of a disease may not be manifested as or coupled with certain histolopathological profiles or display the usual clinical symptoms (Hodges, Strand et al. 2006).

Remarkably, 96% of the detected genes were identified as alternatively spliced between the 12 CNS regions. However, in some cases this pattern reflects differential exon expression usage rather than full isoform detection.

Comparing the mRNA expression and splicing profiles in the cortical regions provided high resolution mapping between these regions, especially in the OCTX. In summary, the splicing patterns in these regions were the main patterns leading to different functions. This suggests that the pathological mechanism of diseases affecting cortical areas can be through the alteration of expression and/or splicing patterns, which leads to the alteration of the expression ratio between different isoforms. One of the examples provided is the *CNTNAP2* gene, which was related to autism, epilepsy, schizophrenia, and mental retardation diseases (Alarcon, Abrahams et al. 2008; Friedman, Vrijenhoek et al. 2008). Considering this gene was highly expressed in all cortical regions (FCTX, TCTX and OCTX), the regions that are mainly affected by these diseases (Tan, Doke et al. 2010) and responsible for language processing in the human CNS, alteration of the expression patterns may contribute to the pathology of the cortical areas. This hypothesis needs further validation to understand the mechanisms underlying the pathology. Similar hypotheses can be applied for *ANO3, CR1* and *BIN1* in relation to AD as they are differentially expressed in different CNS regions, but their splicing patterns have not been investigated in detail.

As a means of moving forward, this approach can be applied to each possible comparison of all other CNS regions to explore and understand how differential expression and splicing patterns will affect gene connection circuits and the biological processes. These differential expression profiles may play an important role in understanding the vulnerability of certain brain regions or cells in neurological disorders.

The results of the alternative splicing patterns suggest some functional importance for certain isoforms in specific CNS regions. *NRXN3* gene encodes for proteins involved in specifying synaptic functions and mediating signalling across the synapse. An alternation in this gene was reported to be associated with autism and cognitive diseases (Sudhof 2008). Different ratios of alpha-*NRXN3* and beta-*NRXN3* isoforms in different brain regions suggest different functions and activities (Craig and Kang 2007). Therefore, any alteration in the right balance of these isoforms may contribute to the pathology of the disease in specific regions. A similar concept can be applied for the *SETX* gene. This gene is not differentially expressed but differentially spliced between CRBL and SPCO regions, which are the main regions to be affected in the AOA2 and ALS4. These results suggest that *SETX* contributes to disease pathology by having different spliced isoforms in the affected regions at different ratios, thus altering gene connectivity and function.

The last example of this group is the *BIN1* gene. A previous study demonstrated that WHMT is one of the regions most involved in AD (Hua, Leow et al. 2008). The results showed an enrichment of the gene in WHMT and different splicing isoforms were also demonstrated and confirmed in this region compared to other regions. The protein product of *BIN1* is involved in regulating synaptic vesicle endocytosis (Wigge and McMahon 1998), so alterations in the isoform balance could lead to misregulation of this process and contribute to the disease aetiology.

The initial results for the expression QTL analysis showed that region specific expression QTL signals accounted for 84% of the total number of expression QTLs. CRBL was the most enriched regions for expression QTLs signals and that is

consistent with the expression pattern in this region confirming the unique functions of the CRBL.

Genome-wide expression QTL mapping was discussed mainly as *cis*-acting signals (69.2%). *Trans*-acting expression QTL signals were reported, but not investigated in this thesis. Although the interpretation of expression QTLs has to be done carefully (Peirce, Li et al. 2006), the results showed significant region specific expression QTL signals which are and related to neurological diseases. This was demonstrated by the examples that were selected in the results section. *CR1* is associated with AD and *LINGO2* is associated with development of PD, providing the evidence that risk SNPs for diseases have a functional impact on expression and on splicing. Therefore a careful investigation of this is an important step forward following the GWAS results. In addition, this will help to focus attention on certain regions for the investigation of specific diseases. However, these findings have to be further validated with different experimental techniques to increase our understanding of the molecular mechanisms underlying the expression QTLs and the impact on cell function.

Furthermore, the expression mechanisms underlying QTLs are not fully understood and there are likely to be more expression QTLs to be found and investigated. More research is required with larger sample sizes to be able to capture more expression QTL signals with smaller effect sizes. This is also true for *trans*-acting expression QTL signals.

In conclusion, these results are considered to be the first of several stages to understand the genetic basis of variation in gene expression in the control human brain, and neurodegenerative diseases. This stage includes generating the data, applying

quality control parameters, performing different ANOVA modules, running association tests, multiple test correction, mining the data to a scalable level, and comprehensive investigation with specific hypotheses for genes of interest. Following on from this work, it is important to undertake further experiments using different samples, cell models, advanced techniques and bioinformatic tools, especially for the expression QTL signals. In this way the functional mechanisms of QTL signals will be understood more clearly.

5 Characterization of *MAPT* expression and splicing in control multiple human brain regions: in relation to neurodegenerative diseases

5.1 Summary

The *MAPT* (microtubule-associated protein tau) locus is one of the most significant genetic markers contributing to some neurodegenerative diseases such as Parksinson's disease (PD). Moreover, the tau protein is recognized as a key pathological feature of neurodegenerative disorders; including progressive supranuclear palsy (PSP), corticobasal degeneration (CBD) and Alzheimer's disease (AD). In addition to the linkage of the *MAPT* locus to disease, the chromosomal region around *MAPT* has been intensely studied because of the polymorphic inversion and complex alternative splicing patterns that make this region of the genome particularly complicated with regard to its genetic architecture. Thus, generating detailed data regarding the expression, splicing and genetic regulation of this transcript within the human brain is of great importance.

In order to generate the required data, 2,011 brain samples originating from 439 control individuals from ten brain regions were used from both UKBEC and NABEC datasets to provide the most reliable and comprehensive results on the regional expression, splicing and regulation of *MAPT* to date.

The results revealed significant regional variation in mRNA expression levels and splicing patterns of *MAPT* in the control human brain. The gene expression level, the regional distribution of mRNA expression and total tau protein expression levels were largely in agreement, appearing to be highly correlated. Finally, the H1/H2 SNP showed an association with expression of exon 3-containing isoforms in the control human brain.

This data suggests that the genetic risk factors for neurodegenerative diseases at the *MAPT* locus are likely to operate through changing mRNA splicing patterns in different brain regions, as opposed to the overall mRNA expression of the *MAPT* gene.

5.2 Introduction

MAPT (microtubule-associated protein tau) locus is one of the most remarkable genes in neurogenetics due not only to its involvement in multiple neurodegenerative disorders such as PSP (Baker, Litvan et al. 1999; Hoglinger, Melhem et al. 2011), CBD (Houlden, Baker et al. 2001), PD (Golbe, Lazzarini et al. 2001; Nalls, Plagnol et al. 2011; Lill, Roehr et al. 2012) and possibly AD (Myers, Kaleem et al. 2005; Gerrish, Russo et al. 2012), but also due to its genetic evolution and complex alternative splicing features. These features are to some extent linked, making the role of *MAPT* in health and disease all the more intriguing (Goedert, Spillantini et al. 1989; Goedert, Spillantini et al. 1989; Stefansson, Helgason et al. 2005).

The evolution of the *MAPT* locus has been comprehensively studied, with an inversion polymorphism on chromosome 17q21. While the majority of individuals inherit this region in the direct orientation, up to 25% of individuals of European-Caucasian-descent have ~970 kb sequence in the inverted orientation (Stefansson, Helgason et al. 2005; Zody, Jiang et al. 2008), introducing a larger ~1.5 Mb region of linkage disequilibrium (Pittman, Myers et al. 2004; Zody, Jiang et al. 2008). This sequence appears, in Europeans at least, to descend from a single founder (Pittman, Myers et al. 2004; Fung, Evans et al. 2005). The common haplotype clades marking the direct (majority) and inverted orientated sequences are termed H1 and H2, respectively. It is not clear which of these sequences is the ancestral orientation because the polymorphism exists in other primate species and the rodent sequence is in the H2 orientation (Stefansson, Helgason et al. 2005). It is worth noting that since this inversion polymorphism excludes recombination over a region of ~1.5 Mb, haplotype-specific polymorphisms have arisen. Genetic studies, including genome-wide

association, have demonstrated the importance of both the inversion polymorphism and inversion-independent polymorphisms in disease. In fact, two distinct types of disease association have been demonstrated, of which the first is the association of the H1 haplotype with an increased risk of PD (odds ratio ~1.7) (Golbe, Lazzarini et al. 2001; Nalls, Plagnol et al. 2011; Lill, Roehr et al. 2012), PSP (odds ratio 5.5) (Baker, Litvan et al. 1999; Hoglinger, Melhem et al. 2011) and CBD (odds ratio ~5) (Houlden, Baker et al. 2001). The second is the association of the H1c haplotype, one of the sub-haplotypes within the H1 clade, with an increased risk of PSP alone (odds ratio ~1.5) (Pittman, Myers et al. 2005; Hoglinger, Melhem et al. 2011)

The importance of the tau protein in neurodegenerative disease has been well documented even before the recognition of the genetic role of *MAPT* genetics in these diseases. Tau is expressed in the adult human central nervous system and tau pathology. Specifically, neurofibrillary tangles (NFTs), are a notable pathological feature of a range of neurodegenerative disorders, including PSP, CBD (Feany, Mattiace et al. 1996) and AD (Goedert, Spillantini et al. 1989). As well as having distinct clinical features, these diseases also have distinct and overlapping profiles of severity of tau pathology within the human brain. These profiles of severity are summarised in **Table 5.2.1** and show that in AD, a disease characterised by early memory loss and difficulties in executive functions, tangle pathology is most prominent in brain regions that are highly associated with these processes, namely the hippocampus, temporal cortex and frontal cortex. These findings have led researchers to investigate whether basal regional expression of *MAPT* mRNA and protein may predispose some brain regions to a higher risk of tau pathology than others and there is evidence in support of this (Feany, Mattiace et al. 1996; Luk, Giovannoni et al. 2009).

Table 5.2.1: Severity distribution of tangle pathology in human brain (Trabzuni, Wray et al. 2012).

Brain region	AD	PSP	CBD	PD
FCTX	3	2	3	1
TCTX	3	1	2	2
OCTX	2	0	0	0
WHMT	0	1	3	0
HIPP	3	2	2	2/3
PUTM	1	2	3	0
THAL	2	1	2	1
HYPO	1	2	2	1
SNIG	1	2	3	0
MEDU	1	2	2	0
CRBL	0	1	1	0
SPCO	0/1	2	2	0

Severity ranking of the distribution of tau pathology within the human brain in Alzheimer's disease (AD), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD) and Parkinson's disease (PD). The data represent average severity of pathology in the following way: 0 = none, 1 = mild, 2 = moderate and 3 = severe (Tamas Revesz, derived from (Braak and Braak 1991; Compta, Parkkinen et al. 2011; Dickson, Hauw et al. 2011)).

However, there is an increasing appreciation of the importance of not only assessing total *MAPT* expression, but also alternative splicing. The *MAPT* gene is extensively spliced to produce 12 mRNA transcripts (according to Ensembl / Havana annotation) and seven protein isoforms. Concentrating on the known protein products within human brain (of which six are recognised), cassette splicing of exons 2 and 3 (E2 and E3), give rise to tau isoforms with 0, 1 or 2 amino-terminal repeats (0N-, 1N-

or 2N-tau) and alternative splicing of exon 10 (E10), yields isoforms with 3- or 4-microtubule binding repeats in the carboxy-terminal half of tau (3R- or 4R-tau) (Goedert, Spillantini et al. 1989). Interestingly, detailed investigations by a number of groups have demonstrated that neurofibrillary tangles in different diseases have a different isoform composition, suggesting that splicing is of key importance in the neuropathological process. Whereas the tangles found in PSP and CBD consist predominantly of 4R-tau (due to exon 10 inclusion), those found in AD contain both 3R- and 4R-tau (Sergeant, Wattez et al. 1999; Goedert 2004). Furthermore, consistent with the findings on total tau, it was previously shown that regions of relatively high 4R-tau in PSP are more susceptible to tau related pathology and neurodegeneration (Feany, Mattiace et al. 1996; Luk, Giovannoni et al. 2009).

Understanding the expression, splicing and regulation of the mRNA and protein isoforms is of key importance to the field. At present, no comprehensive study of this type has been performed. In order to address this, the data from the UKBEC; the largest exon-specific expression data set currently available, were used (Trabzuni, Ryten et al. 2011), containing up to 10 distinct brain regions (including hippocampus and *substantia nigra*): FCTX, *n*= 127, TCTX, *n*= 119, OCTX, *n*= 129, HIPP, *n*= 122, THAL, *n*= 124, PUTM, *n*= 129, SNIG, *n*= 101, MEDU, *n*= 119, CRBL, *n*= 130 and WHMT, *n*= 131 (1,231 Affymetrix exon arrays) (Trabzuni, Ryten et al. 2011). In addition, expression QTL analysis was performed to understand the effects of both the H1 and H1c risk haplotypes on *MAPT* expression and splicing. In order to increase the statistical power, the analysis was performed using a second data set originating from 390 neuropathologically normal individuals from the North American Brain Expression Consortium (NABEC) (See **Table 5.2.2**) (Gibbs, van der Brug et al. 2010).

Table 5.2.2: The number of brains studied for each major type of analysis conducted in both data sets. Table modified from (Trabzuni, Wray et al. 2012).

Analysis	Data set	Number
Regional distribution and splicing of *MAPT* mRNA	UKBEC	**134**
The effect of H1/H2 haplotypes on mRNA expression and splicing		
The effect of the H1c haplotype on *MAPT* mRNA expression	NABEC + UKBEC	**305** **85**

5.3 Results

5.3.1 Regional expression and splicing of *MAPT* mRNA in multiple control human brain regions

The tissue samples from 10 brain regions that were originated from 134 individuals from the UKBEC were profiled on 1,231 Affymetrix Human Exon 1.0 ST arrays. The regional distribution of *MAPT* mRNA expression at the gene level is shown in **Figure 5.3.1**. This demonstrated significant regional differences in *MAPT* mRNA expression with a 1.5 fold difference ($P = 5.7 \times 10^{-49}$) between frontal cortex, the highest *MAPT* expressing region, and the white matter, the lowest (**Figure 5.3.1**). Regional differences in *MAPT* mRNA expression were validated on a subset of 12 individuals in four brain regions (CRBL, OCTX, PUTM and WHMT) using QuantiGene (QG), which showed a similar pattern. The results of QG were shown in detail in Chapter 3 **Figure 3.3.6** (Trabzuni, Ryten et al. 2011) including for *MAPT* mRNA expression.

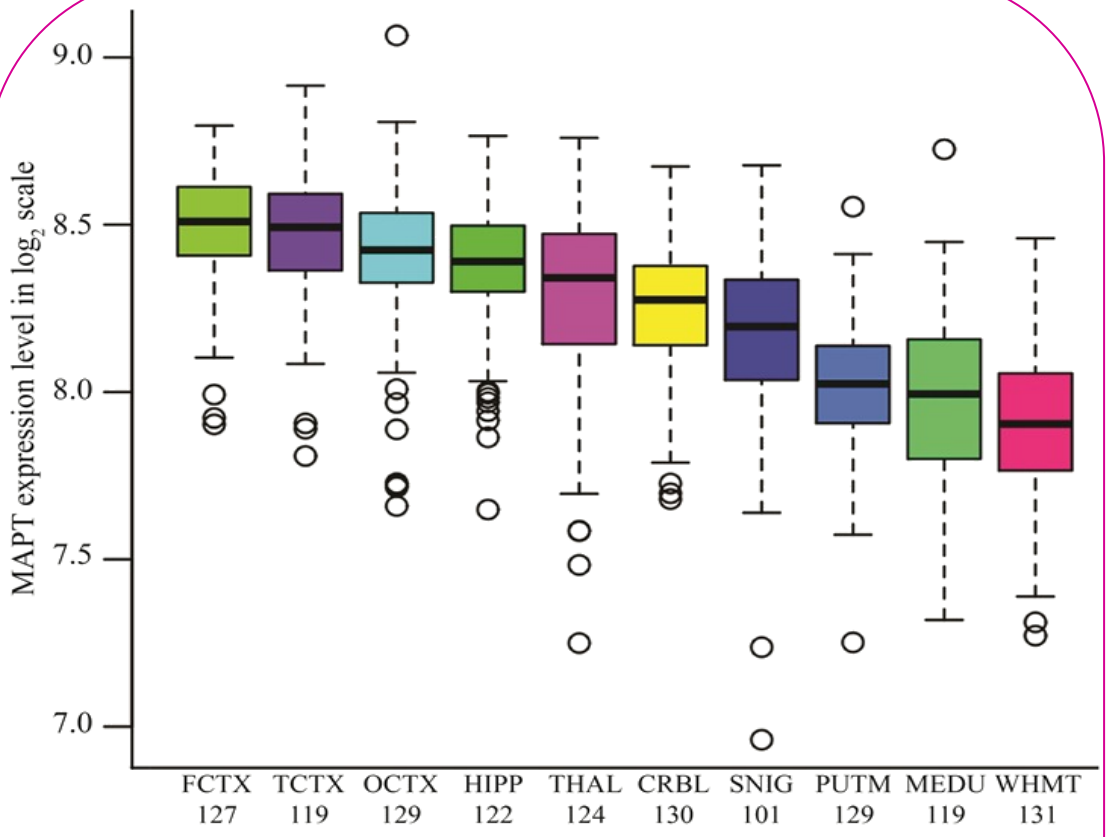

Figure 5.3.1:

Regional Distribution of *MAPT* mRNA expression. FCTX = frontal cortex, TCTX = temporal cortex, OCTX = occipital cortex, HIPP = hippocampus, THAL = thalamus, CRBL = cerebellum, SNIG = *substantia nigra*, PUTM = putamen, MEDU = medulla and WHMT= white matter.

Boxplot of mRNA expression levels for *MAPT* in 10 brain regions, from microarray experiments on a log2 scale (*y*-axis). This plot shows the regional variation in *MAPT* transcript expression across all the regions. Whiskers extend from the box to 1.5 times the inter-quartile range (Trabzuni, Wray et al. 2012).

The unique design of the Affymetrix Exon arrays, with probe sets targeted against individual exons, also allowed us to investigate regional differences in *MAPT* mRNA exon splicing (ANOVA alternative splicing $P < 1\times10^{-45}$, **Figure 5.3.2, C**). This demonstrated a relative reduction in the expression of exon 2 in white matter as compared to other brain regions, suggesting lower expression of exon 2-containing isoforms specifically in this tissue. Similarly, there was a relative increase in the expression of exon 6 in cerebellum as compared to other brain regions, suggesting higher expression of exon 6 containing isoforms in this tissue (**Figure 5.3.2, C**). Alternate splicing of exon 2 in white matter was confirmed by TaqMan assays for 0N (E2-, 3-), 1N (E2+, 3-) and 2N (E2+, 3+) transcripts (**Figure 5.3.3**). These assays demonstrated a selective reduction of exon 2 containing transcripts (2N-tau and 1N-tau) in white matter, whereas 0N-tau isoforms, which do not contain exon 2 or 3, were unchanged between these selected regions.

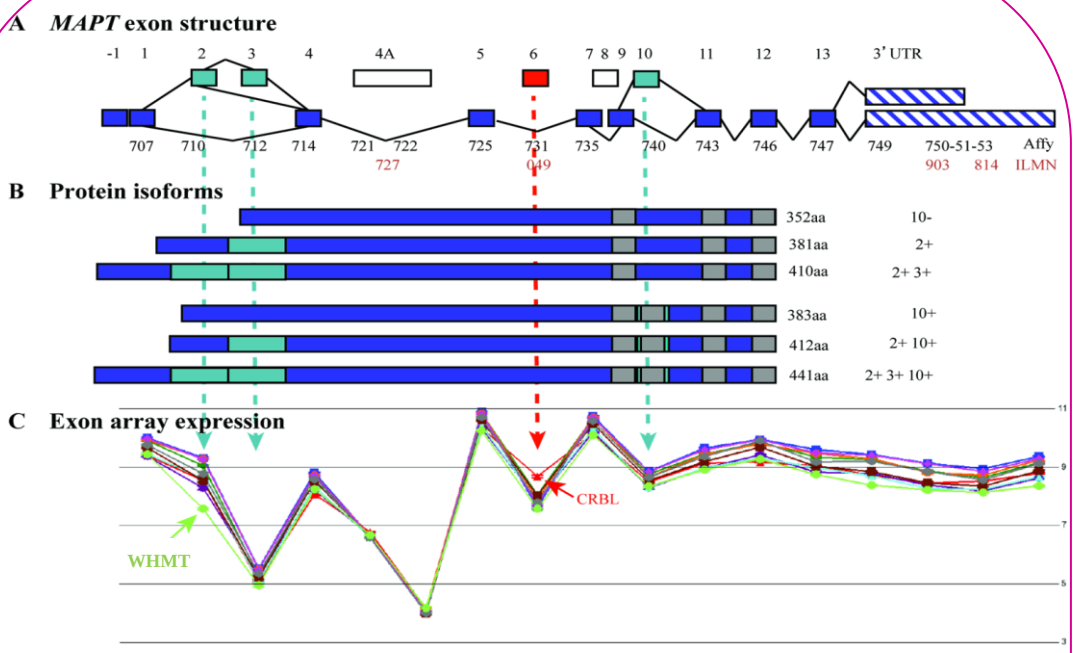

Figure 5.3.2:

***MAPT* Gene/exon Structure, position of expression probes and array expression.** CRBL = cerebellum and WHMT = white matter.

(A, upper panel) The structure of the *MAPT* gene. The position of Affymetrix exon array probe sets (from probe set 3723707 to 3723753) and probe sets of Illumina transcript expression array (ILMN_2298727, ILMN_1800049, ILMN_1710903 and ILMN_2310814) are indicated within the *MAPT* exons. **(B, middle panel)** The tau protein isoforms expressed in human brain. Alternative splicing of exons 2, 3 and 10 generates six protein isoforms of tau with either 0, 1 or 2 N-terminal repeats and 3 or 4 C-terminal microtubule binding repeats. **(C, lower panel)** Exon array data shows the expression of *MAPT* expression at the exon level matching the upper panel of the exon structure. The mean expression levels were (*y*-axis, log2 scale) were plotted for each probe set (*x*-axis) covering the whole transcript. Note that levels are not absolute since they depend upon the hybridisation efficiency of individual probe sets. It is, therefore, not possible to compare the expression between exons. Differences in splicing between brain regions are manifest through the non-parallelism between the expression lines. Two examples suggest evidence of splicing are: at exon 2, the WHMT (light green line) shows less inclusion of this exon than would be expected and at exon 6, CRBL (red line) shows more inclusion (Trabzuni, Wray et al. 2012).

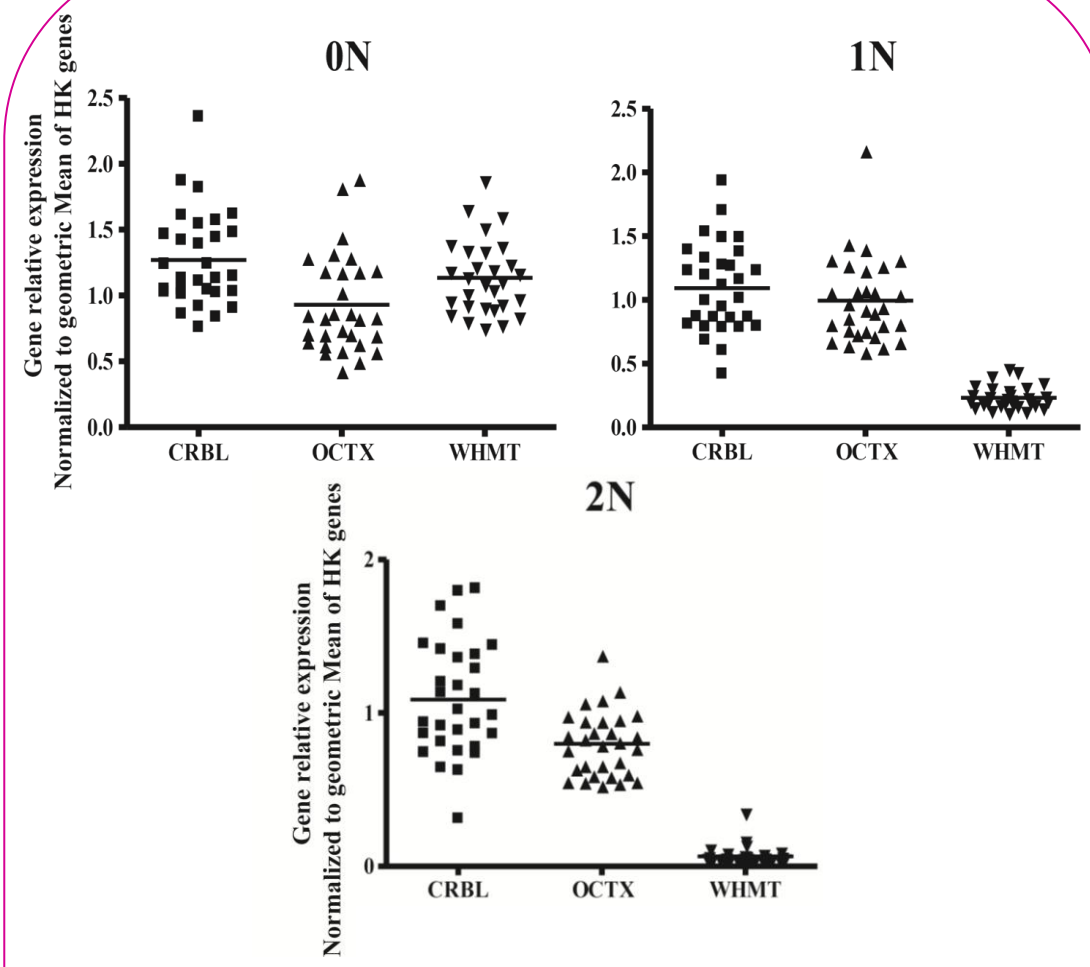

Figure 5.3.3:

***MAPT* qRT-PCR validation using TaqMan assay for different isoforms.** CRBL = cerebellum, OCTX = occipital cortex and WHMT = white matter.

This graph shows the confirmation of the differential splicing pattern of exon 2 in WHMT as compared to CRBL and OCTX. Whereas exon 2 containing isoforms, 1N (exon 2+, 3-) and 2N (exon 2+, 3+), were expressed at significantly lower levels in white matter as compared to occipital cortex and cerebellar cortex, no differences in 0N (exon 2-, 3-) tau transcripts were observed (Trabzuni, Wray et al. 2012) (Data produced by Dr Jana Vandrovcova).

5.3.2 Regional expression of tau protein in multiple control human brain regions

The relationship between mRNA and protein expression is complex. However, establishing the nature of this relationship in the case of *MAPT*/tau is critical in interpreting reported expression quantitative trait loci, which have largely depended on mRNA expression levels alone. The regional variability in total tau protein expression levels in five brain regions (CRBL, FCTX, OCTX, PUTM and WHMT) in 12 individuals were assessed in detail (**Figure 5.3.4, A and B**). This analysis showed that the FCTX has the highest tau protein levels with decreasing levels in the following order: FCTX > OCTX > WHMT > PUTM > CRBL. CRBL tau protein levels were significantly lower than all other regions examined (**Figure 5.3.4**, B, $P < 0.01$) and tau protein levels in PUTM were significantly lower than FCTX and OCTX ($P < 0.05$). These findings are largely in agreement with the mRNA expression results which showed that cortical regions express *MAPT* at the highest levels and cerebellum, putamen and white matter all have lower levels of *MAPT* mRNA.

The levels of each individual isoform of tau protein were determined relative to total tau protein level for the same sample (**Figure 5.3.4, C-H**). 1N-tau protein isoforms formed the majority of total tau (~50%), followed by 0N-tau isoforms (~40%) with 2N-tau isoforms forming the lowest proportion of tau (~10%), and levels of 3R-tau and 4R-tau isoforms were approximately equal, in agreement with previous studies (Boutajangout, Boom et al. 2004). Protein levels of the smallest tau isoform, 0N3R, were significantly lower in cerebellum ($P < 0.05$) compared to other brain regions, but other tau isoforms were not found to vary significantly between brain regions.

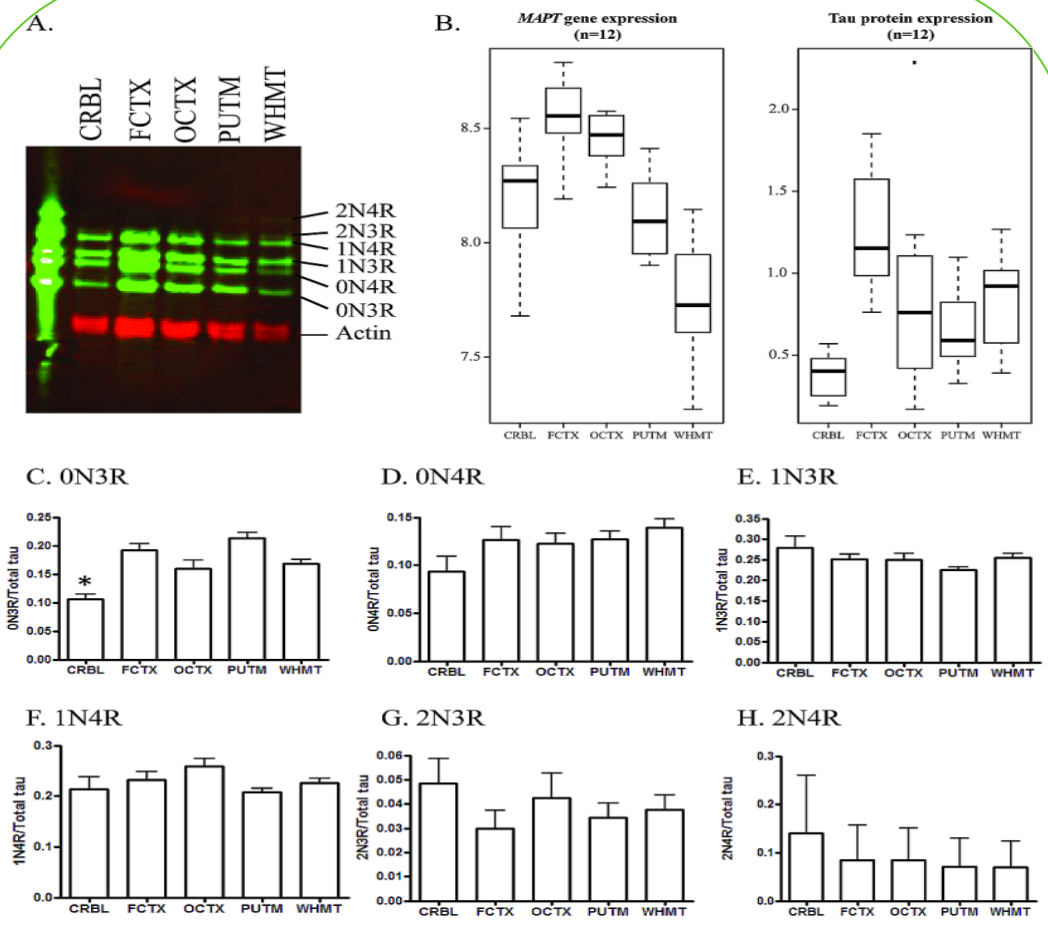

Figure 5.3.4:

Regional expression level of tau protein within multiple human brain regions. CRBL = cerebellum, FCTX = frontal cortex, OCTX = occipital cortex, PUTM = putamen and WHMT = white matter.

Soluble tau extracts were prepared from five different brain regions, separated by SDS-PAGE and detected on Western blots probed with the polyclonal antibody to total tau. **(A)** Specific tau isoforms were identified by comparing to recombinant tau ladder (n=13, representative image shown). **(B)** Comparison of total tau to actin revealed highest tau protein levels in the FCTX and significantly lower tau levels in CRBL (P <0.01) and PUTM (P< 0.05). **(C)** Levels of specific tau isoforms were determined by normalisation to total tau for the same sample and revealed a significantly lower amount of 0N3R in CRBL compared to other regions. **(D-H)** There was no significant variation in other tau isoforms (Trabzuni, Wray et al. 2012) (Data produced by Dr Selina Wray).

The 3R- and 4R-tau protein levels were measured in the same samples using sandwich ELISA assays (Luk, Giovannoni et al. 2009; Luk, Vandrovcova et al. 2010). Both 3R- and 4R-tau are lowest in CRBL and WHMT reflecting total transcript levels (**Figure 5.3.5**). In conclusion, whereas *MAPT* gene-level expression and total tau protein expression follow similar trends between brain regions, the relationship between mRNA and protein isoforms could not be determined.

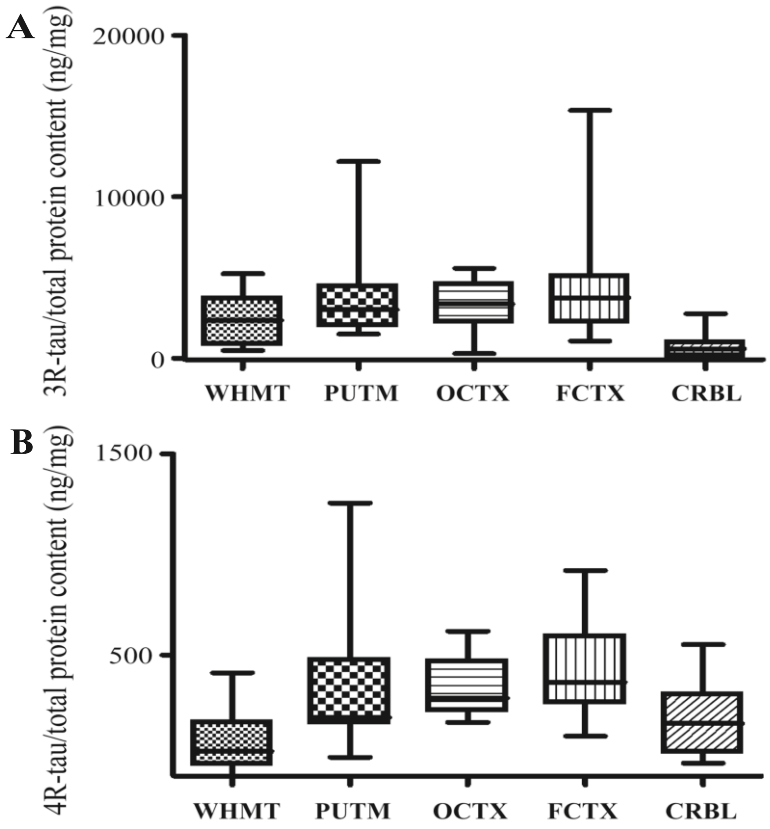

Figure 5.3.5:

The 3R- and 4R-tau protein levels in human brain using ELISA assay. WHMT = white matter, PUTM = putamen, OCTX = occipital cortex, FCTX = frontal cortex and CRBL = cerebellum.

3R and 4R-tau immunoreactivity in five brain regions (*x*-axis) of control brains, measured in ng/mg of tau of total protein (*y*-axis) by sandwich ELISA**.** Using non-parametric Kruskall-Wallis test followed by a posthoc Dunn's multiple comparisons was used to compare brain regions. **(A)** 3R-tau levels show significant differences in of CRBL vs PUTM, OCTX and FCTX (P<0.001) and CRBLs WHMT (P<0.05). **(B)** 4R-tau levels show significant differences of WHMT vs FCTX (P<0.001); WHMT vs OCTX (P<0.01) and FCTX vs CRBL (P<0.05) (Data produced by Dr Connie Luk).

5.3.3 H1/H2 QTLs effect on *MAPT* mRNA expression and splicing

Previous studies have demonstrated the effects of the H1/H2 haplotypes on *MAPT* mRNA expression (Myers, Gibbs et al. 2007; Myers, Pittman et al. 2007; Gibbs, van der Brug et al. 2010; Nalls, Plagnol et al. 2011). However, there are large numbers of haplotype-specific polymorphisms at the *MAPT* locus (both SNP and insertion/deletion polymorphisms), some of which have only recently been identified, and since expression probes have largely been designed against the H1 haplotype, these polymorphisms may have created false expression QTL results (Naiser, Ehler et al. 2008). This can occur because mRNA sequences transcribed from the H2 haplotype will depart from the H1 reference sequence (used to develop the probes), due to both the presence of different nucleotides (i.e. SNPs) and the presence or absence of nucleotides (i.e. indels), and so are likely to exhibit a weaker binding affinity for the probes in question, which are a perfect match for mRNA sequences transcribed from the H1 haplotype. This results in an apparent association between genotype and expression, confounding expression QTL studies which are looking for just such a signal. For this reason, the most recent release of the 1000 Genomes Project (March 2012: Integrated Phase I haplotype release version 3) was used to identify and remove probes on the Affymetrix Exon array that map to *MAPT* and contain SNPs or indels with a minor allele frequency (MAF) of > 1 % in Europeans. This resulted in the removal of 8 of the 25 *MAPT* probe sets from further expression QTL analysis (**Figure 5.3.6** and **Figure 5.3.2 (A)** for positions of probe sets). Exon-specific expression data that were generated from the remaining 17 probe sets (using UKBEC samples) were tested for association against 3,547 SNPs located within ~1 Mb of the transcription start and end site of *MAPT*, and this was performed separately in each brain region.

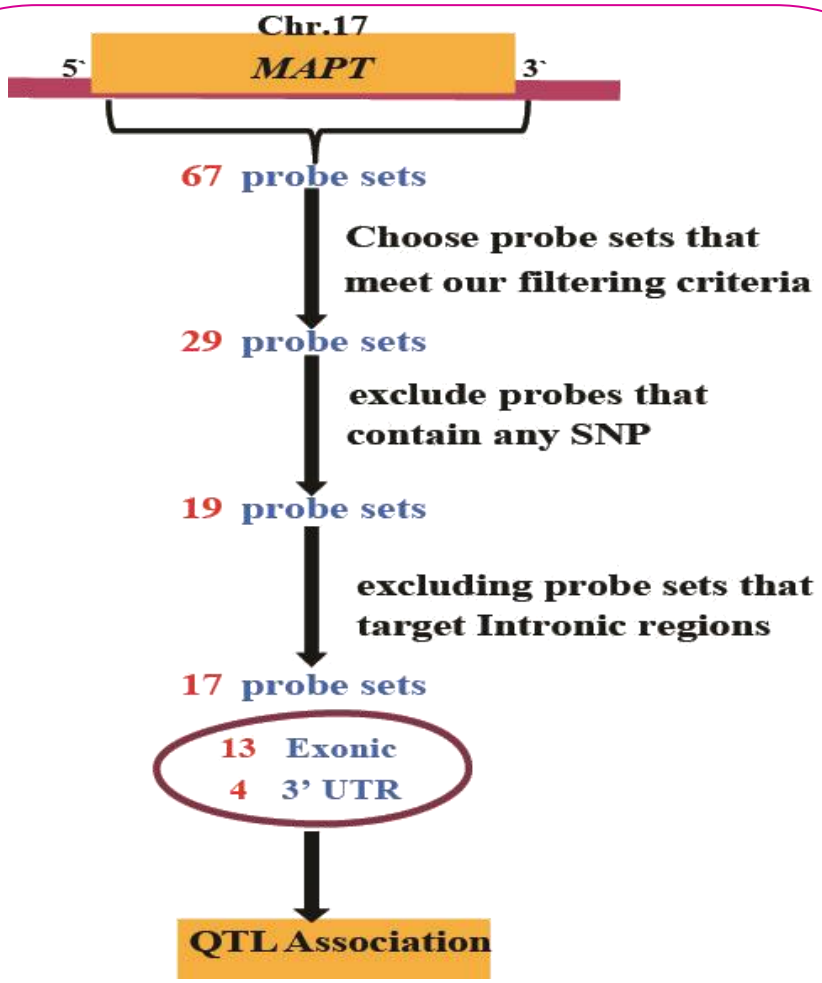

Figure 5.3.6:

Schematic view shows the process of filtering the probe sets for expression QTL association at *MAPT* transcript location.

Comprehensive investigation for the *MAPT* transcript probe sets, the analysis was limited to selected probe sets in order to have reliable and reproducible results.

The most significant expression QTL identified was between SNP rs111522045 (rs118049839) at genomic position chr17:44357157 (250 kilo-bases downstream the 3` end of the *MAPT* gene) from human genome build 19 and exon 3 (Affymetrix probe set 3723712). Since this SNP is in strong linkage disequilibrium (LD) with SNPs that tag the H1/H2 haplotypes, the known H1/H2 tagging SNP, rs17665188 (R^2 = 0.93 with rs111522045) was used to demonstrate the effect of the H2 haplotype on exon 3 expression (as detected by Affymetrix probe set 3723712)**.** The H2 haplotype was associated with higher expression of exon 3 in all brain regions (all region average $P = 5.2\times10^{-13}$) except white matter (**Figure 5.3.7**). The most significant association was observed in frontal cortex ($P = 8.8\times10^{-6}$). In contrast to the alternate splicing of exon 10, the alternate splicing of exons 2 and 3, and more specifically the role of exon 3 has not been extensively studied. However, this data refine a previous report using cell lines (Caffrey, Joachim et al. 2008) which suggested that H2 was associated with increased inclusion of exons 2 and 3, though the former was not observed ($P > 0.01$ for all tissues for exon 2, as measured by Affymetrix probe set 3723710).

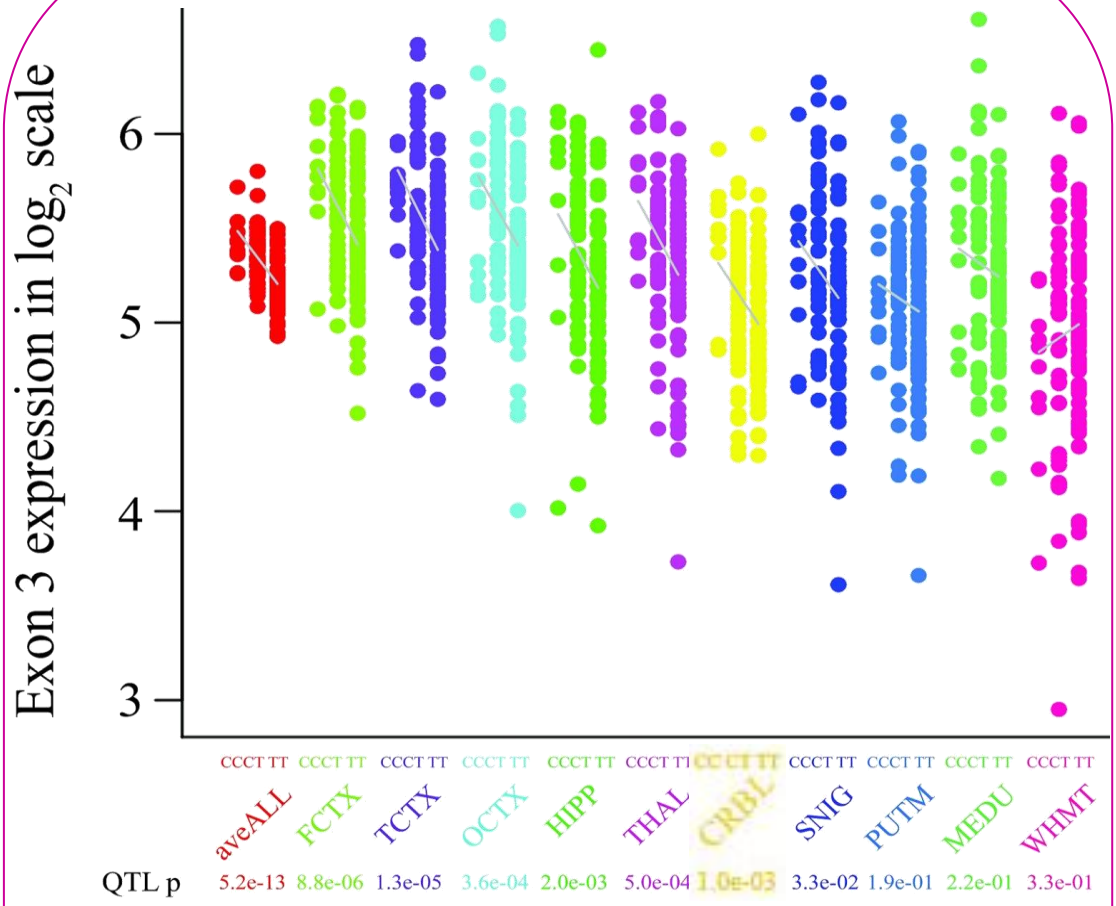

Figure 5.3.7:

The effect of the H1/H2 haplotypes (tagged by rs17665188) on the expression of *MAPT* exon 3. FCTX = frontal cortex, TCTX = temporal cortex, OCTX = occipital cortex, HIPP = hippocampus, THAL = thalamus, CRBL = cerebellum, SNIG = *substantia nigra*, PUTM = putamen, MEDU = medulla and WHMT = white matter.

MAPT exon 3 expression (Affymetrix probe set 3723712) stratified by genotype at rs17665188 for ten brain regions. Increased exon 3 expression was associated with the homozygous major allele (CC). A similar association pattern was observed in all brain regions except for WHMT (P = 0.03). All ten brain regions shown as well as combined mean across region (aveALL) (Trabzuni, Wray et al. 2012).

5.3.4 H1c haplotype (rs242557) effect on *MAPT* mRNA expression

Determining the effect of the H1c haplotype on both *MAPT* mRNA expression and splicing was performed using the Affymetrix exon arrays on the UKBEC data set and NABEC data set combined together in order to increase the number of samples for this analysis. The expression data that were derived from 390 individuals (FCTX and CRBL regions) and generated by the NABEC using the Illumina HumanHT-12 v3 Expression BeadChips were used. This Illumina expression array contains four probes mapping to the *MAPT* transcript, of which only two were detected in control human brain, ILMN_2310814 and ILMN_1710903. As previously described, in order to avoid the detection of false expression QTLs arising from the presence of SNPs within the probe sequences, both Illumina probes for SNPs and indels using the most recent release of the 1000 Genomes Project (March 2012: Integrated Phase I haplotype release version 3) were checked. This detailed quality control process helped to identify a two base pair deletion (rs67759530, MAF = 23% in Europeans) in the target sequence of Illumina probe, ILMN_1710903, specifically within the *MAPT* H2 haplotype.

However, removal of H2 haplotype-carrying individuals did allow using the data generated by both *MAPT* probes to investigate the effects of rs242557 in the remaining 222 H1/H1 individuals. This SNP defines the H1c sub-haplotype within the H1 clade and has also been shown to be a significant risk SNP for PSP (risk allele = A, $P = 9.5\times10^{-18}$ from the relevant GWAS) (Hoglinger, Melhem et al. 2011). In fact, the H1c haplotype (tagged by rs242557) was not significantly associated with increased mRNA expression of *MAPT*, as measured with Illumina probes ILMN_1710903 (*P* =

0.957 in FTCX, P = 0.825 in CRBL) and ILMN_2310814 (P = 0.975 in FTCX, P = 0.768 in CRBL).

5.4 Discussion

This analysis, was based on multiple analyses with a minimum of 780 brain samples used in any single analysis (originating from 390 individuals), provides the most comprehensive dataset on the regional mRNA expression, splicing and regulation of *MAPT* available to date. Thorough quality control steps have been used to ensure that in particular the effects of the H1/H2 and H1c haplotypes on *MAPT* mRNA expression and splicing were as accurate as possible. Thus, the results reveal significant regional variation in *MAPT* mRNA expression and splicing, validated using Quantigene and TaqMan assays, and that at the gene level, *MAPT* mRNA expression and total tau protein are highly correlated. In addition, the effect of the reported H1/H2 effect on *MAPT* mRNA gene level expression is likely to be a technical artefact; as this polymorphism is associated with the expression of exon 3-containing isoforms in control human brain. These findings would suggest that genetic risk factors for neurodegenerative diseases at the *MAPT* locus are likely to operate by changing the balance of mRNA splicing in different brain regions, as opposed to the overall expression of the *MAPT* gene (Caffrey and Wade-Martins 2012).

Profiling mRNA and protein expression in control human brain demonstrated significant regional variation. *MAPT* mRNA expression was 1.5 fold higher in the neocortex, as compared to white matter and cerebellum. Total tau protein expression level had a similar regional pattern of expression to that for mRNA. Since the frontal and temporal cortex are amongst the brain regions most affected by tangle pathology in AD, these findings support the concept that despite high ubiquitous expression of *MAPT* in human brain, regional variation in basal *MAPT* expression might predispose

some brain regions to tangle pathology and explain the regional specificity of disease in AD at least.

Comparable profiling of *MAPT* mRNA splicing and protein isoform expression presents a more complex task. While regional splicing of exon 2 containing mRNA isoforms was confirmed and validated using TaqMan assays, at the protein level this observation was not replicable. The only significant regional difference in tau protein isoform expression was the reduced levels of 0N3R in cerebellum as compared to other regions. This finding was in agreement with previous studies that have observed decreased 0N3R in the cerebellum of control (Boutajangout, Boom et al. 2004) and PSP (Gibb, de et al. 2004) patients. However, the absence of any clear relationship between mRNA splicing and tau protein isoform production is difficult to interpret. One possible reason for this is that altered splicing of exon 2 would be shared across four tau protein isoforms (1N3R, 1N4R, 2N3R and 2N4R), thus a sizeable change would be needed to allow detection by a semi-quantitative technique such as Western blotting.

Finally, the effects of the H1/H2 and H1c haplotypes on *MAPT* mRNA expression and splicing were investigated. In the case of the former, the statistical power was insufficient to use the exon-specific and region-wide expression data provided by the UKBEC. Thus, for the first time the effect of the H1/H2 inversion polymorphism on *MAPT* splicing were assessed. These data showed that the under-represented, protective H2 haplotype is associated, in all grey matter brain regions, with more expression of exon 3 and that this effect is most prominent in cortical regions. This finding is consistent with previous allele-specific expression studies (Caffrey, Joachim et al. 2008). Interestingly, no association was seen in white matter,

where in fact there was a trend in the opposite direction ($P = 0.03$). These data suggest that the inclusion of exon 3 in grey matter is protective in PSP, CBD and PD. Since several physiological roles for the amino-terminal inserts coded by exons 2 and 3 have been suggested, these might help explain this finding. For example, it has been suggested that the amino-terminal inserts could regulate the spacing between microtubules (Chen, Kanai et al. 1992; Frappier, Georgieff et al. 1994). Alternatively, since the amino-terminal region of tau has also been shown to interact with the plasma membrane where it can in turn interact with src-family kinases, it could be involved in signal transduction (Lee, Newman et al. 1998; Lee 2005). Moreover, it has been recently shown that the exon 2- and 10-encoded inserts increase aggregation propensity whereas the exon 3-encoded insert decreases aggregation (Zhong, Congdon et al. 2012; Zhong, Congdon et al. 2012). The latter could in part explain the protective function of H2 with its increased inclusion of exon 3. Interestingly, this function has recently been shown to be important in mediating amyloid-beta toxicity in a mouse model of Alzheimer's disease (Ittner, Ke et al. 2010).

The brain mRNA expression data (FCTX and CRBL) that were generated by the NABEC ($n = 222$) were used to investigate the effects of the H1c haplotype (tagged by rs242557) on *MAPT* expression. Using the most recent release of the 1000 Genomes Project (Interim phase I haplotypes, June 2011), quality control procedures identified a two base pair deletion (rs67759530) in the target sequence of ILMN_1710903 within the H2 haplotype. It is now well recognised that such sequence polymorphisms can result in hybridization artefacts, which in turn cause reporting of false cis-acting expression QTLs (Alberts, Terpstra et al. 2007; Benovoy, Kwan et al. 2008). Since the signals produced by Illumina probe ILMN_1710903 in mixed H1/H2 sample sets are responsible for the widely reported cis-acting *MAPT* expression QTL

(characterised by lower expression of *MAPT* in H2 individuals) (Myers, Gibbs et al. 2007; Gibbs, van der Brug et al. 2010; Nalls, Plagnol et al. 2011), this finding would suggest that this gene level mRNA expression QTL is a technical artefact. However, restricting the analysis to H1 homozygotes to explore the impact of the H1c haplotype on gene expression was performed. Unfortunately, this analysis did not demonstrate a significant association between the H1c haplotype (tagged by rs242557) and *MAPT* mRNA expression, as measured with Illumina probes ILMN_1710903 and ILMN_2310814. This does not exclude the possibility that this haplotype could affect *MAPT* mRNA splicing patterns.

In conclusion, the regional expression patterns of *MAPT* at the exon and gene mRNA and protein levels were examined in this study on the largest control human brain data set. This enhanced the establishment of a valuable reference of normal regional *MAPT* mRNA and protein isoforms expression levels. In addition, this study stresses the importance of exon-specific analysis as indicated by the association of exon-3 inclusion mRNA transcript expression with the H1/H2 inversion polymorphism. Exon-specific analysis may prove particularly significant in elucidating pathogenic mechanisms in neurodegenerative diseases that belong to a similar proteinopathy group but differ in their clinical-pathological profile.

The results of the work described in this chapter were published in *Human Molecular Genetics:*

Trabzuni, D., Wray, S., Vandrovcova, J., Ramasamy, A., Walker, R., Smith, C., Luk, C., Gibbs, J.R., Dillman, A., Hernandez, D.G. *et al.* (2012) MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. *Human Molecular Genetics*, **21**, 4094-4103.

6 Characterization of *LRRK2* mRNA expression and splicing in multiple control human brain regions: its relevance to human disease

6.1 Summary

Common and rare mutations at the *LRRK2* (Leucine-rich repeat kinase 2) locus have been identified in Parkinson's disease (PD), Crohn's disease (CD) and leprosy. However, little is known about the molecular mechanisms underlying these disease associations.

To further characterize this locus, the expression and splicing profiles of UKBEC, 134 control individuals, from ten brain regions were used for this analysis. The regional variation, splicing and regulation of *LRRK2* mRNA and protein levels (MJFF2 antibody) were studied in detail.

The results showed that the mRNA expression levels were highest in cortical regions and lowest in cerebellum. Novel splicing events involving exons 32-33 (COR and kinase domain) and exon 42 (kinase domain) were identified and confirmed. In addition, LRRK2 protein expression was demonstrated in multiple cell types, including neurons, glia and endothelial cells as demonstrated by immunohistochemistry (IHC). An exon quantitative trait locus (QTL) in brain that regulates the splicing of exons 32-33 and which is in strong linkage disequilibrium with CD associated non-synonymous SNP rs3761863 (M2397T) was identified; suggesting a possible molecular mechanism for CD association at this locus.

The results characterize the *LRRK2* locus and highlight the importance and difficulties of fine-mapping and integration of multiple datasets to delineate pathogenic variants and thus develop an understanding of disease molecular mechanisms.

6.2 Introduction

The role of *LRRK2* (Leucine-rich repeat kinase 2) in human disease was first recognised in 2004 when dominant mutations in the *LRRK2* gene were linked to PD (Paisan-Ruiz, Jain et al. 2004; Zimprich, Biskup et al. 2004). Rare genetic variants located in the *LRRK2* gene contribute to a significant fraction of familial clustering of the disease (Ross, Soto-Ortolaza et al. 2011). In particular, heterozygous carriers of the non-synonymous change G2019S have an estimated PD lifetime risk close to 50% (Melrose, Kent et al. 2007; Healy, Falchi et al. 2008), which directly implicates the *LRRK2* gene in the aetiology of PD. In addition, recent GWAS results suggest that common variants with a more modest effect on PD risk also exist at this locus (Nalls, Plagnol et al. 2011). Intriguingly, GWAS have implicated *LRRK2* in the pathogenesis of CD and leprosy (Zhang, Huang et al. 2009; Franke, McGovern et al. 2010; Lewis and Manzoni 2012). However, the mechanisms linking the *LRRK2* gene and human disease remain largely unknown and are the focus of an intense research effort.

The *LRRK2* gene spans 1.4 Mb and is made up of 51 exons. It produces a 2,527 amino acid protein with multiple functional domains, including leucine-rich repeats (LRR), a GTPase domain (Ras of Complex proteins or ROC domain), a domain of unknown function termed the C-terminal of ROC (COR) domain, a kinase domain and a WD40 domain (Cookson 2010). Coding changes causative for PD are located within the enzymatic core of LRRK2, namely the ROC-COR-Kinase triad of domains, and several of the mutations described in this region disrupt the enzymatic activities of the protein, strongly implicating the enzymatic function of LRRK2 in the pathogenesis of PD (West, Moore et al. 2005; Lewis, Greggio et al. 2007; Greggio and Cookson 2009). Existing data demonstrate ubiquitous expression of *LRRK2* in human brain in addition

to liver, kidney and thymus (Paisan-Ruiz, Jain et al. 2004; Zimprich, Biskup et al. 2004). A number of studies have been undertaken to investigate the function, expression and cellular localization of *LRRK2* in the human brain (Galter, Westerlund et al. 2006; Higashi, Biskup et al. 2007; Vitte, Traver et al. 2010; Sharma, Bandopadhyay et al. 2011), however, these findings have been based on relatively small numbers of individuals. Consequently, much of the biology of this gene remains unknown. Very little is known about the pattern of *LRRK2* gene expression across the human brain, to what extent it is the subject of alternative splicing, whether splicing is region-specific, and how the expression of *LRRK2* is regulated. Understanding the mRNA and protein processing is critical for deciphering how *LRRK2* dysfunction results in disease.

A key area of research is the identification of the molecular mechanisms linking variants in the *LRRK2* region with modified risk for PD, CD and leprosy (Barrett, Hansoul et al. 2008; Zhang, Huang et al. 2009; Franke, McGovern et al. 2010; Nalls, Plagnol et al. 2011; Ross, Soto-Ortolaza et al. 2011). The hypothesis is that some of these effects are mediated by genetic control of the expression and/or splicing or *LRRK2* mRNA. This hypothesis can be tested by comparing disease GWAS results with expression datasets and by investigating whether these association signals are compatible with shared causal variants (Plagnol, Smyth et al. 2009). In addition to suggesting causal mechanisms, this analysis can point to tissue types involved in disease pathogenesis. For example, based upon the premise that PD is driven by pathophysiological processes resulting in the death of neuronal cell populations, there is considerable interest in dissecting the genetic basis of PD susceptibility at the *LRRK2* locus by analysing this phenotype in the context of results in parallel with expression QTL studies conducted in multiple human tissues.

Here, the genotyped information that was generated using a custom array (Immunochip) as well as the Illumina Omni-1M Quad chip and imputation techniques (Howie, Fuchsberger et al. 2012) were used to provide dense genetic coverage at the *LRRK2* locus. This data was then compared with gene, exon-specific *LRRK2* mRNA expression profiles of UKBEC from ten brain regions: FCTX, n = 127, TCTX, n = 119, OCTX, n = 129, HIPP, n = 122, THAL, n = 124, PUTM, n = 129, SNIG, n = 101, MEDU, n = 119, CRBL, n = 130 and WHMT, n = 131 (1,231 Affymetrix exon arrays). Of these, SNIG and PUTM are relevant to PD pathophysiology (Wszolek, Vieregge et al. 1997). Moreover, these results were compared with published GWAS and expression QTL findings using imputation techniques (Howie, Fuchsberger et al. 2012) in order to obtain insights into the molecular mechanisms of *LRRK2* disease associations. Using immunohistochemistry, the cell types that express LRRK2 protein in the human brain of controls were also investigated.

6.3 Results

6.3.1 Regional expression of *LRRK2* mRNA in multiple control human brain regions

The array results showed evidence of regional variability in mRNA expression patterns. *LRRK2* mRNA levels were three fold higher in OCTX ($P < 10^{-20}$), the region expressing the highest *LRRK2* levels, as compared to CRBL and WHMT (**Figure 6.3.1**). The *LRRK2* expression level in SNIG and PUTM, the brain regions most implicated in the pathophysiology of PD, were unremarkable (**Figure 6.3.1**). QuantiGene (Chapter 3, **Figure 3.3.7**) and real-time qRT-PCR (**Figure 6.3.2**) were used to confirm the array results and demonstrated similar regional mRNA expression patterns in a subset of brain samples within four selected regions. As previously documented in Chapter 2 (Trabzuni, Ryten et al. 2011), a very good agreement between the different gene expression quantification methods was observed.

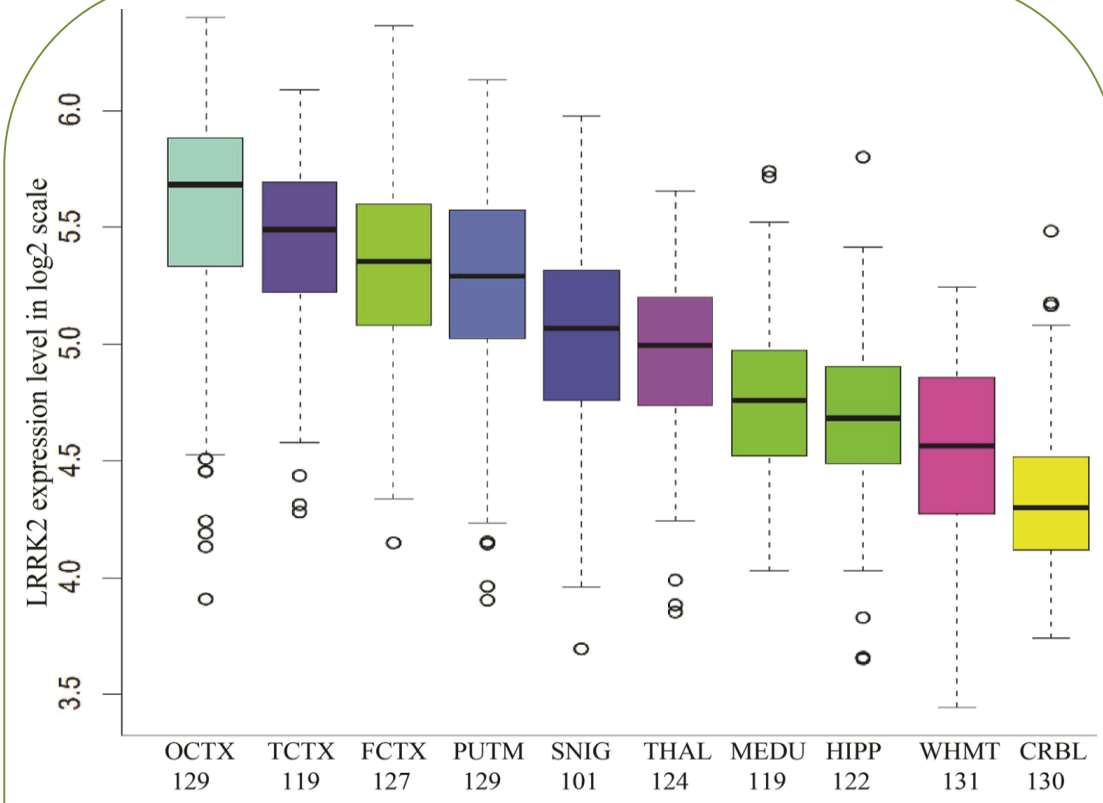

Figure 6.3.1:

Regional Distribution of *LRRK2* mRNA expression. OCTX = occipital cortex, TCTX = temporal cortex, FCTX = frontal cortex, PUTM = putamen, SNIG = *substantia nigra*, THAL = thalamus, MEDU = medulla, HIPP = hippocampus, WHMT = white matter and CRBL = cerebellum.

Boxplot of mRNA expression levels for *LRRK2* in 10 brain regions, from microarray experiments on a log2 scale (*y*-axis). This plot shows the regional variation in *LRRK2* transcript expression across all the regions. Whiskers extend from the box to 1.5 times the inter-quartile range.

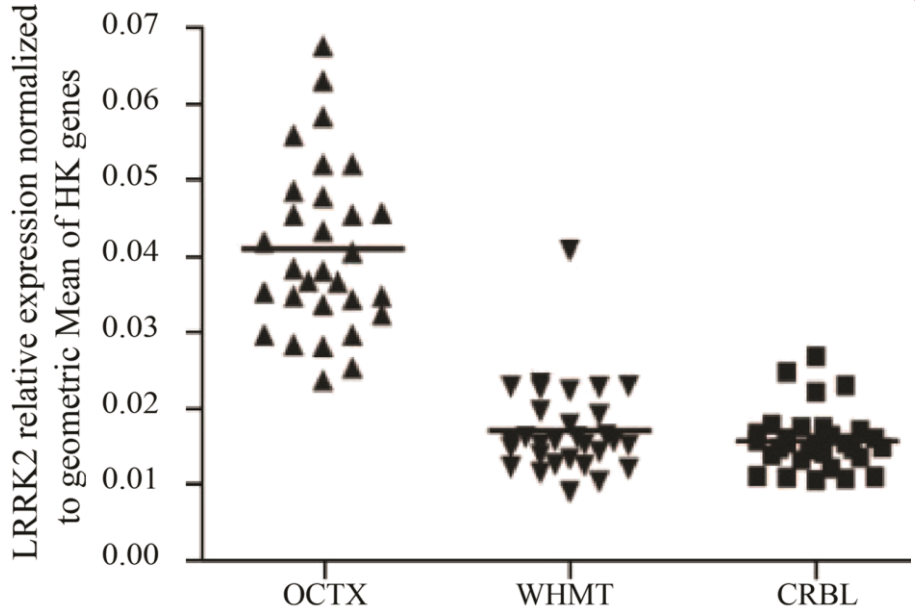

Figure 6.3.2:

Dot plot of mRNA expression levels for *LRRK2* based on TaqMan qRT-PCR validation. OCTX = occipital cortex, WHMT = white matter and CRBL = cerebellum.

The expression levels were normalized to the geometric mean of three housekeeping genes. The graph shows higher expression in OCTX compared with WHMT and CRBL. These data confirm the array results ((Data produced by Dr Jana Vandrovcova).

6.3.2 Regional splicing of *LRRK2* mRNA in multiple control human brain regions

The analysis of the brain expression data at the exon level and across the ten brain regions, demonstrated that exons 32-33 (coding for part of the COR and kinase domains), display the highest expression variability between regions (**Figure 6.3.3**), with lower expression in cerebellum and higher in occipital cortex. A similar pattern for exon 42-43 was also observed (**Figure 6.3.4**), of particular interest owing to its location in the kinase domain (Lewis 2009). This suggestive evidence of alternative splicing of these exons was further confirmed by an analysis of variance on the splicing index (Kang, Kawasawa et al. 2011) of these exons to test for differences across the ten brain regions ($P = 2.46\times10^{-5}$ for exon 32 and $P < 10^{-25}$ for exons 33, 42 and 43).

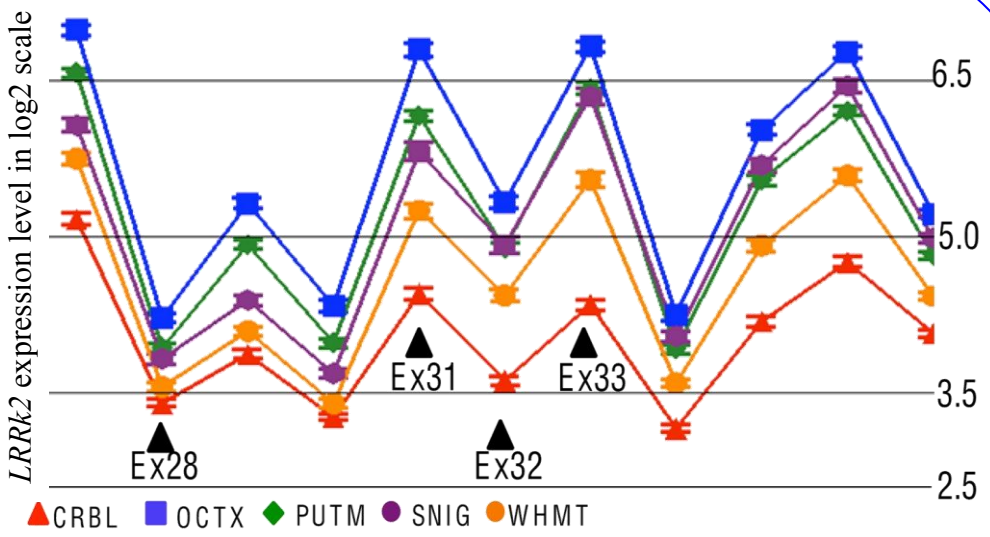

Figure 6.3.3:

Exon array expression data at the probe set level suggesting alternative splicing by brain region of the *LRRK2* transcript. CRBL = cerebellum, OCTX = occipital cortex, PUTM = putamen, SNIG = *substantia nigra* and WHMT = white matter.

Mean expression levels (*y*-axis) plotted for each probe set (*x*-axis) covering exon 27 to 37 of the *LRRK2* gene in CRBL, OCTX, PUTM, SNIG and WHMT. Non-parallel probe set expression levels suggest differential expression and splicing of exons 28, 31, 32 and 33 in different regions. Selected exons are indicated by black arrow (▲). Partek alternative splicing ANOVA *P* value is $< 1.0\times10^{-45}$). Plot was generated in Partek Genomics Suite software.

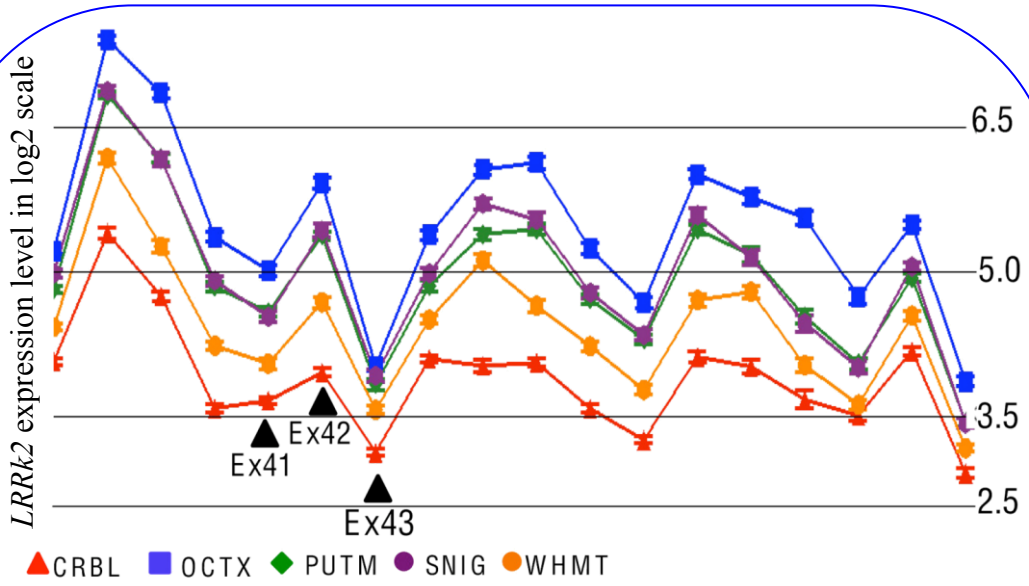

Figure 6.3.4:

Exon array expression data at the probe set level suggesting alternative splicing by brain region of the *LRRK2* transcript. CRBL = cerebellum, OCTX = occipital cortex, PUTM = putamen, SNIG = *substantia nigra* and WHMT = white matter.

Mean expression levels (*y*-axis) plotted for each probe set (*x*-axis) covering exon 38 to 51 of the *LRRK2* gene in CRBL, OCTX, PUTM, SNIG and WHMT. Non-parallel probe set expression levels suggest differential expression and splicing of exons 41, 42 and 43 in different regions. Selected exons are indicated by black arrow (▲). Plot was generated in Partek Genomics Suite software.

Given this evidence, junction and exon-specific primers, and reverse transcriptase PCR (RT-PCR) in 12 brain samples and four brain regions were performed to further explore the splicing patterns of exons 32-33 and 42-43. The splicing of exons 32-33 in SNIG as compared to OCTX, CRBL and MEDU was confirmed (**Figure 6.3.5, A**). This analysis revealed extensive splicing complexity around exons 31-33 (**Figure 6.3.5, A and B**). Exon 42 showed a high expression in OCTX, but low levels of expression or complete splicing out in SNIG and CRBL in 8 out of 12 individuals (**Figure 6.3.5, C**). The corresponding functional domains in LRRK2 protein are shown in **Figure 6.3.5, D**.

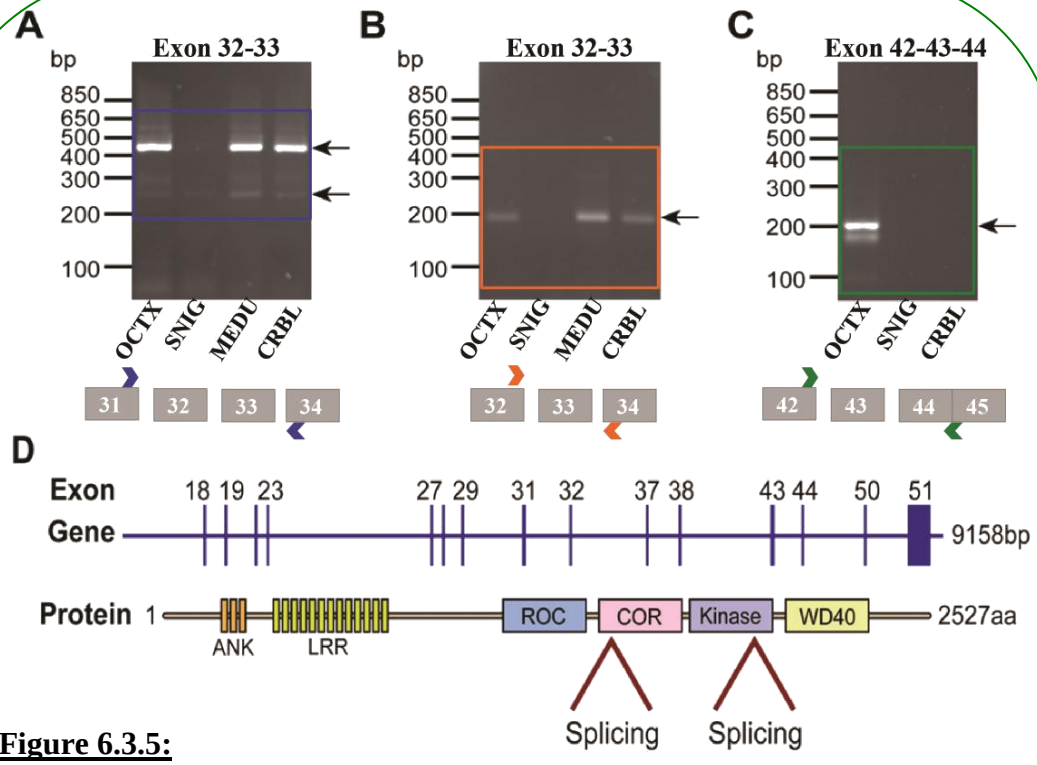

Figure 6.3.5:

RT-PCR results confirming regional differences in the splicing of targeted exons in the *LRRK2* transcript in selected brain regions. OCTX = occipital cortex, SNIG = *substantia nigra*, MEDU = medulla and CRBL = cerebellum.

(A) RT-PCR results confirming the splicing out of exons 31-32 in SNIG, compared with the other brain regions tested. The expected band size for the isoform with exon 31- 32 included is 470 bp, whereas that for the isoform with exon 32 alone spliced out is 270 bp. These results show splicing out of exons 31-32 in SNIG and the existence of an isoform with exon 32 alone spliced out in OCTX, MEDU and CRBL. In addition, the difference in the band intensity between the two observed isoforms, suggests that the isoform with exon 32 spliced out is expressed at lower levels than the isoform with exon 31-32 included. **(B)** RT-PCR results further confirm the splicing out of exon 32 in SNIG. While OCTX, MEDU and CRBL show the expected band size of 195 bp suggesting that exon 32-33 is not spliced out in these regions, SNIG does not. **(C)** RT-PCR results confirming that exons 42-43 are present and have high expression in OCTX in all individuals but are spliced out in SNIG and CRBL in eight out of twelve individuals. The expected band size for the isoform with exon 42-43 included is 200 bp and 100 bp when exon 43 alone is spliced out. OCTX expressed the two isoforms of *LRRK2*, that contain exons 42-43 and the additional isoform with exon 43 spliced out (faint band). Dark blue, orange and green arrows indicate the locations of the exon-specific RT-PCR primers. **(D)** Schematic view shows the functional domains that targeted exons corresponding to in LRRK2 protein.

6.3.3 Regional expression of LRRK2 protein in multiple control human brain regions

In order to investigate whether the trends that were observed at the mRNA level could also be observed at the protein level, Western blot and immunohistochemistry (IHC) techniques were applied using the MJFF2 antibody (validated using tissue *LRRK2* knockout mice (Lin, Parisiadou et al. 2009) (**Figure 6.3.6**). This antibody was raised against recombinant protein covering residues 970-2575. LRRK2 protein levels were shown to exhibit a similar pattern to the mRNA expression levels in five selected regions using Western blots of control tissue from five individuals (**Figure 6.3.7**). A number of specific LRRK2 immunoreactive bands were identified, which may be the product of proteolytic processing or alternative splicing events.

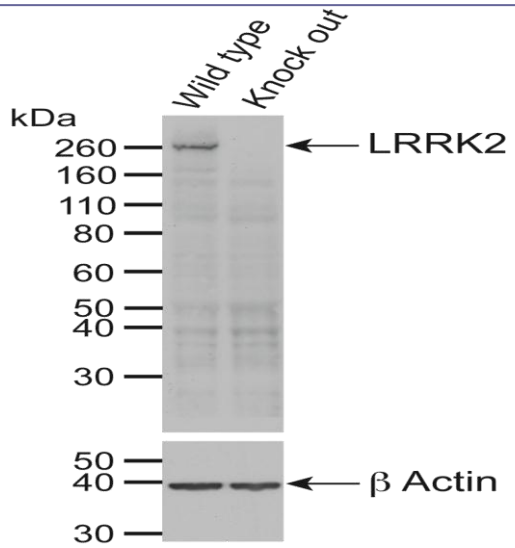

Figure 6.3.6:

Western blot of mouse brain with MJJF2 antibody.

Western blot of wild type and knock out mouse brain probed with anti-LRRK2 MJFF2 antibody (upper band) and beta actin loading control (lower band) demonstrating specificity of MJFF2 antibody (Lin, Parisiadou et al. 2009).

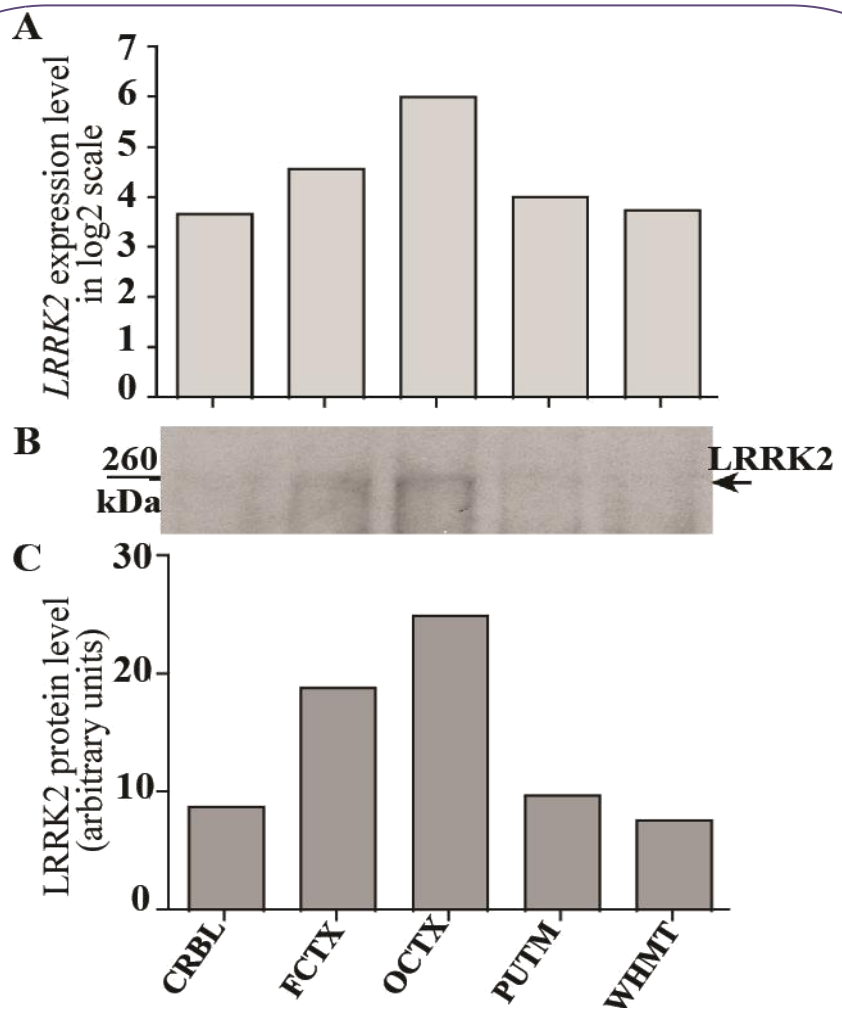

Figure 6.3.7:

LRRK2 protein levels in selected regions using Western blots of control tissue. CRBL = cerebellum, FCTX = frontal cortex, OCTX = occipital cortex, PUTM = putamen and WHMT = white matter.

(A) mRNA expression levels for *LRRK2* in five brain regions, CRBL, FCTX, OCTX, PUTM and WHMT for one individual. **(B)** The Western blot for the LRRK2 protein, showing that the highest expression was in OCTX as a faint band can be observed at the right size of 260 kDa. **(C)** Graph showing the protein level for LRRK2 in the same five brain regions to demonstrate the constant pattern between the array results at the mRNA and the protein level of LRRK2 (Data produced by Dr Patrick Lewis).

IHC was applied to determine the cellular distribution of LRRK2 in control human brain tissue from FCTX, OCTX, CRBL, SNIG and striatum (caudate and PUTM) (**Figure 6.3.8** and summary in **Table 6.3.1**). The results revealed LRRK2 expression to be highest in the cerebellar purkinje cells, both in the neuronal perikarya and also in the processes (Fig 5A and B for negative control pre-absorbed antibody). Low levels of LRRK2 immunoreactivity were detected within neurons in FCTX, in the neuronal perikarya of OCTX (**Figure 6.3.8, C-D**). Low or background levels of LRRK2 staining were noted in the cytoplasm of neuromelanin-positive SNIG neurons (**Figure 6.3.8, E**) and in the medium spiny neurons within the striatum (data not shown). All regions showed LRRK2 immunoreactivity in the smooth muscle and endothelium of blood vessels. Interestingly, some astroglia (**Figure 6.3.8, F**) were positive for LRRK2 and were seen both in the grey and white matter of all regions studied. Although no LRRK2 immunoreactivity could be detected within microglia, it should be noted that the control brain tissue used would not be expected to contain activated microglia, which could have a very different expression profile. A summary of relative staining intensity obtained with the MJFF2 antibody is described in **Table 6.3.1**.

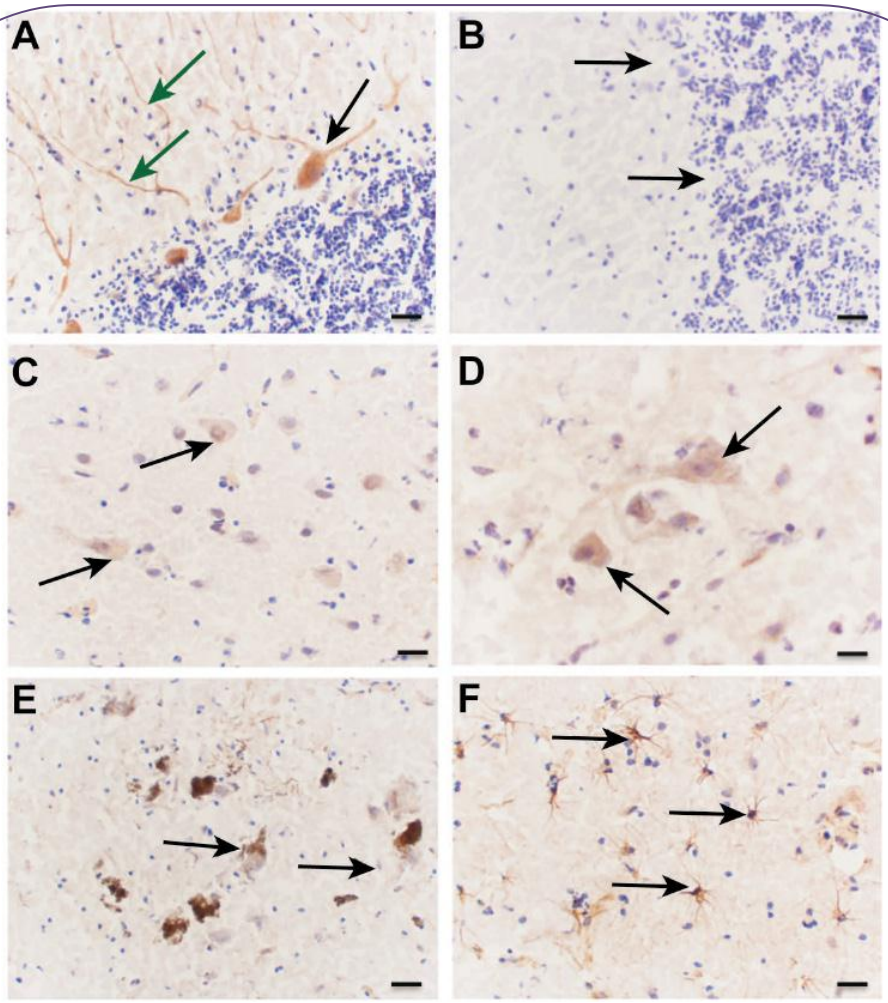

Figure 6.3.8:

Immunohistochemical distribution of LRRK2 in control human brain visualized by rabbit monoclonal MJFF2 antibody.

(A) Cerebellar purkinje neurons (black arrow) and processes (green arrows) display strong labelling for LRRK2. **(B)** Cerebellar section probed with pre-absorbed antibody shows complete absence of LRRK2 labelling in purkinje neurons (black arrows). **(C)** Larger neurons (black arrows) in the deep layer of frontal cortex show low levels of LRRK2 immunoreactivity. **(D)** Neurons in the occipital cortex show low levels of immunoreactivity for LRRK2. **(E)** Melanised nigral neurons from the substantia nigra show minimal LRRK2 immunoreactivity. **(F)** Astrocytes positive for LRRK2 immunoreactivity in the grey matter of frontal cortex. Scale bar: 20 µm in A, B, C, F and 10 µm in D and E (Data produced by Dr Rina Bandopadhyay).

Table 6.3.1: Relative intensity of LRRK2 immunohistochemistry in various cell types of the control human brain.

Region	Neuronal LRRK2	Glial LRRK2	Blood vessels
CRBL	Purkinje cells +++	Astrocytes ++	Endothelial cells ++
CRBL	Granule cells +	±	±
SNIG	DA neurons +	Astrocytes ++	Endothelial cells ++
PUTM	Neurons +	Astrocytes ++	Endothelial cells ++
FCTX	Neurons +	Astrocytes ++	Endothelial cells ++
OCTX	Neurons +	Astrocytes ++	Endothelial cells ++

Key: ±, background expression; + low expression, ++ mild expression; +++ high expression. CRBL = Cerebellum, SNIG = *Substantia nigra*, PUTM = Caudate+Putamen, FCTX = Frontal cortex and OCTX = Occipital cortex.

6.3.4 Relevance of genetic control of expression for PD, CD and leprosy

Using the publically available association data from GWA results for PD, CD and leprosy and UKBEC expression profiles in human brain, the question as to whether the association of *LRRK2* with these diseases is compatible with shared causal expression variant(s) was assessed (**Table 6.3.2**).

The analysis identified an exon expression QTL linking the levels of exons 32 and 33 with SNP rs10784486 (rs118049839) at genomic position chr12:40677029 (In the intron 18-19 of the *LRRK2* gene) (**Table 6.3.3 and Figure 6.3.9**) that was consistent with a shared causal variant. The secondary *LRRK2* CD association (lead SNP rs3761863) was strongly correlated with *LRRK2* mRNA expression of exons 32 and 33 in control human brain (**Figure 6.3.10**). The minor allele T of rs3761863 increases CD risk and is associated with a decrease in *LRRK2* expression (**Table 6.3.2 and Figure 6.3.10**). While the summary statistic level of the latest CD meta-analysis (for the rs3761863/CD exon expression QTL) does not enable a formal test of co-localization between disease and expression QTL signals (Plagnol, Smyth et al. 2009), these results suggest plausible mRNA mechanisms to mediate this secondary CD association signals.

Table 6.3.2: Common variant associations in the *LRRK2* region for PD, CD and leprosy.

	PD	CD, 1st signal	CD, 2nd signal	Leprosy
SNP	rs117762348	rs11564258	rs3761863[a]	rs1491938
Position	chr12:40,597,612	chr12:40,792,300	chr12:40,758,652	chr12:40,645,630
Alleles (major > minor)	A > G	G > A	C > T	T > C
MAF	0.076	0.034	0.32	0.396
Odds ratio for minor allele (95% CI)	0.85 (0.80-0.91)	1.74 (1.55–1.95)	1.1 (1.05-1.15)	0.86 (0.80-0.92)
P-value brain dataset	9.6×10^{-01}	0.081	$\mathbf{2\times10^{-08}}$	8.7×10^{-03}
Expression effect on *LRRK2* mRNA (for the minor allele)	Not detected	Not detected	Decrease	Not detected

For each disease association the *P*-values in the expression datasets were listed.

a: See Figure 6.3.10.

Table 6.3.3: Information on the identified *LRRK2* exon expression QTL in the UKBEC (*n* = 134).

Lead SNP	Position	Alleles	MAF	Probe location	*P*-value	Direction of effect (for minor allele)
rs10784486[a]	chr12:40,677,029	C > A	0.33	exons 32 and 33	5.1E-10	Decrease

For the *P*-value computations in brain samples, *LRRK2* expression values are averaged across all 10 brain regions.

MAF: minor allele frequency. a: See Figure 6.3.9.

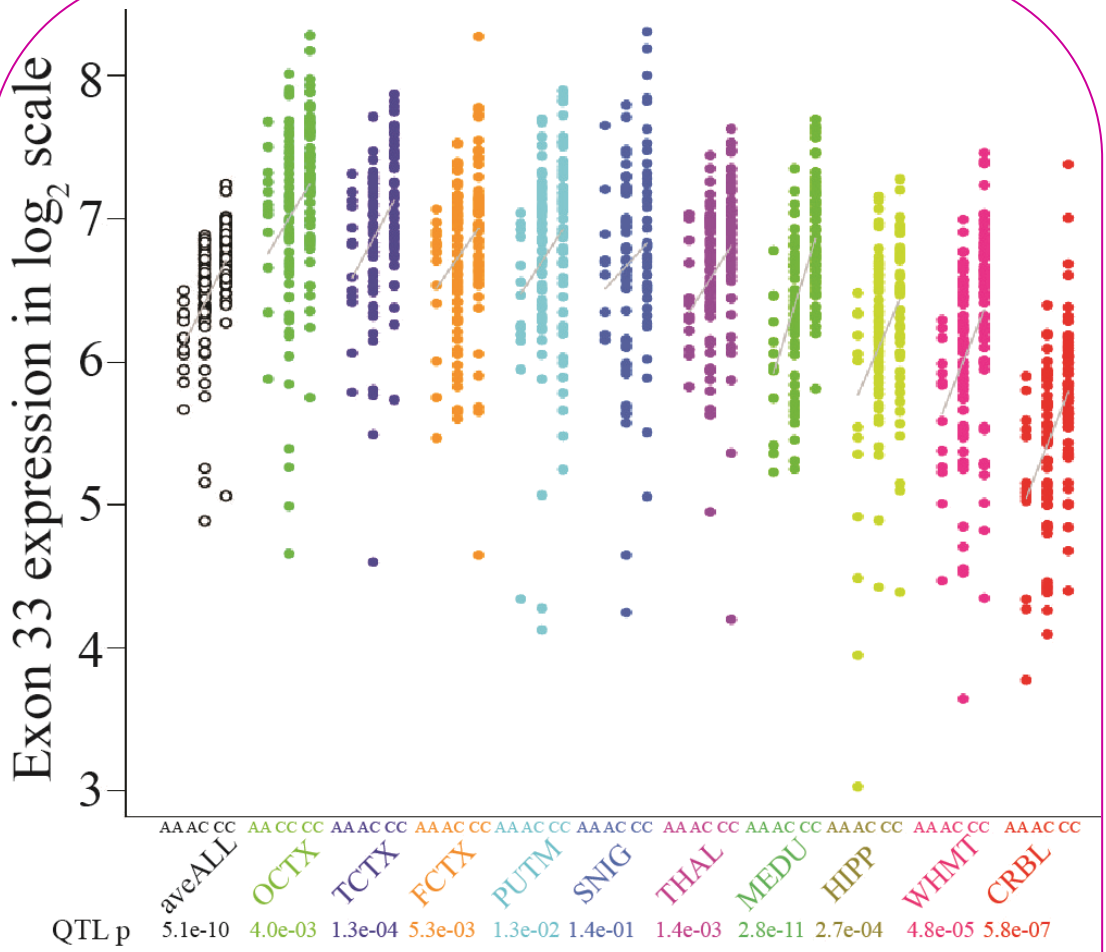

Figure 6.3.9:

The effect of rs10784486 on the expression of *LRRK2* exon 33. OCTX = occipital cortex, TCTX = temporal cortex, FCTX = frontal cortex, PUTM = putamen, SNIG = *substantia nigra*, THAL = thalamus, MEDU = medulla, HIPP = hippocampus, WHMT = white matter and CRBL = cerebellum.

LRRK2 exon 33 expression (Affymetrix probe set 3411379) stratified by rs10784486 for ten brain regions in 134 brain samples. Decreased exon 33 expression was associated with the homozygous minor allele (AA). A similar association pattern was observed in all brain regions. All ten brain regions shown as well as combined mean across region (aveALL).

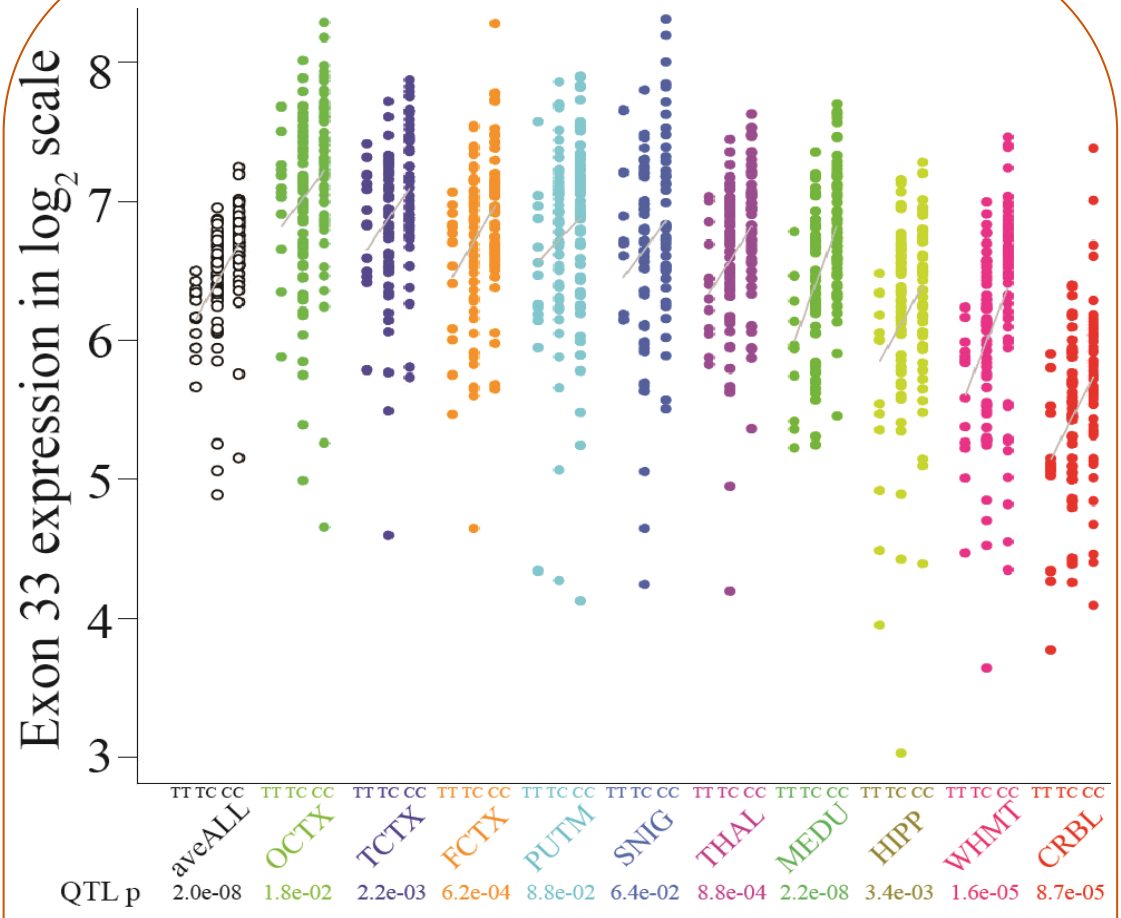

Figure 6.3.10:

The effect of rs3761863 on the expression of *LRRK2* exon 33. OCTX = occipital cortex, TCTX = temporal cortex, FCTX = frontal cortex, PUTM = putamen, SNIG = *substantia nigra*, THAL = thalamus, MEDU = medulla, HIPP = hippocampus, WHMT = white matter and CRBL = cerebellum.

LRRK2 exon 33 expression (Affymetrix probe set 3411379) stratified by rs3761863 for ten brain regions in 134 brain samples. Decreased exon 33 expression was associated with the homozygous minor allele (TT). A similar association pattern was observed in all brain regions. All ten brain regions shown as well as combined mean across region (aveALL).

6.4 Discussion

The key role of the *LRRK2* locus in Parkinson's disease was the motivation for fine-mapping this locus in the UKBEC individuals. The analysis of the gene/exon expression dataset highlighted the complexity of the genetic control of *LRRK2* in the human brain. A significant exon expression QTL was reported in the brain tissue that involves exons 32 and 33 of *LRRK2* but does not co-localize with the PD association.

The most recent meta-analysis results for CD indicated a secondary association for CD, for which the non-synonymous SNP rs3761863 is a leading associated variant ($P = 3.0\times10^{-6}$) (Franke, McGovern et al. 2010). SNP rs3761863 is also strongly associated with the exon 32-33 brain exon expression QTL (SNP rs10784486) that was identified ($P = 2\times10^{-8}$, **Table 6.3.3**). These results suggest a potential mechanism to mediate the CD association. CD may involve any part of the gastrointestinal tract, but most frequently the terminal ileum or any region of the intestine (Bernstein, Wajda et al. 2005) and since the human intestine is heavily innervated with nerve endings, the findings in brain tissue may be relevant to CD. It is therefore possible that the exon/expression correlation may be shared across other tissue types and so analysis of expression in intestinal tissue is warranted. Moreover, additional CD fine-mapping will be required to test whether the CD and brain exon expression QTL datasets are fully consistent with a shared variant. Furthermore, a previous study has associated the nsSNP rs3761863/M2397T with LRRK2 protein stability (Liu, Lee et al. 2011).

These results highlight several factors that complicate the follow-up of associated variants at the *LRRK2* locus in all diseases and in particular PD: the existence of multiple *LRRK2* isoforms, variability in mRNA expression and splicing across brain regions, and localisation of LRRK2 protein to both neuronal and non-

neuronal cell types. On the basis of both exon array and RT-PCR data there is a prediction that at least two further isoforms exist in addition to full-length *LRRK2*: an isoform that lacks exon 32 and an isoform that lacks one or both of exon 42-43. Given that exons 32 and 42-43 all map to functional domains of the LRRK2 protein (COR and kinase domains) these splicing events are likely to have functional consequences upon complex formation and/or enzymatic function. In fact, data from a number of groups have demonstrated that expression of truncated forms of LRRK2 has a major impact on the enzymatic functions of this protein, although these studies have – to date – examined only artificial truncation constructs (Jaleel, Nichols et al. 2007; Luzon-Toro, Rubio de la Torre et al. 2007; Klein, Rovelli et al. 2009). As a result, it is very important to note that future characterization of the regional splicing isoforms of *LRRK2* must include not only mRNA analysis, but also cDNA and amino acid sequencing to ensure that as many variants as possible are captured. Understanding the range of different protein isoforms will be critical to understanding the physiological functions of LRRK2.

The analysis of gene expression in ten brain regions, including PD-relevant tissues such as the *substantia nigra* and putamen, also provided insights into the wide regional variation in gene expression within the human brain. These differences were in some cases large, with a 3.0 fold change in expression between the occipital cortex and cerebellum. Furthermore, regional LRRK2 protein expression measured by Western blotting correlated well with mRNA levels simplifying the interpretation of mRNA-based exon expression QTLs. The use of a recently well-characterised monoclonal antibody (Mandemakers, Snellinx et al. 2012) also showed that LRRK2 expression is not restricted to neurons in the control human brain. Even amongst the neuronal population LRRK2 staining levels were variable, with higher

immunoreactivity in purkinje cells of the cerebellum as compared to neurons of the cortical regions, basal ganglia and dopaminergic *substantia nigra* neurons. These immunohistochemistry results are broadly concordant with previously reported LRRK2 localisation data in rodent and human brain (Greggio, Jain et al. 2006; Higashi, Moore et al. 2007; Melrose, Kent et al. 2007; Vitte, Traver et al. 2010; Sharma, Bandopadhyay et al. 2011). These results were extended by demonstrating that LRRK2 protein expression in glial cells is more widespread than was originally thought. While this may be due to the use of polyclonal antibodies recognising different epitopes of the LRRK2 protein, recent evidence proposes that LRRK2 is expressed in monocytes and in activated microglia implying a role for LRRK2 in the immune system and in PD (Hakimi, Selvanantham et al. 2011; Thevenet, Pescini Gobert et al. 2011; Moehle, Webber et al. 2012).

Furthermore, it is important to realize that the amino acid epitope which the antibody binds to is not known. For example, if the antibody binds to the epitope that is presented by exon 32 or any other spliced exons, a specific protein isoform will not be detected in the Western blot even if it is expressed in the cell/tissue. Therefore, epitope mapping for this antibody is needed to detect and identify different isoforms of LRRK2 protein in further investigations.

In summary, this study provides novel insights into *LRRK2* expression, splicing and regulation with a potential link to the aetiology of CD. It also highlights the relevance of imputation techniques to provide the dense coverage required to integrate multiple disease association and gene expression studies. The localization of the common variant PD association outside of the coding region of *LRRK2* suggests that it is likely that the effect on disease risk is mediated by control of mRNA expression.

These data, together with the recent analysis of the *MAPT* locus, illustrate the complexities of defining precisely how risk loci contribute to disease and illustrate that there is as much work required to dissect a locus as to identify it in the first place (Trabzuni, Wray et al. 2012). In addition, gene expression studies are required to increase the quality and quantity of data to maximize the power to dissect the genetic control of *LRRK2* expression in all those diseases in which it is implicated.

Dr Patrick Lewis and Daniah Trabzuni were awarded the Michael J. Fox foundation for Parkinson's Research (MJFF) grant for "Characterizing region specific splice isoforms of LRRK2." for follow-up investigations of the results presented in this chapter.

The results of the work described in this chapter were submitted to *Plos One* in November*:*

Trabzuni, D., R. Ryten, et al. (2012). "Fine-mapping, gene expression and splicing analysis of the disease associated LRRK2 locus." *PLoS One*.

7 Discussion and future recommendations

The first aim for this thesis was to generate a genome-wide and large scale genetic-phenotypic data set that will advance neuroscience research in the field. The information obtained from the multiple CNS regions that were dissected from 137 control human post-mortem brains will be a publicly available reference data set, to be used to assess specific hypotheses in various projects in the future.

The data set resulting from this study will complement the other data sets that were generated for the same purpose, either from human brain or different human tissues using different sample sizes, different CNS regions and different expression and genotyping platforms (Gardina, Clark et al. 2006; Roth, Hevezi et al. 2006; Myers, Gibbs et al. 2007; Wang, Sandberg et al. 2008; Johnson, Kawasawa et al. 2009; Webster, Gibbs et al. 2009; Gibbs, van der Brug et al. 2010; Kang, Kawasawa et al. 2011; Nica, Parts et al. 2011; Hawrylycz, Lein et al. 2012; Hernandez, Nalls et al. 2012). Data from this study demonstrates that 70 to 80% of transcript/gene expression and splicing profiles from human post-mortem frozen brain regions are reproducibly recoverable to improve our understanding of the genetic regulation of the transcriptome architecture of the human brain in specific tissues, and eventually in specific cell types.

The identification of the expression and splicing patterns in the human CNS was performed to achieve the second aim of this project. Results revealed that the alternative splicing pattern is the main pattern in the human CNS between the 12 regions, particularly in cortical regions (OCTX, FCTX and TCTX). Furthermore, there was a specific expression signature for certain regions such as CRBL and WHMT.

Lastly, this thesis demonstrates that this resource permits the start of the exploration of the functional effects of the genetic variants on the brain transcriptome in specific CNS regions focusing on targeted genes. The primary advantage of this type of data is that further advanced analyses can be undertaken, using different bioinformatics tools and this can be followed up with bench experiments depending on specific hypotheses that the researchers are posing. This is important as new bioinformatic and molecular biology tools become available in the future.

Recently, not only have new techniques been developed in the field to maximize the probability of producing high quality and reliable data from partially degraded RNA and precious tissue samples such as human brain tissues, but also the sensitivity and specificity of isoforms predicting algorithms have been improved (Anton, Aramburu et al. 2010; Chow, Li et al. 2011).

The logical step forward is to perform second stage analyses (Stage II) of the data to address the challenges raised through Stage I analyses of this project. These are outlined as follows.

Firstly, integrating the expression, splicing and expression QTLs resulting from this study into functional databases. The results from this study revealed regional specific differential expression patterns especially in the cerebellum and white matter. Furthermore, the alternative splicing patterns were the leading pattern in the cortical regions and this matches the expression QTL results where the exon level expression QTLs were the dominant signals. The application of weighted gene co-expression network analysis (WGCNA) (Zhang and Horvath 2005) will help to identify network connectivity modules in specific regions or between all regions (Johnson, Kawasawa et al. 2009). Other tools can be used in parallel to integrate the results with the Database

for Annotation, Visualization and Integrated Discovery (DAVID) (Huang da, Sherman et al. 2009) to identify enriched biological themes such as functional and clustering classification using gene ontology (GO). Identification of over-represented transcription factor binding sites (TFBS) in targeted modules is also applicable (Ho Sui, Mortimer et al. 2005). A recent study demonstrated a very specific example for distinct cortical areas in two individuals. By using the combination of some of the mentioned tools and RNA sequences, the link between the cognitive function and the transcriptome of these cortical regions was established (Naumova, Palejev et al. 2012). The use of these bioinformatic tools does not exclude the importance of molecular and cellular biology experiments that are required to validate the results further.

Secondly, it is crucial to follow up and validate the gene and exon expression QTLs using specific cellular modules to perform experiments and advanced bioinformatic tools. To date, GWAS results have provided an extended list of coding and non-coding SNPs that are associated with neurological diseases. Moving beyond that type of information is vital, as the successful identification of variants that contribute functionally to complex traits is highly dependent on the type and reliability of the assay (Stranger, Forrest et al. 2005). In a heterogeneous organ such as the brain, it is challenging to evaluate the effect of the expression QTLs on the transcriptome. The presence of different types of neurons and glial cells at different ratios, in different regions complicate the interpretation of the results. It will be difficult to be certain on which cell type the expression QTL is acting. Moreover, it cannot be confirmed that this expression QTL is the associated biomarker with the investigated disease. At this end to consider these QTLs as functional biomarkers, further validations and integration of these results with different and advanced techniques are important. Induced pluripotent stem (iPS) cell technology is an advanced technique (Takahashi,

Tanabe et al. 2007) that scientists are using to generate specific types of human neurons. This technique is suggested as a future approach to overcome the challenge of mixed cell populations. The Encyclopaedia of DNA Elements (ENCODE) Consortium is also one of the main bioinformatic tools that can be used to identify functional and regulatory elements in the human genome (Dunham, Kundaje et al. 2012; Harrow, Frankish et al. 2012), in addition to all the bioinformatic and molecular biology tools described above.

The experience that was earned from this project has taught us that the distribution patterns and the resulting effects of brain expression QTLs are not fully understood. This was demonstrated by the analysis of two examples, specifically the genes *MAPT* and *LRRK2* that were covered comprehensively in this study. There is still a pressing need to investigate further some of the findings in order to have conclusive results. For *MAPT* for example, further investigations are needed taking into consideration the existence of antisense RNA molecules and how that can affect the splicing and regulate the transcriptional processing of the *MAPT* mRNA in the cell (Peacey, Rodriguez et al. 2012). Regarding *LRRK2*, future work is required to investigate the novel isoform presence in different brain regions by combining cDNA and protein sequences from brain tissues and connecting this to the expression QTLs detected. Moreover, improving the quality of the immune detection method is also required.

Thirdly, RNA sequencing is becoming the method of choice. It will help to overcome some of the limitations of previous technologies by avoiding the biases presented in hybridization, allowing strand specific detection, novel isoforms detection and having a higher resolution than the exon arrays in particular, particularly if it is

applied at the cell level such as neurons and astrocytes. In addition, RNA sequencing will provide information about the non-coding RNA such as microRNA and this is important for neurodegenerative diseases (Baginsky, Hennig et al. 2010; Salta and De Strooper 2012). However, the analyses and the volume of the data that are generated from this complex technique are computationally challenging. To date, there is no standard protocol available to normalize, process the raw data from these experiments in to a manageable format that can then be further interrogated on a personal computer and to align and assemble the data to a reference genome (Auer, Srivastava et al. 2012). RNA sequencing can detect newly discovered RNA species such as antisense 3' RNA that we do not understand the function of yet (Ozsolak and Milos 2011).

Finally, the main challenge in the field of neuroscience after generating and analyzing such genome-wide and large scale data sets, is the ability to integrate different data into higher dimensional data sets. This will be achieved by linking the genetic, functional genomic, network analysis, cellular biology and proteomic data to the central nervous system functions (Geschwind and Konopka 2009; Sanderson 2009; Lewis, Schramm et al. 2010). As a result, novel hypotheses will be generated to be tested in more advanced levels in the field (Hawkins, Hon et al. 2010; Lewis, Schramm et al. 2010) to provide more conclusive answers.

Over the last decade, the fields of genomics and transcriptomics have developed rapidly to routinely use digital platforms, as compared to the progress of other fields of science. For example, the development of biochemical, cellular and functional biology tools are at least two decades behind (Dr Patrick Lewis, personal communication); this may due to the constantly changing factors that influence experiments in these other fields of science.

In conclusion, this project provides a reference dataset and remarkable results to be used as a resource for further functional biology investigations. It can be used to initiate new follow up research areas in various directions and to encourage researchers in the biochemistry, cellular and functional biology fields to develop more advanced and higher throughput tools. This research can be applied in targeted areas and a pipeline established to transfer these results into a solid functional framework. Integrating data from different research areas will ultimately lead to the dissemination of the scientific findings to the health industry and the field of drug discovery to achieve a long term beneficial impact on the clinical management of neurodegenerative diseases in the future.

"Have you noticed how the more you learn, the more you realize how little you know"
Plato

8 Reference

Alarcon, M., B. S. Abrahams, et al. (2008). "Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene." American journal of human genetics **82**(1): 150-159.

Alberts, R., P. Terpstra, et al. (2007). "Sequence polymorphisms cause many false cis eQTLs." PLoS One **2**(7): e622.

Anton, M. A., A. Aramburu, et al. (2010). "Improvements to previous algorithms to predict gene structure and isoform concentrations using Affymetrix Exon arrays." BMC Bioinformatics **11**: 578.

Arikawa, E., Y. Sun, et al. (2008). "Cross-platform comparison of SYBR Green real-time PCR with TaqMan PCR, microarrays and other gene expression measurement technologies evaluated in the MicroArray Quality Control (MAQC) study." BMC Genomics **9**: 328.

Arning, L., J. T. Epplen, et al. (2012). "The SETX missense variation spectrum as evaluated in patients with ALS4-like motor neuron diseases." Neurogenetics. DOI 10.1007/s10048-012-0347-4

Auer, P. L., S. Srivastava, et al. (2012). "Differential expression--the next generation and beyond." Briefings in functional genomics **11**(1): 57-62.

Azevedo, F. A., L. R. Carvalho, et al. (2009). "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain." The Journal of comparative neurology **513**(5): 532-541.

Baginsky, S., L. Hennig, et al. (2010). "Gene expression analysis, proteomics, and network discovery." Plant physiology **152**(2): 402-410.

Bahar, B., F. J. Monahan, et al. (2007). "Long-term stability of RNA in post-mortem bovine skeletal muscle, liver and subcutaneous adipose tissues." BMC Mol Biol **8**: 108.

Baker, M., I. Litvan, et al. (1999). "Association of an extended haplotype in the tau gene with progressive supranuclear palsy." Hum Mol Genet **8**(4): 711-715.

Barbosa-Morais, N. L., M. J. Dunning, et al. (2010). "A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data." Nucleic Acids Res **38**(3): e17.

Barrachina, M., E. Castano, et al. (2006). "TaqMan PCR assay in the control of RNA normalization in human post-mortem brain tissue." Neurochem Int **49**(3): 276-284.

Barrett, J. C., S. Hansoul, et al. (2008). "Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease." Nat Genet **40**(8): 955-962.

Beach, T. G., L. I. Sue, et al. (2008). "The Sun Health Research Institute Brain Donation Program: description and experience, 1987-2007." Cell Tissue Bank **9**(3): 229-245.

Becker, C., A. Hammerle-Fickinger, et al. (2010). "mRNA and microRNA quality control for RT-qPCR analysis." Methods **50**(4): 237-243.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological) **57**(1): 289-300.

Benovoy, D., T. Kwan, et al. (2008). "Effect of polymorphisms within probe-target sequences on olignonucleotide microarray experiments." Nucleic Acids Research **36**(13): 4417-4423.

Bernstein, C. N., A. Wajda, et al. (2005). "The clustering of other chronic inflammatory diseases in inflammatory bowel disease: a population-based study." Gastroenterology **129**(3): 827-836.

Birdsill, A., D. Walker, et al. (2011). "Postmortem interval effect on RNA and gene expression in human brain tissue." Cell and Tissue Banking **12**(4): 311-318.

Black, D. L. (2003). "Mechanisms of alternative pre-messenger RNA splicing." Annual review of biochemistry **72**: 291-336.

Boutajangout, A., A. Boom, 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): 183-189.

Braak, H. and E. Braak (1991). "Neuropathological stageing of Alzheimer-related changes." Acta Neuropathol **82**(4): 239-259.

Brem, R. B., G. Yvert, et al. (2002). "Genetic dissection of transcriptional regulation in budding yeast." Science **296**(5568): 752-755.

Burke, W. J., K. L. O'Malley, et al. (1991). "Effect of pre- and postmortem variables on specific mRNA levels in human brain." Brain Res Mol Brain Res **11**(1): 37-41.

Caffrey, T. M., C. Joachim, et al. (2008). "Haplotype-specific expression of the N-terminal exons 2 and 3 at the human MAPT locus." Neurobiol Aging **29**(12): 1923-1929.

Caffrey, T. M. and R. Wade-Martins (2012). "The role of MAPT sequence variation in mechanisms of disease susceptibility." Biochemical Society transactions **40**(4): 687-692.

Canales, R. D., Y. Luo, et al. (2006). "Evaluation of DNA microarray results with quantitative gene expression platforms." Nat Biotechnol **24**(9): 1115-1122.

Cartegni, L., S. L. Chew, et al. (2002). "Listening to silence and understanding nonsense: exonic mutations that affect splicing." Nat Rev Genet **3**(4): 285-298.

Castle, J. C., C. Zhang, et al. (2008). "Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines." Nature genetics **40**(12): 1416-1425.

Charlesworth, G., V. Plagnol, et al. (2012). "Mutations in Ano3 Cause Dominant Craniocervical Dystonia: Ion Channel Implicated in Pathogenesis." American journal of human genetics **91**(6): 1041-1050.

Chen, J., Y. Kanai, et al. (1992). "Projection domains of MAP2 and tau determine spacings between microtubules in dendrites and axons." Nature **360**(6405): 674-677.

Chervoneva, I., Y. Li, et al. (2010). "Selection of optimal reference genes for normalization in quantitative RT-PCR." BMC Bioinformatics **11**: 253.

Chevyreva, I., R. L. Faull, et al. (2008). "Assessing RNA quality in postmortem human brain tissue." Exp Mol Pathol **84**(1): 71-77.

Chow, M. L., H. R. Li, et al. (2011). "Genome-wide expression assay comparison across frozen and fixed postmortem brain tissue samples." BMC genomics **12**: 449.

Clark, T. A., A. C. Schweitzer, et al. (2007). "Discovery of tissue-specific exons using comprehensive human exon microarrays." Genome biology **8**(4): R64.

Compta, Y., L. Parkkinen, et al. (2011). "Lewy- and Alzheimer-type pathologies in Parkinson's disease dementia: which is more important?" Brain **134**(Pt 5): 1493-1505.

Cookson, M. R. (2010). "The role of leucine-rich repeat kinase 2 (LRRK2) in Parkinson's disease." Nat Rev Neurosci **11**(12): 791-797.

Cookson, W., L. Liang, et al. (2009). "Mapping complex disease traits with global gene expression." Nat Rev Genet **10**(3): 184-194.

Coulson, D. T., S. Brockbank, et al. (2008). "Identification of valid reference genes for the normalization of RT qPCR gene expression data in human brain tissue." BMC Mol Biol **9**: 46.

Craig, A. M. and Y. Kang (2007). "Neurexin-neuroligin signaling in synapse development." Curr Opin Neurobiol **17**(1): 43-52.

de Jonge, H. J., R. S. Fehrmann, et al. (2007). "Evidence based selection of housekeeping genes." PloS one **2**(9): e898.

de Silva, R., T. Lashley, 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.

Dickson, D. W., J.-J. Hauw, et al. (2011). Progressive Supranuclear Palsy and Corticobasal Degeneration. Neurodegeneration: The Molecular Pathology of Dementia and Movement Disorders, Wiley-Blackwell**:** 135-155.

Diehl, H. J., M. Schaich, et al. (1986). "Individual exons encode the integral membrane domains of human myelin proteolipid protein." Proceedings of the National Academy of Sciences of the United States of America **83**(24): 9807-9811.

Dilks, D., H. P. Ling, et al. (1999). "Cloning and expression of the human kv4.3 potassium channel." Journal of neurophysiology **81**(4): 1974-1977.

Dunham, I., A. Kundaje, et al. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Durrenberger, P. F., S. Fernando, et al. (2010). "Effects of antemortem and postmortem variables on human brain mRNA quality: a BrainNet Europe study." J Neuropathol Exp Neurol **69**(1): 70-81.

Emilsson, V., G. Thorleifsson, et al. (2008). "Genetics of gene expression and its effect on disease." Nature **452**(7186): 423-428.

Enard, W., P. Khaitovich, et al. (2002). "Intra- and interspecific variation in primate gene expression patterns." Science **296**(5566): 340-343.

Feany, M. B., L. A. Mattiace, et al. (1996). "Neuropathologic overlap of progressive supranuclear palsy, Pick's disease and corticobasal degeneration." J Neuropathol Exp Neurol **55**(1): 53-67.

Franke, A., D. P. McGovern, et al. (2010). "Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci." Nat Genet **42**(12): 1118-1125.

Franz, H., C. Ullmann, et al. (2005). "Systematic analysis of gene expression in human brains before and after death." Genome biology **6**(13): R112.

Frappier, T. F., I. S. Georgieff, et al. (1994). "tau Regulation of microtubule-microtubule spacing and bundling." J Neurochem **63**(6): 2288-2294.

French, P. J., J. Peeters, et al. (2007). "Identification of differentially regulated splice variants and novel exons in glial brain tumors using exon expression arrays." Cancer research **67**(12): 5635-5642.

Friedman, J. I., T. Vrijenhoek, et al. (2008). "CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy." Molecular psychiatry **13**(3): 261-266.

Fugier, C., A. F. Klein, et al. (2011). "Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy." Nature medicine **17**(6): 720-725.

Fung, H. C., J. Evans, et al. (2005). "The architecture of the tau haplotype block in different ethnicities." Neuroscience letters **377**(2): 81-84.

Galter, D., M. Westerlund, et al. (2006). "LRRK2 expression linked to dopamine-innervated areas." Annals of neurology **59**(4): 714-719.

Gardina, P. J., T. A. Clark, et al. (2006). "Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array." BMC genomics **7**: 325.

Gerrish, A., G. Russo, et al. (2012). "The role of variation at AbetaPP, PSEN1, PSEN2, and MAPT in late onset Alzheimer's disease." J Alzheimers Dis **28**(2): 377-387.

Geschwind, D. H. and G. Konopka (2009). "Neuroscience in the era of functional genomics and systems biology." Nature **461**(7266): 908-915.

Gibb, G. M., S. R. de, 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.

Gibbs, J. R., M. P. van der Brug, et al. (2010). "Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain." PLoS Genet **6**(5): e1000952.

Gilbert, J. M., B. A. Brown, et al. (1981). "The preparation of biologically active messenger RNA from human postmortem brain tissue." J Neurochem **36**(3): 976-984.

Giuditta, A., J. T. Chun, et al. (2008). "Local gene expression in axons and nerve endings: the glia-neuron unit." Physiological reviews **88**(2): 515-555.

Glanzer, J. G., P. G. Haydon, et al. (2004). "Expression profile analysis of neurodegenerative disease: advances in specificity and resolution." Neurochem Res **29**(6): 1161-1168.

Glasel, J. A. (1995). "Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios." Biotechniques **18**(1): 62-63.

Goedert, M. (2004). "Tau protein and neurodegeneration." Semin Cell Dev Biol **15**(1): 45-49.

Goedert, M., M. G. Spillantini, et al. (1989). "Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease." Neuron **3**(4): 519-526.

Goedert, M., M. G. Spillantini, et al. (1989). "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): 393-399.

Golbe, L. I., A. M. Lazzarini, et al. (2001). "The tau A0 allele in Parkinson's disease." Mov Disord **16**(3): 442-447.

Grabowski, P. J. and D. L. Black (2001). "Alternative RNA splicing in the nervous system." Progress in neurobiology **65**(3): 289-308.

Greggio, E. and M. R. Cookson (2009). "Leucine-rich repeat kinase 2 mutations and Parkinson's disease: three questions." ASN Neuro **1**(1).

Greggio, E., S. Jain, et al. (2006). "Kinase activity is required for the toxic effects of mutant LRRK2/dardarin." Neurobiol Dis **23**(2): 329-341.

Grohmann, K., M. Schuelke, et al. (2001). "Mutations in the gene encoding immunoglobulin mu-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1." Nature genetics **29**(1): 75-77.

Gu, J. and X. Gu (2003). "Induced gene expression in human brain after the split from chimpanzee." Trends Genet **19**(2): 63-65.

Hakimi, M., T. Selvanantham, et al. (2011). "Parkinson's disease-linked LRRK2 is expressed in circulating and tissue immune cells and upregulated following recognition of microbial structures." J Neural Transm **118**(5): 795-808.

Hall, J. S., H. S. Leong, et al. (2011). "Exon-array profiling unlocks clinically and biologically relevant gene signatures from formalin-fixed paraffin-embedded tumour samples." Br J Cancer **104**(6): 971-981.

Hanger, D. P., G. M. Gibb, et al. (2002). "The complex relationship between soluble and insoluble tau in tauopathies revealed by efficient dephosphorylation and specific antibodies." FEBS Lett. **531**(3): 538-542.

Hardy, J. and A. Singleton (2009). "Genomewide association studies and human disease." N Engl J Med **360**(17): 1759-1768.

Hardy, J., D. Trabzuni, et al. (2009). "Whole genome expression as a quantitative trait." Biochem Soc Trans **37**(Pt 6): 1276-1277.

Hardy, J. A., P. Wester, et al. (1985). "The patients dying after long terminal phase have acidotic brains; implications for biochemical measurements on autopsy tissue." J Neural Transm **61**(3-4): 253-264.

Harrison, P. J., P. R. Heath, et al. (1995). "The relative importance of premortem acidosis and postmortem interval for human brain gene expression studies: selective mRNA vulnerability and comparison with their encoded proteins." Neurosci Lett **200**(3): 151-154.

Harrow, J., A. Frankish, et al. (2012). "GENCODE: the reference human genome annotation for The ENCODE Project." Genome research **22**(9): 1760-1774.

Haverty, P. M., Z. Weng, et al. (2002). "HugeIndex: a database with visualization tools for high-density oligonucleotide array data from normal human tissues." Nucleic Acids Research **30**(1): 214-217.

Hawkins, R. D., G. C. Hon, et al. (2010). "Next-generation genomics: an integrative approach." Nat Rev Genet **11**(7): 476-486.

Hawrylycz, M. J., E. S. Lein, et al. (2012). "An anatomically comprehensive atlas of the adult human brain transcriptome." Nature **489**(7416): 391-399.

Healy, D. G., M. Falchi, et al. (2008). "Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study." Lancet neurology **7**(7): 583-590.

Heinzen, E. L., D. Ge, et al. (2008). "Tissue-specific genetic control of splicing: implications for the study of complex traits." PLoS biology **6**(12): e1.

Hernandez, D. G., M. A. Nalls, et al. (2012). "Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain." Neurobiology of disease **47**(1): 20-28.

Higashi, S., S. Biskup, et al. (2007). "Localization of Parkinson's disease-associated LRRK2 in normal and pathological human brain." Brain Res **1155**: 208-219.

Higashi, S., D. J. Moore, et al. (2007). "Expression and localization of Parkinson's disease-associated leucine-rich repeat kinase 2 in the mouse brain." Journal of neurochemistry **100**(2): 368-381.

Hindorff, L. A., J. MacArthur, et al. (2012). "A Catalog of Published Genome-Wide Association Studies. ." **Available at: www.genome.gov/gwastudies [Accessed 6th September 2012].**

Hindorff, L. A., P. Sethupathy, et al. (2009). "Potential etiologic and functional implications of genome-wide association loci for human diseases and traits." Proceedings of the National Academy of Sciences of the United States of America **106**(23): 9362-9367.

Hishimoto, A., Q. R. Liu, et al. (2007). "Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms." Human molecular genetics **16**(23): 2880-2891.

Ho Sui, S. J., J. R. Mortimer, et al. (2005). "oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes." Nucleic Acids Res **33**(10): 3154-3164.

Hodges, A., A. D. Strand, et al. (2006). "Regional and cellular gene expression changes in human Huntington's disease brain." Human molecular genetics **15**(6): 965-977.

Hoglinger, G. U., N. M. Melhem, et al. (2011). "Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy." Nature Genetics **43**(7): 699-705.

Hollingworth, P., D. Harold, et al. (2011). "Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease." Nature genetics **43**(5): 429-435.

Houlden, H., M. Baker, et al. (2001). "Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype." Neurology **56**(12): 1702-1706.

Howie, B., C. Fuchsberger, et al. (2012). "Fast and accurate genotype imputation in genome-wide association studies through pre-phasing." Nature genetics **44**(8): 955-959.

Hua, X., A. D. Leow, et al. (2008). "Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: an MRI study of 676 AD, MCI, and normal subjects." NeuroImage **43**(3): 458-469.

Huang da, W., B. T. Sherman, et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.

Imbeaud, S., E. Graudens, et al. (2005). "Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces." Nucleic Acids Res **33**(6): e56.

International Parkinson's Disease Genomics Consortium and Wellcome Trust Case Control Consortium 2 (2011). "A two-stage meta-analysis identifies several new loci for Parkinson's disease." PLoS Genet **7**(6): e1002142.

Irizarry, R. A., B. Hobbs, et al. (2003). "Exploration, normalization, and summaries of high density oligonucleotide array probe level data." Biostatistics **4**(2): 249-264.

Ittner, L. M., Y. D. Ke, et al. (2010). "Dendritic function of tau mediates amyloid-beta toxicity in Alzheimer's disease mouse models." Cell **142**(3): 387-397.

Jaleel, M., R. J. Nichols, et al. (2007). "LRRK2 phosphorylates moesin at threonine-558: characterization of how Parkinson's disease mutants affect kinase activity." Biochem J **405**(2): 307-317.

Jin, W., R. M. Riley, et al. (2001). "The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster." Nature genetics **29**(4): 389-395.

Johnson, M. B., Y. I. Kawasawa, et al. (2009). "Functional and evolutionary insights into human brain development through global transcriptome analysis." Neuron **62**(4): 494-509.

Johnston, N. L., J. Cervenak, et al. (1997). "227 - Measurement of RNA from 89 postmortem human brains; A multivariate statistical analysis of pre- and postmortem effects on the yields of GAPdH as measured by RT-PCR." Schizophrenia Research **24**(1–2): 87.

Kamboh, M. I., F. Y. Demirci, et al. (2012). "Genome-wide association study of Alzheimer's disease." Translational psychiatry **2**: e117.

Kang, H. J., Y. I. Kawasawa, et al. (2011). "Spatio-temporal transcriptome of the human brain." Nature **478**(7370): 483-489.

Karaman, M. W., M. L. Houck, et al. (2003). "Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts." Genome research **13**(7): 1619-1630.

Karch, C. M., A. T. Jeng, et al. (2012). "Expression of Novel Alzheimer's Disease Risk Genes in Control and Alzheimer's Disease Brains." PLoS One **7**(11): e50976.

Khaitovich, P., B. Muetzel, et al. (2004). "Regional patterns of gene expression in human and chimpanzee brains." Genome research **14**(8): 1462-1473.

Klein, C. L., G. Rovelli, et al. (2009). "Homo- and heterodimerization of ROCO kinases: LRRK2 kinase inhibition by the LRRK2 ROCO fragment." Journal of neurochemistry **111**(3): 703-715.

Krawczak, M., J. Reiss, et al. (1992). "The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences." Human genetics **90**(1-2): 41-54.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lee, C. J. and K. Irizarry (2003). "Alternative splicing in the nervous system: an emerging source of diversity and regulation." Biological psychiatry **54**(8): 771-776.

Lee, G. (2005). "Tau and src family tyrosine kinases." Biochim Biophys Acta **1739**(2-3): 323-330.

Lee, G., S. T. Newman, et al. (1998). "Tau interacts with src-family non-receptor tyrosine kinases." J.Cell Sci. **111 (Pt 21)**: 3167-3177.

Leonard, S., J. Logel, et al. (1993). "Biological stability of mRNA isolated from human postmortem brain collections." Biol Psychiatry **33**(6): 456-466.

Lewis, N. E., G. Schramm, et al. (2010). "Large-scale in silico modeling of metabolic interactions between cell types in the human brain." Nature biotechnology **28**(12): 1279-1285.

Lewis, P. A. (2009). "The function of ROCO proteins in health and disease." Biol Cell **101**(3): 183-191.

Lewis, P. A., E. Greggio, et al. (2007). "The R1441C mutation of LRRK2 disrupts GTP hydrolysis." Biochem Biophys Res Commun **357**(3): 668-671.

Lewis, P. A. and C. Manzoni (2012). "LRRK2 and Human Disease: A Complicated Question or a Question of Complexes?" Sci Signal **5**(207): pe2.

Li, J. Z., F. Meng, et al. (2007). "Sample matching by inferred agonal stress in gene expression analyses of the brain." BMC Genomics **8**: 336.

Li, Y., C. Willer, et al. (2009). "Genotype imputation." Annu Rev Genomics Hum Genet **10**: 387-406.

Li, Y., C. J. Willer, et al. (2010). "MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes." Genet Epidemiol **34**(8): 816-834.

Lill, C. M., J. T. Roehr, et al. (2012). "Comprehensive Research Synopsis and Systematic Meta-Analyses in Parkinson's Disease Genetics: The PDGene Database." PLoS Genet **8**(3): e1002548.

Lin, X., L. Parisiadou, et al. (2009). "Leucine-rich repeat kinase 2 regulates the progression of neuropathology induced by Parkinson's-disease-related mutant alpha-synuclein." Neuron **64**(6): 807-827.

Liss, B., O. Franz, et al. (2001). "Tuning pacemaker frequency of individual dopaminergic neurons by Kv4.3L and KChip3.1 transcription." The EMBO journal **20**(20): 5715-5724.

Liu, Z., J. Lee, et al. (2011). "The kinase LRRK2 is a regulator of the transcription factor NFAT that modulates the severity of inflammatory bowel disease." Nat Immunol **12**(11): 1063-1070.

Llinas RR, W. K., Lang EJ. (2004). "Ch. 7 Cerebellum". In Shepherd GM. . The Synaptic Organization of the Brain., New York: Oxford University Press.

Luk, C., G. Giovannoni, et al. (2009). "Development of a sensitive ELISA for quantification of three- and four-repeat tau isoforms in tauopathies." Journal of neuroscience methods **180**(1): 34-42.

Luk, C., J. Vandrovcova, et al. (2010). "Brain tau isoform mRNA and protein correlation in PSP brain." Translational Neuroscience **1**(1): 7.

Luzon-Toro, B., E. Rubio de la Torre, et al. (2007). "Mechanistic insight into the dominant mode of the Parkinson's disease-associated G2019S LRRK2 mutation." Hum Mol Genet **16**(17): 2031-2039.

Mandemakers, W., A. Snellinx, et al. (2012). "LRRK2 expression is enriched in the striosomal compartment of mouse striatum." Neurobiology of disease **48**(3): 582-593.

Manolio, T. A. (2010). "Genomewide association studies and assessment of the risk of disease." The New England journal of medicine **363**(2): 166-176.

Manolio, T. A., F. S. Collins, et al. (2009). "Finding the missing heritability of complex diseases." Nature **461**(7265): 747-753.

McCarthy, M. I., G. R. Abecasis, et al. (2008). "Genome-wide association studies for complex traits: consensus, uncertainty and challenges." Nat Rev Genet **9**(5): 356-369.

Melrose, H. L., C. B. Kent, et al. (2007). "A comparative analysis of leucine-rich repeat kinase 2 (Lrrk2) expression in mouse brain and Lewy body disease." Neuroscience **147**(4): 1047-1058.

Mexal, S., R. Berger, et al. (2006). "Brain pH has a significant impact on human postmortem hippocampal gene expression profiles." Brain Res **1106**(1): 1-11.

Milenkovic, V. M., M. Brockmann, et al. (2010). "Evolution and functional divergence of the anoctamin family of membrane proteins." BMC evolutionary biology **10**: 319.

Millar, T., R. Walker, et al. (2007). "Tissue and organ donation for research in forensic pathology: the MRC Sudden Death Brain and Tissue Bank." J Pathol **213**(4): 369-375.

Mirnics, K., F. A. Middleton, et al. (2001). "Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse." Trends in neurosciences **24**(8): 479-486.

Mistry, M. and P. Pavlidis (2010). "A cross-laboratory comparison of expression profiling data from normal human postmortem brain." Neuroscience **167**(2): 384-395.

Mizuno, Y., K. Takahashi, et al. (1995). "[Cerebellin in the cerebellum in spinocerebellar degeneration]." No to shinkei = Brain and nerve **47**(11): 1069-1074.

Moehle, M. S., P. J. Webber, et al. (2012). "LRRK2 inhibition attenuates microglial inflammatory responses." The Journal of neuroscience : the official journal of the Society for Neuroscience **32**(5): 1602-1611.

Monks, S. A., A. Leonardson, et al. (2004). "Genetic inheritance of gene expression in human cell lines." American journal of human genetics **75**(6): 1094-1105.

Monoranu, C. M., M. Apfelbacher, et al. (2009). "pH measurement as quality control on human post mortem brain tissue: a study of the BrainNet Europe consortium." Neuropathol Appl Neurobiol **35**(3): 329-337.

Moreira, M. C., S. Klur, et al. (2004). "Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2." Nature genetics **36**(3): 225-227.

Morley, M., C. M. Molony, et al. (2004). "Genetic analysis of genome-wide variation in human gene expression." Nature **430**(7001): 743-747.

Myers, A. J., J. R. Gibbs, et al. (2007). "A survey of genetic human cortical gene expression." Nature Genetics **39**(12): 1494-1499.

Myers, A. J., M. Kaleem, et al. (2005). "The H1c haplotype at the MAPT locus is associated with Alzheimer's disease." Hum Mol Genet **14**(16): 2399-2404.

Myers, A. J., A. M. Pittman, 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): 561-570.

Naiser, T., O. Ehler, et al. (2008). "Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: comparison of single base mismatches and base bulges." BMC Biotechnol **8**: 48.

Naj, A. C., G. Jun, et al. (2011). "Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease." Nature genetics **43**(5): 436-441.

Nalls, M. A., V. Plagnol, 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): 641-649.

Naumova, O. Y., D. Palejev, et al. (2012). "Age-related changes of gene expression in the neocortex: Preliminary data on RNA-Seq of the transcriptome in three functionally distinct cortical areas." Development and psychopathology **24**(4): 1427-1442.

Nica, A. C., L. Parts, et al. (2011). "The architecture of gene regulatory variation across multiple human tissues: the MuTHER study." PLoS genetics **7**(2): e1002003.

Ozsolak, F. and P. M. Milos (2011). "RNA sequencing: advances, challenges and opportunities." Nat Rev Genet **12**(2): 87-98.

Paisan-Ruiz, C., S. Jain, et al. (2004). "Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease." Neuron **44**(4): 595-600.

Peacey, E., L. Rodriguez, et al. (2012). "Targeting a pre-mRNA structure with bipartite antisense molecules modulates tau alternative splicing." Nucleic Acids Research **40**(19): 9836-9849.

Peirce, J. L., H. Li, et al. (2006). "How replicable are mRNA expression QTL?" Mammalian genome : official journal of the International Mammalian Genome Society **17**(6): 643-656.

Pitt, M., H. Houlden, et al. (2003). "Severe infantile neuropathy with diaphragmatic weakness and its relationship to SMARD1." Brain : a journal of neurology **126**(Pt 12): 2682-2692.

Pittman, A. M., A. J. Myers, et al. (2005). "Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration." Journal of Medical Genetics **42**(11): 837-846.

Pittman, A. M., A. J. Myers, et al. (2004). "The structure of the tau haplotype in controls and in progressive supranuclear palsy." Hum.Mol.Genet. **13**(12): 1267-1274.

Plagnol, V., D. J. Smyth, et al. (2009). "Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13." Biostatistics **10**(2): 327-334.

Prescott, G. R. and L. H. Chamberlain (2011). "Regional and developmental brain expression patterns of SNAP25 splice variants." BMC neuroscience **12**: 35.

Purcell, S., B. Neale, et al. (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." American journal of human genetics **81**(3): 559-575.

Qin, W., V. Haroutunian, et al. (2009). "PGC-1alpha expression decreases in the Alzheimer disease brain as a function of dementia." Archives of Neurology **66**(3): 352-361.

Roghani, A., C. Welch, et al. (1996). "Assignment of the mouse vesicular monoamine transporter genes, Slc18a1 and Slc18a2, to chromosomes 8 and 19 by linkage analysis." Mammalian genome : official journal of the International Mammalian Genome Society **7**(5): 393-394.

Ross, D. T., U. Scherf, et al. (2000). "Systematic variation in gene expression patterns in human cancer cell lines." Nature genetics **24**(3): 227-235.

Ross, J. (1995). "mRNA stability in mammalian cells." Microbiol Rev **59**(3): 423-450.

Ross, O. A., A. I. Soto-Ortolaza, et al. (2011). "Association of LRRK2 exonic variants with susceptibility to Parkinson's disease: a case-control study." Lancet neurology **10**(10): 898-908.

Roth, R. B., P. Hevezi, et al. (2006). "Gene expression analyses reveal molecular relationships among 20 regions of the human CNS." Neurogenetics **7**(2): 67-80.

Sajdel-Sulkowska, E. M., R. E. Majocha, et al. (1988). "The postmortem Alzheimer brain is a source of structurally and functionally intact astrocytic messenger RNA." J Neurosci Methods **23**(2): 173-179.

Salta, E. and B. De Strooper (2012). "Non-coding RNAs with essential roles in neurodegenerative disorders." Lancet neurology **11**(2): 189-200.

Sanderson, C. M. (2009). "The Cartographers toolbox: building bigger and better human protein interaction networks." Briefings in functional genomics & proteomics **8**(1): 1-11.

Schadt, E. E., C. Molony, et al. (2008). "Mapping the genetic architecture of gene expression in human liver." PLoS biology **6**(5): e107.

Schadt, E. E., S. A. Monks, et al. (2003). "Genetics of gene expression surveyed in maize, mouse and man." Nature **422**(6929): 297-302.

Schreiber, R., I. Uliyakina, et al. (2010). "Expression and function of epithelial anoctamins." The Journal of biological chemistry **285**(10): 7838-7845.

Schroeder, A., O. Mueller, et al. (2006). "The RIN: an RNA integrity number for assigning integrity values to RNA measurements." BMC Mol Biol **7**: 3.

Schultz, W. (1998). "Predictive reward signal of dopamine neurons." Journal of neurophysiology **80**(1): 1-27.

SeattleSNPs. (2012). "NHLBI Program for Genomic Applications, SeattleSNPs, Seattle, WA." **Available at: http://pga.gs.washington.edu [Accessed 1st September 2012].**

Sergeant, N., A. Wattez, et al. (1999). "Neurofibrillary degeneration in progressive supranuclear palsy and corticobasal degeneration: tau pathologies with exclusively "exon 10" isoforms." J Neurochem **72**(3): 1243-1249.

Serodio, P., E. Vega-Saenz de Miera, et al. (1996). "Cloning of a novel component of A-type K+ channels operating at subthreshold potentials with unique expression in heart and brain." Journal of neurophysiology **75**(5): 2174-2179.

Shabalin, A. A. (2012). "Matrix eQTL: Ultra fast eQTL analysis via large matrix operations." Bioinformatics **28**(10): 1353-8.

Sharma, S., R. Bandopadhyay, et al. (2011). "LRRK2 expression in idiopathic and G2019S positive Parkinson's disease subjects: a morphological and quantitative study." Neuropathology and Applied Neurobiology **37**(7): 777-790.

Sherwood, K. R., M. W. Head, et al. (2011). "RNA integrity in post mortem human variant Creutzfeldt-Jakob disease (vCJD) and control brain tissue." Neuropathology and Applied Neurobiology **37**(6): 633-642.

Simon, H. H., H. Saueressig, et al. (2001). "Fate of midbrain dopaminergic neurons controlled by the engrailed genes." The Journal of neuroscience : the official journal of the Society for Neuroscience **21**(9): 3126-3134.

Son, C. G., S. Bilke, et al. (2005). "Database of mRNA gene expression profiles of multiple human organs." Genome research **15**(3): 443-450.

Srinivasan, K., L. Shiue, et al. (2005). "Detection and measurement of alternative splicing using splicing-sensitive microarrays." Methods **37**(4): 345-359.

Stan, A. D., S. Ghose, et al. (2006). "Human postmortem tissue: what quality markers matter?" Brain Res **1123**(1): 1-11.

Stefansson, H., A. Helgason, et al. (2005). "A common inversion under selection in Europeans." Nat.Genet. **37**(2): 129-137.

Stranger, B. E., M. S. Forrest, et al. (2005). "Genome-Wide Associations of Gene Expression Variation in Humans." PLoS Genet **1**(6): e78.

Su, A. I., T. Wiltshire, et al. (2004). "A gene atlas of the mouse and human protein-encoding transcriptomes." Proceedings of the National Academy of Sciences of the United States of America **101**(16): 6062-6067.

Sudhof, T. C. (2008). "Neuroligins and neurexins link synaptic function to cognitive disease." Nature **455**(7215): 903-911.

Suraweera, A., Y. Lim, et al. (2009). "Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation." Human molecular genetics **18**(18): 3384-3396.

Tabuchi, K. and T. C. Sudhof (2002). "Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing." Genomics **79**(6): 849-859.

Takahashi, K., K. Tanabe, et al. (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-872.

Tamhane, A. C. and D. D. Dunlop (2000). Statistics and data analysis : from elementary to intermediate. Upper Saddle River, NJ ; London, Prentice Hall, pages 473-474.

Tan, G. C., T. F. Doke, et al. (2010). "Normal variation in fronto-occipital circuitry and cerebellar structure with an autism-associated polymorphism of CNTNAP2." NeuroImage **53**(3): 1030-1042.

Thevenet, J., R. Pescini Gobert, et al. (2011). "Regulation of LRRK2 expression points to a functional role in human monocyte maturation." PloS one **6**(6): e21519.

Tollervey, J. R., Z. Wang, et al. (2011). "Analysis of alternative splicing associated with aging and neurodegeneration in the human brain." Genome research **21**(10): 1572-1582.

Tomita, H., M. P. Vawter, et al. (2004). "Effect of agonal and postmortem factors on gene expression profile: quality control in microarray analyses of postmortem human brain." Biol Psychiatry **55**(4): 346-352.

Trabzuni, D., M. Ryten, et al. (2011). "Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies." Journal of neurochemistry **119**(2): 275-282.

Trabzuni, D., S. Wray, et al. (2012). "MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies." Human Molecular Genetics **21**(18): 4094-4103.

Vaags, A. K., A. C. Lionel, et al. (2012). "Rare deletions at the neurexin 3 locus in autism spectrum disorder." American journal of human genetics **90**(1): 133-141.

Vennemann, M. and A. Koppelkamm (2010). "mRNA profiling in forensic genetics I: Possibilities and limitations." Forensic Sci Int **203**(1-3): 71-75.

Vitte, J., S. Traver, et al. (2010). "Leucine-rich repeat kinase 2 is associated with the endoplasmic reticulum in dopaminergic neurons and accumulates in the core of Lewy bodies in Parkinson disease." J Neuropathol Exp Neurol **69**(9): 959-972.

Wade, C. M., E. J. Kulbokas, 3rd, et al. (2002). "The mosaic structure of variation in the laboratory mouse genome." Nature **420**(6915): 574-578.

Wang, E. T., R. Sandberg, et al. (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature **456**(7221): 470-476.

Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-562.

Webster, J. A., J. R. Gibbs, et al. (2009). "Genetic control of human brain transcript expression in Alzheimer disease." Am J Hum Genet **84**(4): 445-458.

Weis, S., I. C. Llenos, et al. (2007). "Quality control for microarray analysis of human brain samples: The impact of postmortem factors, RNA characteristics, and histopathology." J Neurosci Methods **165**(2): 198-209.

West, A. B., D. J. Moore, et al. (2005). "Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity." Proc Natl Acad Sci U S A **102**(46): 16842-16847.

Whistler, T., C. F. Chiang, et al. (2010). "The comparison of different pre- and post-analysis filters for determination of exon-level alternative splicing events using Affymetrix arrays." Journal of biomolecular techniques : JBT **21**(1): 44-53.

Wigge, P. and H. T. McMahon (1998). "The amphiphysin family of proteins and their role in endocytosis at the synapse." Trends in neurosciences **21**(8): 339-344.

Workman, C., L. J. Jensen, et al. (2002). "A new non-linear normalization method for reducing variability in DNA microarray experiments." Genome biology **3**(9): research0048.

Wszolek, Z. K., P. Vieregge, et al. (1997). "German-Canadian family (family A) with parkinsonism, amyotrophy, and dementia - Longitudinal observations." Parkinsonism.Relat Disord. **3**(3): 125-139.

Wu, Y. W., K. M. Prakash, et al. (2011). "Lingo2 variants associated with essential tremor and Parkinson's disease." Human genetics **129**(6): 611-615.

Wu, Z. and R. A. Irizarry (2004). "Preprocessing of oligonucleotide array data." Nature biotechnology **22**(6): 656-658; author reply 658.

Xi, L., A. Feber, et al. (2008). "Whole genome exon arrays identify differential expression of alternatively spliced, cancer-related genes in lung cancer." Nucleic Acids Research **36**(20): 6535-6547.

Yamaguchi, N., A. Okui, et al. (2002). "Spinesin/TMPRSS5, a novel transmembrane serine protease, cloned from human spinal cord." The Journal of biological chemistry **277**(9): 6806-6812.

Yeo, G., D. Holste, et al. (2004). "Variation in alternative splicing across human tissues." Genome biology **5**(10): R74.

Zeller, T., P. Wild, et al. (2010). "Genetics and beyond--the transcriptome of human monocytes and disease susceptibility." PLoS One **5**(5): e10693.

Zhang, B. and S. Horvath (2005). "A general framework for weighted gene co-expression network analysis." Statistical applications in genetics and molecular biology **4**: Article17.

Zhang, F. R., W. Huang, et al. (2009). "Genomewide association study of leprosy." N Engl J Med **361**(27): 2609-2618.

Zhong, Q., E. E. Congdon, et al. (2012). "Tau isoform composition influences rate and extent of filament formation." The Journal of biological chemistry **287**(24): 20711-20719.

Zimprich, A., S. Biskup, et al. (2004). "Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology." Neuron **44**(4): 601-607.

Zody, M. C., Z. Jiang, et al. (2008). "Evolutionary toggling of the MAPT 17q21.31 inversion region." Nat.Genet. **40**(9): 1076-1083.

Zou, F., M. M. Carrasquillo, et al. (2010). "Gene expression levels as endophenotypes in genome-wide association studies of Alzheimer disease." Neurology **74**(6): 480-486.