Functional genomic characterisation of animal models of AD: relevance to human dementia

Submitted by
Isabel Castanho
to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Medical Studies
in March 2019

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: 

Abstract

The onset and progression of Alzheimer's disease (AD) is characterised by increasing intracellular aggregation of hyperphosphorylated tau protein and the accumulation of amyloid beta (Aβ) in the neocortex. Despite recent success in identifying genetic risk factors for AD, the transcriptional and epigenomic mechanisms involved in disease progression are not fully understood. The main aim of this project was to evaluate transcriptional and epigenomic differences associated with the development of tau and amyloid pathology. To achieve this, I used transgenic mice harbouring human tau (rTg4510) and amyloid precursor protein (J20) mutations. I profiled transcriptional and epigenomic variation in brains from rTg4510 and J20 mice, collected at four time points carefully selected to span from early to late stages of neuropathology in each model. I identified robust gene expression and methylomic changes in both models, including genes associated with familial AD from genetic studies of human patients, and genes annotated to both common and rare variants identified in genome-wide association and exome-sequencing studies of late-onset sporadic AD. I quantified neuropathological burden across multiple brain regions in the same individual mice, identifying genomic changes paralleling the development of tau pathology in rTg4510 mice and amyloid pathology in J20 mice. Furthermore, I compared gene co-expression networks identified in my rTg4510 and J20 samples to those identified in AD human brains, finding considerable overlap with disease-associated co-expression modules (or clusters of genes) identified in the human cortex. In summary, this project represents the most systematic analysis of transcriptional and methylomic variation in mouse models of tau and amyloid pathology, providing further support for an immune-response component in the accumulation of AD-associated neuropathology, and highlighting novel molecular pathways involved in AD progression.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my first supervisor, Prof. Jonathan Mill, one of the most brilliant people I have known, for his guidance, continuous support and encouragement, for investing and believing in me, and for keeping his sense of humour even when I was about to lose mine. I would also like to thank my second supervisor, Prof. Katie Lunnon, a truly inspirational female researcher, for her insightful comments and feedback, and for her continuous availability and motivation. I could have never imagined I would learn this much from my two supervisors in the last three and a half years; I really feel I hit the jackpot when I decided to accept this PhD, supervised by such an amazing duo.

To all my fellow lab mates in the Complex Disease Epigenetics Group – the best research group I could ever have joined for my PhD – thank you. A special thanks to Dr Eilis Hannon, whose statistical and analytical help and input were absolutely priceless; to Dr Joana Viana and Dr Aaron Jeffries for all their guidance and contribution both analytical and laboratory-wise; to Dorothea Seiler Vellame and Emma Walker for lending me their skills and giving me some help with the analysis during more desperate times; and to Hedley Baulf for being such a dedicated and involved student.

My sincere thanks to Dr Michael O'Neill, who provided me an opportunity to join his team as a placement student, giving me access to the laboratory and research facilities at Eli Lilly. I would like to thank everyone at the Molecular Pathology laboratories at Eli Lilly, particularly Tracey K. Murray and Zeshan Ahmed, as well as Joshua Harvey, Mark Ward, Katherine Sung, Alice Fisher, and Claire Cella – thank you for all the technical assistance with the histopathology and thank you for making me feel so welcome and part of your 'family'.

A special thanks to Prof. Andrew Randall from the Exeter Applied Neurophysiology group – the director of the Alzheimer's Society Doctoral Training Centre at the University of Exeter, that funds my PhD – for funding, for lending his immense knowledge, and for his insightful comments.

I would also like to thank the team from the Sequencing Service and Computational core facilities at the University of Exeter, particularly Dr Karen Moore, Audrey Farbos, and Paul O'Neill.

To João, my partner, my best friend, and the person that supports me the most, believing more in me than anyone, including myself. Firstly, I would like to thank him for rescuing me way too many times with coding/programming. Secondly, I would like to thank him for all his unconditional love, endless patience and understanding, comforting cuddles and for taking such good care of me. I would not have done it without him (at least not with my sanity intact); so, from the bottom of my heart: thank you for being there throughout the roller-coaster journey that was my PhD.

To my family and friends – thank you for putting up with me all these years and for supporting me so much. A special thanks to my dad for his constant concern and for making me see things in perspective; to Alexandra for always being present despite the physical distance; to Dr Anais Kahve for all the conversations over lunch or tea.

Finally, I dedicate this thesis to myself. Today I am very proud of my achievements and I recognise all the hard work and the difficulties that I overcame throughout my life – all contributors to get me to where I am today – but this was not always the case. I am now very grateful for my life and for myself, and I am very proud of this thesis.

List of Contents

List of Figures

List of Tables

Publications arising from this thesis

Chapter 1

CASTANHO, I. & LUNNON, K. 2019. Epigenetic processes in Alzheimer's disease. *In:* BINDA, O. (ed.) *Chromatin Signaling and Neurological Disorders.* Elsevier.

Chapter 4

CASTANHO, I., MURRAY, T. K., HANNON, E., JEFFRIES, A., WALKER, E., LAING, E., BAULF, H., HARVEY, J., RANDALL, A., MOORE, K., O'NEILL, P., LUNNON, K., COLLIER, D. A., AHMED, Z., O'NEILL, M. J. & MILL, J. 2019. Transcriptional signatures of progressive neuropathology in transgenic tau and amyloid mouse models. *bioRxiv***,** 548578.

Chapter 5

CASTANHO, I., BAULF, H., HARVEY, J., MURRAY, T. K., LUNNON, K., O'NEILL, M. J., AHMED, Z. & MILL, J. in prep. Global DNA modifications profiling in the brain of tau and amyloid mouse models.

Chapter 6

CASTANHO, I., SEILER VELLAME, D., MURRAY, T. K., VIANA, J., HANNON, E., MOORE, K., O'NEILL, P., LUNNON, K., COLLIER, D. A., AHMED, Z., NEILL, M. J. & MILL, J. in prep. Genome-wide methylation profiling of tau and amyloid mice.

Selected Conference Proceedings

2019	Castanho I. et al. Mapping genomic consequences of Alzheimer's disease pathology in amyloid and tau mouse models. **Poster presentation** at the AD/PD conference; 2019. Lisbon, Portugal.
2018	Castanho I. et al. Mapping genomic consequences of Alzheimer's disease pathology in amyloid and tau mouse models. **Poster presentation** at the Alzheimer's Association International Conference (AAIC); 2018. Chicago, USA.
	Castanho I. et al. Amyloid and tau mouse models: mapping genomic consequences. **Oral communication** at the Alzheimer's Society Research Conference; 2018. London, UK.
	Castanho I. et al. Transcriptional and epigenomic profiling in the entorhinal cortex in amyloid and tau mouse models of Alzheimer's disease. **Oral communication** at the Alzheimer's Research UK (ARUK) annual conference; 2018. London, UK.
2017	Castanho I. et al. Transcriptional and epigenomic profiling in the entorhinal cortex in amyloid and tau mouse models of Alzheimer's disease. **Poster presentation** at the Alzheimer's Association International Conference (AAIC); 2017. London, UK.
	Castanho I. et al. Transcriptional and epigenomic profiling in the entorhinal cortex in a mouse model of Alzheimer's disease. **Poster presentation** at the Conference of the Portuguese Society for Neuroscience; 2017. Braga, Portugal.
	Castanho I. et al. Transcriptional and epigenomic profiling in the entorhinal cortex in a mouse model of Alzheimer's disease. **Poster presentation** at the Alzheimer's Society Research Conference; 2017. Aberdeen, UK.
2016	Castanho I. et al. Transcriptional and epigenomic profiling of the entorhinal cortex in a mouse model of Alzheimer's disease. **Poster presentation** at the Epigenomics of Common Diseases conference; 2016. Wellcome Trust Centre, Cambridge, UK.
	Castanho I. et al. A functional genomic analysis of rodent models of Alzheimer's disease: relevance to human dementia. **Poster presentation** at the Alzheimer's Society Research Conference; 2017. Bristol, UK

Grants, Awards and Achievements

2018	ISTAART Student Volunteer Lead award for the Alzheimer's Association International Conference (AAIC) 2018
	British Neuroscience Association (BNA) Local Group Funding for public engagement (Brain Awareness Week)
2017	ISTAART Student Volunteer award for the Alzheimer's Association International Conference (AAIC) 2017
	Travel award for the Portuguese Society for Neuroscience (SPN) Meeting 2017
2016	Registration bursary award for the Epigenomics of Common Diseases conference 2016 in Cambridge, UK
	Alzheimer's Research UK (ARUK) South West Network Centre pilot grant

Declarations

All mouse samples used in this thesis were obtained from Eli Lilly & Co. Ltd., Windlesham (United Kingdom).

All laboratory work and associated analyses were conducted by me, with the following exceptions:

- Sample collection (**Chapter 2**, **Section 2.3**) was performed by Dr Tracey K. Murray at Eli Lilly, who supervised all animal work.
- The initial steps of processing (formalin fixation and paraffin embedding) of the right brain hemispheres (**Chapter 2**, **Section 2.4.2**) for histology experiments were performed by Joshua Harvey, and Mark Ward, at Eli Lilly.
- The remaining steps of processing of the right brain hemisphere (described in **Chapter 2**, **Section 3.2.1**), and all histopathology experiments for **Chapter 3** and **Chapter 5** were performed by me with assistance from Mark Ward, Katherine Sung, and Joshua Harvey, at Eli Lilly.
- I had support from Joshua Harvey, Katherine Sung, Alice Fisher, and Claire Cella from Eli Lilly for the quantification of positivity described in **Chapter 3** (**Section 3.2.5**) and **Chapter 5** (**Section 5.2.8**).
- Strand-specific mRNA sequencing libraries described in **Chapter 4** were prepared by me with the help of Audrey Farbos at the University of Exeter Sequencing Service.
- Sequencing of the RNA libraries for **Chapter 4** and of the RRBS libraries for **Chapter 6** was carried out by Audrey Farbos and Dr Karen Moore at the University of Exeter Sequencing Service.
- Hedley Baulf, an undergraduate student from the University of Exeter that I directly supervised, performed the LUMA experiments under my guidance (**Chapter 5**).

All bioinformatics and statistical analyses were conducted by me with a few exceptions, where I had support from others:

- Demultiplexing of the raw sequencing files, contamination check and spike-in internal controls quality check for **Chapter 4** and **Chapter 6** were performed by Paul O'Neill at the University of Exeter Sequencing Service.
- Emma Walker from the University of Exeter and Emma Laing from Eli Lilly provided me with assistance for the comparison of my RNA-seq data with the Mouseac data and with human data (**Chapter 4**).
- Quality control, trimming, alignment and statistical analysis of the RRBS data (**Chapter 6**) was achieved in collaboration with Dorothea Seiler Vellame from the University of Exeter.

Abbreviations

5fC – 5-formylcytosine

5hmC – 5-hydroxymethylcytosine

5mC – 5-methylcytosine

AD – Alzheimer's disease

APP – Amyloid precursor protein

Aβ – Amyloid beta

bp – base pairs

CA (CA1, CA3) – *Cornu Ammonis*

cDNA – complementary deoxyribonucleic acid

CNS – central nervous system

CpG – Cytosine-guanine dinucleotide

CSF – Cerebrospinal fluid

DEG – Differentially expressed gene

DG – Dentate gyrus

DLB – Dementia with Lewy bodies

DMP – Differentially methylated position

DMR – Differentially methylated region

DNA – Deoxyribonucleic acid

DNMT – DNA methyltransferase

ECX – Entorhinal cortex

EDTA – Ethylenediamine tetraacetic acid

EOAD – Early onset Alzheimer's disease

ERCC – External RNA Controls Consortium

EWAS – Epigenome-wide association studies

FC – Fold-change

FDA – Familial Alzheimer’s disease

FDR – False discovery rate

FTD – Frontotemporal dementia

GO – Gene ontology

GWAS – Genome-wide association studies

HDCA – Histone deacetylase

HIP – Hippocampus

IHC – Immunohistochemistry

kb – kilobases

LOAD – Late onset Alzheimer’s disease

LUMA – LUminometric Methylation Assay

M1 – Primary motor cortex

M2 – Secondary motor cortex

MAPT – Microtubule associated protein tau, or tau

mRNA – Messenger ribonucleic acid

NFTs – Neurofibrillary tangles

PBS – Phosphate-buffered saline

PCR – Polymerase chain reaction

PD – Parkinson’s disease

PSEN – presenilin

QC – Quality control

qPCR – Quantitative polymerase chain reaction

QQ – Quantile-quantile

RIN – Ribonucleic acid integrity number

RNA – Ribonucleic acid

RNA-seq – RNA sequencing

RRBS – Reduced representation bisulfite sequencing

RT-PCR – Real-time polymerase chain reaction

SAD – Sporadic Alzheimer’s disease

SNP – Single nucleotide polymorphism

TET – Ten-eleven translocation

TG – Transgenic

TPM – Transcripts per million

UV – Ultraviolet

VaD – Vascular dementia

WGCNA – Weighted correlation network analysis

WT – wild type

Chapter 1. General Introduction

This chapter represents a modified version of a soon-to-be published book chapter that I wrote during my PhD (Castanho and Lunnon, 2019).

1.1. Alzheimer's disease

Alzheimer's disease (AD), first described by the German psychiatrist Alois Alzheimer in 1898 (Alzheimer, 1898, Alzheimer, 1907), is the most common type of dementia. AD is characterised by memory loss, cognitive decline, behavioural changes and, ultimately, death (Blennow et al., 2006), which are consequences of progressive neurodegeneration and synaptic loss (Crews and Masliah, 2010, Selkoe, 2011). AD contributes significantly to the global burden of disease, affecting more than 850,000 people in the UK (Prince et al., 2014) and about 26 million people worldwide (Prince et al., 2016). The high prevalence of AD is not only devastating for affected families, but also inflicts a high economic burden; the cost of caring for patients with AD was estimated to be £26.3 billion per year in 2014 in the United Kingdom alone (Prince et al., 2016, Prince et al., 2014). Although potentially promising new drugs are currently being tested in clinical trials, at present there are no available disease-modifying treatments, and existing medications only partially alleviate the symptoms of AD (Alzheimer's-Association, 2018).

People affected by AD display brain atrophy, restricted to cortical and limbic regions, accompanied by enlargement of the ventricles, reflecting differential neurodegeneration and presence of neuropathology; some regions of the brain are particularly affected by AD pathology, while other regions are relatively resistant (Blennow et al., 2006) (**Figure 1.1**). The hippocampus and the entorhinal cortex, for example, both involved in memory formation and recall, are particularly affected from the earliest stages of disease; in contrast, the cerebellum is relatively spared from pathology even in advanced AD (Serrano-Pozo et al., 2011). Other characteristic features of AD include the activation of astrocytes and microglia (De Strooper and Karran, 2016).

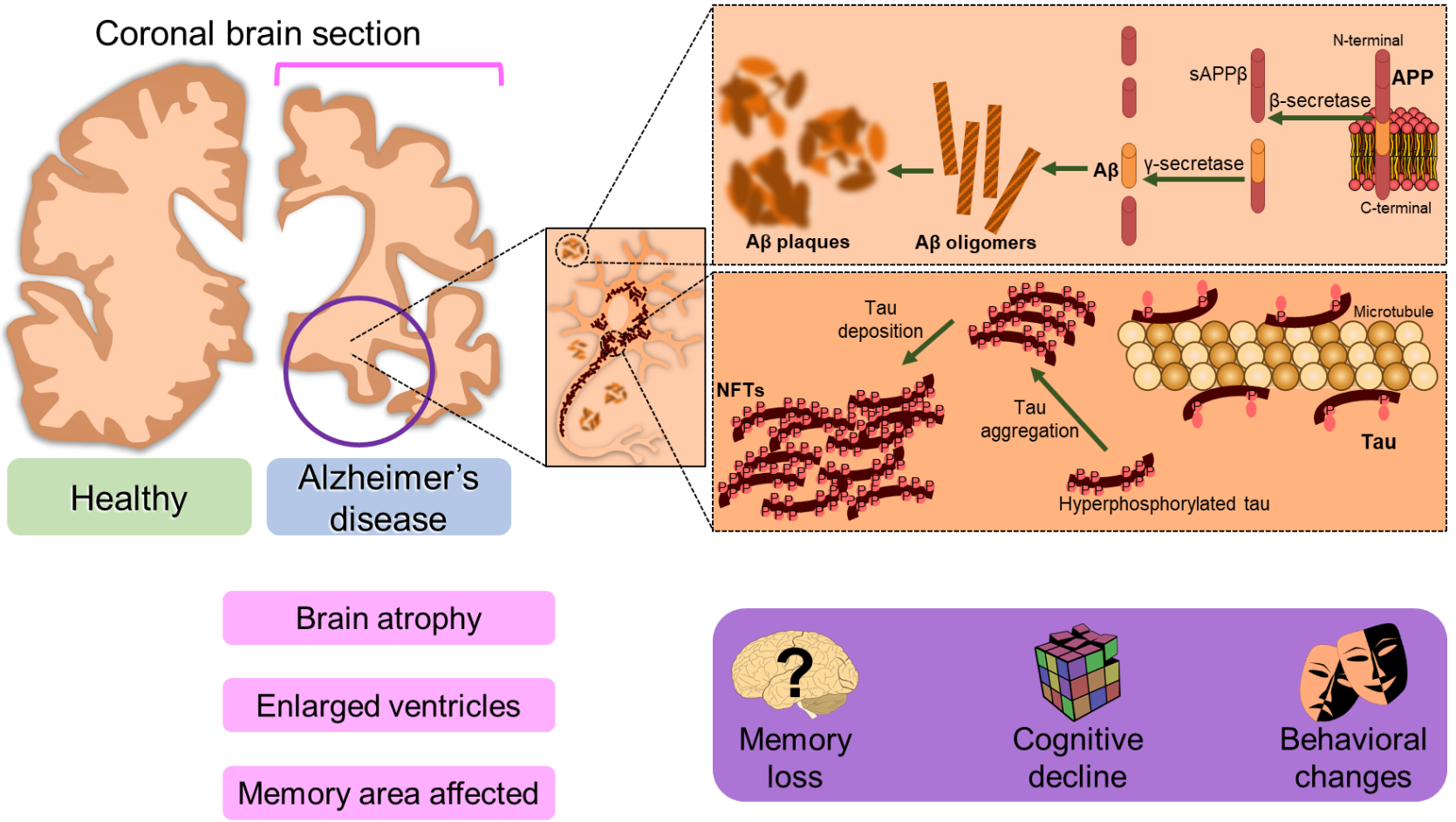


Figure 1.1 – Pathological hallmarks of Alzheimer's disease.
On a microscopic level, Alzheimer's disease (AD) is characterised by the presence of extracellular amyloid beta (Aβ) plaques and intracellular neurofibrillary tangles (NFTs). On a macroscopic level the brain of an individual with AD is characterised by extensive brain atrophy across limbic and cortical regions, and ventricular enlargement. Figure from Castanho and Lunnon (2019).

AD neuropathology is believed to start decades before the actual manifestation of clinical symptoms (Rajan et al., 2015), and is characterised by two key histopathological hallmarks: extracellular amyloid beta (Aβ) plaques, and intracellular neurofibrillary tangles (NFTs) (O'Brien and Wong, 2011, Wang and Mandelkow, 2016). Amyloid plaques are aggregates of Aβ peptides that result from the amyloidogenic cleavage of amyloid precursor protein (APP) (**Figure 1.1**), a transmembrane protein highly expressed in the brain and involved in neurite outgrowth and synapse formation (O'Brien and Wong, 2011). APP is a single-pass transmembrane protein, with a large extracellular domain, and is metabolised in a rapid and highly complex fashion by a series of sequential proteases. The sequential cleavage of APP occurs by two pathways: 1) nonamyloidogenic, and 2) amyloidogenic. The amyloidogenic processing of APP involves β-secretase (or BACE1), followed by the γ-secretase complex, and results in the production of Aβ (**Figure 1.1**). NFTs are deposits of hyperphosphorylated and aggregated microtubule associated protein tau (MAPT) (**Figure 1.1**), a crucial protein for microtubule assembly and stabilization with important functions in neurons (e.g. axonal growth and transport). Aggregation of hyperphosphorylated tau into NFTs characterises a wide range of neurodegenerative diseases known as tauopathies, including AD. The longest tau isoform (2N4R) contains 85 potential phosphorylation sites (80 Ser or Thr, and 5 Tyr); in AD and other tauopathies 17 Thr-Pro or Ser-Pro motifs are targeted by several signal-transducing proline-directed serine/threonine kinases, and are abnormally hyperphosphorylated, with the phosphorylation sites of tau clustering in the flanking regions. Other phosphorylation sites in or near the repeat domain can be phosphorylated by alternative kinases (e.g. microtubule affinity-regulating kinases (MARKs; also known as PAR1 kinases), cyclic AMP-dependent protein kinase (PKA) and Ca2+- or calmodulin-dependent protein kinase II (CaMKII)). Tau can also be phosphorylated by tyrosine kinases such as lymphocyte-specific protein tyrosine kinase (Lck), spleen tyrosine kinase (Syk) and tyrosine-protein kinase Fyn (Fyn) (SRC family members) at Tyr18, and by Abelson-related gene enzyme (Arg, also known as ABL2) and Abelson murine leukemia viral oncogene homolog 1 (ABL1) (ABL family members) at Tyr394 (Wang and Mandelkow, 2016).

As neuropathology precedes clinical symptoms, AD can only be clinically diagnosed relatively late in the disease process, with a definitive diagnosis only

being possible post-mortem (Blennow et al., 2006). According to the microscopic presence of the aforementioned neuropathological hallmarks, AD can be categorised into different stages. Considering the spread of NFTs, AD can be classified into six groups known as Braak stages (Braak et al., 2006, Braak and Braak, 1991): tau pathology starts in the transentorhinal and entorhinal regions (Braak I and II), extending progressively to the limbic system (e.g. hippocampus) and adjoining neocortex (Braak III and IV), before later spreading to other neocortical regions (Braak V and VI) (Jouanne et al., 2017, Murphy, 2019) (**Figure 1.2a**). Similarly, according to the progressive deposition of Aβ, AD can also be categorised into five groups known as Thal stages (Thal et al., 2002): β-amyloidosis spreads from the neocortex (Thal I), progresses into allocortical regions such as the entorhinal area, hippocampus, and striatum (Thal II and III), until it reaches subcortical structures, the brain stem and finally the cerebellum (Thal IV and V) (Jouanne et al., 2017, Thal et al., 2002) (**Figure 1.2b**).

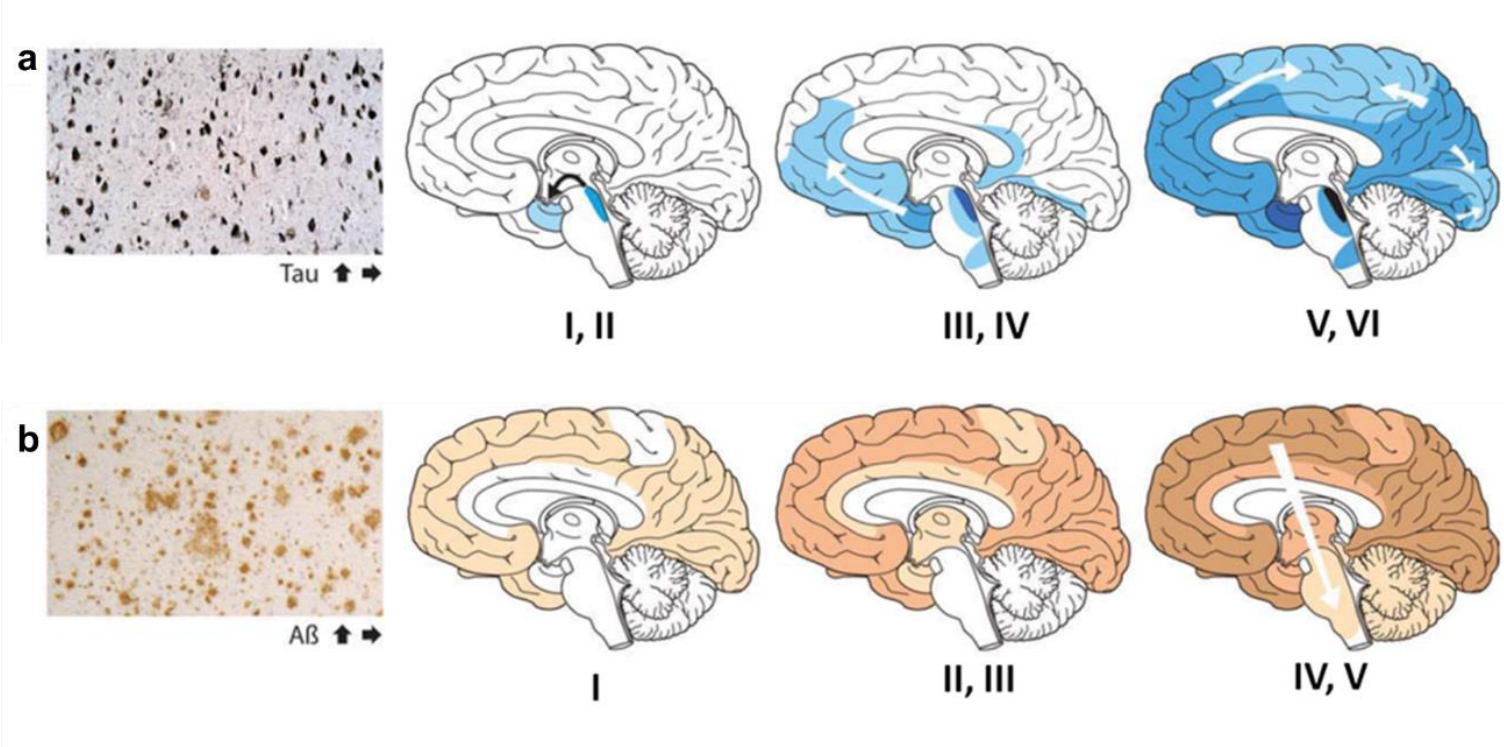


Figure 1.2 – Spatial-temporal progression of the pathological hallmarks of AD according to Braak and Thal.
(a) Tau neurofibrillary tangles (NFTs) progression. **(b)** Amyloid (Aβ) progression. Figure and legend adapted from Jouanne et al. (2017).

1.2. Alzheimer's disease: the genomic era

AD can be divided into two subtypes: familial AD (FAD) and sporadic AD (SAD). FAD is usually early-onset (EOAD), generally occurring in individuals <65 years, while SAD is mainly late-onset (LOAD), affecting people >65 years (**Figure 1.3**). FAD accounts for less than 5% of disease incidence and is caused by one or more autosomal dominant mutations in the *APP* gene, or in the genes that encode for presenilin-1 or -2 (*PSEN1* or *PSEN2*), which are components of the γ-secretase enzyme that is involved in the cleavage of APP (Guerreiro et al., 2012). Although the role of genetic variation in SAD is less well-defined compared to FAD, it has been shown to be highly heritable, exemplified by the larger monozygotic twin concordance rate for the disease (probandwise concordance = ~50%), compared to dizygotic twin concordance (probandwise concordance ~30%) (Gatz et al., 2006, Plomin et al., 1994).

To date, the most robust genetic risk factor identified for SAD is in the apolipoprotein E (*APOE*) gene, more specifically the presence of the ε4 allele (Corder et al., 1993, Strittmatter et al., 1993). However, many common, low penetrance genetic variants, and many rare but high penetrant genetic variants, have been shown to be involved in the disease, suggesting that SAD has a large polygenic component, where many variants influence disease. In addition to *APOE*, recent large genome-wide association studies (GWAS) and subsequent meta-analyses have identified numerous common single nucleotide polymorphisms (SNPs) that individually confer a relatively small risk for disease (Kunkle et al., 2019). Results from these studies have identified variants annotated to numerous genes, including clusterin (*CLU*), phosphatidylinositol binding clathrin assembly protein (*PICALM*), complement receptor 1 (*CR1*) (Carrasquillo et al., 2010, Corneveaux et al., 2010, Harold et al., 2009, Jun et al., 2010, Lambert et al., 2009, Zhang et al., 2010), bridging integrator 1 (*BIN1*), cluster of differentiation 2-associated protein (*CD2AP*), cluster of differentiation 33 (*CD33*), erythropoietin-producing hepatoma receptor A1 (*EPHA1*), and ATP-binding cassette subfamily A member 7 (*ABCA7*), amongst others (Hollingworth et al., 2011, Lambert et al., 2013). With advances in genomic sequencing technology, researchers have also identified a number of rarer variants conferring

a moderate increase in disease risk, including variants annotated to the genes for triggering receptor expressed on myeloid cells 2 protein (*TREM2*) (Guerreiro et al., 2013, Jonsson et al., 2013, Sims et al., 2017), and phospholipase D3 (*PLD3*) (Cruchaga et al., 2014). The variants identified from genetic studies suggest specific biological pathways are altered in AD, including synaptic pathways, immune regulation, cholesterol transport and lipid metabolism, endocytosis, ubiquitination, and protein folding (Smith et al., 2016a). Recently, three large studies (GWAS and meta-analyses) have provided further support for a role of the immune system and lipid metabolism in AD by confirming the association of previously identified loci with AD. More importantly, these reports collectively identified 15 novel loci, of which one, ADAM metallopeptidase domain 10 (*ADAM10*), was common to all three studies (Jansen et al., 2019, Kunkle et al., 2019, Marioni et al., 2018).

Of note, very few of the risk variants identified in GWAS of AD directly alter protein structure, with the majority residing in noncoding regions of the genome (Guerreiro et al., 2012); Kunkle et al., for example, reported only 2% of AD-associated variants in their study to be located in exons, with the majority (58%) being located in introns (Kunkle et al., 2019). AD risk variants are predicted to influence gene regulation rather than directly affect the coding sequences of transcribed proteins, emphasising the importance of studying transcriptional (**Section 1.3**) and epigenomic (**Section 1.4**) mechanisms in disease (**Figure 1.4**).

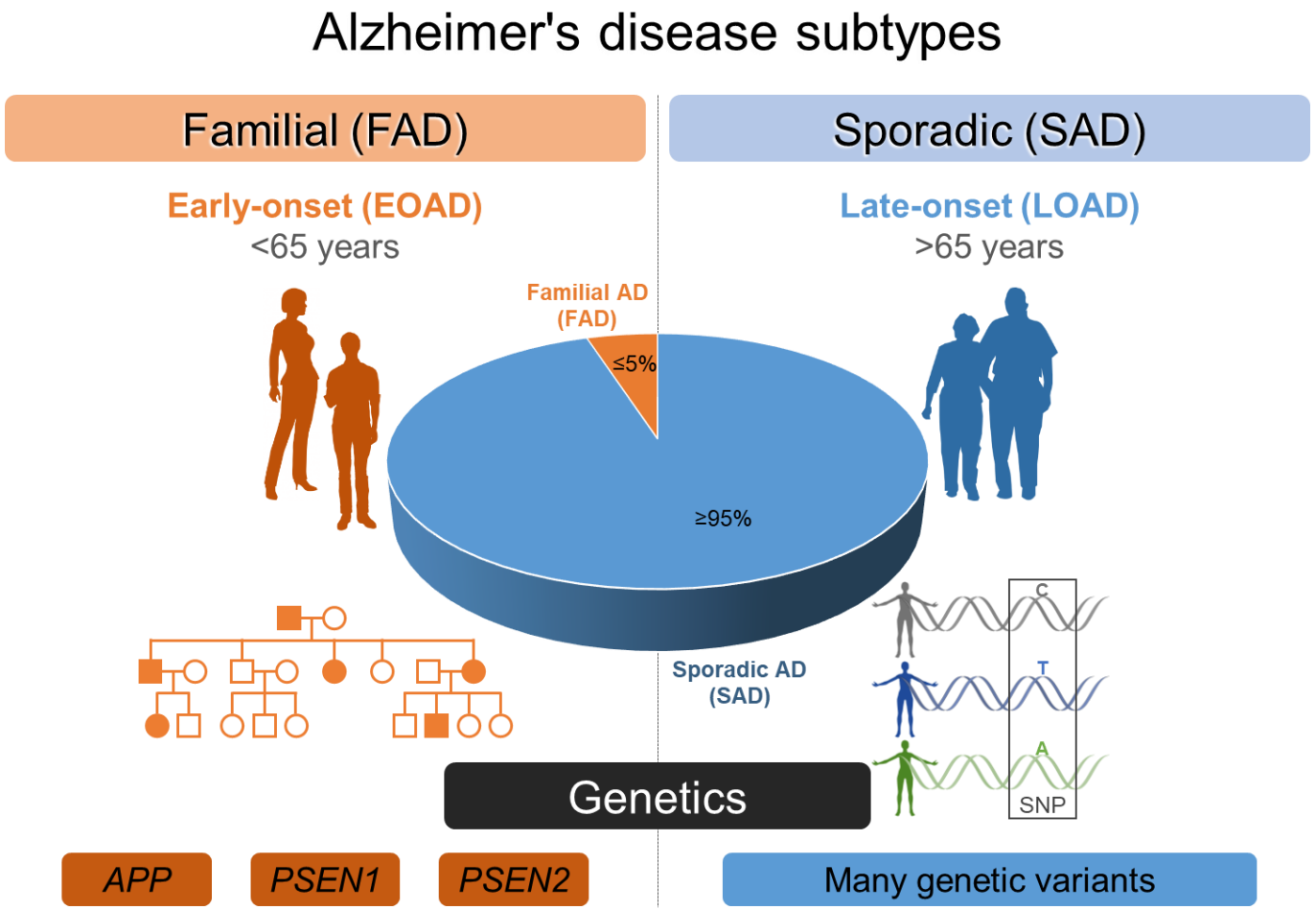


Figure 1.3 – AD subtypes and their genetics.
FAD accounts for <5% of AD cases and is caused by autosomal dominant mutations in the *APP*, *PSEN1* or *PSEN2* genes. SAD accounts for >95% of AD cases and involves the cumulative effect of many common genetic variants. Figure from Castanho and Lunnon (2019).

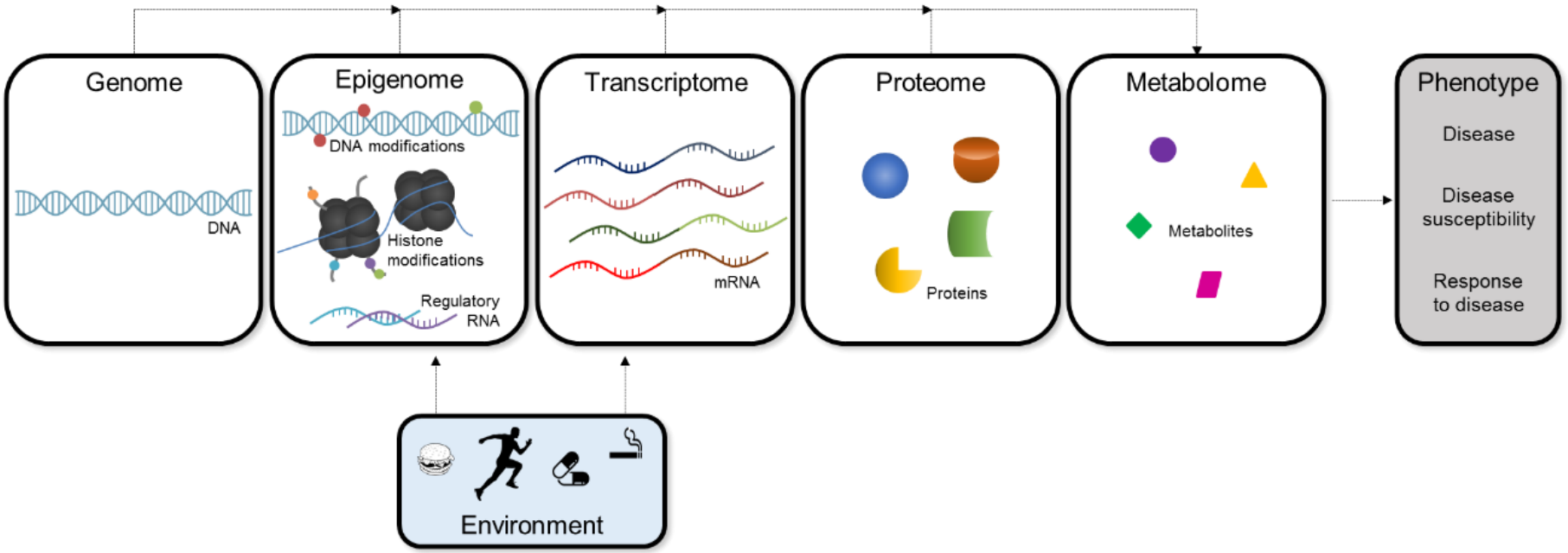


Figure 1.4 – Different levels of gene regulation.
From the set of genes present in an organism (genome) to the observable characteristics of an individual (phenotype), there are several levels of gene regulation, including the epigenome and transcriptome, which can also be influenced by factors in the environment. The epigenome corresponds to the set of chemical modifications and regulatory RNA molecules that can control gene activity. The transcriptome refers to all RNA transcripts expressed from the genes of an organism.

1.3. Evidence for differential gene expression in the Alzheimer's disease brain

The transcriptome refers to the complete set of RNA transcripts expressed in a specific type of cell or tissue, at a given time. A number of studies have quantified patterns of gene expression in brain tissue identifying messenger RNA (mRNA) changes associated with AD pathological processes. The majority of these genome-wide gene expression studies have used microarrays to investigate gene expression alterations in AD (Blalock et al., 2004, Bossers et al., 2010, Emilsson et al., 2006, Hokama et al., 2014, Horesh et al., 2011, Katsel et al., 2007, Miller et al., 2008, Miller et al., 2013, Tan et al., 2010, Xu et al., 2006, Youn et al., 2007, Zhang et al., 2013), and the results of these studies are summarised in **Table 1.1**. Of note, many of these studies are characterised by design limitations; they have often been performed on relatively small numbers of samples and there are a number of potential confounding effects in these analyses that make the results generally hard to interpret. More recently, the advent of highly-parallel RNA sequencing (RNA-seq) (**Table 1.2**) has enabled more systematic AD expression studies to be performed, profiling larger sample cohorts, and using more sophisticated bioinformatic approaches (Logsdon et al., 2019, Mostafavi et al., 2018, Raj et al., 2018). One issue is that these analyses generally assess 'bulk' brain tissue comprising a mix of different cell-types; a few studies have used laser capture microdissection (LCM) to isolate and profile neurons or astrocytes from AD brain tissue (Dunckley et al., 2006, Liang et al., 2008, Mastroeni et al., 2017, Sekar et al., 2015, Simpson et al., 2011) (**Table 1.3**).

Table 1.1 – Gene expression studies in AD using array-based technologies.

Reference	Tissue	Samples	Method	Key findings
Blalock et al. (2004)	Hippocampus (CA1)	22 AD (varying severity) versus 9 controls	Microarrays	Altered biological processes included: • Up-regulation of genes involved in proliferation and differentiation, including tumour suppressors, oligodendrocyte growth factors, and protein kinase A modulators (correlated with incipient AD); • Up-regulation of adhesion, apoptosis, lipid metabolism, and initial inflammation processes; • Down-regulation of protein folding/metabolism/transport and some energy metabolism and signalling pathways.
Xu et al. (2006)	Hippocampus (CA1–CA4)	16 AD versus 4 controls 5 AD *APOE* ε4/ε4 versus 5 AD *APOE* ε3/ε3	Microarrays	<u>AD versus controls</u> • Genes with increased expression were associated with negative regulation of cell proliferation, signal transduction, stress response, regulation of transcription, regulation of cell cycle, modulation of the cytoskeleton, immune response, and iron ion homeostasis (e.g. *NDUFC2*, *NDUFA1, COX6C*, *COX6B*, *COX7C*, *COX8, ATP5J2*, *ATP5L*, *ATP5H*, *ATP5A1*, *ATP5G2*, *ATP5B, MXI1*, *BS69*, *DUSP1*, *NSEP1, GNB1, EDG2, WASF3*, *SNX3, HSP90AA1*, *DNAJA1, DNAJB1, SOD3, PGF, GTF2I, RGC32, ACTR3, SPTBN1, MYO10*); • Genes with reduction in expression were involved in mitochondrial function, cytoskeletal structure, synaptic plasticity, neuronal cell adhesion, neurogenesis, and vesicle neurotransmission (*APOE*, *C3, TUBB2*, *TUBB4*, *TUBB5*, *TUBA1*, *TUBB*, *TUBA3*, *TUBA6, VAMP2*, *BIN1, NRGN, PLD3*). <u>AD *APOE* ε4/ε4 versus AD *APOE* ε3/ε3</u> • Genes with increased expression were associated with cell growth suppression, signal transduction, myelinogenesis, cell adhesion and migration, heavy metal metabolism and detoxification (*BIN1, RASSF2, FYN, ITPKB, S100A1, MYO10*); • Genes with reduction in expression were involved in mitochondrial metabolism, and synaptic processes (*ATP1A1, ATP5H, COX6B, LDHA, VAMP2*).

Emilsson et al. (2006)	Frontal cortex	61 AD versus 53 controls	Microarrays and RT-PCR	Three genes showing large and consistent expression differences: *ITPKB* (increased), *RAB3A* (decreased), and *RGS4* (decreased).
Younet al. (2007)	Hippocampus	19 AD versus 15 controls	Microarrays	Decreased expression of *ITPKB* and *KALRN;* increased expression of *RGS4.* No significant differences between AD and control specimens for cerebellum.
Katsel et al. (2007)	17 brain regions	19 controls, 34 mild AD, 16 moderate AD, 48 severe AD	Microarrays	Dysregulation of lipid metabolism, including expression changes of key enzymes involved in sphingolipid metabolism, particularly in temporal and frontal cortices.
Miller et al. (2008)	Hippocampus (CA1)	Data from Blalock et al., 2004	Co-expression network analysis	AD progression-associated gene modules related to mitochondrial metabolism ('hub' genes: *VDAC1*, *VDAC3*, *ATP5F1*), and synaptic processes ('hub' genes: *WDR7*, *SYNJ1*, *STXBP1*, *SNAP91*).
Tan et al. (2010)	Temporal cortex	12 AD (Braak III/IV-V/VI) versus 8 controls (Braak 0-III/IV).	Exon microarrays	*AQP1* as the transcript with the highest increase in expression by fold-change; other gene expression changes included reduction in expression of *RPH3A*, *NRN1*, *SV2A*, *SV2B*, *SYT11*, *SYT12*, *SYT13*, suggestive of synaptic dysfunction and perturbed neurotransmission.
Bossers et al. (2010)	Prefrontal cortex	7 AD individuals for each Braak stage (6 Braak stages) versus 7 controls	Microarrays	Gene expression changes associated with AD progression (Braak staging), with the most significant changes identified between Braak stages II and III: • Increased expression of genes involved in the exocytosis of neurotransmitters, and potassium voltage-gated channels (*SNAP25, CPLX1, VAMP7, SYT1*, *SYT3*, *SYT4, NAPB, SV2C, KCNS3*, *KCNB1*, *KCNA1* and *KCNAB1*); • Decreased expression of *SYT6* in Braak stage II. In late Braak stages: • Increased expression of *HSPA2*, *HSPB2, MT1B,* and *MT1G;* • Reduction of expression of *SST.*
Horesh et al. (2011)	Multiple brain regions	55 AD, 28 schizophrenia, 22 controls	Microarrays	Common transcriptomic changes in AD and schizophrenia, particularly in genes involved in regulation of autophagy (e.g. *ULK2, ATG3*) and inositol phosphate metabolism (e.g. *PIK3R4*).

Zhang et al. (2013)	Dorsolateral prefrontal cortex, visual cortex, and cerebellum	376 LOAD versus 173 controls (total of 1,647 samples)	Microarrays; co-expression network analysis	Constructed multi-tissue co-expression networks and compared gene connectivity, identifying an immune- and microglia-specific (with increased expression) as the top-ranked module: • *TYROBP* as key causal regulator; • Other genes exhibiting increased expression: *CTSC, HCK, SERPINA1, S100A11, LY86, DOCK2, FCER1G.*
Miller et al. (2013)	Hippocampus (CA1 and CA3)	17 LOAD (moderate and severe) versus 16 controls	Microarrays; co-expression network analysis	AD progression-associated genes: • initial enrichment of increase in expression of genes in CA1 (over-represented genes related to signal transduction, immune response, and cell motility); • initial enrichment of genes with decreased expression in CA3 (overrepresented genes relevant for synaptic transmission, cytoskeletal protein binding, and cholesterol biosynthesis). More dramatic differences in CA1 than CA3. CA1-only gene expression changes: • genes with increased expression related to cell death and cell proliferation (top genes: *S100A6*, *PFKFB3*, *SERPINA3*); • genes with reduction in expression involved in synaptic transmission and cell-cell signalling processes (top genes: *SEPT5*, *5CS*, *WFDC1*, *KCNIP1*). Co-expression network analysis identified a microglia-associated module (with increased expression in controls of Braak stage II).
Hokama et al. (2014)	Frontal cortex, temporal cortex, and hippocampus	26 AD (Braak I-VI) versus 62 non-AD	Microarrays	Expression changes in genes: • Involved in noninsulin-dependent diabetes mellitus and obesity; • Previously associated with AD and other psychiatric disorders, both in AD brains and hippocampal samples from an AD mouse model (see **Table 1.7**). The most significant changes identified in the hippocampus: • Reduction in expression of transcripts associated with neuronal dysfunction (*MET, PCSK1, PTPN3, SERPINF1, VEGFA, GABRA1*, *GABRA4*, *GABRA5*, *GABRG2, SYT4*, *SYT5*, *SYT7, STX1B, KCNJ6, RGS4*, *RGS7*); • Increased expression: *AEBP1, TXNIP, VCAM1, ANGPT1.*

Table 1.2 – Gene expression studies in AD using RNA-seq.

Reference	Tissue	Samples	Method	Key findings
Magistri et al. (2015)	Hippocampus	4 LOAD (Braak V-VI) versus 4 controls (Braak I-II)	RNA-seq	Reduction in expression of *TAC1*; increase in expression of *SERPINE1.* Pathway analysis: dysregulation in neural communication, cerebral vasculature, and Aβ clearance.
Raj et al. (2018)	Dorsolateral prefrontal cortex	268 AD,182 controls (total of 450 individuals from 2 aging cohorts)	RNA-seq	Expression changes identified in: • Novel transcripts (e.g. *AP2A1, AP2A2,* FUS, *MAP1B, TBC1D7*); • Genes previously linked to AD (e.g. *SPI1, PTK2B, CLU, MTCH2, PICALM*).
Mostafavi et al. (2018)	Dorsolateral prefrontal cortex	478 individuals from 2 longitudinal cohorts (280 with pathological AD; 185 with clinical AD)	Co-expression network analysis	Identified modules of co-expressed genes associated with both cognitive decline and AD neuropathology (amyloid pathology). The module (increased expression) most strongly associated was enriched for genes involved in regulation of the cell cycle and chromatin modification (e.g. *PLXNB1* and *INPPL1*).
Logsdon et al. (2018)	Multiple brain regions	2114 samples across 7 brain regions and 3 research studies	Various co-expression methods	Five consensus clusters consistent across studies, methods, and samples. AD genetic associations, previously studied AD-related biological processes: • AD targets under active investigation were enriched in only three of these five clusters; • The remaining two clusters demonstrated strong heterogeneity between males and females.

Table 1.3 – Gene expression studies in AD at the single cell level.

Reference	Tissue	Samples	Method	Key findings
Dunckley et al. (2006)	Neurons from entorhinal cortex	19 AD, 14 controls AD neurons with NFT versus adjacent non-NFT-bearing neurons (from 19 AD)	Microarrays	225 differentially expressed genes exhibiting progressively increased or decreased expression from AD NFT neurons, AD non-NFT neurons, and controls. NFT-associated elevated gene expression levels included *APOJ*, *TIMP3*, and *IRAK1*.
Liang et al. (2008)	Normal (non-NFT) cortical neurons from multiple brain regions	34 AD, 14 controls	Microarrays	Regional differential expression in AD compared to controls, including in genes previously implicated in AD: • Increased expression of *ITPKB* in neurons from the entorhinal cortex and reduction in expression of *GSK3B* in hippocampal neurons; • Decreased expression of *MAPT* in neurons from the entorhinal cortex, hippocampus, middle temporal gyrus, and posterior cingulate cortex.
Simpson et al. (2011)	Astrocytes from temporal cortex	6 Braak 0–II, 6 Braak III–IV, 6 Braak V–VI	Microarrays	Altered biological pathways in early versus late stages of AD: • Actin cytoskeleton, proliferation, apoptosis, and ubiquitin-mediated proteolysis dysfunction at Braak stages I–II; • Dysregulation of intracellular signalling pathways such PI3K/MAPK pathways associated with Braak stages V-VI; in individuals with APOE ε4 these pathways were already altered in Braak I–II.
Sekar et al. (2015)	Astrocytes from posterior cingulate cortex	10 AD, 10 controls	RNA-seq; co-expression network analysis	Changes in the expression of mitochondria-related genes (e.g. *TRMT61B*, *FASTKD2*, and *NDUFA4L2*) and immune response genes (e.g. *CLU*, *C3*, and *CD74*).
Mastroeni et al. (2017)	Microglia, astrocytes and neurons from hippocampus (CA1)	6 AD, 6 controls	RNA-seq	Increased expression of *ANK1* in AD microglia, but not in neurons or astrocytes from the same individuals.

Taken together, the transcriptomic studies of AD performed to date nominate the dysregulation of genes involved in various pathways including synaptic and neural function, calcium signalling, neuroinflammation, lipid and cholesterol metabolism, and mitochondrial dysfunction. One issue with these studies is that they have largely focused on identifying differential gene expression but have largely ignored looking at different transcript variants, isoforms and alternative splicing; these analyses require ultra-deep sequencing and are unfeasible for large studies. In the future, it will be important to examine whether specific transcript variants are altered in AD, particularly since changes at the level of total gene expression do not always result into changes at the protein level. Some research has been conducted to assess this (Malik et al., 2013, Raj et al., 2018, Raj et al., 2014, Tollervey et al., 2011), including reported AD-associated splicing alterations in several genes, including CD33 (Malik et al., 2013, Raj et al., 2018, Raj et al., 2014), *PICALM*, *CLU*, and *PTK2B* (Raj et al., 2018, Raj et al., 2014).

1.4. Alzheimer's disease from an epigenetic perspective

Despite the increased understanding of the role of genetic variation in AD, recent studies have highlighted that common variants only explain about 30% of disease incidence (Ridge et al., 2013), with much of the remaining heritability for disease still being unclear. A number of features of AD are suggestive of an epigenetic contribution to disease aetiology, for example monozygotic twin discordance (Mastroeni et al., 2009, Plomin et al., 1994) for both the development of disease (Gatz et al., 2006) as well as age of onset (Nee and Lippa, 1999). Aging, considered the greatest risk factor for AD, has been shown to be associated with significant epigenetic changes (Pal and Tyler, 2016, Sen et al., 2016), including in the *APP* and *MAPT* genes (Tohgi et al., 1999a, Tohgi et al., 1999b). Furthermore, many of the non-genetic risk factors identified for AD are related to lifestyle and/or the environment, including low education levels, hypertension, obesity, smoking, depression and/or stress, physical inactivity, social isolation, and diabetes (Alzheimer's-Association, 2018, Prince et al., 2016); of note, these exposures may influence gene regulation via so-called 'epigenetic' alterations (**Figure 1.4**). Taken together, it has been hypothesised that a combination of both genetic and epigenetic factors is likely involved in AD.

Epigenetic processes that control gene regulation can be divided into three major categories: 1) DNA modifications, 2) histone modifications, and 3) RNA-based mechanisms (**Figure 1.5**), which together comprise the 'epigenome'. Epigenetic mechanisms regulate gene activation/repression in a dynamic and reversible fashion, independently of genetic variation, and have been widely hypothesised to play a role in AD (Chouliaras et al., 2010, Lunnon and Mill, 2013, Mill, 2011). Because epigenetic changes are potentially reversible, targeting disease-associated epigenetic changes makes them potentially attractive therapeutic options for the treatment of AD. Although, previous and ongoing studies in AD have investigated many types of epigenetic modifications in AD, most of the research has looked at DNA modifications (**Section 1.4.1**), which are one focus of my PhD research.

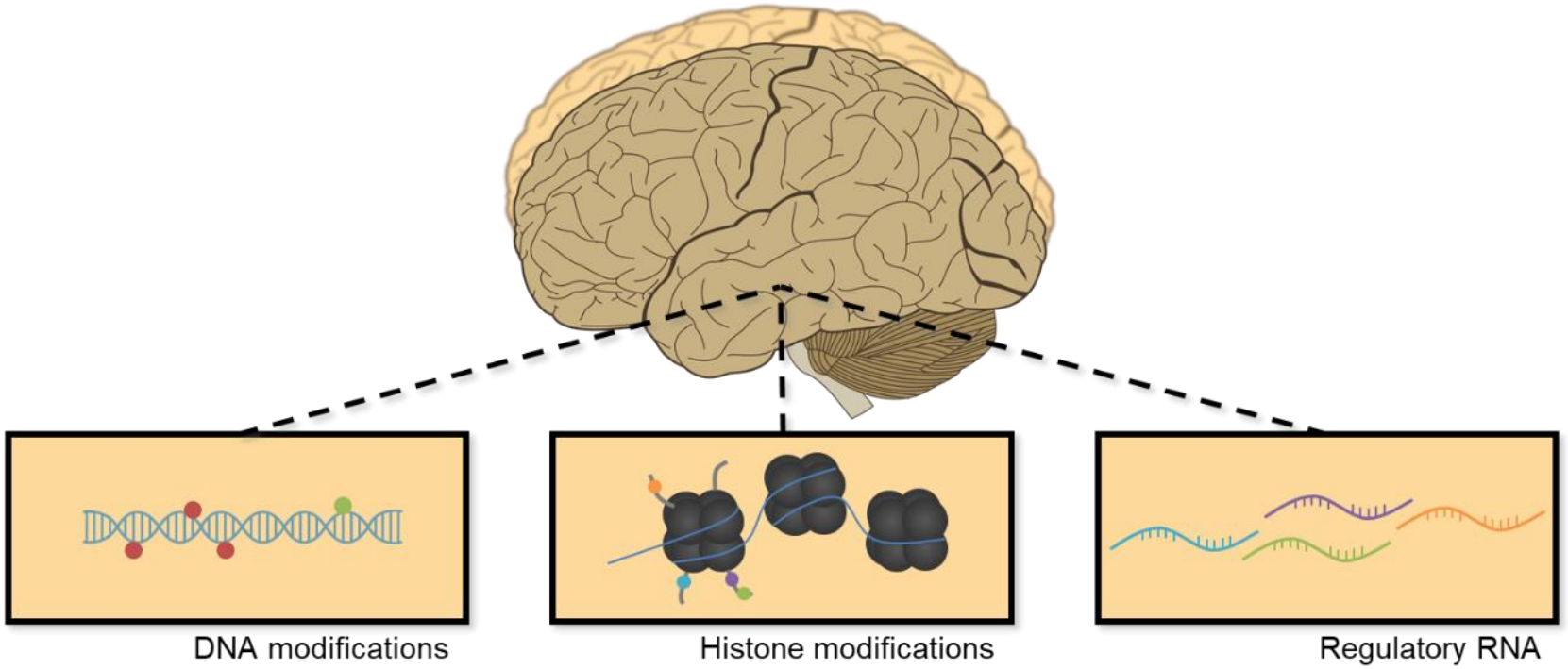


Figure 1.5 – Epigenetic processes hypothesised to play a role in AD.
There are three broad categories of epigenetic processes influencing gene regulation: DNA modifications, histone modifications and RNA-based mechanisms. Figure from Castanho and Lunnon (2019).

1.4.1. DNA modifications

Methylation of DNA at the 5th carbon of the cytosine nucleotide is the most-studied epigenetic modification, mainly due to its relative stability as an epigenetic modification and because numerous robust methods have been developed for its quantification across the genome. Accordingly, it has been the primary focus of most AD-related epigenetic studies to date (Smith and Lunnon, 2017). DNA methylation is established and maintained by DNA-methyltransferases (DNMTs) resulting in the formation of 5-methylcytosine (5mC). It can be demethylated by oxidation into 5-hydroxymethylcytosine (5hmC), another DNA modification, via the action of the ten-eleven translocation (TET) family of enzymes (Yong et al., 2016).

1.4.1.1. Global DNA modification changes in AD

Several early epigenetic studies of AD used immunoreactive approaches to evaluate 'global' variation in DNA methylation, with varying results. Some studies identified disease-associated hypomethylation, other studies reported hypermethylation, and others found no changes in AD samples (Bradley-Whitman and Lovell, 2013, Chouliaras et al., 2013, Condliffe et al., 2014, Coppieters et al., 2014, Lashley et al., 2015, Mastroeni et al., 2010, Mastroeni et al., 2009) (**Table 1.4**). These discrepancies potentially reflect the fact that different studies explored distinct brain regions and/or different subtypes of AD, with a very limited number of samples being evaluated. It is important to note that these immunoassay approaches can only evaluate DNA modifications at a global level; as such, it provides they no insight into the exact biological pathways or specific loci that are epigenetically altered in disease. Furthermore, immunostaining-based methods are not very sensitive for the detection of DNA methylation changes; a big change at just one locus would not be detected using this approach.

Table 1.4 – Global DNA modification studies in AD using immunostaining.

Reference	Brain region(s)	Samples	Key findings
Mastroeni et al. (2009)	Temporal neocortex	1 pair of monozygotic twins, discordant for AD	Hypomethylation (5mC) in the twin with AD.
Mastroeni et al. (2010)	Entorhinal cortex	20 AD, 20 controls	Hypomethylation (5mC) in AD compared to cognitively and neurologically healthy controls.
Chouliaras et al. (2013)	Hippocampus	10 AD, 10 controls	Hypomethylation (5mC and 5hmC) in AD compared to controls.
Bradley-Whitman and Lovell (2013)	Hippocampus	7 late-stage AD, 5 preclinical AD, 5 controls	Hypermethylation (5mC and 5hmC) in late-stage AD and preclinical AD compared to age-matched controls.
Coppieters et al. (2014)	Middle frontal gyrus and middle temporal gyrus	29 AD, 29 controls	Hypermethylation (5mC and 5hmC) in AD compared controls.
(Condliffe et al. (2014)	Entorhinal cortex, cerebellum	13 AD, 8 controls	Decreased 5hmC levels in AD compared to controls in both brain regions; however, no differences in 5mC detected.
Lashley et al. (2015)	Entorhinal cortex	12 AD, 14 controls	No differences in 5mC and 5hmC.

1.4.1.2. Methylation changes in AD candidate genes

A number of studies have investigated DNA methylation across regulatory regions associated with specific AD candidate genes. The first published study assessing epigenetic dysregulation in the context of AD used an enzymatic digestion followed by southern-blotting to evaluate DNA methylation at the *APP* gene (West et al., 1995). In this study, West *et al.* compared an AD patient to a non-demented subject, reporting absent DNA methylation at a site in the *APP* gene in the temporal lobe of the AD individual. A more recent candidate-based study used bisulfite-pyrosequencing to measure aberrant CpG DNA methylation in the *APP*, *MAPT* and glycogen synthase kinase 3 beta (*GSK3B*) genes in post-mortem AD brain samples (Iwata et al., 2014). Interestingly, in their report, Iwata *et al.* suggest that the hypermethylation and hypomethylation changes that they identified in *APP* and *MAPT*, respectively, were associated with increased expression of both genes. In another recent publication, gene specific DNA methylation in the vicinity of *MAPT* was evaluated by pyrosequencing in FAD brains from patients bearing a mutation in *PSEN1* (which results in higher production and accumulation of Aβ), revealing no changes in DNA methylation in the visual cortex, but a significant decrease in methylation levels in the cerebellum of AD patients compared to age-matched controls (Coupland et al., 2015). One study used real-time polymerase chain reaction (RT-PCR) to measure gene-specific changes in DNA methylation in post-mortem frontal cortex of AD patients when compared to age-matched controls (Rao et al., 2012). The authors assessed the methylation status of the promoter regions for a number of genes, reporting hypomethylation of cyclooxygenase-2 (*PTGS2*) and nuclear transcription factor kappa B (*NF-kB*), and hypermethylation of brain-derived neurotrophic factor (*BDNF*), transcription factor cAMP response element-binding protein (*CREB*) and synaptic protein synaptophysin (*SYP*) in AD.

DNA methylation changes in transposable elements (repetitive elements) such as Alu sequences and long interspersed nuclear elements (LINEs) have also been suggested to be altered in AD (Bollati et al., 2011). However, studies exploring this relationship are very limited and further investigations are necessary, particularly in brain tissue.

1.4.1.3. Genome-wide DNA methylation studies in AD

The first study to take a more systematic genome-wide approach used the Illumina 27K methylation array to assess DNA methylation at >27,000 CpG sites in the prefrontal cortex of LOAD patients compared to control individuals (Bakulski et al., 2012). The promoter region of transmembrane protein 59 (*TMEM59*) was the CpG site most strongly associated with AD status, and gene ontology analysis suggested hypermethylation of genes related to transcription and DNA replication, with hypomethylation in regions annotated to membrane transporters. The authors also reported hypomethylation in one site of the *PSEN1* gene in AD cases compared to controls.

The successor to the 27K array, the Illumina 450K methylation array, has more recently been used to assess DNA methylation at >485,000 CpG sites. Two independent epigenome-wide association studies (EWAS) were published together in 2014 and used this method to compare DNA methylation profiles in post-mortem brain samples from AD cases and low pathology controls (De Jager et al., 2014, Lunnon et al., 2014). Both studies highlighted neuropathology-associated differential methylation in ankyrin 1 (*ANK1*), ribosomal protein L1 (*RPL13*), rhomboid 5 homolog 2 (*RHBDF2*) and cadherin-related family member 23 (*CDH23*) (Lord and Cruchaga, 2014). Disease-associated *ANK1* hypermethylation was observed in cortical samples from four independent study cohorts, and interestingly brain regions most affected in the disease (e.g. the entorhinal cortex, superior temporal gyrus and frontal cortex) showed robust differences, whilst the cerebellum and pre-mortem blood showed no alterations in *ANK1* methylation (Lunnon et al., 2014). The study by de Jager and colleagues also highlighted two additional genes, *ABCA7* and *BIN1* that overlapped with previous genetic studies (De Jager et al., 2014). A follow up study assessing levels of DNA methylation in known AD risk genes within this data-set highlighted neuropathology-associated differential DNA methylation in sortilin-related receptor 1 (*SORL1*), *ABCA7*, major histocompatibility complex class II DR beta 5 (*HLA-DRB5*), solute carrier family 24 member 4 (*SLC24A4*) and *BIN1* (Yu et al., 2015). Another follow-up study, using the Lunnon *et al.* (2014) data-set, highlighted AD-associated hypermethylation at a CpG site within the *TREM2* gene in the superior temporal gyrus, which was validated in two other cohorts (Smith et al., 2016b). More recently, other AD EWAS have been published;

Watson *et al.* (2016) identified 479 AD-associated differential methylated regions (DMRs) in the superior temporal gyrus, the majority of which were hypermethylated in AD cases compared to non-demented controls and enriched for brain-specific histone signatures and for binding motifs of transcription factors, finding a significant correlation between their data-set and the Lunnon *et al.* (2014) data-set (Watson et al., 2016), whilst Smith *et al.* (2018) highlighted AD-associated hypermethylation within the homeobox A (*HOXA*) gene cluster in the frontal cortex and superior temporal gyrus, which spanned 48kb (Smith et al., 2018). A recent study found that *ANK1* DNA methylation changes in the entorhinal cortex are specific to certain neurodegenerative diseases, with hypermethylation observed in AD, Huntington's disease (HD) and to a lesser extent Parkinson's disease (PD). Interestingly, in individuals with vascular dementia (VaD) or dementia with Lewy bodies (DLB), disease-associated hypermethylation was only observed in individuals with co-existing AD pathology (Smith et al., 2019). Another recent study using fluorescence-activated cell sorting (FACS) investigated AD-associated DNA methylation changes in sorted neuronal and non-neuronal (glial) nuclei from post-mortem human occipital cortex samples, in addition to "bulk" tissue from the frontal and temporal cortex (Gasparoni et al., 2018). Gasparoni *et al.* provided evidence that sorting cells improves the detection of AD-associated DNA methylation changes and reported cell-type-specific DNA methylation signatures associated with Braak stage, including loci annotated to MCF.2 cell line derived transforming sequence like (*MCF2L*), *ANK1*, microtubule associated protein 2 (*MAP2*), volume-regulated anion channel subunit LRRC8B (*LRRC8B*), serine/threonine-protein kinase 32C (*STK32C*) and *S100B*. A meta-analysis combining the data of significant cell sorted differentially methylated regions from neuronal and glial populations identified Braak stage-associated changes in known AD genes including *APP*, *HOXA3* and *ADAM17* (Gasparoni et al., 2018). Notably, the study also validated DNA methylation changes previously identified in AD EWAS in genes such as *HOXA3* and *ANK1*, assigning them to neurons and glia, respectively.

1.4.1.4. DNA hydroxymethylation

Recent studies have also started to profile site-specific variation in 5hmC in AD; of note, this modification has been shown to be enriched in the brain, particularly in synaptic genes (Khare et al., 2012), is found at different levels across different brain regions (Lunnon et al., 2016), and is altered during brain development (Spiers et al., 2017). To date, only one study has quantified DNA hydroxymethylation at the genome-wide level in AD brains (Zhao et al., 2017). This study used a biotin tag capturing method followed by high-throughput sequencing and reported both hyperhydroxymethylation and hypohydroxymethylation changes in the dorsolateral prefrontal cortex from patients with a high burden of AD pathology compared to individuals with low AD neuropathology. The study nominated ~300 loci at which variable 5hmC was associated with high amyloid burden, 50 loci that were associated with high tau neuropathology, and four genes that were associated with both amyloid and tau pathology: 4-aminobutyrate aminotransferase (*ABAT*), calcium/calmodulin dependent protein kinase (*CAMK1D*), high-temperature requirement factor A3 (*HTRA3*), and leucine rich repeat neuronal 1 (*LRRN1*). Using network analyses to explore co-hydroxymethylation patterns, the authors identified dysregulation of modules containing genes involved in ion channel activity, neuron differentiation and development, neurotransmitter transport, glutamatergic synapse, and calcium ion-dependent exocytosis of neurotransmitter. Of note, the authors highlight that the low resolution of the sequencing used in their study means they might not have been able to fully discriminate 5hmC from 5mC (Zhao et al., 2017). As such, additional studies using other methodological approaches are hotly awaited by the field.

1.5. Modelling Alzheimer's disease: mouse models as powerful tools

Despite highlighting pathways potentially involved in the pathogenesis of AD, molecular epidemiological studies using human post-mortem tissue are potentially confounded by many secondary factors associated with the disease itself (including environmental exposures, medication effects, diet, potential stressors, and reverse causation); it may therefore be difficult to conclude much about underlying causal processes in disease pathology (Mill and Heijmans, 2013).

The use of mouse models over the last few decades has enabled researchers to better comprehend diseases at a molecular level, identifying important pathways affected by and/or involved in pathology, establishing and validating biomarkers, and identifying novel therapeutic targets (Gotz and Ittner, 2008, Sasaguri et al., 2017, Gotz et al., 2018). Other key advantages of using mouse models include the tight control of potentially confounding factors (e.g. differential exposure to environmental factors, reverse causation, and genetic differences), easy access to specific regions of the brain, and the ability to investigate changes at multiple stages / across development (e.g. through the development of brain pathology).

Age-related disorders, such as AD, for which disease-associated changes can take decades to manifest in humans, rarely occur spontaneously in animals used for research, mainly because their lifespan may not be long enough to observe them. Therefore, the most used animal models are genetically engineered mice that recapitulate many aspects of human diseases. Animal models in general, and mouse models in particular, represent valuable tools for probing the pathogenic mechanisms involved in a wide range of diseases, but especially those that affect inaccessible tissues such as the brain.

1.5.1. Currently available AD mouse models

A diverse range of mouse models have been used to study aspects of AD pathology. Most currently used mouse models (**Figure 1.6**) overexpress human *APP*, either alone (e.g. Tg2576, J20) or in conjunction with presenilin (*PSEN*) overexpression (e.g. APP/PSEN1, 5xFAD), bearing mutations identified in FAD (**Table 1.5**). These mice show some of the characteristics of AD, particularly the formation of Aβ plaques.

Table 1.5 – Examples of AD mouse models overexpressing human APP.

Mouse model	Mutation(s)	Reference(s)
Tg2576	*APP* K670N/M671L	Hsiao et al. (1996)
J20	*APP* KM670/671NL, *APP* V717F	Mucke et al. (2000)
APP/PSEN1	*APP* K670N/M671L, *PSEN1* M146V	Borchelt et al. (1996) Borchelt et al. (1997)
5xFAD	*APP* KM670/671NL, *APP* I716V, *APP* V717I, *PSEN1* M146L, *PSEN1* L286V	Oakley et al. (2006)

A number of tau mouse models are also available (**Figure 1.6**), overexpressing human *MAPT* (e.g. hTau.P301S, rTg4510) and bearing mutations identified in people with familial frontotemporal dementia (FTD) (**Table 1.6**). These mice recapitulate NFT neuropathology.

Table 1.6 – Examples of AD mouse models overexpressing human tau.

Mouse model	Mutation(s)	Reference(s)
hTau.P301S	MAPT P301S	Allen et al. (2002)
rTg4510	*MAPT* P301L	Santacruz et al. (2005) (Ramsden et al., 2005)

Finally, mice that develop plaques and tangles due to overexpressing mutated *APP*, *PSEN* and *MAPT* (e.g. the 3xTg-AD mouse model with *APP* KM670/671NL, *MAPT* P301L, and *PSEN1* M146V) (Oddo et al., 2003) are also widely used as well (**Figure 1.6**).

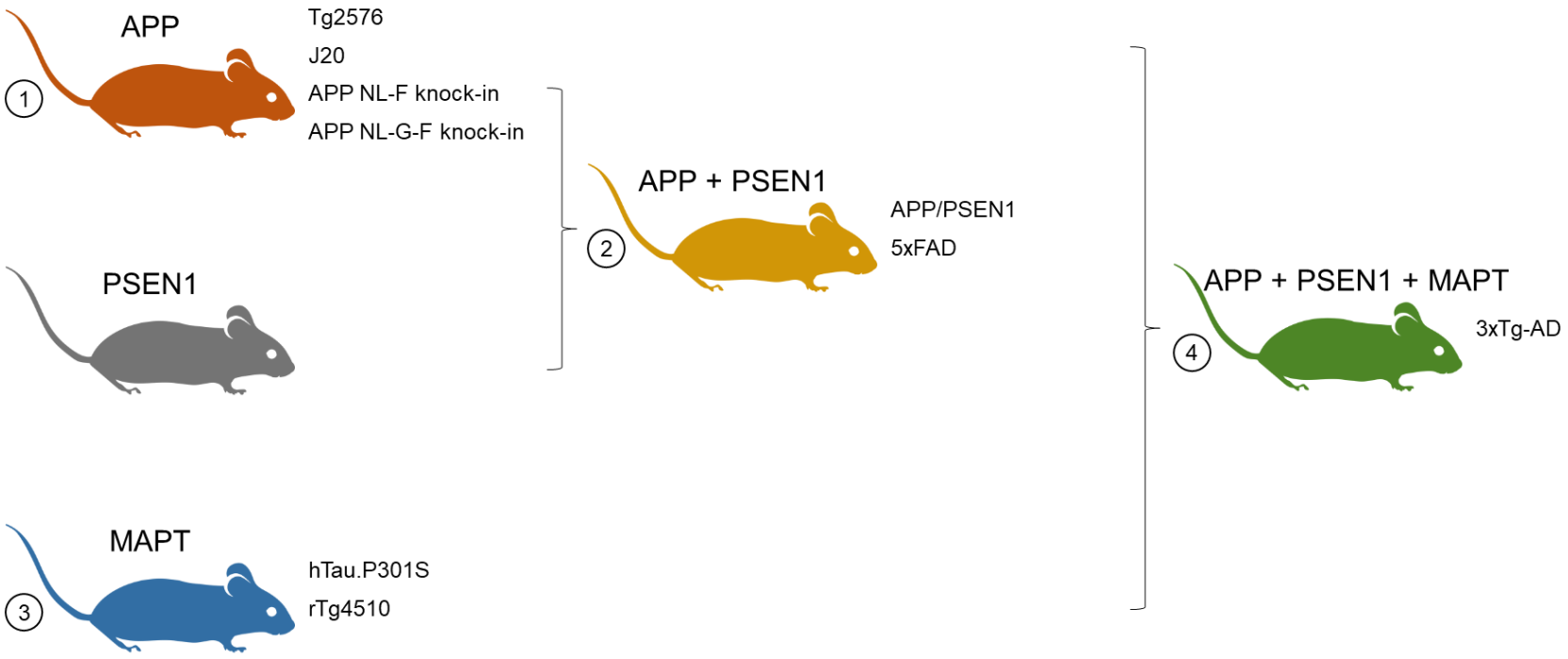


Figure 1.6 – Most common AD mouse models.
Most widely used AD mouse models either express 1) mutant human APP alone (examples: Tg2576, J20, APP knock-in mice) or 2) in conjunction with mutated PSEN (examples: APP/PSEN1, 5xFAD); 3) mutated human MAPT that codes for tau (examples: hTau.P301S, rTg4510); and 4) mutated APP, PSEN and MAPT (example: 3xTg-AD).

The insertion of the transgenes in these models has been shown to disrupt several endogenous mouse genes (Goodwin et al., 2017, Tosh et al., 2017). More recently, APP knock-in lines, such as the $App^{NL\text{-}G\text{-}F/NL\text{-}G\text{-}F}$ (*APP* KM670/671NL, *APP* I716F, APP E693G) and $App^{NL\text{-}F/NL\text{-}F}$ (*APP* KM670/671NL, *APP* I716F) mouse models (Saito et al., 2014, Sasaguri et al., 2017), have been developed, which are characterised by Aβ accumulation without some of the additional effects observed in first-generation APP over-expression models.

It is important to note that the validity and credibility of different mouse models of AD sparks robust debate. Although no model fully recapitulates the pathology and symptoms of human AD and their pathology is driven by the introduction of human mutations identified in familial forms of AD or FTD, there are specific advantages and disadvantages to the various genetic models that are currently available. Despite the inherent limitations in many of the models, studying mouse models can still provide a means to understanding processes mediating the early changes and progression of AD pathology. Furthermore, the use of laboratory animals is highly regulated, including the control of their environment and the general conditions they are exposed to, allowing the exclusion of potential confounders common in human research.

In my research for this thesis, I have used the rTg4510 and J20 mouse models, to specifically profile progressive changes in gene regulation associated with the development of either tau (rTg4510) or amyloid (J20) pathology. More details about these two models will be given in the next sections (**Section 1.5.1.1** and **Section 1.5.1.2**).

1.5.1.1. Modelling tau neuropathology: the rTg4510 mouse model

rTg4510 mice (Ramsden et al., 2005, Santacruz et al., 2005) overexpress a human mutant (P301L) form of tau identified in familial FTD, under the calcium-calmodulin-dependent kinase II (*Camk2a*) promoter, with transgene expression being largely restricted to the forebrain by the *Camk2a* promoter. Of note, the transgene also contains exons 2-3 of the mouse prion protein gene (*Prnp*) untranslated sequence. These mice develop age-dependent tauopathy, neurofibrillary tangle formation and associated neuron loss in the forebrain, as well as cognitive, behavioural and electrophysiological deficits (Blackmore et al., 2017, Booth et al., 2016, Ramsden et al., 2005, Santacruz et al., 2005) (**Figure 1.7**). Sex differences in the manifestation of these phenotypes have been reported, with females being affected earlier and more aggressively than males (Blackmore et al., 2017, Yue et al., 2011).

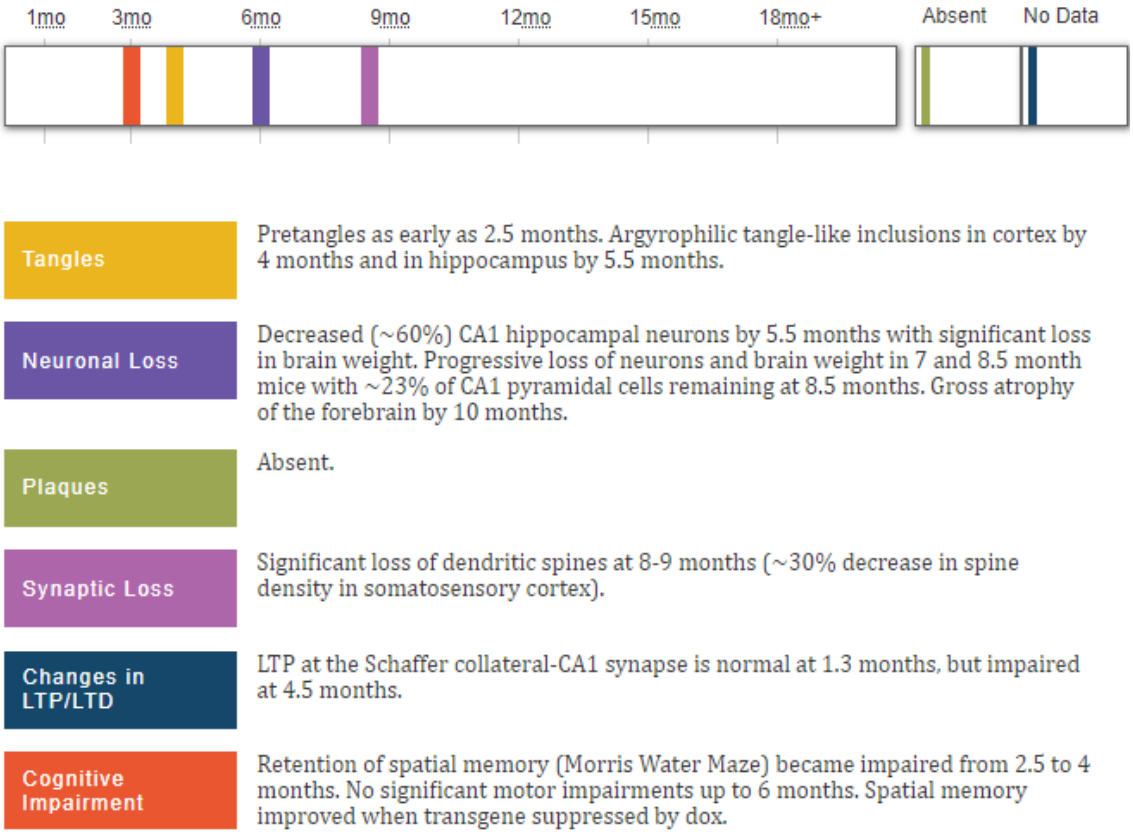


Figure 1.7 – Phenotype characterisation of rTg4510 mice.
Figure taken from Alzforum (www.alzforum.org; accessed 7th March 2019). LTP – long-term potentiation; LTD – long-term depression.

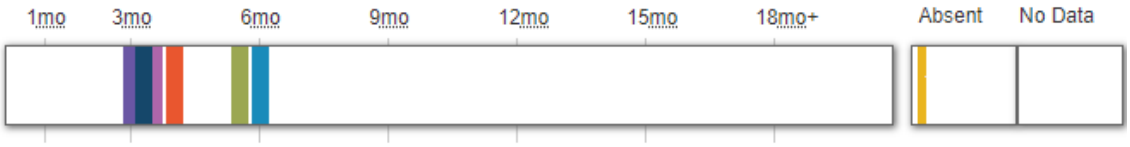
rTg4510 mice were originally described as developing pretangles around 2.5 months of age, with neurofibrillary tangle pathology starting in the neocortex and progressing rapidly into the hippocampus and limbic structures with increasing age (Ramsden et al., 2005, Santacruz et al., 2005), a profile which has been replicated in more recent studies (Sahara et al., 2014). The spread of neuropathology in this mouse model recapitulates the Braak stages in human AD brains, classified according to the spread of NFTs (Braak et al., 2006, Braak and Braak, 1991) (see **Figure 1.2**).

The “r” in rTg4510 refers to “regulatable”, as these mice result from crossing the 4510 ‘responder’ line (carrying human *MAPT* P301L cDNA downstream of a tetracycline operon–responsive element), with an ‘activator’ line (expressing a tetracycline-controlled transactivator (tTA) under control of the *Camk2a* promoter) (see **Chapter 2**, **Table 2.1**). Bi-transgenic progeny (rTg4510 carrier/carrier or CC, as homozygous mice are not viable) express human tau constitutively, which can be inactivated by administration of doxycycline, providing an option for temporal control of mutant tau transgene expression. Several studies have shown that neuronal death and cognitive decline can be stopped and even reversed following transgene suppression with doxycycline (Blackmore et al., 2017, Santacruz et al., 2005, Spires et al., 2006).

Of note, the genome integration sites for both the *Camk2a*-tTA and *MAPT* transgenes have been shown to disrupt endogenous mouse genes, resulting in deletions affecting *Wdr60*, *Esyt2*, *Ncapg2*, *Ptprn2*, and *Fgf14* (Goodwin et al., 2017).

1.5.1.2. Modelling amyloid neuropathology: the J20 mouse model

J20 mice overexpress a mutant form of the human amyloid precursor protein (hAPP) with two mutations identified in FAD – the Swedish (K670N/M671L) and Indiana (V717F) mutations – under control of the platelet-derived growth factor subunit B (PDGFB) promoter. These mice are characterised by progressive amyloid deposition in the forebrain, abnormal neuronal morphology and function, as well as cognitive deficits (Mucke et al., 2000, Palop et al., 2007). Specifically, progressive deposition of amyloid plaques in J20 mice has been described to occur in the hippocampus and neocortex at 5-7 months of age, with all transgenic mice exhibiting plaques by 8-10 months of age (Harris et al., 2010, Mucke et al., 2000), resembling amyloid deposition in individuals with AD (Thal et al., 2002) (see **Figure 1.2**). Neuronal loss has been reported in J20 transgenic mice in the CA1 region of the hippocampus at 12, 24, and 36 weeks of age, with no neuronal loss observed in the CA3 region up to 36 weeks of age (Wright et al., 2013). The transgene integration site in the J20 model has been shown to disrupt the mouse *Zbtb20* gene, affecting its expression in young J20 transgenic mice, with no difference reported at the protein level (Tosh et al., 2017).



Tangles	Absent.
Neuronal Loss	Cell loss varies by brain region. No significant neuronal loss was observed in the CA3 region of the hippocampus at 6, 12, 24 and 36 weeks of age nor in the CA1 region at 6 weeks; however, at 12, 24, and 36 weeks significant neuronal loss was observed in the CA1 region compared to age-matched wild-type animals (Wright et al., 2013).
Plaques	At 5-7 months of age diffuse amyloid-β plaques deposit in the dentate gyrus and neocortex. Amyloid deposition is progressive with widespread plaques by 8-10 months. Aβ puncta are deposited in the hippocampus as early as 1 month (Hong et al., 2016).
Gliosis	At 24 and 36 weeks a significant increase in the number of reactive GFAP+ astrocytes and CD68+ microglia was observed in the hippocampi of J20 mice compared to age-matched wild-type controls. No significant difference was observed at 6 and 12 weeks (Wright et al., 2013).
Synaptic Loss	Age-dependent loss of synaptophysin, synaptotagmin, PSD-95, and homer immunoreactivity in the hippocampus by 3 months; synapse loss was confirmed by electron microscopy. No significant difference was seen at 1 month (Hong et al., 2016).
Changes in LTP/LTD	Basal synaptic transmission is impaired between 3-6 months; extracellularly recorded field EPSPs at the Schaffer collateral to CA1 synapse in acute hippocampal slices were on average smaller in amplitude than those seen in wild-type mice. Significant deficits in LTP at the Schaffer collateral–CA1 synapse compared with control mice at 3-6 months (Saganich et al., 2006).
Cognitive Impairment	Deficits in spatial memory and learning appear as the mice age. By 4 months, J20 mice demonstrate spatial reference memory deficits as measured by the radial arm maze (Wright et al., 2013) and Morris water maze (Cheng et al., 2007).

Figure 1.8 – Phenotype characterisation of J20 mice.
Figure taken from Alzforum (www.alzforum.org; accessed 7th March 2019). EPSP – excitatory postsynaptic potential; LTP – long-term potentiation.

1.5.2. Genomic alterations in mouse models of AD

Even with the development of validated animal models of AD and the widespread utility of model organisms as a valuable tool for understanding the molecular processes involved in disease progression (Gotz and Ittner, 2008, Sasaguri et al., 2017), limited work has been undertaken to assess regulatory genomic changes associated with neuropathology in non-human models of AD.

1.5.2.1. Transcriptomic studies in AD mouse models

Recent studies have identified widespread gene expression differences in transgenic mice harbouring a diverse range of AD-associated mutations (**Table 1.7** and **Table 1.8**). Many of these studies have assessed individual brain regions from a single mouse model, with most analyses performed on relatively small numbers of animals. Furthermore, the vast majority of these studies have not directly related transcriptional alterations to the progressive burden of pathology in the same mice, and only limited attempts to relate observed variation to human AD data have been undertaken.

Similarly to AD studies in human (see **Section 1.3**), initial gene expression studies in AD mouse models used microarray-based approaches (**Table 1.7**). One study in particular found that the progression of Aβ plaques in amyloid transgenic mice was associated with immune gene expression, and that the presence of neurofibrillary tangles was negatively correlated with the expression of synaptic genes, in cortical and hippocampal samples, with larger differences identified in the hippocampus (Matarin et al., 2015). In a follow-up study, Salih et al. (Salih et al., 2018) used RNA-seq to reanalyse these hippocampal mouse samples, reporting similar results to their previous study, particularly for genes involved in innate immune regulation (**Table 1.8**).

Table 1.7 – Gene expression studies in bulk brain tissue from AD mouse models using microarray-based methods.

Reference	Model(s)	Tissue	Age(s)	Key findings
Hokama et al. (2014)	3xTg-AD	Hippocampus	14 months	Significant expression alterations in genes involved in noninsulin-dependent diabetes mellitus and obesity in both AD brains (see **Table 1.1**) and 3xTg-AD samples: • Transcripts with decreased expression included *Srd5a1*, *Mlh1*, *Cdkn1b*, *Pcsk1*, *Camk2d*, *Cplx3*, *Vgf*, *Chrna4*, *Pygb*, *Pik3cg*, and *Pla2g16;* • Transcripts with increased expression included *Cst7*, *Ide*, *Apobec3b*, *Ldlr*, and *Il18bp.*
Landel et al. (2014)	5xFAD	Neocortex and hippocampus	1, 4, 6 and 9 months	Shift in gene expression patterns between 1 and 4 months of age. Region-specific variation in gene expression at 1 month (with hippocampus showing many more changes), but a large proportion of common gene expression changes in both brain regions in all other age stages (4, 6 and 9 months), after onset of plaque deposition. Common transcripts with increased expression in the two brain regions (e.g. *Clec7a*, *Cst7*, *Itgax* (or Cd11c)), were predominantly involved in inflammatory and immune processes, complement, major histocompatibility complexes, and toll-like receptors, and were suggestive of microglial activation. Other transcripts showing increased expression included *Igf1*, *Osmr* and *Grn*/*Pgrn*, suggestive of dysregulation of insulin signalling pathways.
Wes et al. (2014)	rTg4510	Cortex and hippocampus	1.9, 4.7, to 6.1 months	Used Camk2a-tTA transgenic mice instead of double negative wildtype mice as controls. Identified age-dependent (1.9, 4.7, to 6.1 months) changes in the expression of transcripts such as *C4b*, *Gfap*, *Trem2*, and *Tyrobp*, as well as *Itgax* and *Cd68*, further suggesting *Tyrobp* and *Cd68* as major hub genes.

Matarin et al. (2015)	APP (K670N/M671L) PSEN1 (M146V), APP/PSEN1 (APP: K670N/M671L, PSEN1: M146V) CaMKII-MAPT (P301L)	Cortex, hippocampus and cerebellum	2, 4, 8 and 18 months	Gene expression differences associated with amyloid and tau pathology, with larger effects identified in APP/PSEN1 mouse samples: • Changes of immune-related genes positively corelated with amyloid plaques; • The presence of neurofibrillary tangles was negatively correlated with the expression of synaptic genes.
Castillo et al. (2017)	$App^{NL\text{-}G\text{-}F/NL\text{-}G\text{-}F}$ 3xTg-AD	Cortex	12 months	Expression changes in neuroinflammatory genes in cortex tissue from 12 months-old mice highly correlated with amyloidosis, with higher changes in $App^{NL\text{-}G\text{-}F/NL\text{-}G\text{-}F}$ compared to 3xTg mice. Transcripts with increased expression in $App^{NL\text{-}G\text{-}F/NL\text{-}G\text{-}F}$ compared to control non-transgenic mice included *C4a*/*C4b, Cd74, Ctss, Gfap, Nfe2l2, Phyhd1, S100b, Tf, Tgfbr2, Vim, Apoe, Abi3*, and *Cd33, Bin2*, *Ctsc*, *Dock2*, *Fcer1g*, *Frmd6*, *Hck*, *Inpp5D*, *Ly86*, *Plcg2*, *Trem2*, and *Tyrobp*. Only *Trem2* showed increased expression in both models, and *Frmd6* exhibited a decrease in expression in both mouse models compared to non-transgenic controls.

Table 1.8 – Gene expression changes in bulk brain tissue from AD mouse models using RNA-seq.

Reference	Model(s)	Tissue	Age(s)	Key findings
Lee et al. (2018)	5xFAD/TREM2	5xFAD	2, 4, and 7 months	Partial rescue of cortical transcriptional dysregulation with increased TREM2 gene dosage.
Salih et al. (2018)	APP (K670N/M671L) PSEN1 (M146V) APP/PSEN1 (APP: K670N/M671L, PSEN1: M146V)	Hippocampus	2, 4, 8 and 18 months	A microglia-associated expression module was correlated with amyloid pathology and contained the mouse orthologs of AD-relevant previously identified loci, particularly from GWAs studies, such as *Trem2*, *Spi1*, *Ms4a6d*, *Abi3*, and *Cd33*, as well as *Tyrobp*, identified in gene expression studies in the human AD brain. Other transcripts that have been previously associated with amyloid pathology, such as *Laptm5*, *Itgam* (or Cd11b), *Lilrb4a* and *Cxcl10* were also fdisrupted.
Rothman et al. (2018)	Tg2576 TgCRND8	Cortex	Tg2576: 3, 6, 9, 12, 15 months TgCRND8: 1.5, 3, 4.5, 6, 10 months	Age-dependent transcriptional changes when comparing transgenic male mice to wildtype age-matched littermate controls. Larger changes were found in TgCRND8 mice: • Increase in expression of *Trem2*, *Tyrobp*, *Cd68*, *Clec7a*, *Tspo*, *Itgfax*, and *C4b.* In laser captured plaque-enriched regions versus non-plaque regions from 6-months-old TgCRND8 mice the majority of transcriptional changes identified were enriched within plaques.
Swarup et al. (2019)	MAPT P301S	Cortex, hippocampus, brain stem, and cerebellum	3 and 6 months	Expression changes of the top genes with increased expression in the cerebellum had an opposite profile in the cortex. Using co-expression analysis, two modules were found to be correlated highly with phosphorylated tau in MAPT P301S mice: • One module, showind decreased expression at 6 months in all brain regions except the cerebellum, was enriched for synaptic pathways; • One module, with increased expression at 6 months in all brain regions except the cerebellum, was enriched for immune and inflammatory categories; • Preservation of these modules was identified in four independent datasets.

At the single cell level, Keren-Shaul et al. (2017) identified a specific subset of microglia in 5xFAD transgenic individuals that was absent in wildtype animals, which they called disease-associated microglia (DAM). DAM were characterised by reduction in expression of microglia homeostatic genes (e.g. *P2ry12/P2ry13*, *Cx3cr1*, *Tmem119*) and increased expression of *Apoe*, *Cst7*, cathepsin D (*Ctsd*), lipoprotein lipase (*Lpl*), *Tyrobp*, and *Trem2*. Gene ontology analysis of DAM-specific genes revealed a significant involvement of transcripts associated with lysosomal/phagocytic pathways, endocytosis, and regulation of the immune response. The authors further report that these DAM surround plaques and require TREM2 in order to be activated. Additional analysis by Keren-Shaul et al. identified DAM in a mouse model of amyotrophic lateral sclerosis (SOD1-G93A mice) (Keren-Shaul et al., 2017), and a similar microglia cell state was identified in a different mouse model of neurodegeneration (CK-p25 mice) at later stages of disease progression in another study (Mathys et al., 2017), with expression profiles of these late-response microglia characterised by increased expression of common genes to DAM described by Keren-Shaul et al. (e.g. *Itgax*, *Clec7a*, *Ctsd, Cd63, Lpl*, *Cst7*, *Apoe*). Together, the results from these two studies suggest that microglia expressing a core set of genes may respond to various causes of brain homeostasis disruption. In addition, a study looking at microglia isolated from the forebrain of rTg4510 mice reported temporal transcriptomic alterations in microglia from transgenic mice compared to controls (2, 4, 6, and 8 months), including an increase in the expression of genes involved in innate immunity and a reduction in the expression of genes involved in the glutamatergic synapse (Wang et al., 2018).

1.5.2.2. Epigenetic studies in AD mouse models

Despite their widespread utility, limited work has been undertaken to assess epigenomic changes in AD-relevant mouse models. Most studies carried out to date have focused either on global levels of epigenetic modifications, particularly histone modifications, or at a small number of loci (Francis et al., 2009, Gstir et al., 2014, Marques et al., 2012, Sanchez-Mut et al., 2013, Walker et al., 2013, Zhang et al., 2014).

Very few studies assessing DNA modifications in AD mouse models have been performed to date (**Table 1.9**). The first studies used genome-wide DNA methylation arrays (Cong et al., 2014, Sanchez-Mut et al., 2013), and two subsequent studies have assessed global methylation using immunohistochemistry (Cadena-del-Castillo et al., 2014, Lardenoije et al., 2018). Interestingly, hypomethylation of the *Mapt* promoter has been reported in 5-month-old J20 transgenic mice, which display little or no plaque pathology at this time point (see **Section 1.5.1.2**), compared to controls (Coupland et al., 2014). All these studies have profiled a very low number of mice (1-5 animals per group), and it is currently difficult to make any firm conclusions about changes in DNA modifications associated with AD pathology in any transgenic mouse model.

Table 1.9 – Studies of DNA modifications in AD mouse models.

Reference	Model(s)	Tissue	Method	Key findings
Sanchez-Mut et al. (2013)	APP/PSEN1 (12 months) 3xTg-AD (18 months)	Frontal cortex	Custom designed mouse brain DNA methylation array (covering 384 genes), and pyrosequencing	Hypermethylation of promoter CpG sites annotated to *Tbxa2r*, *F2rl2*, *Sorbs3*, and *Sptbn4* in APP/PSEN1 transgenic mice compared to controls. This was related to increased promoter hypermethylation for *TBXA2R*, *SORBS3* and *SPTBN4* in AD human samples.
Cong et al. (2014)	APP/PSEN1 (11 months)	Cortex	Mouse DNA methylation promoter array	Hypermethylation in 485 genes, with several differentially methylated genes from the same gene family (e.g. olfactory receptor (OLFR), solute carrier, interleukin, keratin, cytochrome P450, and nuclear receptor family members). Top hypermethylated genes included *Serpinb10*, *Tgfb1*, *Scnn1a*, *Thoc7*, *Optc*, *Myh4*, and *Prnp*.
Coupland et al. (2014)	J20 (5 months)	Hippocampus	Bisulfite-pyrosequencing	Lower DNA methylation in J20 transgenic mice compared to controls in the *Mapt* promoter.
Cadena-del-Castillo et al. (2014)	3xTg-AD mice (1, 3, 8, and 17 months)	Cortex and hippocampus	Immunohistochemistry	Increase in global 5hmC levels in the cortex of transgenic mice compared to controls. No statistically significant differences in global levels of 5mC were observed.
Lardenoije et al. (2018)	3xTg-AD (5, 14, 17, 27 months) J20 (4, 8, 16, 34 months) APP/PSEN1 (6, 16, 17, 18, 34, 27 months)		Immunohistochemistry	Age-associated increase in global 5mC in the DG and CA1-2 hippocampal sub-regions in 3xTg-AD mice (from 5 to 25 months old), with no differences in global 5hmC identified in this mouse model. Age-related decrease in 5mC in the DG, a decrease in the ratio of 5mC and 5hmC in the DG and CA3, and a negative correlation between plaque load and 5mC in the DG, in J20 transgenic mice compared to controls. No significant age-associated alterations identified in APP/PSEN1 mice.

1.6. Aims and objectives

Considering the literature described in the previous sections, the overall hypothesis to be tested in my PhD was that the accumulation of tau and amyloid pathology in the brain is associated with progressive transcriptomic and epigenomic dysregulation in the entorhinal cortex. My goal was to describe transcriptional mechanisms involved in i) the accumulation of tau and amyloid, ii) managing the neuroinflammatory response to this pathology, as well as iii) the biological consequences of these toxic protein species. Using two well-characterised rodent models of AD neuropathology, the rTg4510 (**Section 1.5.1.1**) and J20 (**Section 1.5.1.2**) mouse lines, the aim of this project was to evaluate transcriptional and epigenomic consequences of progressive tau or amyloid pathology.

The main objectives of my PhD project were:

1) to characterise the neuropathological burden of tau and amyloid mice (**Chapter 3**);
2) to identify changes in gene transcription associated with the progression of tau or amyloid pathology, integrating these changes with neuropathological information from the same mice (**Chapter 4**);
3) to identify global changes in DNA modifications associated with the progression of tau or amyloid pathology (**Chapter 5**);
4) to identify site-specific variation in DNA methylation associated with tau or amyloid pathology (**Chapter 6**).

The overall study design for my thesis is shown in **Figure 1.9**, and the approach used to achieve my proposed objectives, including the methods used across this thesis, is shown in **Figure 1.10**.

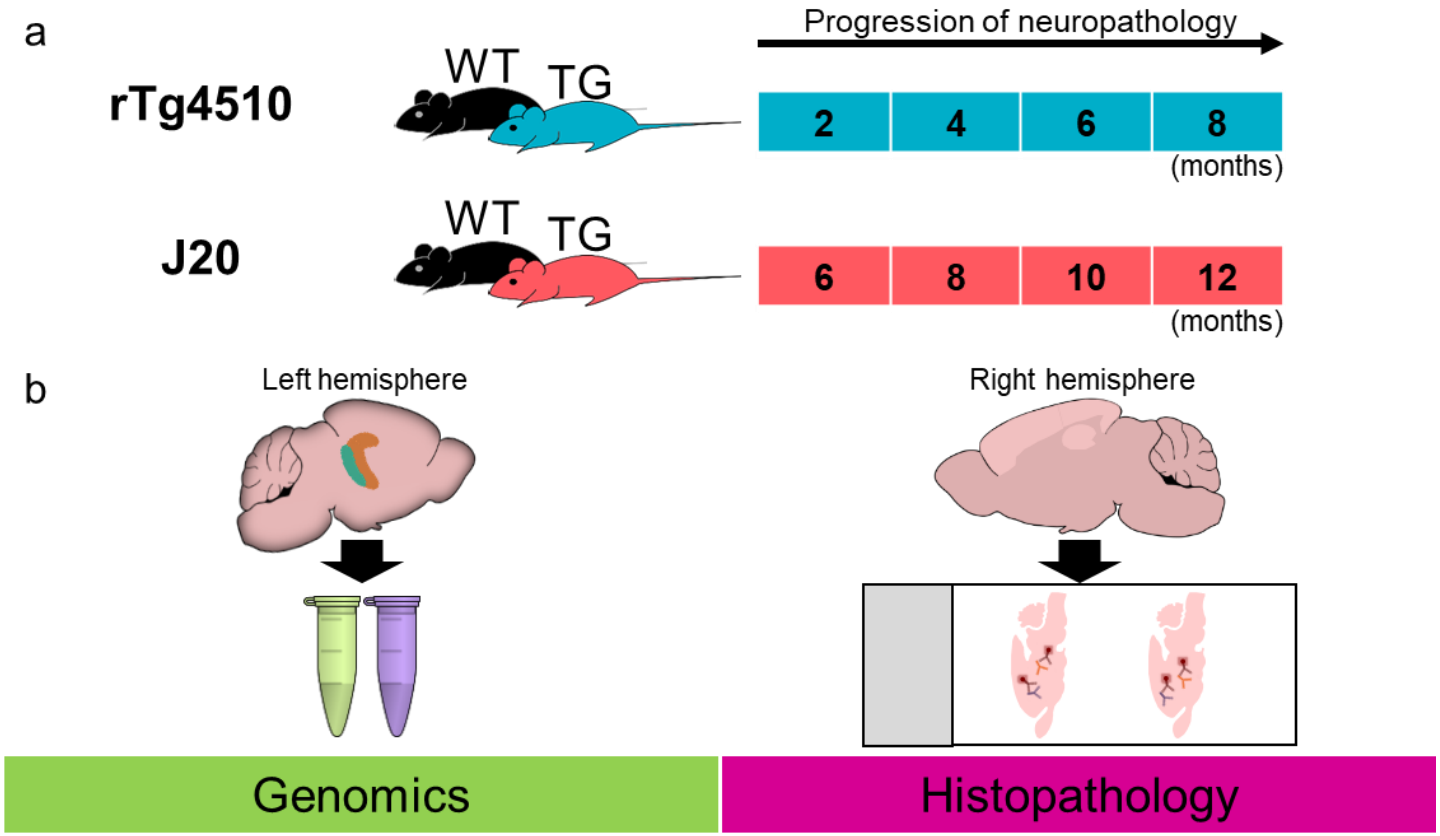


Figure 1.9 – Study design.
(a) To investigate genomic signatures of tau pathology, I used the rTg4510 model, which overexpresses a human mutant (P301L) form of the microtubule-associated protein tau (MAPT). To investigate amyloid pathology, I used the J20 mouse line, which expresses a mutant (K670N/M671L and V717F) form of the human amyloid precursor protein (APP). Tissue was collected from transgenic (TG) and wildtype (WT) control mice at four time points. **(b)** From each mouse, the entorhinal cortex (ECX, dark green) and hippocampus (HIP, orange) were dissected from the left hemisphere and used for genomic studies. The right hemisphere was used for histopathology experiments.

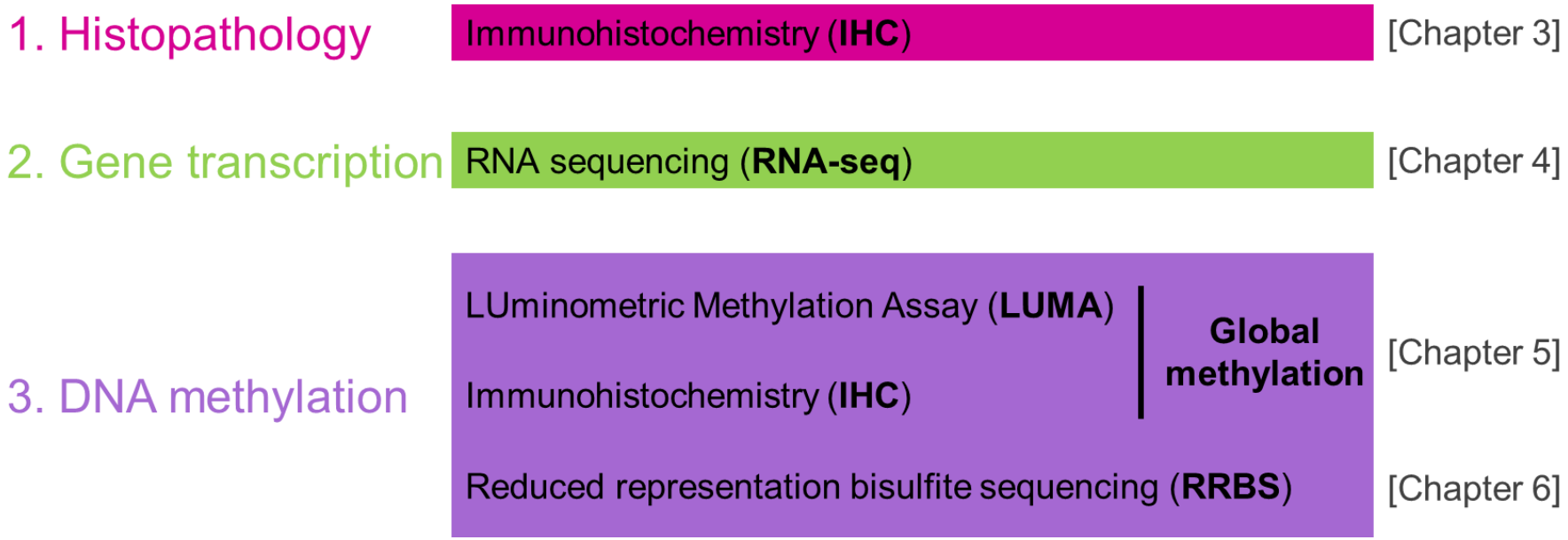


Figure 1.10 – Approach.
Levels of investigation, including genomic interrogation, and the methods used to profile them.

1.7. Originality and contribution of my PhD

Our group and others have previously identified differences at the transcriptomic and epigenomic level in human post-mortem tissue from AD brains (see **Section 1.3** and **Section 1.4**). As discussed above (**Section 1.5**), despite highlighting pathways potentially involved in disease pathogenesis, studies using human post-mortem tissue are confounded by many secondary factors; the use of animal models can therefore play a major role in understanding the processes mediating the initial stages and progression of disease. Nonetheless, limited work has been undertaken to assess regulatory genomic changes associated with neuropathology in AD-relevant animal models (see **Section 1.5.2**). My PhD builds up on the literature described in the previous sections and develops our understanding about the brain response to the production of toxic protein species – Aβ or tau – and the gradual deposition of plaques or tangles, respectively. It provides information about the biological systems involved in fighting the accumulation of plaques and tangles (including microgliosis and neuroinflammatory responses in general, astrocytosis, autophagy, and lipid metabolism pathways), as well as the biological consequences of toxic protein species (neurotoxicity and apoptosis, for example) in the mouse brain. Additionally, using systems-level analyses, loci characterised by transcriptomic changes associated with the progression of amyloid and/or tau pathology have been subsequently explored in human brain data generated from ongoing studies of AD patients and matched controls.

To my knowledge, this project represents the most systematic analysis of transcriptional and methylomic variation in mouse models of tau and amyloid pathology and is the first to specifically focus on changes in the entorhinal cortex, a key region of the brain implicated early in the pathogenesis of AD (see **Section 1.1**). Compared to previous studies of transcriptional variation in transgenic mouse models of AD (see **Section 1.5.2**), I profiled a relatively large number of samples spanning multiple time-points selected to encompass the development of pathology; this study was therefore well powered to identify gene expression differences associated with both genotype and the progression of AD pathology. My detailed immunohistochemical analyses also allowed me to directly compare

transcriptional variation with measures of tau and amyloid pathology from the same individual mice.

Chapter 2. General Methods

This chapter contains general methods used for the experiments and results described in multiple chapters (**Chapters 3-6**). Specific experimental procedures relevant to individual results chapters are described in detail in the methods section of each chapter.

2.1. Animal models

2.1.1. Brief overview

Numerous mouse models of AD pathology have been described (see **Section 1.5** in **Chapter 1** for an overview). To investigate changes in transcriptomic and epigenomic regulation associated with tau or amyloid pathology, in this thesis I used the rTg4510 (rTg(tet-o-TauP301L)4510) and J20 (B6.Cg-Zbtb20Tg(PDGFB-APPSwInd)20Lms/2Mmjax) mouse models, respectively. rTg4510 mice overexpress a pathogenic human mutant form (TauP301L, with absence of an N-terminal segment, 4R0N) of the microtubule-associated protein tau (MAPT) under the *Camk2a* promoter. Transgene expression is highest in the forebrain, with mice characterised by an age-dependent formation of neurofibrillary tangles (Ramsden et al., 2005, Santacruz et al., 2005) (see **Section 1.5.1.1**). The J20 mouse line expresses a mutant form (with the Swedish K670N/M671L and Indiana V717F mutations) of the human amyloid precursor protein (APP) under the platelet-derived growth factor beta (PDGFB) promoter. These animals have an age-dependent distribution of amyloid pathology, which is highest in the hippocampus and neocortex (Harris et al., 2010, Mucke et al., 2000) (see **Section 1.5.1.2**). Littermate female mice were used in this study (initial n = 9-10 animals per group).

2.1.2. Breeding strategies

All animal procedures were carried out at Eli Lilly where I did an industrial placement (~3.5 months) during my PhD (from May 2017 to August 2017), in

accordance with the UK Animals (Scientific Procedures) Act 1986 and with approval of the local Animal Welfare and Ethical Review Board.

rTg4510 mice, licensed from the Mayo Clinic (Jacksonville, FL, United States), were bred on a mixed 129S6/SvEvTac + FVB/NCrl background (heterozygous tau responder x heterozygous tTA effector). Six breeding colonies are required to produce rTg4510 transgenic (bigenic) mice in addition to the stock colony: two parental inbred background lines (Background Line 1: 129S6/SvEvTac from Taconic, and Background Line 2: FVB/NCrl from Charles River); two transgenic parental lines (Parental line 1: Responder line tau P301L on a FVB/NCrl background, and Parental line 2: tTA Effector line Camk2a-tTa (Tetr transgenic carrier) on a 129S6/SvEvTac background); two heterozygous (HET) transgenic lines (offspring from Parental line 1: HET Tau Responder line (these mice express low levels of human tau), and offspring from Parental line 2: HET tTA Effector line (these animals only express tTa)) – **Table 2.1**.

Table 2.1 – rTg4510 breeding procedures.

Breeding method	Line	Strain background	Breeding scheme	Offspring
1	Background Line 1	129S6/SvEvTac (Taconic)	129S6/SvEvTac x 129S6/SvEvTac	129S6/SvEvTac
2	Background Line 2	FVB/NCrl (Charles River)	FVB/NCrl x FVB/NCrl	FVB/NCrl
3	Tau Responder Parental Line 1	FVB/NCrl	Tau P301L x FVB/NCrl	Heterozygous Tau Responder
4	tTa Effector Parental Line 2	29S6/SvEvTac	Camk2a-tTa x 129S6/SvEvTac	Heterozygous tTA Effector
5	rTg4510	Mixed	HET Tau Responder x HET tTA Effector	rTg4510 wildtype control (tTA WT;Tau WT) rTg4510 bigenic transgenic (tTA;Carry Tau transgene)

J20 mice, licensed from Gladstone Institute (San Francisco, California, United States), with founder mice purchased from MMRRC at The Jackson Laboratory (Bar Harbor, Maine, United States), were bred on a C57BL/6JOlaHsd background (parental generation: hemizygous male x C57BL/6J wild type female). Hemizygous (here identified as TG) females and littermate controls (WT), 6, 8, 10 and 12 months of age (n = 9-10 animals per group), were used for this study.

All mice were bred and delivered to Eli Lilly and Company (Windlesham, UK) by Envigo (Loughborough, UK).

2.1.3. Housing conditions

At Eli Lilly, animals were housed under standard conditions (constant temperature and humidity) with a 12h light/dark cycle in individually ventilated cages (up to 5 animals per cage), with free access to food (Teklad irradiated global rodent diet (Envigo, United Kingdom)) and water.

2.2. Genotyping

To confirm identity and validate transgene status, all mice were genotyped following instructions from original suppliers. Briefly, genomic DNA was extracted from tail tip or ear biopsy using DNAeasy Tissue Kit (Qiagen) and subsequently genotyped by Envigo (Loughborough, UK). Genotyping of rTg4510 mice was achieved using two polymerase chain reaction (PCR) experiments to genotype the tau transgene (primers for Human Tau Exon 1, Human Tau Exon 5, and Mouse T-cell receptor as internal control), and Camk2a-tTA (primers for Camk2a-tTa and T cell receptor delta chain as internal control), following instructions from the Mayo Clinic (Jacksonville, FL, USA). Genotyping of J20 mice was performed using PCR (with primers for the transgene and internal positive control) and quantitative PCR (qPCR) to distinguish hemizygous from homozygous transgenic animals, following instructions from the Jackson Laboratory (Bar Harbor, Maine, United States). Ear biopsy samples from mice that exhibited unexpected pathology (or unexpected lack of pathology) in my immunohistochemistry

analyses (see **Chapter 3**) were sent to Envigo to be re-genotyped; mice with genotype discrepancies were removed from further analyses.

2.3. Study design and sample collection

A more detailed overview of the study design (briefly described in **Section 1.6** of **Chapter 1**) is shown in **Figure 2.1**.

For the rTg4510 model, samples from bi-transgenic (CC, here referred as TG) female mice and littermate controls (WW, here identified as WT), 2, 4, 6 and 8 months old (n = 9-10 animals per group), were collected –**Figure 2.1a** (blue). For J20 mice, samples from hemizygou (here identified as TG) females and littermate controls (WT), 6, 8, 10 and 12 months of age (n = 9-10 animals per group), were collected (**Figure 2.1a** (red)).

Mice were terminally anaesthetised with pentobarbital (intraperitoneal injection) and transcardially perfused with phosphate-buffered saline (PBS).

The whole brain was collected, weighed, and the two brain hemispheres were separated and processed as described below.

2.4. Sample processing

2.4.1. Left brain hemisphere

The entorhinal cortex (ECX) and hippocampus (HIP) were dissected from the left-brain hemisphere on wet ice as described by Heffner et al. (1980) (**Figure 2.1b** left). Dissected regions were snap-frozen, and transferred to Exeter on dry ice for subsequent nucleic acid extractions and downstream genomic assessments (**Chapter 4**, **Chapter 5**, and **Chapter 6**). All tissue samples were stored at -80ºC until processing.

2.4.2. Right brain hemisphere

The right hemisphere was immersed in 10% buffered formalin for fixation (7-8 days) and processed for subsequent immunohistochemistry (**Chapter 3** and **Chapter 4**) assessments (**Figure 2.1b** right).

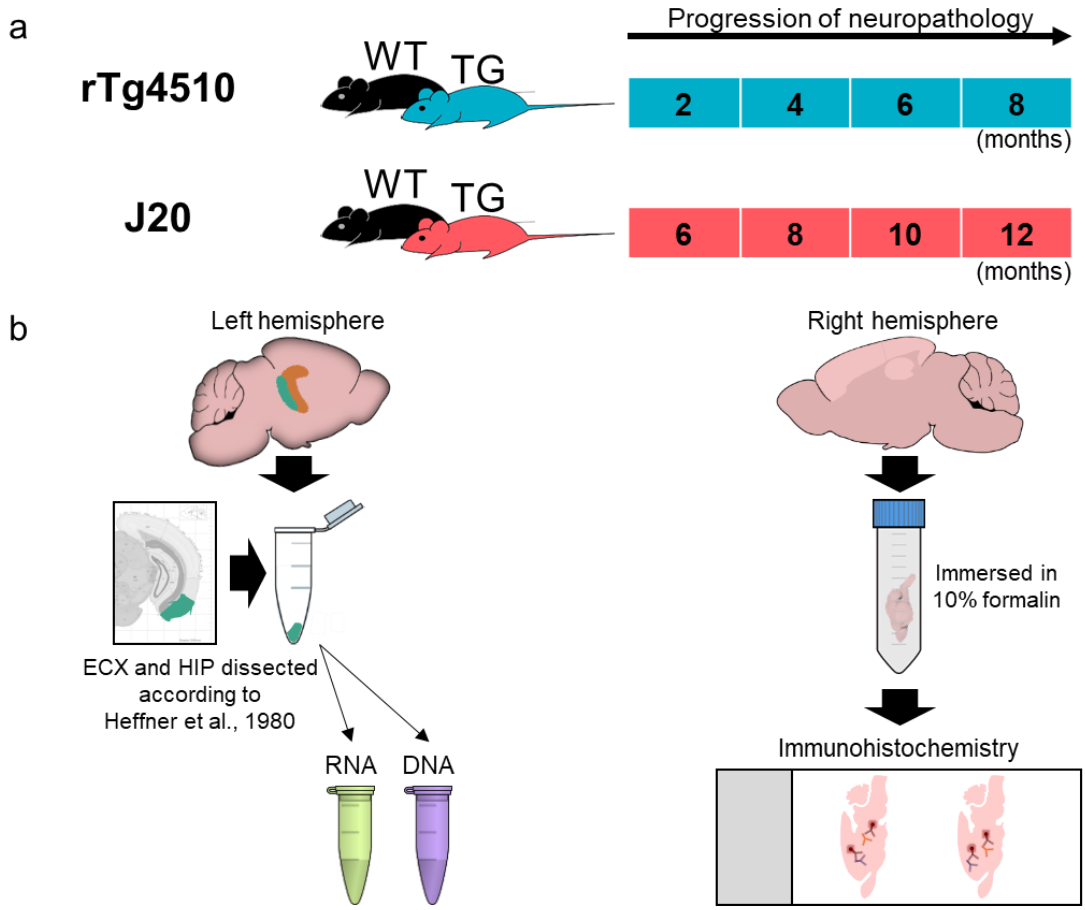


Figure 2.1 – An overview of the experimental design.
(a) To investigate genomic signatures of tau pathology I used the rTg4510 model, and to investigate amyloid pathology I used the J20 mouse line. Tissue was collected from transgenic (TG) and wildtype (WT) control mice at four time points. **(b)** From each mouse, the entorhinal cortex (ECX, dark green) and hippocampus (HIP, orange) were dissected from the left hemisphere and used for RNA and DNA isolation. The right hemisphere was immersed in 10% formalin, and further processed for immunohistochemistry experiments.

2.5. Nucleic acid isolation

2.5.1. Selection of the optimal method for DNA and RNA isolation from mouse brain

Given the limited amount of brain tissue available from each animal, and the desire to isolate both DNA and RNA from the same tissue sample, I compared several extraction protocols using 'test' samples (hippocampus) from control mice (**Table 2.2**).

The optimal yield and quality for both DNA and RNA (see **Section 2.5.3**) were obtained using the AllPrep DNA/RNA Mini Kit (Qiagen) (highlighted in green in **Table 2.2**, **Figure 2.5**, **Figure 2.9** and **Figure 2.10**), therefore the elected method for extraction of DNA and RNA from rTg4510 and J20 brain samples used in this study.

2.5.2. Isolation of DNA and RNA from mouse brain samples

Samples were labelled with anonymised ID codes and processed in batches, blinding me to genotype and age for individual samples. Tissue samples from each mouse model were processed separately and individual samples were randomised to ensure that each group was equally represented in each processing batch.

Total genomic DNA and RNA from all samples was isolated from the entorhinal cortex (**Chapter 4** and **Chapter 6**) and hippocampus (**Chapter 5**) in a laminar flow hood (class II), using the AllPrep DNA/RNA Mini Kit (Qiagen) with minor modifications to the original protocol. An overview of the major steps for the AllPrep DNA/RNA Mini Kit can be seen in **Figure 2.2** and a more detailed description is provided in the following subsections.

All reagents were provided with the kit unless otherwise stated, with all reagents and consumables used being sterile and DNase and RNase free.

Before starting, every surface and all equipment were thoroughly cleaned using 70% ethanol followed by RNaseZap (Thermo Fisher Scientific), and 100% ethanol (molecular biology grade, Sigma-Aldrich) was added to the RPE and RWT buffers as indicated in the manufacturer's instructions.

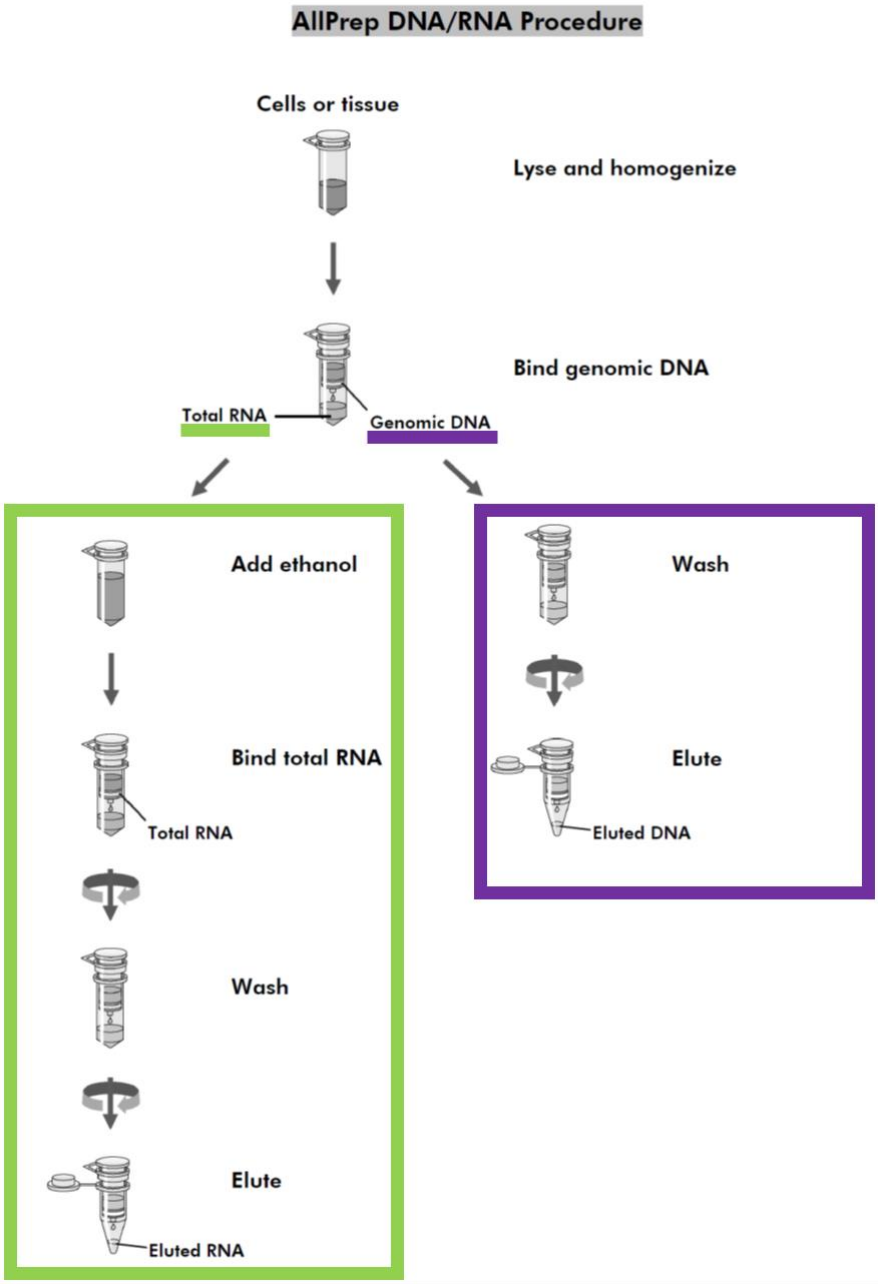


Figure 2.2 – Overview of the experimental procedures for isolation of total RNA and genomic DNA using the AllPrep DNA/RNA Mini Kit.
Figure taken from the AllPrep DNA/RNA Mini Kit Handbook (Qiagen, 2005).

2.5.2.1. Tissue disruption and homogenization

Briefly, I added 350 µL of lysis buffer (prepared fresh by adding β-mercaptoethanol (Sigma-Aldrich) in a concentration of 1 µL of β-mercaptoethanol to 1 mL of buffer RLT Plus) to each tissue sample (up to 30 mg), disrupted the tissue using a DNAse free sterile plastic pestle and homogeniser, and further homogenised the lysate using a P1000 pipette. The lysate was centrifuged (3 minutes at 16,000 x g) and the supernatant carefully removed and transferred to an AllPrep DNA spin column.

2.5.2.2. Isolation of RNA

The AllPrep DNA spin column (in a collection tube) was centrifuged (30 seconds at 10,000 x g), and the flow-through was used for RNA purification whilst the AllPrep DNA spin column was put aside for DNA isolation (**Section 2.5.2.3**).

The flow-through was mixed with freshly prepared 70% ethanol (in a 1:1 ratio) and transferred it to the RNeasy spin column which was then centrifuged (15 seconds at 10,000 x g) and the flow-through discarded. The RNA in the column was then washed with Buffer RW1 (350 µL) and centrifuged (15 seconds at 10,000 x g), treated with DNase (80 µL for 15 minutes, Qiagen), washed with RW1 and centrifuged again (15 seconds at 10,000 x g), washed with Buffer RPE (500 µL) twice and centrifuged (15 seconds at 10,000 x g for the first wash, and 2 minutes at 10,000 x g for the second wash).

Finally, the RNeasy spin column was transferred into a new tube and centrifuged (1 minute at 16,000 x g), before being transferred to a collection tube and eluted in RNase-free water (50 µL) and centrifuged (1 minute at 10,000 x g).

2.5.2.3. Isolation of genomic DNA

The DNA spin column was kept at room temperature during RNA purification (**Section 2.5.2.2**).

I added buffer AW1 (500 µL) to the AllPrep DNA spin column and centrifuged (30 seconds at 10,000 x g), followed by the addition of buffer AW2 (500 µL) and centrifugation (2 minutes at 16,000 x g).

I then carefully placed the column into a new tube (making sure not to transfer any AW2 buffer) and eluted the DNA in buffer EB (50 µL) provided in the kit, incubated at room temperature for 1 minute, and centrifuged (1 minute at 10,000 x g), repeating this process again (total volume of 100 µL).

2.5.2.4. Storage of isolated samples

Aliquots of eluted RNA were stored at -20ºC for short-term use (e.g. RIN score assessment described in **Section 2.5.3.3.**, and RNA-seq experiments described in **Chapter 4**), with the remainder being transferred to -80ºC for long-term storage. Genomic DNA was stored at -20ºC.

2.5.3. Assessment of quality and quantity of isolated nucleic acids

2.5.3.1. Nanodrop

Following extraction from brain tissue, DNA and RNA samples were quantified and checked for purity using a NanoDrop 8000 spectrophotometer (Thermo Scientific). 2µL of a 'blank' solution (elution buffer for DNA and RNase-free water for RNA) and 2µL from each DNA/RNA sample were used.

Nucleic acids absorb UV light at a wavelength of 260nm, whereas proteins absorb UV light at a wavelength of 280nm and other compounds, such as ethylenediamine tetraacetic acid (EDTA), carbohydrates and phenol, absorb UV light at a wavelength of ~230nm. Therefore, the absorbency ratios of 260/280 and 260/280 indicate the presence of protein and other contaminants in the DNA and RNA samples. A 260/280 ratio of ~1.8 (for DNA) and ~2.0 (for RNA), and a 260/230 ratio between 1.8 and 2.2 is generally accepted as indicative of pure samples (Matlock, 2015). A typical pure nucleic acid nanodrop profile is shown in **Figure 2.3**. I initially obtained a suboptimal 260/230 ratio for the DNA testing samples (**Table 2.2**), suggestive of presence of contaminants that absorb at 230 nm; however when I processed the samples used for this thesis I was careful in the final steps of the experimental procedure not to contaminate the samples with reagents containing phenol (from previous steps) which resulted in greatly

improved 260/230 ratios (~1.8) (example shown in **Figure 2.4a**). Representative DNA and RNA spectrographs from one of the mouse brain samples used in this study (**Chapter 4**, **Chapter 5** and **Chapter 6**) are shown as an example in **Figure 2.4**.

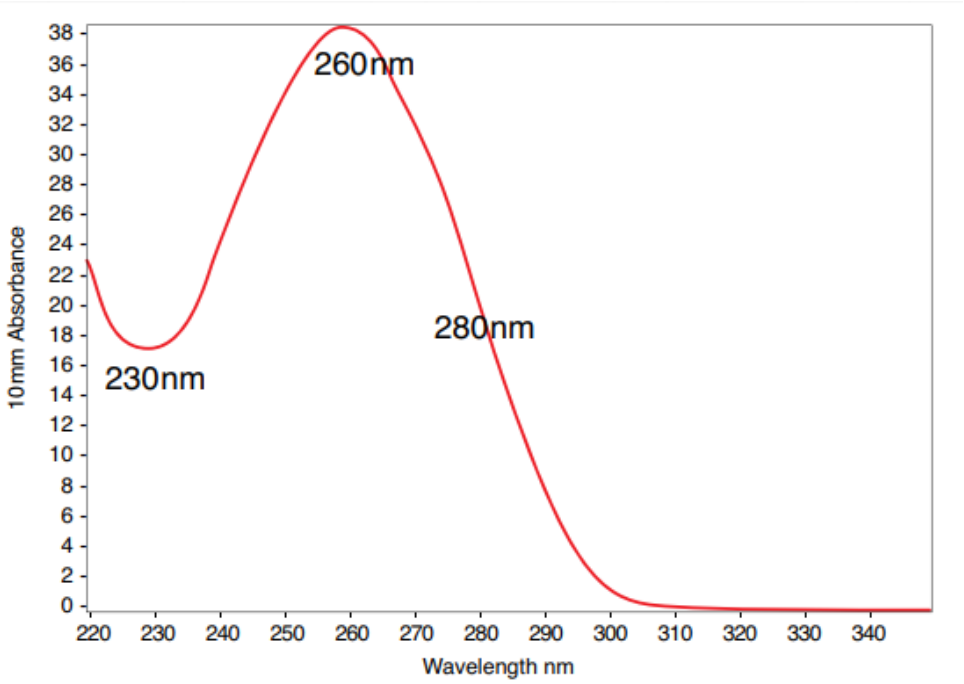
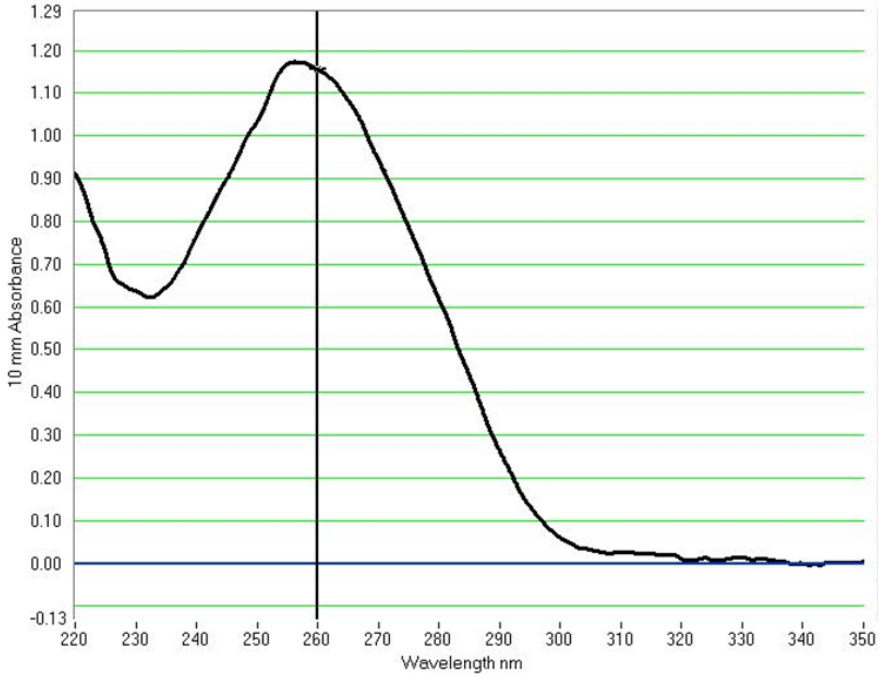


Figure 2.3 – Typical nucleic acid spectrum.
Image taken from the NanoDrop Nucleic Acid Handbook (Thermo Fisher Scientific, 2010).

Table 2.2 – Description and Nanodrop results obtained for all methods tested for the isolation of DNA and RNA from test mouse brain samples.

Extraction method	Sample ID	Tissue weight	Total DNA	Nanodrop (DNA)		Total RNA	Nanodrop (RNA)	
				260/280	260/230		260/280	260/230
AllPrep DNA/RNA Mini Kit (Qiagen)	LH3 (Lilly)	19.1 mg	4.3 µg	1.96	0.69	5.8 µg	2.10	1.82
	LH4 (Lilly)	19.6 mg	4.8 µg	1.92	0.55	1.3 µg	1.99	0.32
RNA/DNA/Protein Purification Plus kit (Norgen)	LH1 (Lilly)	19.7 mg	1.4 µg	1.99	0.56	17.0 µg	2.12	1.96
	RH1 (Lilly)	16.7 mg	1.2 µg	1.84	0.58	16.1 µg	2.11	1.96
Quick RNA MiniPrep (Zymo Research)	RH5 (Lilly)	13.3 mg	1.6 µg	1.83	0.14	0.5 µg	1.45	0.65
Manual extraction with TRIzol (Ambion)	LH2 (Lilly)	14.4 mg	0.8 µg	0.91	0.29	8.1 µg	1.87	0.44
	LH5 (Lilly)	~7.8 mg	0.7 µg	0.81	0.18	6.0 µg	1.86	0.31
Direct-zol RNA MiniPrep (Zymo Research)	RH2 (Lilly)	18.6 mg	-	-	-	6.7 µg	2.02	2.16
	RH3 (Lilly)	23.6 mg	-	-	-	6.2 µg	2.00	2.17
Manual extraction	RH4 (Lilly)	22.3 mg	56.6	1.98	0.72	-	-	-

a



b

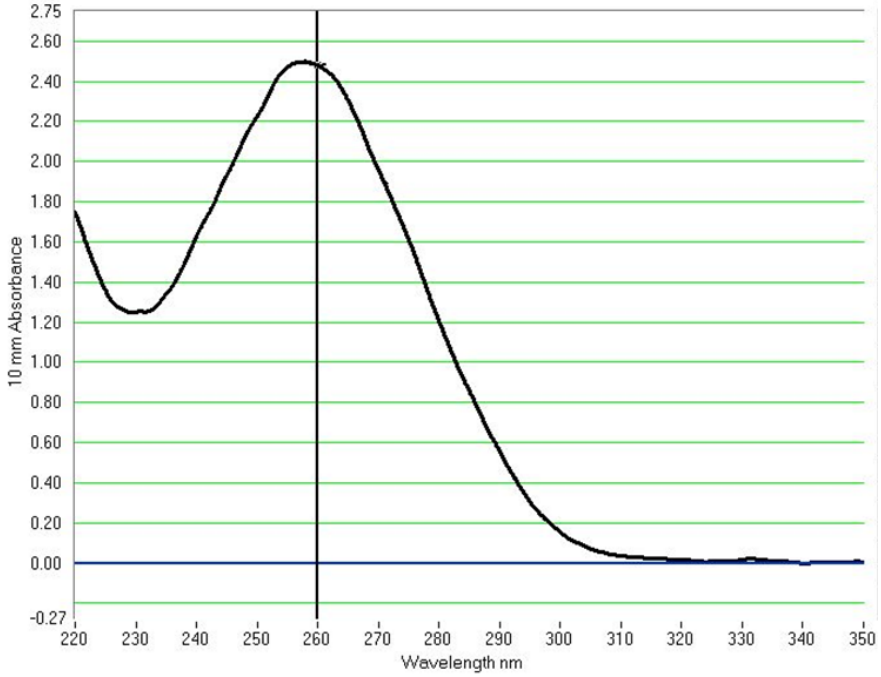


Figure 2.4 – UV spectrographs from one representative sample used in this study.
(a) DNA spectrum (concentration: 57.71ng/µL; 260/280 ratio: 1.86; 260/230 ratio: 1.81).
(b) RNA spectrum (concentration: 99.24ng/µL; 260/280 ratio: 2.06; 260/230 ratio: 1.98).

2.5.3.2 Agarose gel electrophoresis

The integrity of high molecular weight DNA was further evaluated by running the testing samples through a 1% agarose gel by electrophoresis (**Figure 2.5**). To test smaller fragments (e.g. restriction digests, such as described in **Section 5.2.3.2** in **Chapter 5**) a lower concentration (0.8%) gel was used. This method relies on the separation of DNA fragments based on their charge and size by running the DNA through a three-dimensional agarose matrix. The DNA moves due to an electrical charge that forces the negatively charged DNA to migrate towards the positive end of the system. Because different sized molecules move at different speeds within the agarose matrix pores, with smaller molecules moving faster, fragmented DNA can be easily identified. Good quality genomic DNA is expected to produce a single band that should not move quickly due to its high molecular weight. The DNA is then viewed using compounds that intercalate into the DNA structure, allowing it to be visualised under UV light (e.g. ethidium bromide and Syto 60).

Briefly, 2g of agarose powder (Sigma) was added to 200mL 1x tris-borate EDTA (TBE) buffer (prepared from 10x TBE buffer (Fisher Scientific)) in an Erlenmeyer flask. The mixture was heated in a microwave with frequent agitation until complete dissolution of the agarose and cooled under raining tap water. 1µL of ethidium bromide (Sigma-Aldrich) was added to the cooled agarose solution which was then mixed well by agitation, poured into a gel electrophoresis tray and left to solidify for about 30 minutes. During this time, 10µL of each extracted DNA sample was mixed with 2µL of loading buffer (Orange G, Sigma-Aldrich) in a 0.2 mL tube (PCR strips). The set gel was submerged in 1X TBE solution in an electrophoresis tank and 6µL of each sample-loading buffer solution was loaded into an individual well in the agarose gel. A molecular-weight size marker (1kb DNA ladder, New England BioLabs) was added to the first well for approximating the mass of DNA in comparably intense samples of similar size. The gel was left to run for 35min at 120v.

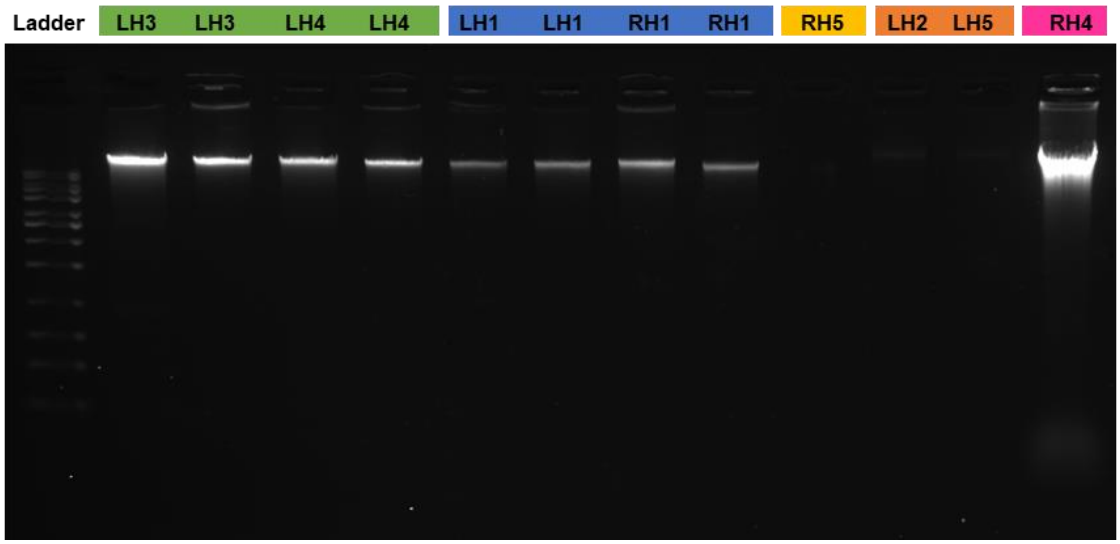


Figure 2.5 – Agarose gel to asses quality of DNA isolated from the test samples described in Table 2.2.
Colours match **Table 2.2**. High molecular weight genomic DNA was obtained for each method that was tested.

2.5.3.3. Calculating RNA integrity number (RIN)

The purity and integrity of RNA isolated from the test samples (**Table 2.2**) was further assessed using an Agilent 2100 Bioanalyzer Instrument (Agilent Technologies) in conjunction with the Agilent RNA 6000 Nano Kit (Agilent Technologies). This system uses a fluorescent dye that binds to RNA on a gel electrophoresis chip. The Agilent Bioanalyzer Software analyses the resulting electrophoretic separation (**Figure 2.6** and **Figure 2.7**) and calculates an RNA integrity number (RIN), that can range between 1 (suggesting that the RNA is highly degraded) and 10 (indicating the RNA has perfect integrity) (**Figure 2.8**).

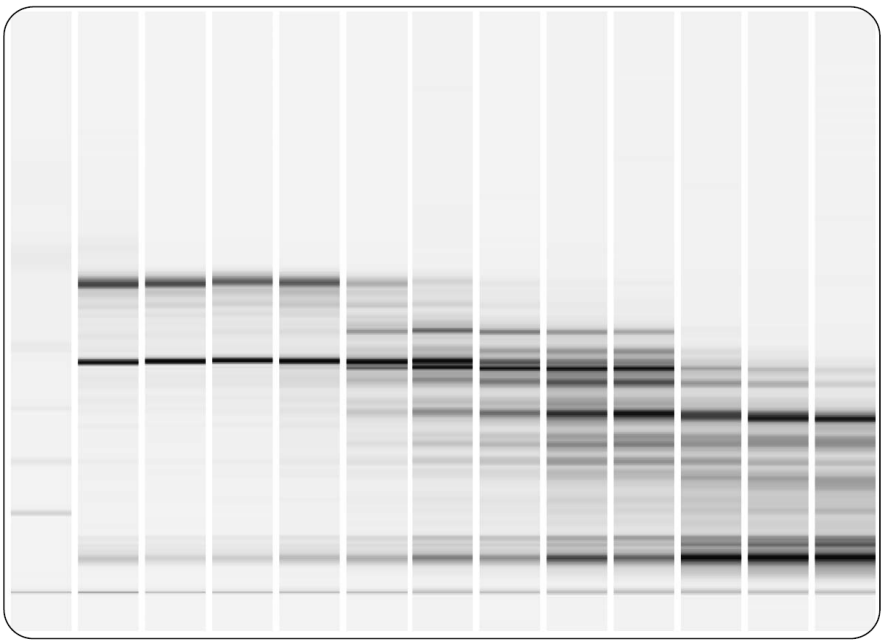


Figure 2.6 – A total RNA sample was degraded for varying times and the resulting samples were analysed on the Agilent 2100 Bioanalyzer System using the Eukaryote Total RNA Nano assay.
A shift towards shorter fragment sizes can be observed with progressing degradation. Figure and legend taken from Mueller et al. (2016).

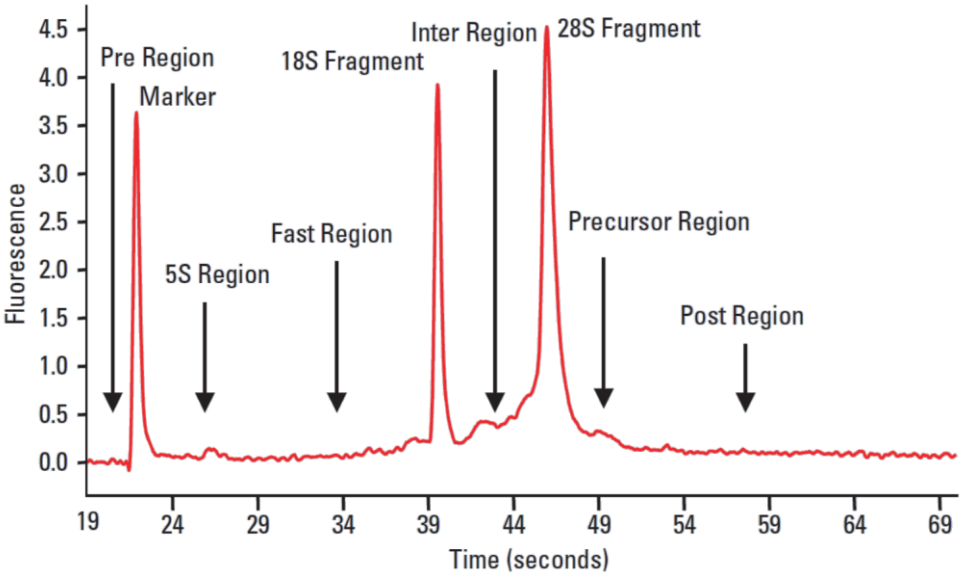


Figure 2.7 – Electropherogram (Bioanalyzer) detailing the regions that are indicative of RNA quality.
Figure and legend adapted from Mueller et al. (2016).

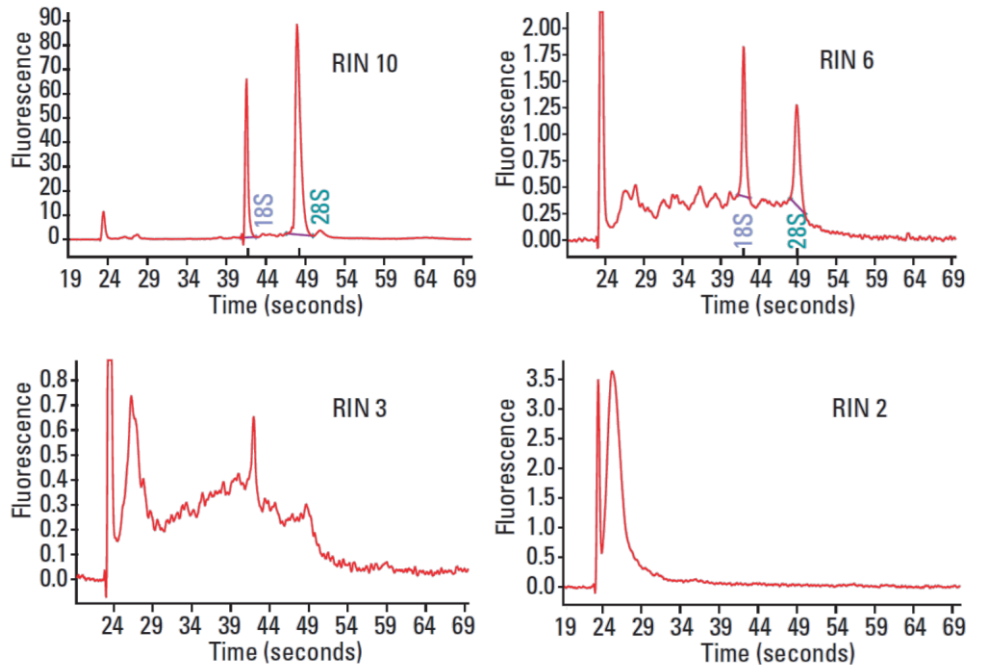


Figure 2.8 – Examples of Bioanalyzer electropherograms from samples ranging from intact (RIN = 10) to very degraded (RIN = 2) RNA.
Figure and legend adapted from Mueller et al. (2016).

Following the Agilent RNA 6000 Nano Assay Protocol from the Agilent RNA 6000 Nano Kit Guide (Agilent Technologies), I set up the chip priming station and the bioanalyzer to make sure they were ready to use, decontaminated the electrodes with RNaseZAP (Invitrogen), prepared the gel, and prepared and loaded the gel-dye mix. I then pipetted 5 µL of the RNA 6000 Nano marker into each well, followed by 1 µL of RNA ladder into its corresponding well and 1 µL of each sample into each of the remaining wells. The chip was then vortexed for 60 seconds at 2400 rpm in an IKA vortex mixer, inserted in the Agilent 2100 Bioanalyzer and left to run.

The gel and the resulting electropherograms are shown in **Figure 2.9** and **Figure 2.10**, respectively, where it can be seen that the RNA isolated with the AllPrep DNA/RNA Mini Kit (Qiagen, in green in **Figure 2.9** and **Figure 2.10**) and with the RNA/DNA/Protein Purification Plus kit (Norgen, in blue in **Figure 2.9** and **Figure 2.10**) had very good quality (RIN = ~8-10).

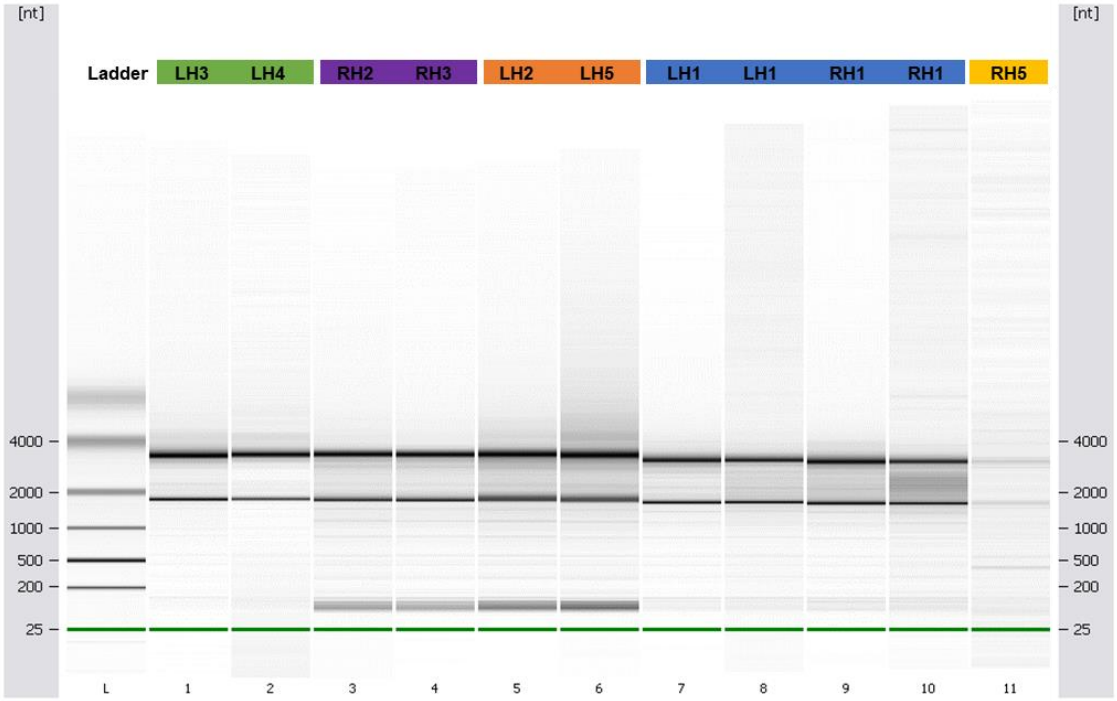
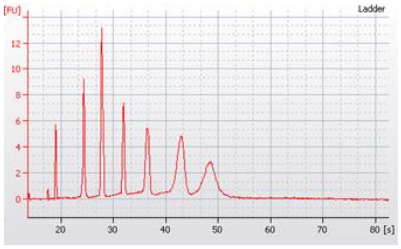
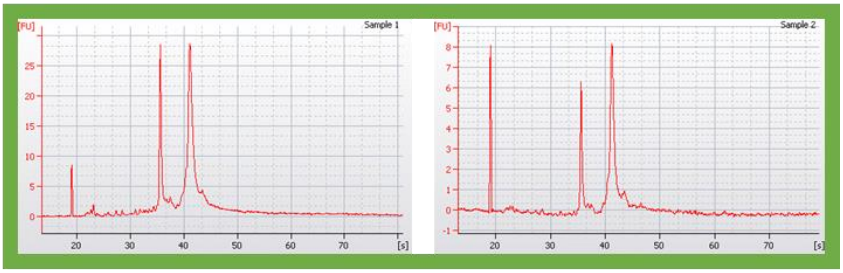


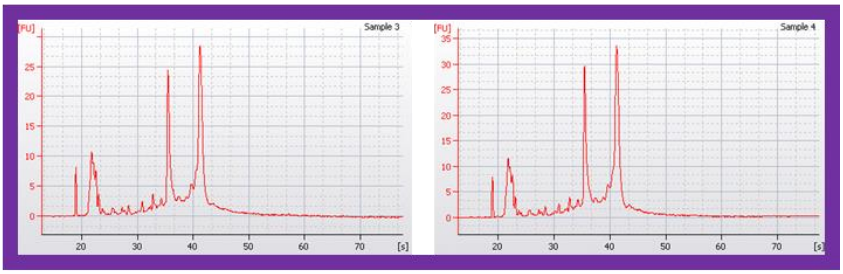
Figure 2.9 – Bioanalyzer electrophoresis to assess integrity for RNA isolated from testing samples described in Table 2.

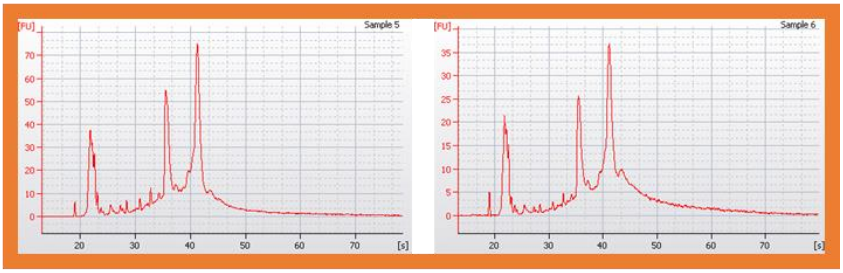
Colours match Table 2.2. Results obtained using the Agilent RNA 6000 Nano Kit and Agilent 2100 Bioanalyzer Instrument (Agilent Technologies).

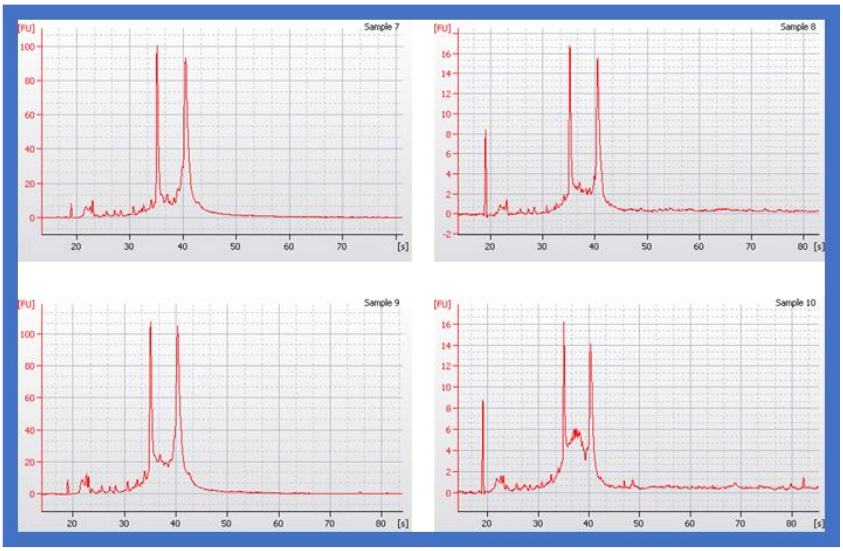
Ladder


Ladder
[FU]
[s]


Sample 1
Sample 2
[FU]
[s]


Sample 3
Sample 4
[FU]
[s]


Sample 5
Sample 6
[FU]
[s]


Sample 7
Sample 8
Sample 9
Sample 10
[FU]
[s]

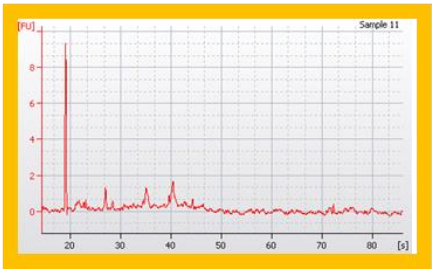

Sample 11
[FU]
[s]

Figure 2.10 – Bioanalyzer electropherograms to assess integrity for RNA isolated from test samples described in Table 2.
Colours match **Table 2.2** and sample numbers match **Figure 2.9**. Results obtained using the Agilent RNA 6000 Nano Kit and Agilent 2100 Bioanalyzer Instrument (Agilent Technologies).

The purity and integrity of RNA for the samples used in this thesis (obtained from mouse entorhinal cortex and hippocampus) was further assessed using a 2200 TapeStation System (Agilent Technologies), in conjunction with the RNA ScreenTape Assay (Agilent Technologies). Similarly to the Bioanalyzer, the TapeStation System uses a fluorescent dye that binds to RNA on a gel electrophoresis chip, carrying out electrophoretic separation of the RNA. Agilent 2200 TapeStation Software also evaluates the resulting electrophoretic separation and calculates the RIN for each sample.

An overview of the experimental procedures for this part is given in **Figure 2.11**.

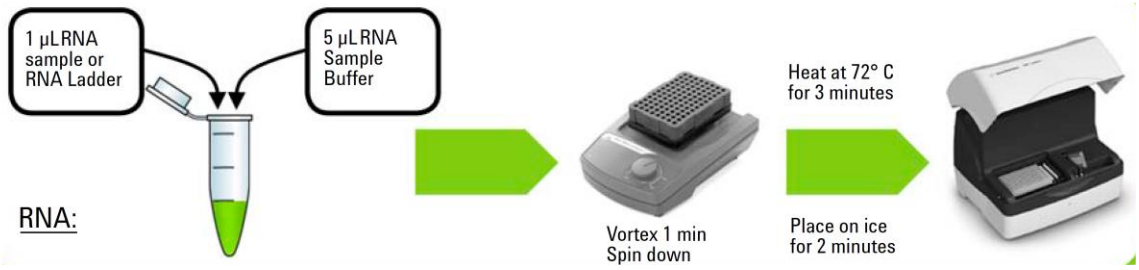


Figure 2.11 – Overview of the experimental procedures for checking RNA concentration and integrity using the RNA ScreenTape Assay and 2200 TapeStation System.
Figure taken from the Agilent 2200 TapeStation User Manual (Agilent, 2013).

Briefly, following the manufacturer's instructions, in a 95-well plate, I added 5µL of RNA Sample Buffer (provided with the kit) to 1µL of each sample or 1µL of RNA Ladder (provided with the kit), vortexed using IKA vortexer and adaptor at 2000 rpm for 1 min and performed a spin down to bring the samples to the bottom of the wells. I then heated the samples at 72ºC for 3 min in a thermocycler, placed the samples on ice for 2 min, and centrifuged to bring the samples to the bottom of the wells again.

After preparing the 2200 TapeStation instrument, including placing the sample block and loading tips, I inserted the ScreenTape into the TapeStation, placed the samples into the sample block, and started the run. When the run was finished, I checked the RNA quantity and quality (integrity) for each sample in the TapeStation Software.

Figure 2.12 shows two representative examples from the RNA samples isolated from mouse entorhinal cortex for the purpose of the work for this thesis. The RNA profiles for the vast majority of the samples isolated had a similar profile to the example shown in **Figure 2.12a**, which corresponds to a sample that was used for downstream analysis (**Chapter 4**), and represented very high-quality RNA. One sample, shown in **Figure 2.12b**, had low RNA quality (RIN = 3.9) and was therefore not used for further analysis. All RNA samples used in my thesis were between RIN 8-10.

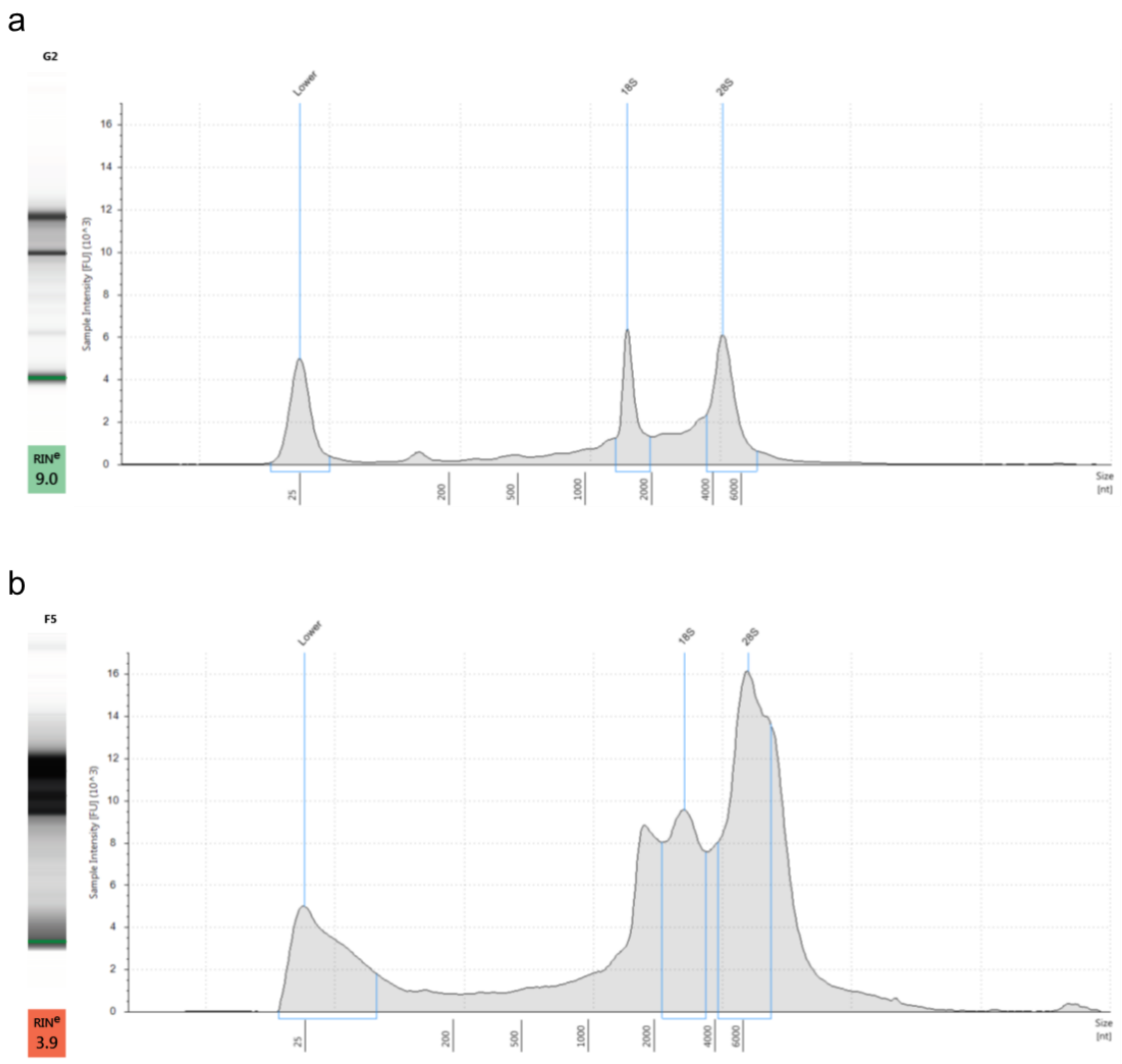


Figure 2.12 – TapeStation gel electrophoresis and corresponding electropherograms from RNA isolated from two mouse entorhinal cortex samples. **(a)** Example of a good quality sample (RIN = 9.0). **(b)** Example of a bad quality sample (RIN =3.9). Results obtained with the RNA ScreenTape Assay and 2200 TapeStation System (Agilent Technologies).

Chapter 3. Characterising the progression of neuropathology in transgenic mouse models of Alzheimer's disease

3.1. Introduction

Transgenic mice expressing mutant human *MAPT* and *APP* (see **Chapter 1**, **Section 1.5**) are characterised by progressive neuropathology in the brain. Previous studies have shown that rTg4510 mice (see **Section 1.5.1.1**) develop pretangles around 2.5 months of age, with neurofibrillary tangle (NFT) pathology starting in the neocortex and progressing rapidly into the hippocampus and limbic structures with increasing age (Ramsden et al., 2005, Sahara et al., 2014, Santacruz et al., 2005). In J20 mice (see **Section 1.5.1.2**), progressive deposition of amyloid plaques in the hippocampus and neocortex at 5-7 months, and ubiquitous plaque pathology by 8-10 months of age, has been described (Harris et al., 2010, Mucke et al., 2000).

Immunohistochemistry, the most common application of immunostaining, relies on the selective identification of antigens within tissue sections by means of specific antibodies. Antigen-antibody binding is then visualised using a coloured histochemical reaction, visible by light microscopy (Ramos-Vara and Miller, 2014).

This chapter describes my use of immunohistochemistry to quantify the progression of neuropathology across multiple regions of the mouse brain (**Figure 3.3** and **Figure 3.4**) using right hemisphere tissue sections from transgenic (TG) and wild-type (WT) littermate control mice (see **Chapter 2**). These experiments complement the genomic (gene expression and DNA methylation) analyses undertaken on the left hemisphere of the same brain samples (see **Chapter 4**, **Chapter 5** and **Chapter 6**). All experiments for this chapter were performed in the Molecular Pathology laboratories at Eli Lilly & Co. Ltd. (Windlesham, United Kingdom), where I spent ~3.5 months during my PhD

(from May 2017 to August 2017). I prepared the sections with the help of Mark Ward (Eli Lilly). Mark Ward and Katherine Sung (Eli Lilly) assisted me with the immunostaining for tau and amyloid pathology, respectively. Katherine Sung, Josh Harvey, and Alice Fisher (Eli Lilly) provided me with training and assistance for the visualisation and quantification of immunostaining. Experiments were performed in a blinded and randomised fashion to avoid any biases in quantification and analysis.

3.2. Methods

In this section the specific experimental procedures used for the evaluation of tau and amyloid neuropathology in brain sections (described in **Chapter 2**, section **2.4.**) from rTg4510 and J20 mice (**Chapter 2**, **section 2.1.**) using immunohistochemistry are described in detail. A general overview of the immunohistochemistry procedures is given in **Figure 3.1**.

3.2.1. Tissue preparation

As described in **Section 2.2** (**Chapter 2**), following the sacrifice of rTg4510 and J20 mice, the right brain hemispheres were immediately immersed in 10% buffered formalin for fixation (for 7-8 days) (**Table 2.1b** right, and **Figure 3.1** in light green). Samples were subsequently processed using the Tissue TEK® VIP processor (GMI Inc), embedded in paraffin wax and stored for sectioning.

Next, 6 µm serial sagittal mouse brain sections were prepared (from lateral 0.84 to 1.08 mm from bregma – **Figure 3.2**) using rotary microtomes (HM 200 from Ergostar and HM 355S from Thermo Scientific). Sections were subsequently mounted on glass slides (two sections per slide) for immunohistochemistry assessments. These procedures were performed during my placement at Eli Lilly (with technical assistance from the Molecular Pathology department).

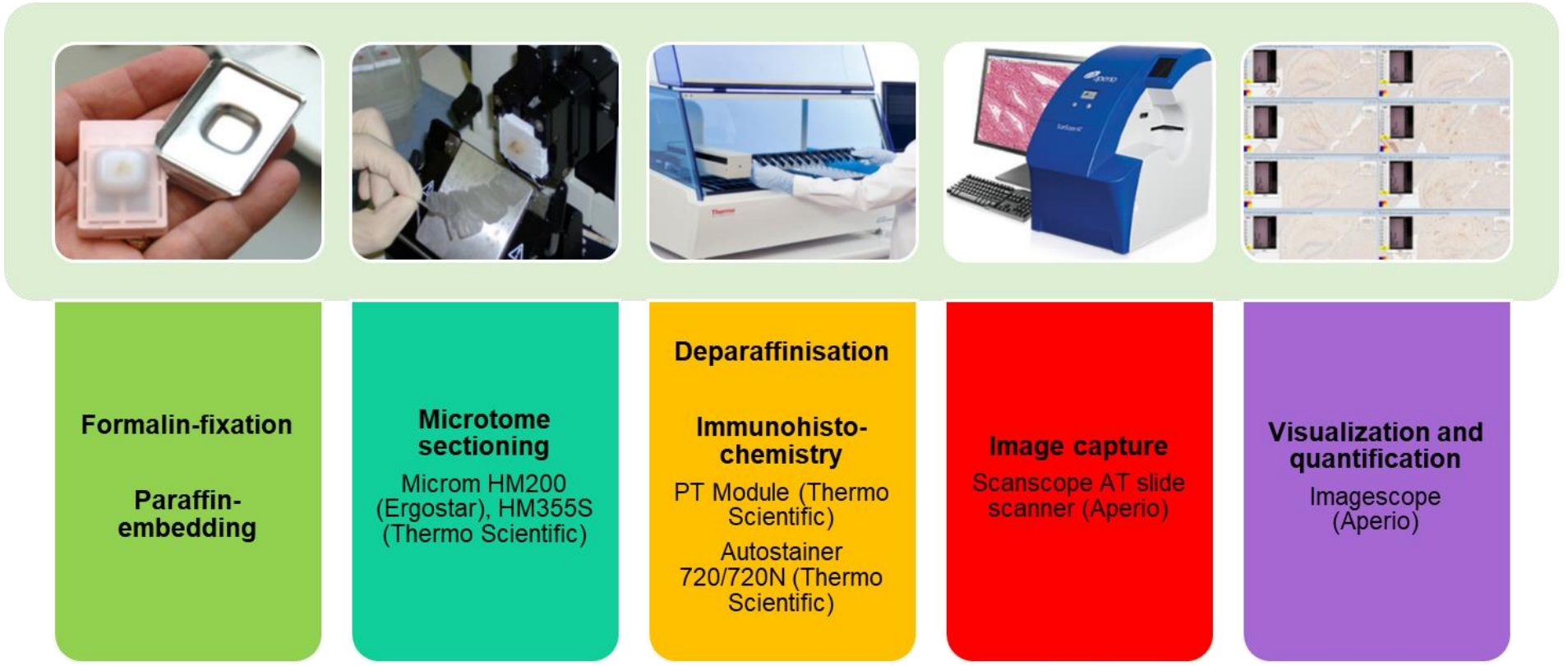


Figure 3.1 – An overview of the experimental steps used for the assessment of tau and amyloid pathology in brain sections from rTg4510 and J20 mice, respectively, by immunohistochemistry.

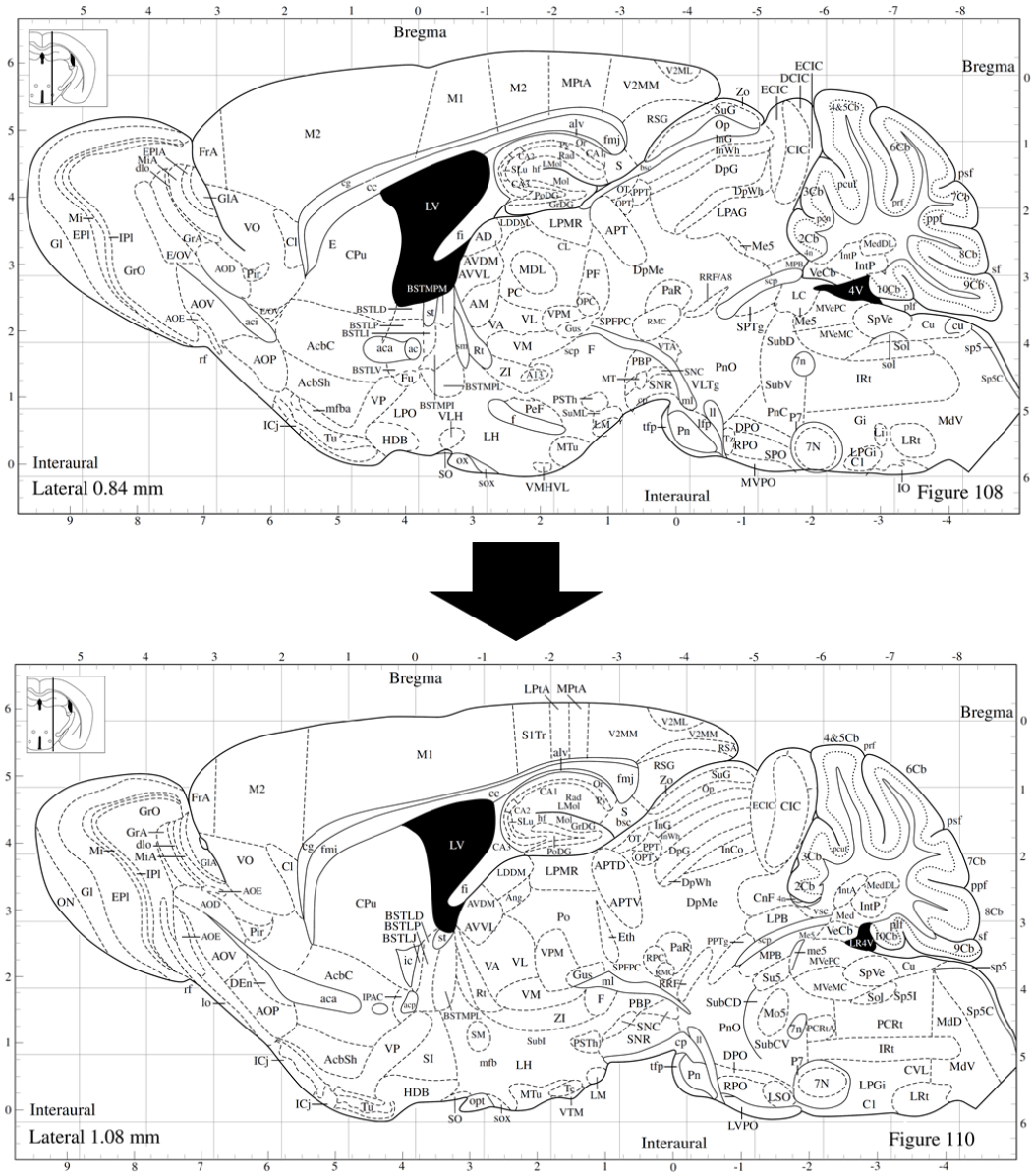


Figure 3.2 – Sagittal sections from paraffin embedded mouse brain right hemispheres were prepared from 0.84 to 1.08 mm from bregma.
Figures taken from Paxinos and Franklin (2001).

3.2.2. Deparaffinisation

Deparaffinisation of the brain sections was achieved using xylene (Fisher Scientific), followed by 70% ethanol (industrial methylated spirit, Fisher Scientific) and deionised water for rehydration of the sections.

3.2.3. Immunohistochemistry

3.2.3.1. Antigen retrieval and blocking

Heat induced antigen retrieval was performed in an antigen retrieval device (PT Module, Thermo Scientific) containing citrate buffer (dilution 1:100). Samples were blocked using normal goat serum (Vector labs, catalogue number S-1000).

3.2.3.2. Antigen-antibody labelling

To assess tau pathology, we used mouse monoclonal PG-5 (provided by Peter Davies from Albert Einstein College of Medicine, Bronx, NY, USA) (Jicha et al., 1999) as the primary antibody (diluted 1:8000), which recognises tau phosphorylated at Ser409 (p-tau). We used biotinylated goat anti-mouse IgG (Vector labs, catalog number BA-9200, lot number 2B0324) as the secondary antibody (diluted 1:200), as previously described (Ahmed et al., 2014). A sample with known tau pathology was used as a positive control, and one slide containing no primary antibody and one slide containing no secondary antibody were included as negative controls.

To assess amyloid pathology, we used mouse monoclonal biotinylated 3D6 (b3D6, provided by Eli Lilly (Demattos et al., 2012), diluted 1:1000), which binds both soluble and insoluble $A\beta_{1-42}$, specifically targeting the amino acids 1-5 in Aβ. A sample with known amyloid pathology was used as a positive control, and a slide containing no biotinylated antibody was included as a negative control.

All samples for each mouse model (rTg4510: n = 80 samples; J20: n = 79 samples) were immunostained (two sections in each glass slide) in a single batch in an autostainer (Autostainer 720 for PG-5 and 720N for b3D6, Thermo Scientific).

3.2.3.3. Detection

For the detection of antibody staining we undertook enzymatic labelling using peroxidase (Vectastain Elite ABC HRP Reagent, Vector Laboratories) and DAB substrate (Vector Laboratories).

3.2.3.4. Counterstain

Tissues were counterstained with haematoxylin to provide contrast and dehydrated with ethanol and xylene to prepare for mounting.

3.2.4. Mounting and image capture

Samples were mounted in an automated coverslipper (ClearVue Coverslipper from Thermo Scientific) using ClearVue mountant. Images were digitised with Scanscope AT slide scanner (Aperio) at 20x magnification, and saved in *svs* (ScanScope Virtual Slide) format on an Eli Lilly network server. Images were subsequently accessed using the eSlide Manager system (Aperio).

3.2.5. Visualisation and quantification

Visualisation of the digitised tissue sections stored in the network server and delineation of the anatomical regions of the brain of interest (or regions of interest (ROI)) was achieved using the software *ImageScope* (version 12.2.1.5005; Aperio). ROI were selected to reflect the known progression of amyloid and tau pathology in the two models and to overlap with previous and ongoing studies at Eli Lilly in the same mouse models.

For rTg4510, the ROI delineated and quantified were whole hippocampus (subdivided into CA1, CA3 and dentate gyrus (DG)), whole cortex (subdivided into secondary motor cortex (M2), primary motor cortex (M1), visual cortex, and retrosplenial cortex), and thalamus, which is expected to exhibit much lower tau pathology that is confined to late stages (**Figure 3.3**).

In the J20 mice, the ROI delineated and quantified were whole hippocampus, whole cortex (further subdivided into rostral cortex and caudal cortex), and thalamus, which is not expected to exhibit amyloid pathology (**Figure 3.4**).

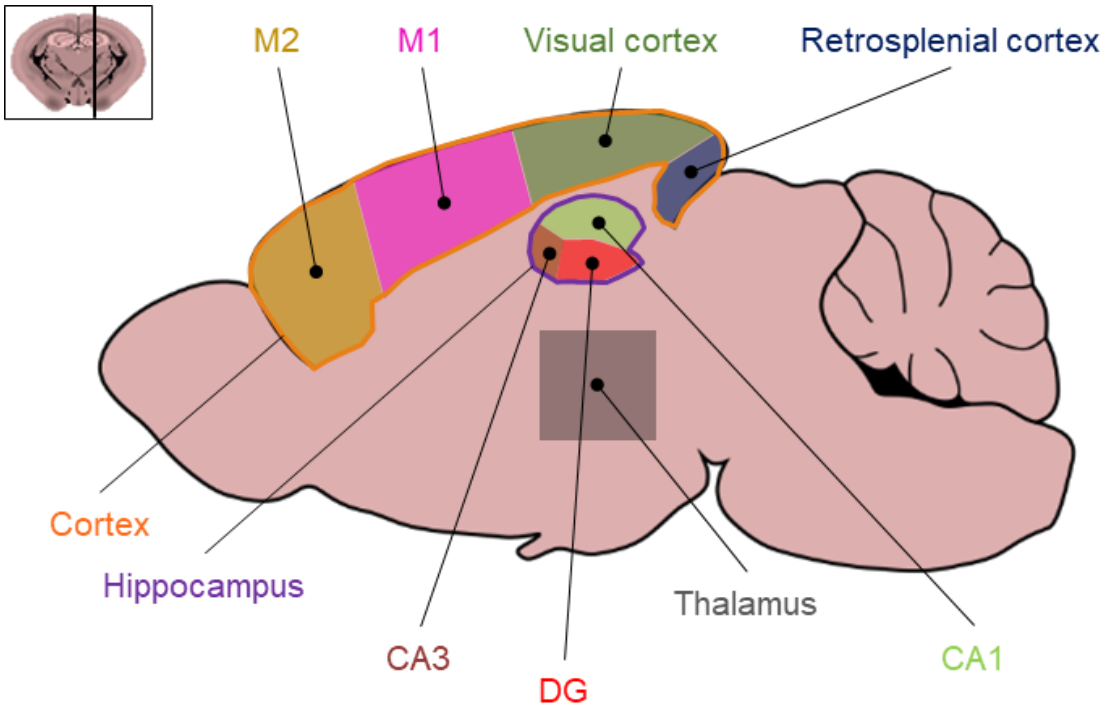


Figure 3.3 – Anatomical regions of the brain examined in rTg4510 mice.
Brain regions tested for tau pathology: whole hippocampus (purple line), CA1 sub-region of the hippocampus (light green area within the purple line), CA3 sub-region of the hippocampus (brown area within the purple line), dentate gyrus (DG) sub-region of the hippocampus (light red area within the purple line), whole cortex (orange line), secondary motor cortex (M2, dark yellow area within the orange line), primary motor cortex (M1, magenta area within the orange line), visual cortex (dark green area within the orange line), retrosplenial cortex (dark blue are within the orange line), thalamus (dark grey square).

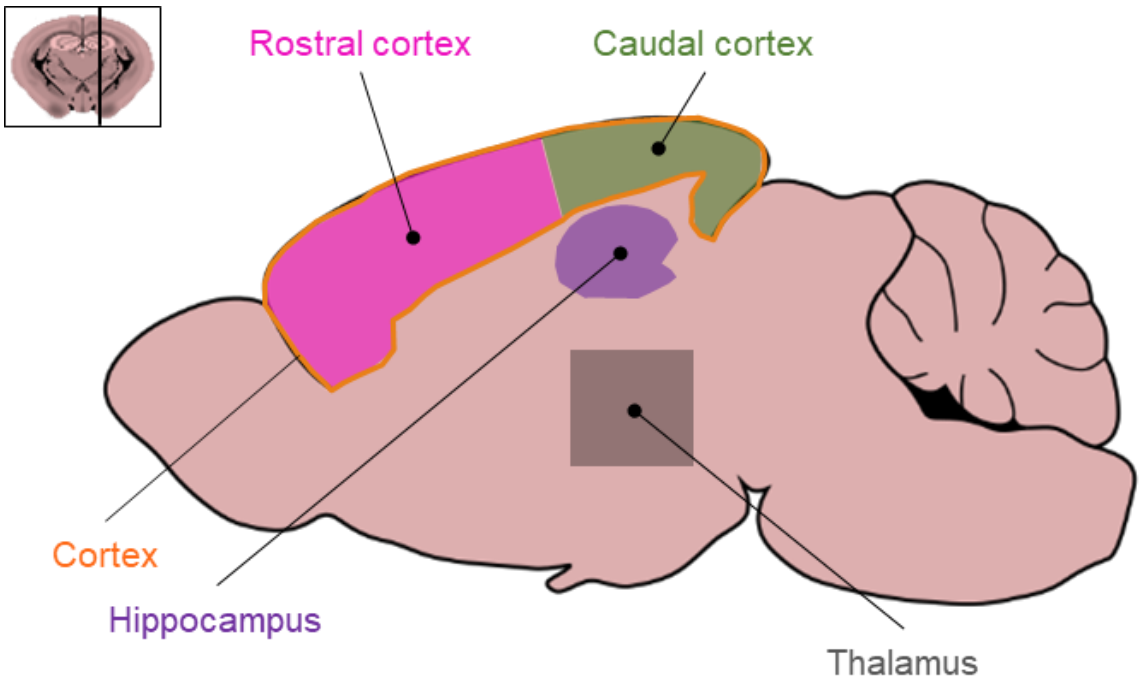


Figure 3.4 – Anatomical regions of the brain examined in J20 mice.
Brain regions tested for amyloid pathology: whole hippocampus (purple area), whole cortex (orange line), rostral cortex (magenta area within the orange line), caudal cortex (green area within the orange line), thalamus (dark grey square).

Positivity was quantified automatically using the Aperio Positive Pixel Count Algorithm (Aperio), applied to the images using *ImageScope* (Aperio). The algorithm quantifies the amount of a specific stain present in a scanned slide image, based on user specification of colour (across a range of hues and saturation). Specifically, the number and intensity of pixels that fall within the parameters are quantified, which can be classified into intensity ranges (weak, positive, and strong) as defined by the user. For pixels which satisfy the colour specification, the algorithm counts the number and intensity-sum in each intensity range, along with three additional quantities: average intensity, ratio of strong/total number, and average intensity of weak and positive pixels. Pixels which are stained, but do not fall into the positive-colour specification, are considered negative stained pixels and contribute to determine the fraction of positive to total stained pixels. The values for the algorithm input parameters (**Table 3.1** and

Table 3.2), calibrated to ignore non-specific staining, were based on optimisation experiments and previous/ongoing studies conducted at Eli Lilly using the same stains.

Table 3.1 – Algorithm inputs for PG-5 (tau pathology).

Version	9.1
View Width (Pixels)	1000
View Height (Pixels)	1000
Overlap Size	0
Image Zoom	1
Classifier	None
Classifier Neighbourhood	0
Pixel Area (mm^2)	2.46E-07
Hue Value (Centre)	0.1
Hue Width	0.1
Colour Saturation Threshold	0.3
Intensity Threshold WEAK (Upper Limit)	255
Intensity Threshold WEAK (Lower Limit)	175
Intensity Threshold MEDIUM (Upper Limit)	175
Intensity Threshold MEDIUM (Lower Limit)	100
Intensity Threshold STRONG (Upper Limit)	100
Intensity Threshold STRONG (Lower Limit)	0
Intensity Threshold Negative Pixels	-1

Table 3.2 – Algorithm inputs for b3D6 (amyloid pathology).

Version	9.1
View Width (Pixels)	1000
View Height (Pixels)	1000
Overlap Size	0
Image Zoom	1
Classifier	None
Classifier Neighbourhood	0
Pixel Area (mm^2)	2.46E-07
Hue Value (Centre)	0.1
Hue Width	0.1
Colour Saturation Threshold	0.3
Intensity Threshold WEAK (Upper Limit)	240
Intensity Threshold WEAK (Lower Limit)	175
Intensity Threshold MEDIUM (Upper Limit)	175
Intensity Threshold MEDIUM (Lower Limit)	100
Intensity Threshold STRONG (Upper Limit)	100
Intensity Threshold STRONG (Lower Limit)	0
Intensity Threshold Negative Pixels	-1

Hue Value is a metric of colour defined from 0-1 (e.g Blue=0.66). Hue Width is the range around the Hue Value which will also satisfy positive quantification. Colour Saturation Threshold is a metric of how much of the positive colour is present within the tested pixel; defined between 0-1, with 1 corresponding to 100% saturation of the pixel with the positive colour. Pixels with saturation under the threshold are not counted as positive. All of the Intensity Threshold values are defined using the same units, which is a measurement of the brightness of the pixels using the formula R+G+B/3. It has a possible range of 0-255 where 0=completely black and 255=bright white. All the other parameters relate to how the algorithm runs on the analysis server and less to what is quantified; these were kept as default.

A representative image showing how the algorithm distinguished tau and amyloid pathology in my samples is shown in **Figure 3.5** and **Figure 3.6**, respectively. The burden of tau or amyloid pathology was expressed as percentage of positive pixels.

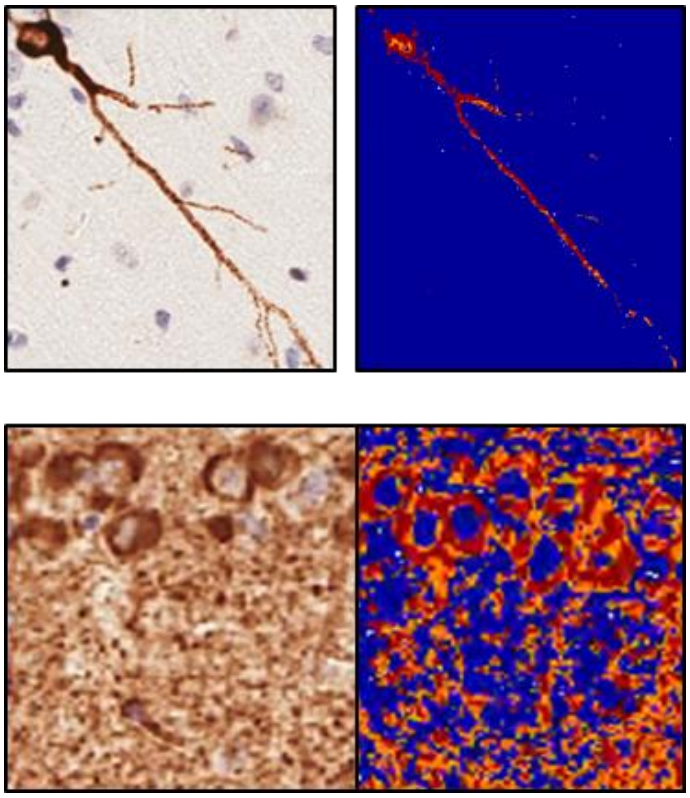


Figure 3.5 – Representative immunohistochemistry images from the hippocampus (CA1 sub-region) of a rTg4510 transgenic mouse, showing how the positive pixel count algorithm recognised the positive stain for quantification of tau pathology.

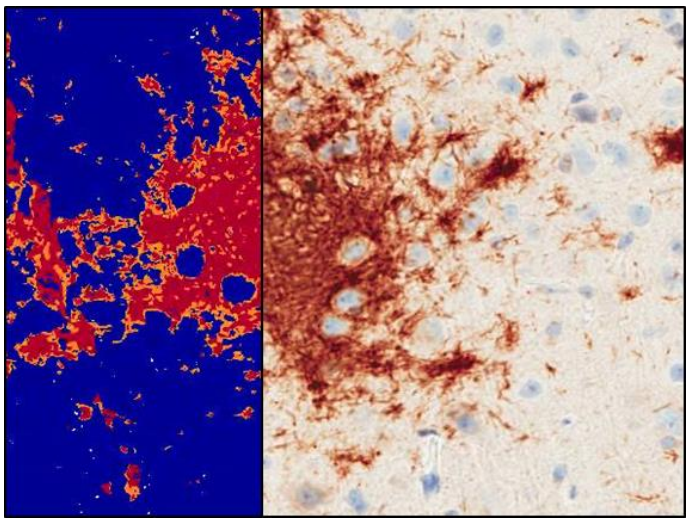


Figure 3.6 – Representative immunohistochemistry image from the hippocampus of a J20 transgenic mouse, showing how the positive pixel count algorithm recognised the positive stain for quantification of tau pathology.

3.2.6. Statistical analysis

Statistical analysis (two-way factorial ANOVA) was performed using Microsoft Excel 2013. The following statistical model was used:

$$Pathology = Genotype + Age + Genotype * Age$$

Plots were produced in R (version 3.4.3).

3.3. Results

3.3.1. Tau pathology in rTg4510 mice

In rTg4510 mice we measured levels of phosphorylated tau (p-tau, using the antibody PG-5) at ages 2, 4, 6 and 8 months comparing them to WT controls at the same ages (n = 7-10 animals per group, total n = 74). This paralleled the collection of genomic data (gene expression and DNA methylation) on the same animals. In **Figure 3.7** representative pictures for one sample from each rTg4510 group are shown, demonstrating the clear accumulation of tau pathology in the hippocampus and cortex of rTg4510 TG mice, with tau pathology absent in WT mice at the same ages. Images depicting the negative and positive controls included in the experiment (see **Section 3.2.3.4**) are also shown. Tau accumulation in the hippocampus of rTg4510 TG mice, specifically the CA1 hippocampal sub-region, is shown in detail in **Figure 3.8**. Overall, I observed intracellular accumulation of phosphorylated tau in neurons at the initial stages (upper panel in **Figure 3.8**), and an overall distribution of increased levels of phosphorylated tau in affected brain tissues in later stages (bottom panel in **Figure 3.8**). Following the quantification of the images, we detected a dramatic accumulation of tau pathology in the hippocampus and each of the cortex regions tested from rTg4510 TG mice (**Table 3.3** and **Figure 3.9**), reflecting findings from previous studies (Ramsden et al., 2005, Sahara et al., 2014, Santacruz et al., 2005). We also quantified neuropathology in the thalamus (**Figure 3.9k**), which is relatively protected from aggressive tau pathology. As expected, this region showed negligible levels of tau pathology relative to the other brain regions tested; despite these low levels, we observed statistically significant increases in TG mice indicating that there is some accumulation of phosphorylated tau in this brain region in later stages (**Table 3.3**).

Table 3.3 – Statistical results (two-way ANOVA) for PG-5 quantification in tested brain regions from rTg4510 mice to assess progression of tau pathology.

Brain region	N	Genotype*Age	
		F-value	P-value
Whole hippocampus	73	F(3,66) = 69.76	1.96E-20
CA1	74	F(3,67) = 68.86	1.96E-20
CA3	73	F(3,66) = 64.22	1.51E-19
DG	73	F(3,66) = 67.42	4.58E-20
Whole cortex	72	F(3,65) = 53.33	1.65E-17
M2	74	F(3,67) = 44.90	5.01E-16
M1	73	F(3,66) = 29.28	3.74E-12
Visual cortex	73	F(3,66) = 55.22	5.64E-18
Retrosplenial cortex	73	F(3,66) = 38.38	1.79E-14
Thalamus	74	F(3,67) = 111.48	5.41E-26

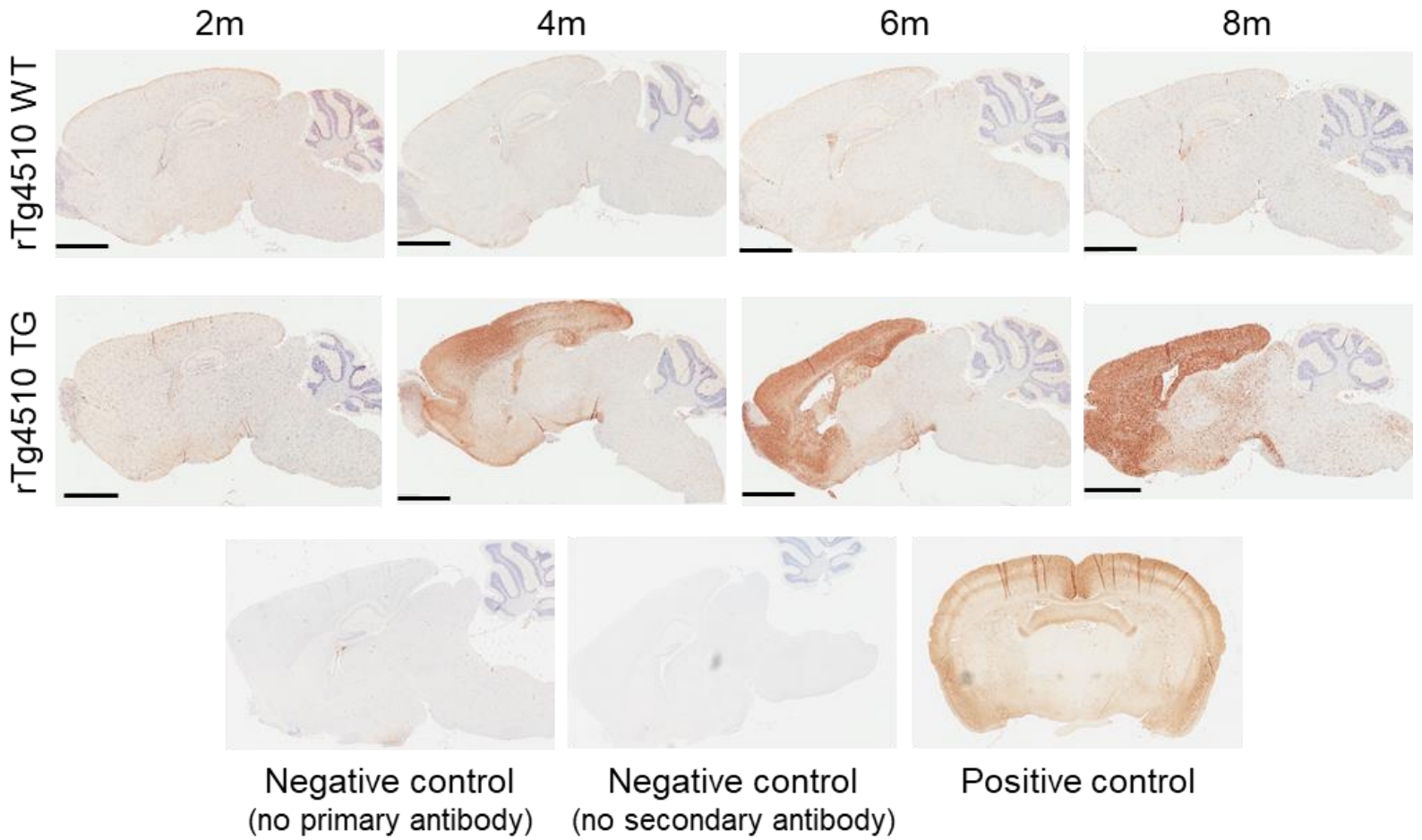


Figure 3.7 – Representative immunohistochemistry images showing the dramatic accumulation of tau pathology in the hippocampus and cortex of rTg4510 mice.

Sagittal sections showing the dramatic accumulation of tau pathology (PG-5) in the brains of rTg4510 transgenic (TG) mice (second row) compared to wild-type control (WT) mice (top) at 2, 4, 6 and 8 months of age. Primary antibody: PG-5 (1:8000); secondary antibody: biotinylated goat anti-mouse (1:200). Black bar = 2 mm.

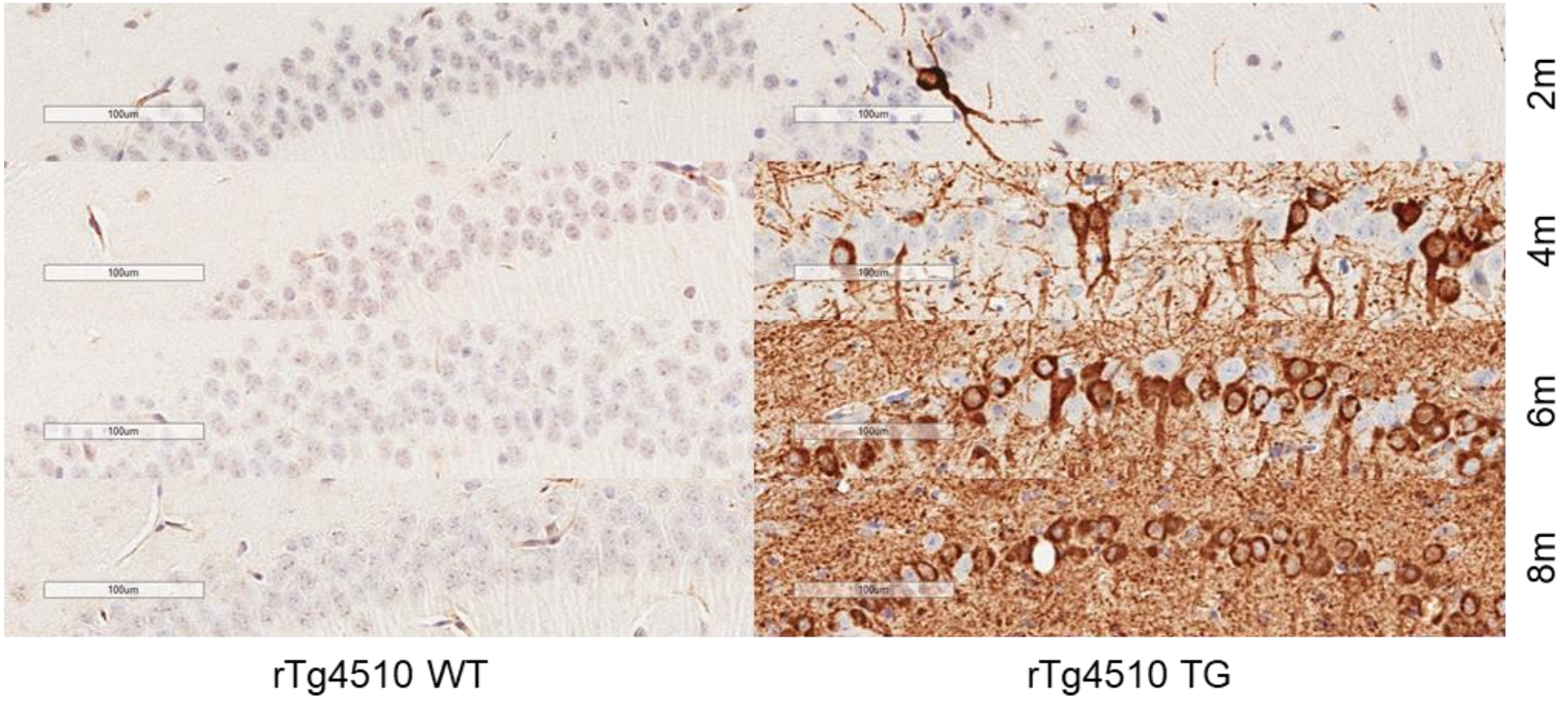


Figure 3.8 – Representative immunohistochemistry images showing the dramatic accumulation of tau pathology in the hippocampus (CA1) of rTg4510 transgenic mice.

WT: wild type. TG: transgenic. Primary antibody: PG-5 (1:8000); secondary antibody: biotinylated goat anti-mouse (1:200). Bar = 100 µm.

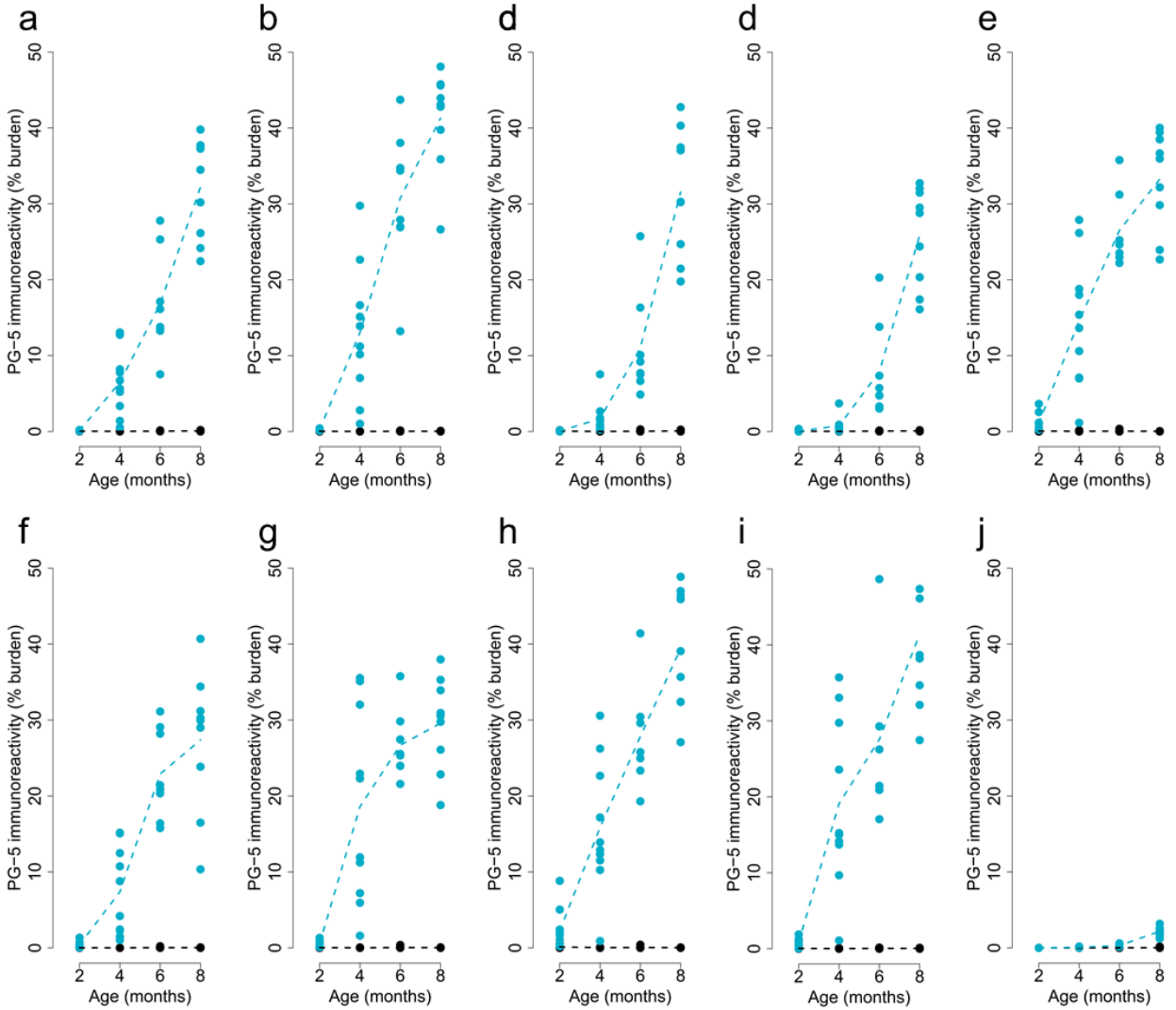


Figure 3.9 – Transgenic mice expressing mutant human MAPT exhibit progressive neuropathology across the hippocampus and cortex.
Quantification of PG-5 immunoreactivity for each of the tested brain regions: **(a)** whole hippocampus, **(b)** CA1 **(c)** CA3, **(d)** DG, **(e)** whole cortex, **(f)** M2, **(g)** M1, **(h)** visual cortex, **(i)** retrosplenial cortex, **(j)** thalamus. Dashed lines represent mean paths of pathological burden across the four age groups. n = 7-10 animals per group.

3.3.2. Amyloid pathology in J20 mice

We quantified levels of amyloid (using the antibody b3D6) in J20 TG mice at ages 6, 8, 10 and 12 months comparing them to WT controls at the same ages (n = 9-10 animals per group, total n = 73). **Figure 3.10** shows representative images for one sample from each J20 group, demonstrating the accumulation of amyloid pathology in the hippocampus and cortex of J20 TG mice, with no pathology in WT mice. Pictures of the negative and positive controls included in the experiment (see **Section 3.2.3.4**) are also shown. Amyloid accumulation in the hippocampus of J20 TG mice is shown in detail in **Figure 3.11**.

Following quantification of the images, we identified dramatic increases in amyloid pathology in the hippocampus, with a highly significant accumulation of amyloid also observed in each of the cortical regions tested (**Table 3.4** and **Figure 3.12**), reflecting the results from other studies of this model (Harris et al., 2010). No amyloid pathology was detected in thalamus in either TG or WT mice.

Table 3.4 – Statistical results (two-way ANOVA) for b3D6 quantification in tested brain regions from J20 mice to assess progression of amyloid pathology.

Brain region	N	Genotype*Age	
		F-value	P-value
Hippocampus	77	F(3,68) = 66.85	3.00E-20
Whole cortex	73	F(3,66) = 24.75	7.63E-11
Rostral cortex	73	F(3,66) = 10.63	8.65E-06
Caudal cortex	74	F(3,67) = 28.62	5.02E-12
Thalamus	73	F(3,66) = 1.	0.20

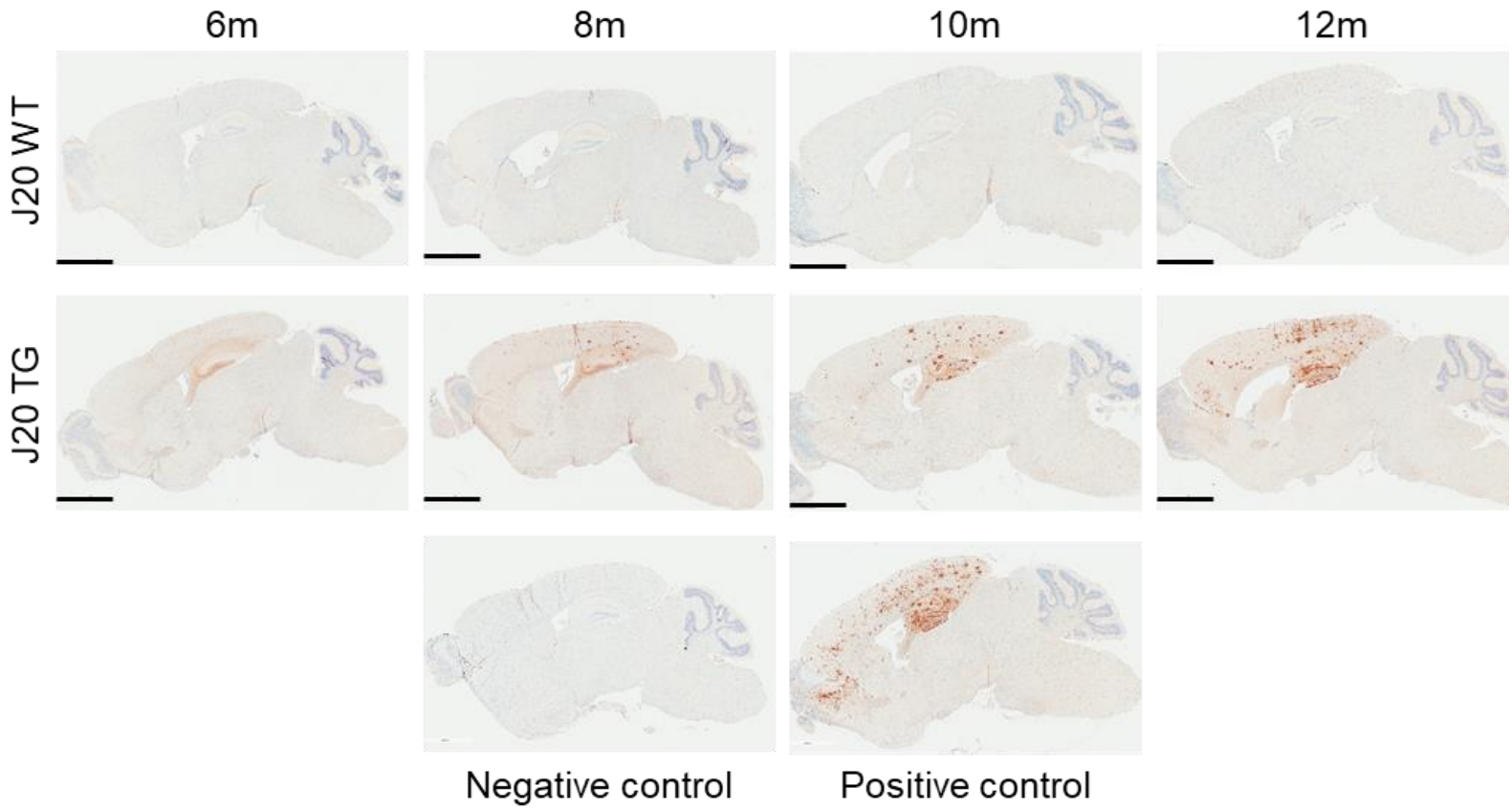


Figure 3.10 – Representative immunohistochemistry images showing accumulation of amyloid pathology in the hippocampus and cortex of J20 TG mice.

Sagittal sections showing the dramatic accumulation of amyloid pathology (b3D6) in the brains of a J20 transgenic (TG) mouse (bottom) compared to a wild-type control (WT) mouse (top) at 6, 8, 10 and 12 months of age. Primary antibody: mouse monoclonal biotinylated 3D6 (1:1000); no secondary antibody. Black bar = 2 mm.

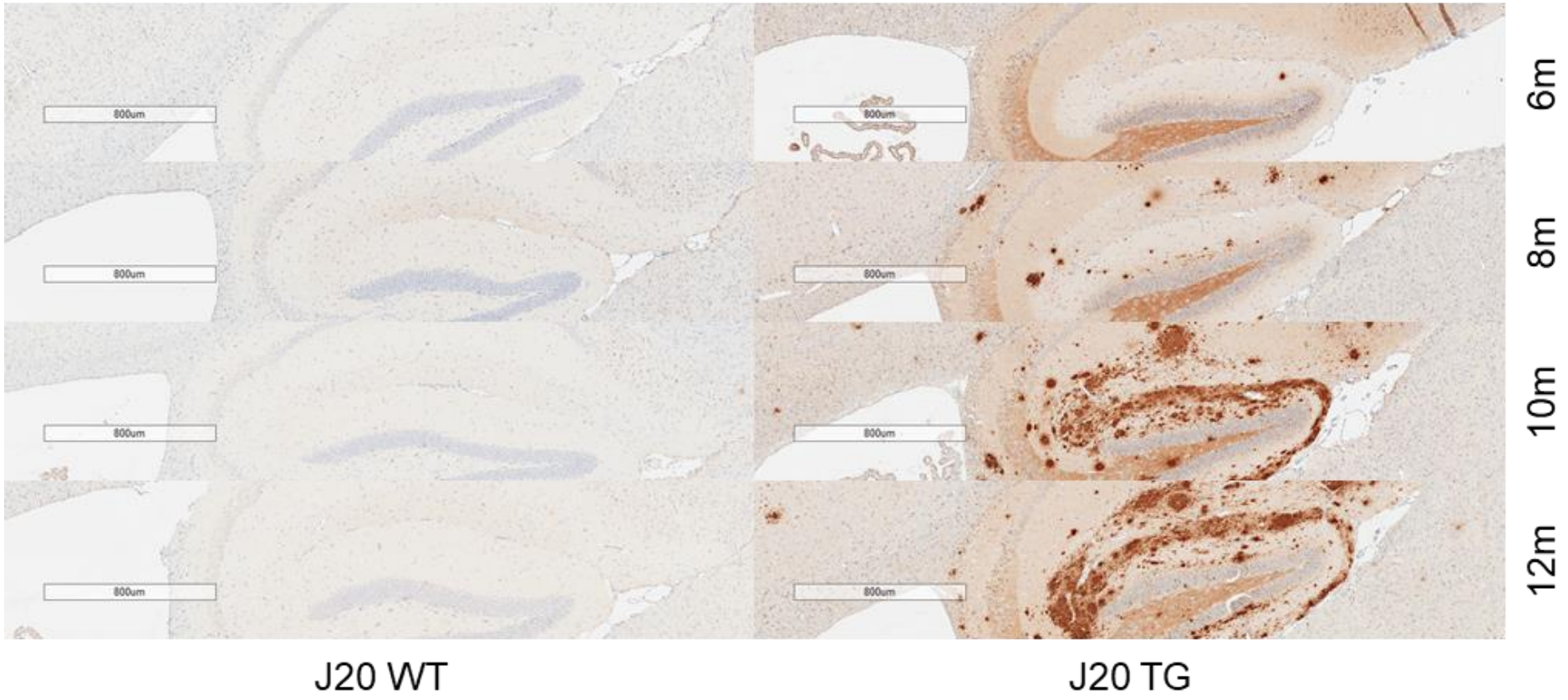


Figure 3.11 – Representative immunohistochemistry images showing accumulation of amyloid pathology in the hippocampus of J20 transgenic mice.

WT: wild type. TG: transgenic. Primary antibody: mouse monoclonal biotinylated 3D6 (1:1000); no secondary antibody. Bar = 800µm.

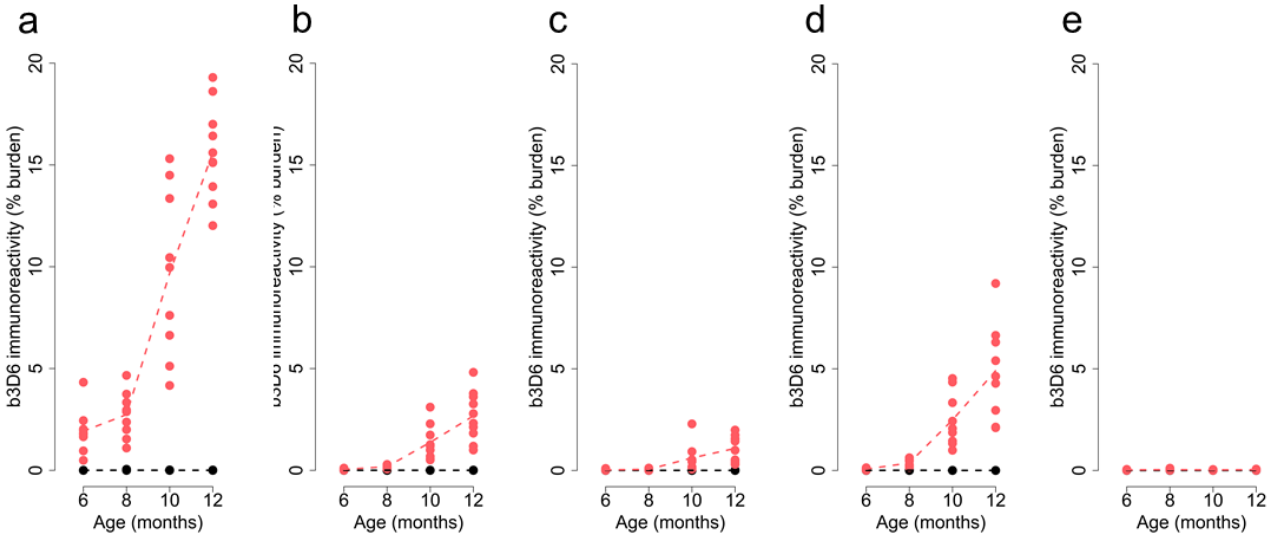


Figure 3.12 – Transgenic mice expressing mutant human APP exhibit progressive neuropathology across the hippocampus and cortex.
Quantification of b3D6 immunoreactivity for each of the tested brain regions: **(a)** hippocampus, **(b)** whole cortex, **(c)** rostral cortex, **(d)** caudal cortex, **(e)** thalamus. Dashed lines represent mean paths of pathological burden across the four age groups. n = 8-10 animals per group.

3.4. Discussion

In this Chapter, I identified progressive accumulation of tau and amyloid neuropathology in brain regions from transgenic mice overexpressing mutated human *MAPT* (rTg4510 mouse model) and human *APP* (J20 mouse model).

I measured levels of phosphorylated tau (using the antibody PG-5) in rTg4510 mice at 2, 4, 6 and 8 months, comparing them to WT controls at the same ages. I observed highly significant increases in phosphorylated tau in the hippocampus and cortex of rTg4510 TG mice (including in specific hippocampal sub-regions and each of the cortical regions that I quantified), reflecting data from several parallel cohorts of mice from our rTg4510 colony profiled in previous and ongoing studies at Eli Lilly using antibodies beyond PG-5 (Blackmore et al., 2017, Wang et al., 2018). My findings were also in accordance with those from previous studies by others, including the first studies using this mouse model, which reported progressing NFT pathology with age in rTg4510 TG mice, starting in the neocortex and progressively affecting the hippocampus and limbic structures (Ramsden et al., 2005, Sahara et al., 2014, Santacruz et al., 2005). The spread of tau pathology in rTg4510 mice therefore reflects the spread of NFTs with increasing Braak stage in AD (Braak et al., 2006, Braak and Braak, 1991) (see **Section 1.1** and **Figure 1.2**). Of note, in rTg4510 TG mice we also detected significant differences in the thalamus, particularly in late stages. The magnitude of these differences was much smaller, reflecting the later and less aggressive accumulation of tau in this region of the brain.

I also quantified levels of amyloid pathology (using the antibody b3D6) in J20 mice at ages 6, 8, 10 and 12 months, comparing them to WT controls at the same ages. Again, I identified dramatic increases in pathology in the hippocampus, with a highly significant accumulation of amyloid also observed in each of the cortical regions examined. These results concur with previous data highlighting progressive deposition of amyloid plaques in the hippocampus and neocortex of J20 mice at 5-7 months, and ubiquitous plaque pathology by 8-10 months of age (Harris et al., 2010, Mucke et al., 2000). The diffusion of amyloid pathology in J20 mice is therefore similar to the progressive deposition of amyloid seen in humans with AD (Thal et al., 2002) (see **Section 1.1** and **Figure 1.2**).

In addition to confirming dramatic and progressive accumulation of pathology in the rTg4510 and J20 mouse models, the results from these experiments provide accurate pathology burden data for each individual animal used for genomic profiling. This means that pathological burden can be related directly to the transcriptomic (**Chapter 4**) and methylomic data (**Chapter 5** and **Chapter 6**) generated on each individual animal.

Despite the potential challenges inherent when performing experiments using antibodies, for which immunohistochemistry-based assays are no exception (e.g. non-specific binding), the antibodies used in this chapter (PG-5 and 3D6) have been extensively validated (Blackmore et al., 2017, Demattos et al., 2012, Holmes et al., 2016). I carried out the immunocytochemistry experiments described in this chapter at Eli Lilly, where these stains are routinely undertaken. At Eli Lilly, my collaborators have assessed a variety of antibodies to stain tau and Aβ and optimised the conditions for them. Using paraffin embedded tissue, PG-5 and 3D6 showed the best results at staining tau inclusions and amyloid plaques, respectively.

PG-5 labels p-tau species that are mainly present in NFTs, therefore being a good measure of aggregated tau burden (Blackmore et al., 2017, Holmes et al., 2016). There are many other antibodies available that can also be used to stain for p-tau in tissue sections, but most are not as good at detecting aggregates or NFTs as PG-5 or result in high residual background staining. In general, other antibodies available, such as AT-8 (stains p-tau, phosphorylate at Ser202 and Thr205) and MC-1 (stains aggregated tau), show similar patterns of pathology as PG-5, but often the staining is fainter, and the NFT-like inclusions are not as clearly observed, with increasing antibody concentrations often resulting in high levels of background staining across the whole section. Of note, this is obviously dependent from other experimental conditions, such as the use of free floating or frozen cryostat sections, instead of paraffin-embedded tissue sections. Furthermore, my collaborators at Eli Lilly compared the results of PG-5 with MC-1 and AT-8, as well as with the Gallyas method (a silver impregnation technique highly sensitivity for NFTs (Kuninaka et al., 2015)), observing that PG-5 correlated well with all the other stains. Interestingly, PG-5 staining has been

shown to be reduced when rTg4510 TG mice were treated with doxycycline to switch off the transgene expression (Blackmore et al., 2017, Holmes et al., 2016).

The antibody 3D6 binds to both soluble and insoluble Aβ at the extreme amino terminus ($Aβ_{1-5}$). Interestingly, if the first amino acid is missing, the antibody affinity to Aβ drops dramatically, i.e., it is amino-terminal specific for full-length Aβ peptide. Other studies from our collaborators at Eli Lilly have compared 3D6 to other antibodies for amyloid staining, such as mE8, observing similar staining between 3D6 and mE8, with the labelling being more intense and widespread for the 3D6 antibody (Demattos et al., 2012).

Finally, it is worth mentioning that soluble tau/ Aβ species were not assessed in this study, which have been hypothesised to have a major impact in AD-associated toxicity (Koss et al., 2016). Other approaches, such as western blot and mass spectrometry to evaluate soluble toxic species would be useful complimentary experiments for future studies.

Chapter 4. Transcriptional signatures of progressive neuropathology in transgenic tau and amyloid mouse models

This chapter represents a modified version of the primary manuscript arising from my thesis, currently available as a preprint and under review for publication (Castanho et al., 2019).

For information about data availability and accessibility to supplementary tables see **Supplementary Information** (**page 399**).

4.1. Introduction

In this chapter I describe my analysis of transcriptional changes in the entorhinal cortex associated with the development of both tau and amyloid pathology. Using highly-parallel RNA sequencing (RNA-seq), I profiled transcriptional variation in rTg4510 and J20 mice, identifying changes in gene expression paralleling the development of tau or amyloid pathology measured using the immunohistochemistry data generated on the same individual mice (**Chapter 3**). Finally, I used systems-level analyses to identify gene co-expression networks associated with the progressive accumulation of tau or amyloid pathology and explored overlaps between these networks and AD-associated co-expression modules identified in the human cortex.

Recent studies have identified widespread gene expression differences in transgenic mice harbouring a diverse range of mutations associated with aspects of AD pathology (Castillo et al., 2017, Landel et al., 2014, Matarin et al., 2015, Rothman et al., 2018, Swarup et al., 2019, Wes et al., 2014) (discussed in more detail in **Section 1.5.2.1** in **Chapter 1**). However, most analyses to date have been undertaken on relatively small numbers of animals and have not attempted to directly relate transcriptional alterations to the progressive burden of pathology in the same mice.

In recent years, highly-parallel RNA sequencing (RNA-seq) has replaced gene expression arrays as the method of choice for transcriptional profiling. RNA-seq for strand-specific transcript sequencing was first described by Lister et al. (2008), and first reported in mammals, specifically mouse tissues, by Mortazavi et al. (2008). In addition to enabling the quantification of gene expression levels, RNA-seq allows profiling and discovery of new transcripts – an obvious advantage compared to microarrays which are limited to pre-existing probes. Furthermore, if sufficient sequencing depth is achieved, the assessment of alternative splicing is also possible from RNA-seq data (Mortazavi et al., 2008). RNA-seq library preparation involves the initial conversion of RNA into complementary DNA (cDNA), followed by addition of adapters that prepare the libraries for high-throughput sequencing (**Figure 4.1**). Illumina sequencers, such as the HiSeq 2500 System used in my research, are powerful high-throughput sequencing systems that use *Sequencing by Synthesis* technology, with fluorescently-labelled nucleotides to detect single bases as they are incorporated into growing DNA strands (**Figure 4.2**). The fragments are attached to the surface of a flow cell and amplified. The DNA molecule is sequenced by adding polymerase and fluorescently-labelled reversible terminator nucleotides (with each base having a different colour associated to it). A nucleotide is incorporated in each cycle, which is detected by fluorescence. After each cycle, the fluorescent label is removed from the terminator and the 3'-OH is then free for a new base to be incorporated. The resulting RNA-seq reads can then be individually mapped to a reference genome and quantified in order to obtain the number of reads corresponding to mRNA from each known exon, splice event or new candidate gene, which can be compared between samples (Mortazavi et al., 2008).

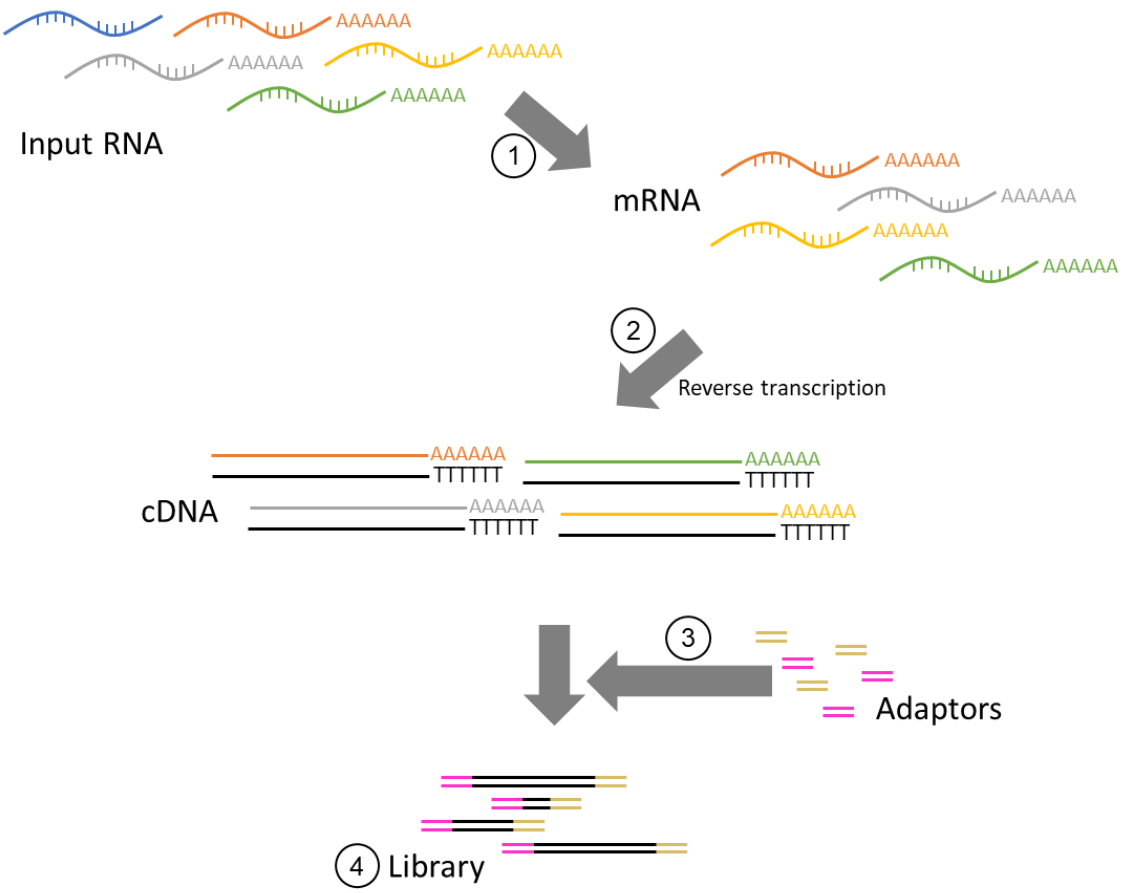


Figure 4.1 – Schematic representation of a typical RNA-seq library preparation protocol.
Messenger RNA (mRNA)-specific library preparation is shown. (1) Poly(A) selection. (2) Conversion of mRNA into complementary DNA (cDNA) by reverse transcription. (3) Adaptor ligation. (4) The final library ready for sequencing. RNA fragmentation

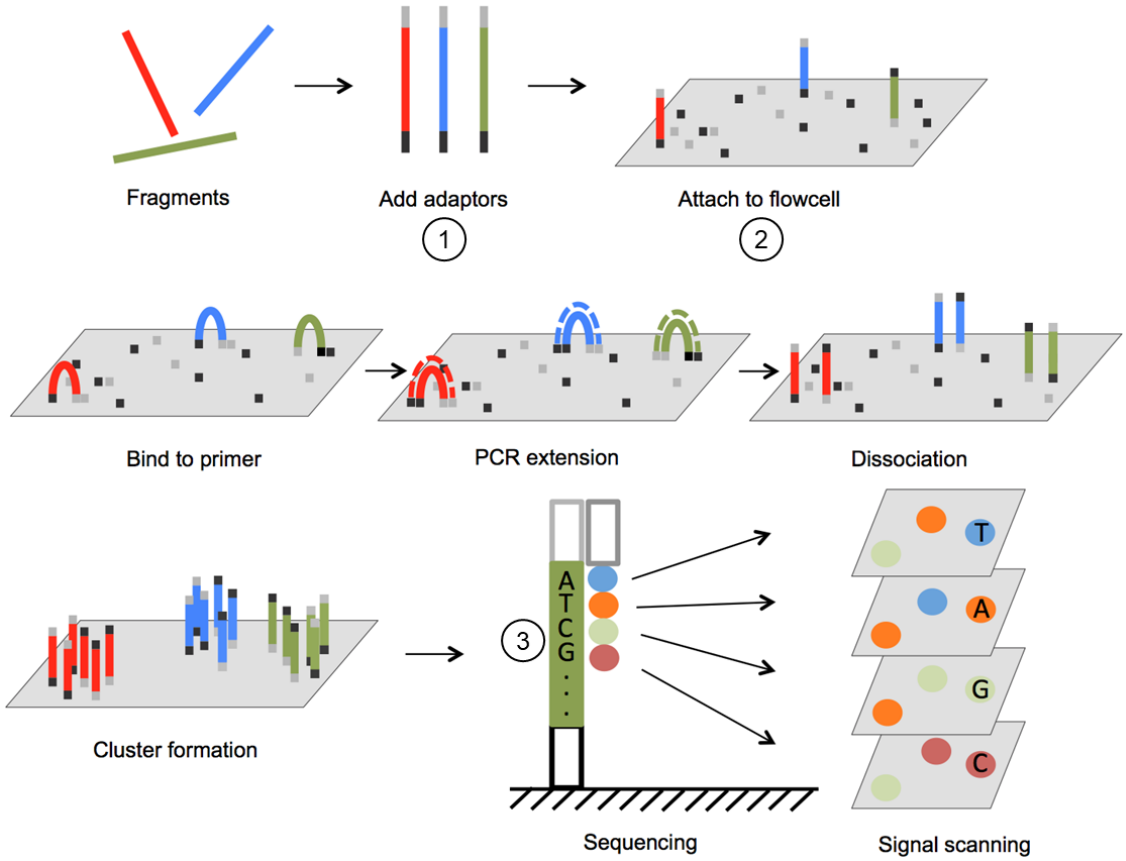


Figure 4.2 – Outline of Illumina sequencing by synthesis.
(1) Adaptors are annealed to the ends of sequence fragments. (2) Fragments bind to primer-loaded flow cell and bridge PCR reactions amplify each bound fragment to produce clusters of fragments. (3) During each sequencing cycle, one fluorophore attached nucleotide is added to the growing strands. A laser excites the fluorophores in each fragment being sequenced and an optic scanner collects the signals from each fragment cluster. Finally, the sequencing terminator is removed, and the next sequencing cycle starts. Figure and legend adapted from Lu et al. (2015).

4.2. Methods

In this section, the experimental procedures used for the evaluation of genome-wide transcriptional profiles using RNA-seq in entorhinal cortex (ECX, described in **Chapter 2**, **Section 2.4** and **Section 2.5**) RNA samples from rTg4510 and J20 mice (**Chapter 2**, **Section 2.1**) are described in detail. An overview of the laboratory and analytical procedures performed is given in **Figure 4.3**.

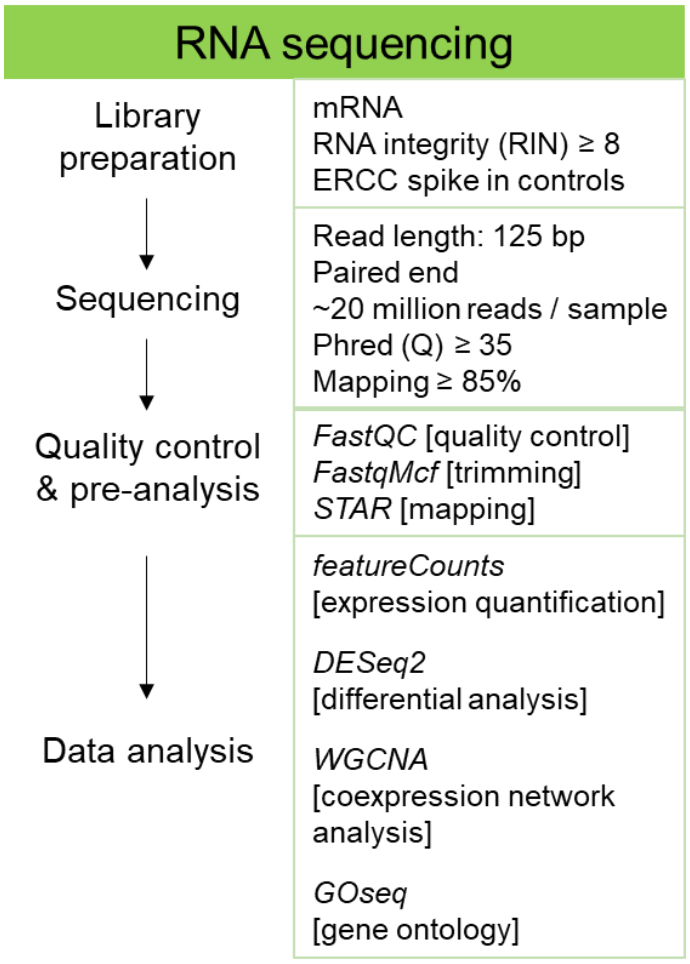


Figure 4.3 – Overview of RNA-seq laboratory and analytical procedures used for transcriptional profiling of entorhinal cortex RNA samples from rTg4510 and J20 mice.

Samples were separated into two batches, with the two mouse models processed and sequenced independently – 64 samples were profiled in each batch (64 samples for the rTg4510 model, and 63 samples plus a negative control for the J20 model). Experiments were performed in a blinded and randomised fashion to avoid any biases in quantification and analysis.

When I first started the analysis of my data, there was no standard protocol described for identifying progressive changes in gene expression whilst controlling for changes due to the presence of the transgenes (i.e., genotype-associated differences). Therefore, before obtaining the final results, I optimised the analysis of my RNA-seq data by testing several approaches, including more than one analytical protocol (or "pipeline"), before selecting the most appropriate statistical strategy to address our biological questions. **Section 4.2.4** and **Section 4.2.5** cover both analytical pipelines and the respective bioinformatics software (or "packages) that I tested; the final packages used are shown in the bottom section of **Figure 4.3**.

4.2.1. Assessment of RNA quality and concentration

All RNA samples used for my RNA-seq experiments were quantified and checked for purity using a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific), as described in **Section 2.5.3.1** (**Chapter 2**). Before RNA-seq library preparation, RNA quality and quantity for all samples was checked using RNA ScreenTape (Agilent), as described in **Section 2.5.3.3** (**Chapter 2**).

4.2.2. Library preparation

Stranded-specific mRNA sequencing libraries were prepared using the TruSeq Stranded mRNA Sample Prep Kit (Illumina) using the Bravo Automated Liquid Handling Platform (Agilent). An overview of the library prep workflow using this kit is given in **Figure 4.4**. I performed the experiments, with technical assistance from Audrey Farbos and supervision by Dr Karen Moore of the Sequencing Service at the University of Exeter. There was no need for optimisations as the

TruSeq Stranded mRNA Sample Prep Kit (Illumina) has been extensively validated and routinely used at the Exeter Sequencing Service.

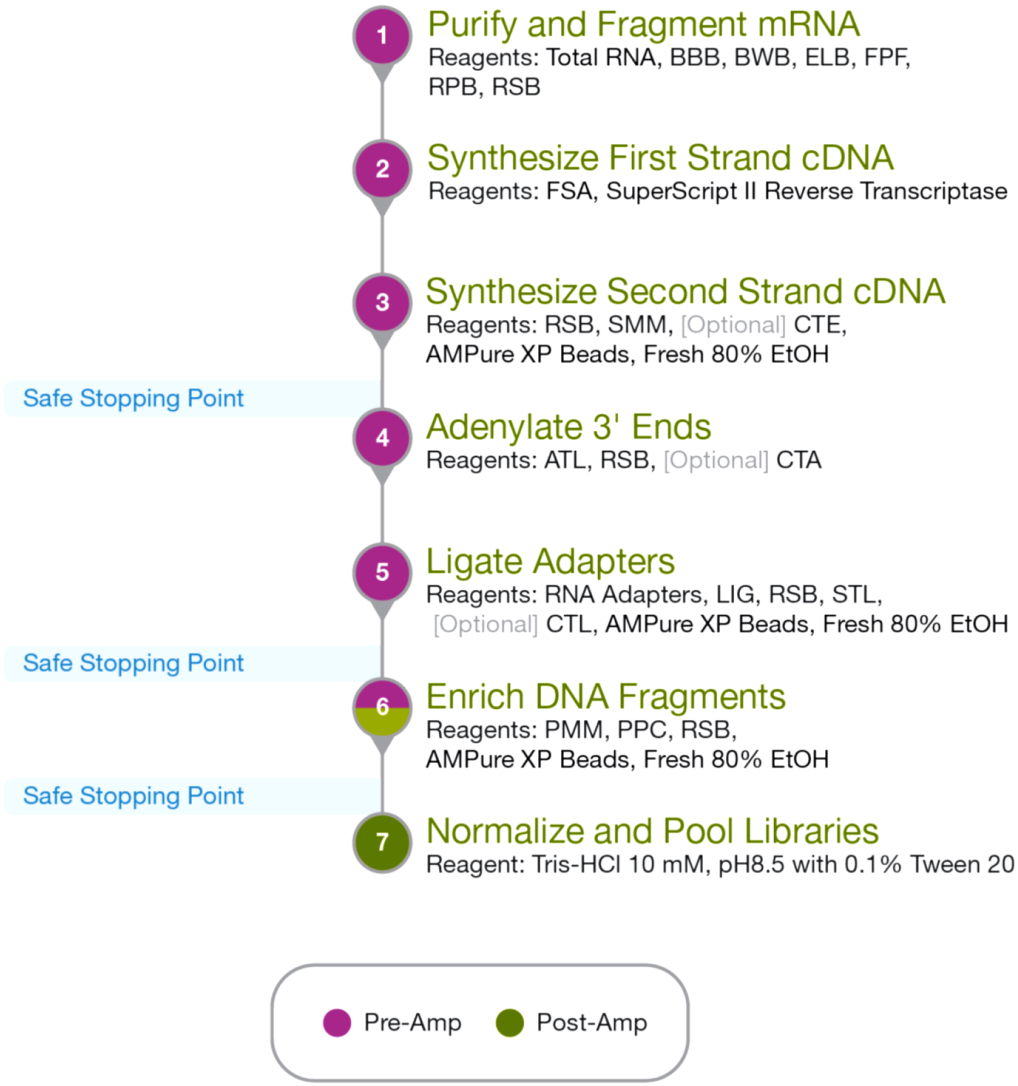


Figure 4.4 – TruSeq Stranded mRNA Library Prep Workflow.
BBB: Bead Binding Buffer; BWB: Bead Washing Buffer; ELB: Elution Buffer; FPF: Fragment, Prime, Finish Mix; RPB: RNA Purification Beads; RSB: Resuspension Buffer; FSA: First Strand Synthesis Act D Mix; SMM: Second Strand Marking Master Mix; CTE: End Repair Control; EtOH: ethanol; ATL: A-Tailing Mix; CTA: A-Tailing Control; LIG: Ligation Mix; STL: Stop Ligation Buffer; CTL: Ligation Control; PMM: PCR Master Mix; PPC: PCR Primer Cocktail. Figure and taken legend adapted from the TruSeq Stranded mRNA Reference Guide (Illumina, 2017).

All reagents were provided with the kit unless otherwise stated, with all reagents and consumables being sterile and DNase/RNase free. External RNA Controls Consortium (ERCC) spike-in synthetic RNA controls – a control mixture of known transcripts which allows the evaluation of successful library preparation (Risso et al., 2014) – were added to all samples.

The complete protocol for library preparation for the 64 samples from each batch (each mouse model processed separately, as described above) took four days, with the protocol repeated twice (total number of samples = 128 samples, 64 samples for each mouse model / batch). All volume calculations (e.g. digestion master mix) were performed in advance. All reagents were prepared in advance, as indicated in the manufacturer's instructions. On each day, before starting, every surface and all equipment were thoroughly cleaned using RNaseZap (Thermo Fisher Scientific).

4.2.2.1. Sample dilutions (day 1)

The eight most-optimal high-quality RNA samples for each group (RIN ≥ 8; mean RIN rTg4510 = 8.9 (SD = 0.2), mean RIN J20 = 8.6 (SD = 0.3); **Supplementary Table 1** and **Supplementary Table 2**) were selected for transcriptional profiling (total n = 128 samples; two models (rTg4510/J20) x two groups (TG/WT) x four time-points x eight individual animals per group).

Libraries were prepared from 450ng of total RNA plus ERCC spike-in synthetic RNA controls (Ambion, dilution 1:100). 450 ng of DNA from each sample, and 0.9 µL of ERCC spike-in controls was transferred to a total volume of 100 µL of RNase-free water (Qiagen, see **Section 2.5.2.3** of **Chapter 2**) in a 96-well plate, starting in well A01 and proceeding in columns. The plate was stored at -20°C until further processing.

4.2.2.2. Purify and fragment mRNA (day 2)

In this step, the polyA mRNA molecules are purified using magnetic beads containing oligo-dT. Two rounds of purification are performed, with the RNA being fragmented and primed for cDNA synthesis during the second elution (Illumina, 2017). A schematic illustration of this step is shown in **Figure 4.5**.

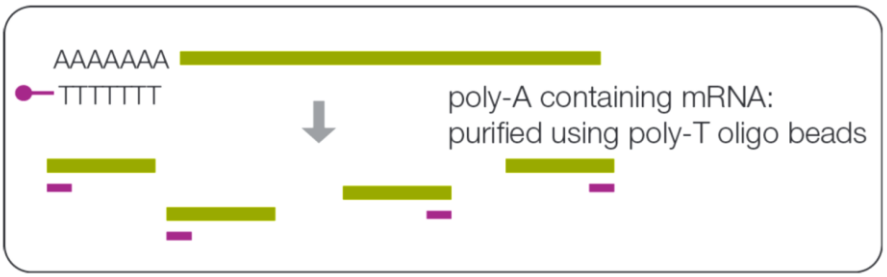


Figure 4.5 – Purifying and fragmenting mRNA.
The Poly-A containing mRNA molecules are purified using poly-T oligo attached magnetic beads. Following purification, the mRNA is fragmented into small pieces using divalent cations under elevated temperature. Figure and legend taken from the TruSeq stranded mRNA reference guide (Illumina, 2017).

The total RNA was diluted in nuclease-free ultrapure water to a final volume of 50 µl and added to each well of a RNA Purification Bead (RBP) plate. The RBP tube was brought to room temperature and vortexed vigorously to resuspend the oligo-dT beads. Using the Bravo Automated Liquid Handling Platform (Agilent), 50 µl of RBP was added to each well of the RBP plate to bind the polyA RNA to the beads, and the entire volume was gently pipetted to mix thoroughly. The plate was placed on a pre-programmed thermocycler, and left to incubate as follows:

65°C	5 minutes
4°C	∞

When it reached 4°C, the plate was removed from the thermocycler, and left to incubate at room temperature for 5 minutes to allow the RNA to bind to the beads. The plate was then placed on a magnet stand at room temperature for at least 5 minutes (until the liquid was clear) to separate the polyA RNA bound beads from the solution. The supernatant from each well was removed and discarded, and the tubes were removed from the magnetic stand. In order to remove unbound RNA, the beads were washed by adding 200 µl of Bead Washing Buffer to each well of the RBP plate, gently pipetting the entire volume to mix thoroughly, and

placing the plate on the magnetic stand at room temperature for 5 minutes. During incubation, the thawed Elution Buffer was centrifuged at 600 × g for 5 seconds. The supernatant from each well, containing most of the ribosomal and other non-messenger RNA, was removed and discarded, and the plate was removed from the magnetic stand. 50 µl of Elution Buffer was added to each well, and the entire volume was gently pipetted to mix thoroughly. The plate was placed on a pre-programmed thermocycler, and the following program was used:

80°C	2 minutes
25°C	∞

When it reached 25°C, the plate was removed from the thermocycler and placed on the bench at room temperature. 50 µl of Bead Binding Buffer was added to each well, and the plate was incubated at room temperature for 5 minutes. After incubation, the tubes were placed on a magnetic stand at room temperature for at least 5 minutes, and the supernatant from each well was removed and discarded. The plate was removed from the magnetic stand, and the beads were washed by adding 200 µl of Bead Washing Buffer to each well, and gently pipetting the entire volume to mix thoroughly. The plate was placed on the magnetic stand at room temperature for 5 minutes, and the supernatant from each well of the RBP plate was removed and discarded. After removing the plate from the magnetic stand, 19.5 µl of Fragment Prime Finish Mix was added to each well, and the entire volume was gently pipetted to mix well. The plate was placed on a pre-programmed thermocycler, and the following program was used:

94°C	8 minutes
4°C	∞

When it reached 4°C, the plate was removed from the thermocycler and placed on the magnetic stand at room temperature for 5 minutes. 17 µl of the supernatant from each well was transferred to a new plate.

4.2.2.3. Synthesise first strand cDNA (day 2)

In this step the cleaved RNA fragments primed with random hexamers are reverse transcribed into first strand cDNA. First Strand Synthesis Act D Mix (FSA) contains Actinomycin D (a toxin) that prevents spurious DNA-dependent synthesis, while allowing RNA-dependent synthesis, and improving strand specificity (Illumina, 2017). A schematic illustration of this step is shown in **Figure 4.6**.

Figure 4.6 – Synthesising first strand cDNA.
Cleaved RNA fragments are copied into first strand cDNA using reverse transcriptase and random primers. Adding Actinomycin D to FSA (First Stand Synthesis Act D mix) prevents spurious DNA-dependent synthesis, while allowing RNA-dependent synthesis, improving strand specificity. Figure and legend taken from the TruSeq stranded mRNA reference guide (Illumina, 2017).

This step was performed manually. Before starting, the thawed First Strand Synthesis Act D Mix (FSA) tube was centrifuged at 600 × g for 5 seconds. The First Strand Synthetic + SuperScript II mix was prepared at a ratio of 1 µl SuperScript II to 9 µl FSA. 8 µl of the FSA plus SuperScript II mix was added to each well of the plate containing the libraries, which was placed in a thermocycler and left to run using the following program:

25°C	10 minutes
42°C	15 minutes
70°C	15 minutes
4°C	∞

4.2.2.4. Synthesise second strand cDNA (day2)

In this stage of the experiment the RNA template is removed, a replacement strand is synthesised, and dUTPs are incorporated in place of dTTPs to generate

double-stranded cDNA. The incorporation of dUTP quenches the second strand during amplification., and the magnetic beads separate the double-stranded cDNA from the second strand reaction mix. The result is blunt-ended cDNA. A schematic illustration of this step is shown in **Figure 4.7**.

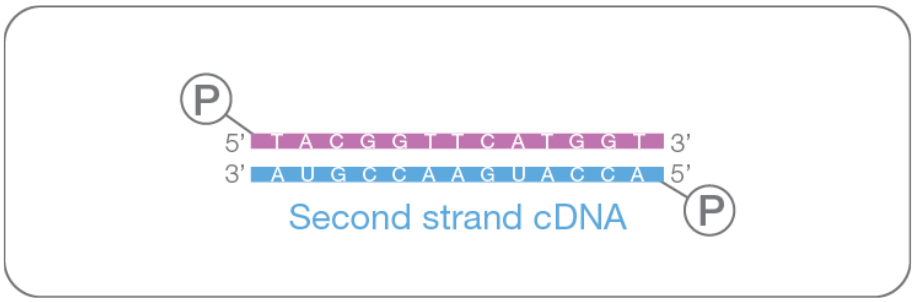


Figure 4.7 – Synthesising second strand cDNA.
Strand specificity is achieved by replacing dTTP with dUTP in the SMM (Second Strand Marking Mix), followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. The incorporation of dUTP in second strand synthesis quenches the second strand during amplification. Figure and legend taken from the TruSeq stranded mRNA reference guide (Illumina, 2017).

5 µl of diluted (1:50 dilution) End Repair Control was added to each well, followed by 20 µl of thawed Second Strand Marking Master Mix. All wells in the plate were mixed well by pipetting and the plate was incubated for one hour at 16°C.

Using the Bravo Automated Liquid Handling Platform (Agilent), 90 µl of MgNa magnetic beads brought to room temperature was added to each well, mixed by pipetting, and incubated at room temperature for 5 minutes. The plate was placed on the magnetic stand for at least 5 minutes (until the beads were captured and the liquid was clear), and the liquid was carefully removed without disturbing the beads. With the plate on the magnetic stand, the beads were gently washed twice by adding 200 µl 80% ethanol, incubating at room temperature for 30 seconds, and carefully removing and discarding the ethanol. The plate was then removed from the magnetic stand and the beads were allowed to dry at room temperature for 3 minutes. The dried beads were resuspended in 17.5 µl of Resuspension Buffer, incubated at room temperature for 2 minutes, and the plate was placed on the magnetic stand for at least 5 minutes (until the beads were captured and the liquid was clear) once again. 15 µl of the supernatant from each well was transferred to a new plate and stored at -20°C until the following day.

4.2.2.5. Adenylate 3' ends (day 3)

This step is characterised by the addition of one adenine (A) nucleotide to the 3' ends of the blunt fragments to prevent them from ligating to each other during the adapter ligation reaction. One corresponding thymine (T) nucleotide on the 3' end of the adapters provides a complementary overhang for ligating the adapter to the fragment (Illumina, 2017). A schematic illustration of this step is shown in **Figure 4.8**.

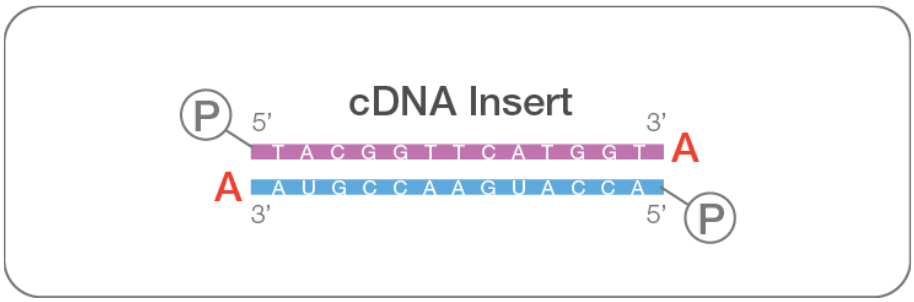


Figure 4.8 – Adenylating 3' ends.
A single 'A' nucleotide is added to the 3' ends of the blunt fragments to prevent them from ligating to each other during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragment. This strategy ensures a low rate of chimera (concatenated template) formation. Figure and legend taken from the TruSeq stranded mRNA reference guide (Illumina, 2017).

2.5 µl of diluted A-Tailing Control (1:100 dilution) was added to each well, followed by 12.5 µl of thawed A-Tailing Mix. Everything was mixed well by pipetting and the plate was placed in a thermocycler and incubated as indicated below:

37°C	30 minutes
70°C	5 minutes
4°C	∞

4.2.2.6. Ligate adapters (day 3)

In this stage of the protocol multiple indexing adapters are ligated to the ends of the double-stranded cDNA fragments, preparing them for hybridization onto a flow cell (Illumina, 2017). A schematic illustration of this step is shown in **Figure 4.9**.

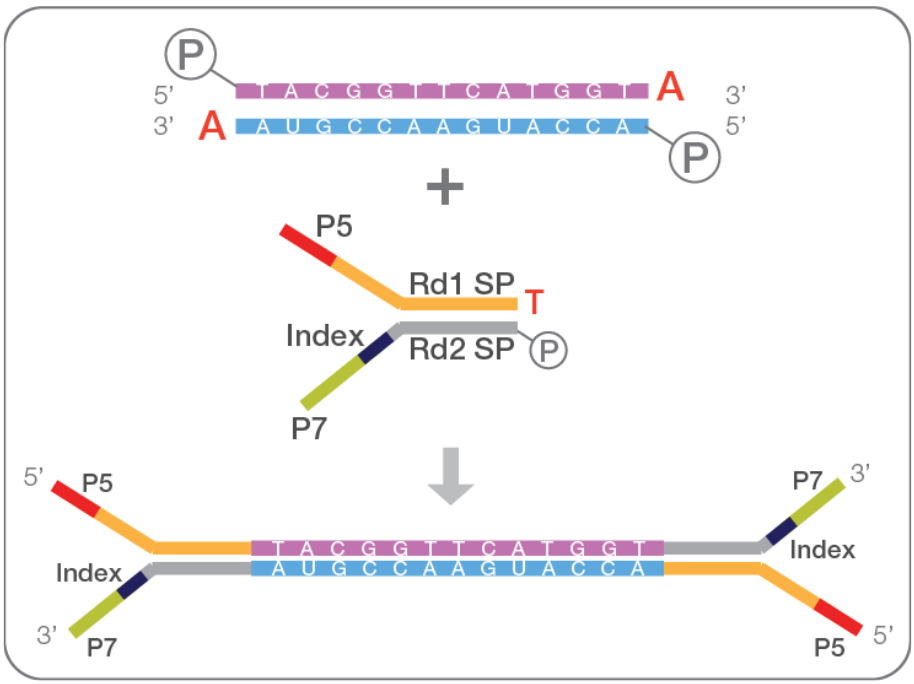


Figure 4.9 – Ligating adapters.
Adapter ligation prepares the ds cDNA for hybridization onto a flow cell. Figure and legend taken from the TruSeq stranded mRNA reference guide (Illumina, 2017).

A Barcode Plate, containing the 24 Adaptor Indexes (Illumina) described in **Table 4.1**, was used. Adaptors are short chemically synthesised nucleotide sequences that can be added to the ends of DNA molecules, allowing DNA fragments to bind to a next generation sequencing flow cell, and for PCR enrichment of adapter-ligated DNA fragments. The adaptors include indexes (or "barcodes", such as the ones described in **Table 4.1**) that permit multiple samples (or libraries) to be mixed together (or pooled) and sequenced simultaneously (multiplex sequencing).

Table 4.1 – Index adapters used for the RNA-seq library preparation.
Truseq LT single indexing.

Index	Sequence
TruSeq index 1	ATCACG
TruSeq index 2	CGATGT
TruSeq index 3	TTAGGC
TruSeq index 4	TGACCA
TruSeq index 5	ACAGTG
TruSeq index 6	GCCAAT
TruSeq index 7	CAGATC
TruSeq index 8	ACTTGA
TruSeq index 9	GATCAG
TruSeq index 10	TAGCTT
TruSeq index 11	GGCTAC
TruSeq index 12	CTTGTA
TruSeq index 13	AGTCAA
TruSeq index 14	AGTTCC
TruSeq index 15	ATGTCA
TruSeq index 16	CCGTCC
TruSeq index 18	GTCCGC
TruSeq index 19	GTGAAA
TruSeq index 20	GTGGCC
TruSeq index 21	GTTTCG
TruSeq index 22	CGTACG
TruSeq index 23	GAGTGG
TruSeq index 25	ACTGAT
TruSeq index 27	ATTCCT

Using the Bravo Automated Liquid Handling Platform (Agilent), 2.5 µl of diluted Ligation Control (1:100 dilution) was added to each well of the plate containing the libraries, followed by 2.5 µl of Ligation Mix, and 2.5 µl of RNA Adapter Index (alternating between the different sets of adaptors – important for sample pooling described in **Section 4.2.2.8**). Everything was mixed well by pipetting and the plate was incubated at 30°C for 10 minutes. Following incubation, 5 µl of Stop Ligation Buffer was added to each well and mixed by pipetting. 42 µl of MgNa magnetic beads was added to each well, mixed by pipetting, and incubated at room temperature for 5 minutes. The plate was placed on the magnetic stand for 5 minutes (until the beads were captured and the liquid was clear), and the liquid was carefully removed without disturbing the beads. With the plate on the magnetic stand, the beads were gently washed twice by adding 200 µl of 80% ethanol, incubating at room temperature for 30 seconds, and removing the ethanol carefully with a pipette. The plate was removed from the magnetic stand and the beads allowed to dry at room temperature for 3 minutes. The dried beads were then resuspended in 52.5 µl of Resuspension Buffer, and after incubating the plate at room temperature for 2 minutes, the place containing the libraries was place on the magnetic stand for 5 minutes (until the beads were captured and the liquid was clear). 50 µl of the supernatant from each well was transferred to a new plate, and 50 µl of MgNa beads were added and mixed by pipetting for a second wash. The new plate was incubated at room temperature for 5 minutes, place on the magnetic stand for 5 minutes (until the beads were captured and the liquid was clear), and the supernatant was carefully removed without disturbing the beads. Once again, with the plate on the magnetic stand, the beads were gently washed twice by adding 200 µl of 80% ethanol, incubating the plate at room temperature for 30 seconds, and removing the ethanol carefully with a pipette. The plate was then removed from the magnetic stand, and the beads were left to dry at room temperature for 3 minutes. The dried beads were resuspended in 22.5 µl of Resuspension Buffer, and after incubation at room temperature for 2 minutes, the plate was placed on the magnetic stand for 5 minutes (until the liquid was clear) and 20 µl of the supernatant from each well was transferred to a new plate. The plate was stored at -20°C until the next day.

4.2.2.7. Enrich DNA fragments (day 4)

In this stage of the protocol, the DNA fragments that have adapter molecules on both ends are selectively enriched / amplified by PCR. A schematic illustration of this step is shown in **Figure 4.10**.

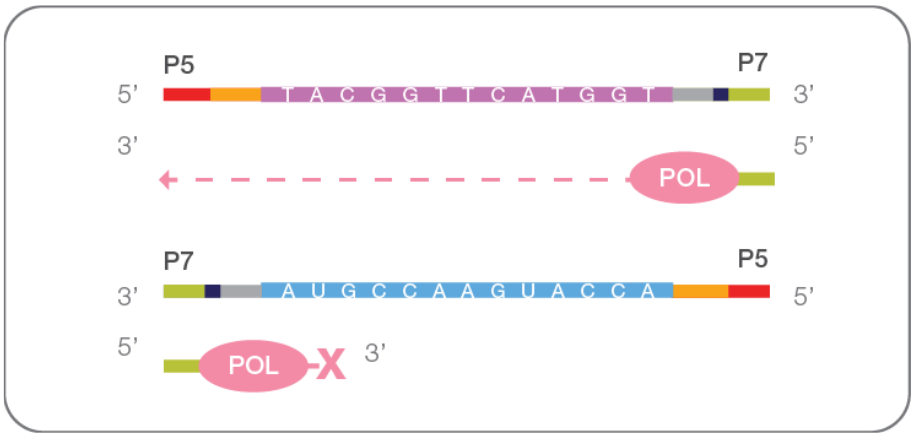


Figure 4.10 – Enriching DNA Fragments.
Polymerase used in the assay does not incorporate past dUTP. Therefore, the second strand is effectively quenched during amplification. The products are enriched with PCR and purified to create the final cDNA library. Figure and legend taken from the TruSeq stranded mRNA reference guide (Illumina, 2017).

5 µl of the PCR Primer Cocktail and 25 µl of the PCR Master Mix was added to each well, mixed by pipetting, and the plate was placed in a thermocycler and left to run using the following program:

Temperature	Time	Cycles
98°C	30 seconds	1
98°C	10 seconds	15
60°C	30 seconds	
72°C	5 minutes	
4°C	∞	

50 µl of MgNa beads at room temperature were added to each well, mixed by pipetting, and left to incubate for 5 minutes. The plate was placed on the magnetic stand for 5 minutes (until the beads were captured and the liquid was clear), and the liquid was carefully removed (without disturbing the beads). With the plate on the magnetic stand, the beads were gently washed twice by adding 200 µl of 80%

ethanol, incubating the plate at room temperature for 30 seconds, and carefully removing the ethanol with a pipette. The plate was removed from the magnetic stand and the beads were left to dry at room temperature for 3 minutes. The dried beads were then resuspended in 32.5 µl of Resuspension Buffer and the plate was incubated at room temperature for 2 minutes. Finally, the plate was placed on the magnetic stand for 5 minutes (until the beads were captured and the liquid was clear), and 30 µl of the supernatant was transferred to a new plate.

4.2.2.8. Sample pooling and quality control (day 4)

Library concentrations were determined using the D1000 ScreenTape System (Agilent) (as described in **Section 2.5.3.4** of **Chapter 2**). An overview of the experimental procedures for the High Sensitivity D1000 Screentape, designed for analysing DNA molecules from 35-1000 bp, is given in **Figure 4.11**.

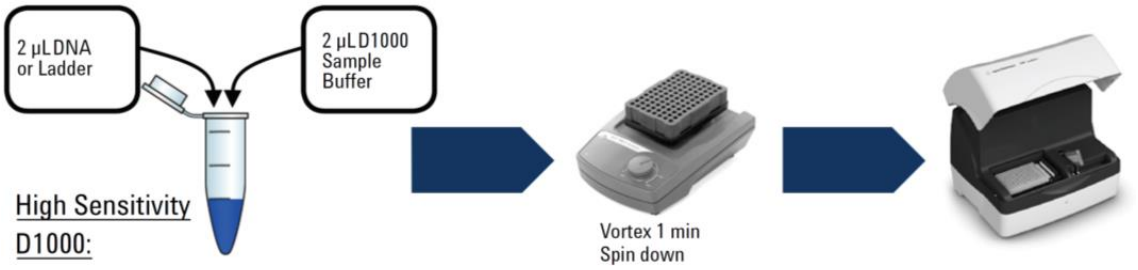


Figure 4.11 – Overview of the experimental procedures for DNA analysis using the High Sensitivity D1000 Screentape Assay and 2200 TapeStation System.
Figure taken from the Agilent High Sensitivity D1000 ScreenTape System Quick Guide (Agilent, 2015).

Following the manufacturer's instructions, I added 2µL of D1000 Sample Buffer (provided with the kit) to 2µL of each pool or 2µL of Ladder (provided with the kit) into strip-tube wells. The mix was vortexed using an IKA vortexer and adaptor at 2000 rpm for 1 min, and the tubes spun-down to bring the samples to the bottom. After preparing the 2200 TapeStation instrument, including placing the sample block and loading tips, I inserted the ScreenTape into the TapeStation, placed the samples into the sample block, and started the run.

Samples were pooled together to a 2nM concentration, for subsequent sequencing (three pools of 22 samples for J20 samples and one pool of 64 samples for rTg4510 samples).

Before sequencing, pooled libraries were quantified using a Qubit Fluorometer (Thermo Fisher Scientific), Tapestation HS ScreenTape System (Agilent Technologies), and qPCR, which were all carried out at the University of Exeter Sequencing Service.

4.2.3. Sequencing

Final library pools (containing fragments with ~300 bp in size) were distributed across twelve HiSeq2500 (Illumina) lanes (six lanes for each mouse model) and subjected to 125bp paired-end sequencing yielding a mean untrimmed read depth of ~18-20 million reads/sample (**Supplementary Table 1** and **Supplementary Table 2**). Paired-end sequencing was used as it enables both ends of the DNA fragment to be sequenced, which improves the alignment of the resulting RNA-seq reads significantly.

4.2.4. Sequencing data processing

4.2.4.1. Data pre-processing

All sequencing data processing was performed on a Unix-based operating system server. Initial demultiplexing of the raw sequencing files into *FASTQ* files, contamination checks and ERCC spike-in control quality checks were performed in collaboration with Paul O'Neill at the University of Exeter Sequencing Service. Raw RNA-seq data (*FASTQ*) has been deposited in GEO under accession number GSE125957.

4.2.4.2. Quality control of raw data using FastQC

The randomised *FASTQ* files (Phred score (Q) ≥ 35, **Supplementary Table 1** and **Supplementary Table 2**) underwent quality control (QC) assessments using *FastQC* (Andrews, 2010) (version 0.11.4). We obtained a mean of 18.81 (SD =

3.37) million sequencing reads per sample for the rTg4510 dataset (**Supplementary Table 1**), and a mean of 22.05 (SD = 2.88) million sequencing reads per sample for the J20 dataset (**Supplementary Table 2**).

FastQC provides quality metrics for each read and for each sample, and summarises them in an HTML report. A representative example of the quality measures reported graphically by *FastQC* for one sample is shown in **Figure 4.14** to **Figure 4.19**.

“Per base sequence quality” (**Figure 4.14**), expressed as a Phred score (Q), represents the distribution of read quality scores at each position, i.e., the confidence in the base call (Korpelainen et al., 2014). To calculate Q, the logarithm to base 10 (log10) of the probability that the base is wrong is taken and multiplied by -10 (e.g. if there is a chance that 1 base in 100 is wrong, Q = -10 x log10(0.01) = 20). Q typically ranges from 0 to 40, with Q = 10 meaning that there is a 1 in 10 chance of the base being called incorrectly, Q= 20 meaning there is a 1 in 100 chance, Q = 30 meaning 1 in 1000 (Datta and Nettleton, 2014). Typically, base quality values start to decrease towards the end of a read, which is easily detected in the box plots for the base sequence quality along the reads. Moreover, the forward reads have usually higher quality towards the ends of the reads compared to the reverse reads. These are characteristics of all second-generation sequencing data (Korpelainen et al., 2014). Reads containing low quality bases can be filtered and/or trimmed (see **Section 4.2.4.3**).

“Per sequencing quality scores” (**Figure 4.15**) give the average quality per read. They inform about the distribution of the read’s mean quality, detecting if a subset of the sequences has universally poor quality (Andrews, 2010).

“Per base sequence content” (**Figure 4.16**) plots base composition along the reads, expressing the proportion of DNA bases (A, C, T, G or N base calls) at each position over all reads. In a random library it is expected that there would be little to no difference in the base composition for the different bases, therefore the lines in this plot should run parallel with each other. Some types of libraries (including most RNA-seq libraries) produce base composition bias (caused by random hexamer priming during conversion of RNA into cDNA), normally at the start of the read (Hansen et al., 2010, Korpelainen et al., 2014, van Gurp et al.,

2013). Despite generating a technical bias, this can be corrected by trimming off the first positions of every read (**Section 4.2.4.3**).

“Per sequence GC content” (**Figure 4.17**) plots the GC content across the whole length of each sequence compared to a modelled normal distribution of GC content. In a normal diverse library an approximate normal distribution of GC content is expected, with the central peak corresponding to the overall GC content of the underlying genome (Andrews, 2010).

“Sequence length distribution” (**Figure 4.18**) plots the distribution of fragment size, showing the relative amounts of each different size of sequence fragment (Andrews, 2010). In a properly diverse library most sequences should fall into the far left of the plot in both the red and blue lines.

In the “Sequence duplication levels” (**Figure 4.19**) plot, the proportion of the library which is made up of sequences in each of the different duplication level bins is shown (Andrews, 2010). As reads are ends of random fragments, most reads should be unique, and in a random library it is expected that most sequences occur only once in the final set, i.e. most sequences fall into the far left of the plot in both the red and blue lines. For other next-generation sequencing approaches, high levels of identical reads (which tend to produce peaks towards the right of the plot) suggest PCR overamplification or the presence of low complexity contaminants; however, in RNA-seq data, duplicates are often associated with highly expressed transcripts. In order to detect lowly expressed transcripts, it is common to over-sequence highly expressed transcripts, which potentially creates large sets of duplicates. For these reasons, duplicates should not be removed for differential expression analysis, but their presence should be explored in cases where a tower of reads in one position is observed, which most likely indicates PCR artefacts (Andrews, 2010, Korpelainen et al., 2014).

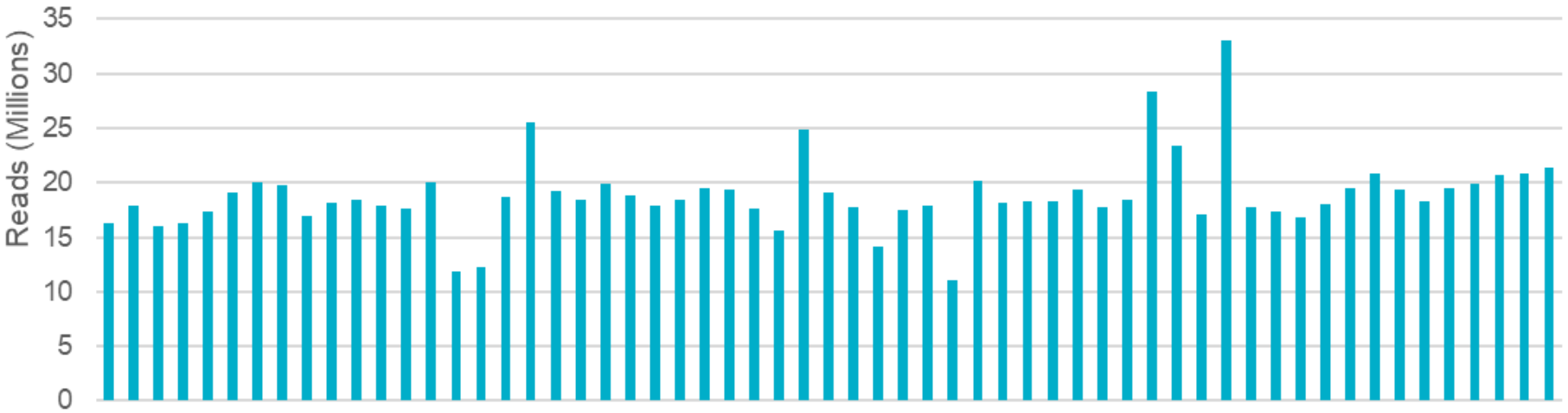


Figure 4.12 – Number of raw RNA-seq reads for each sample in the rTg4510 dataset.

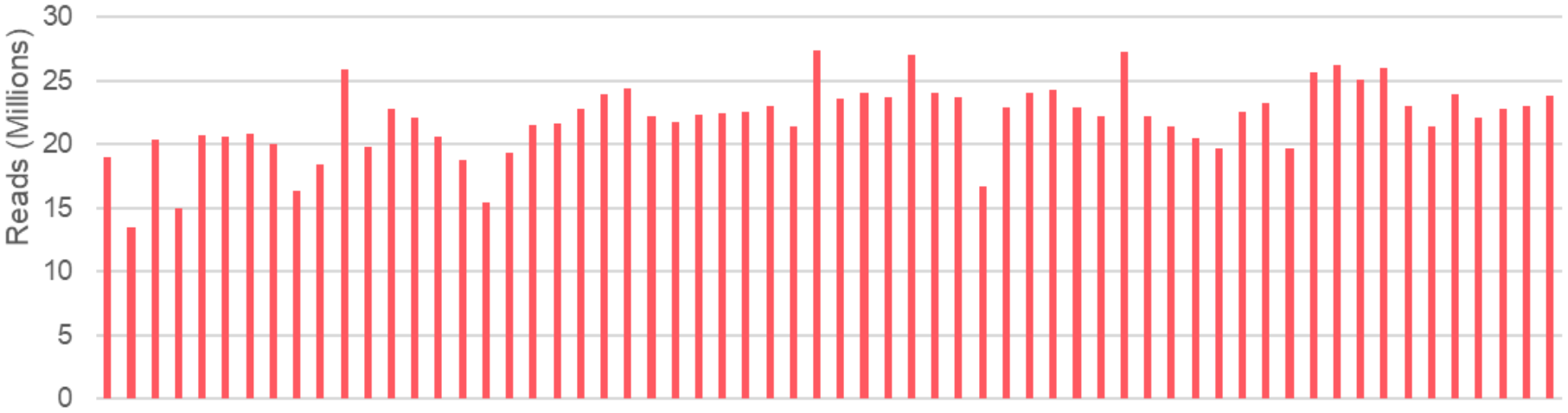


Figure 4.13 – Number of raw RNA-seq reads for each sample in the J20 dataset.

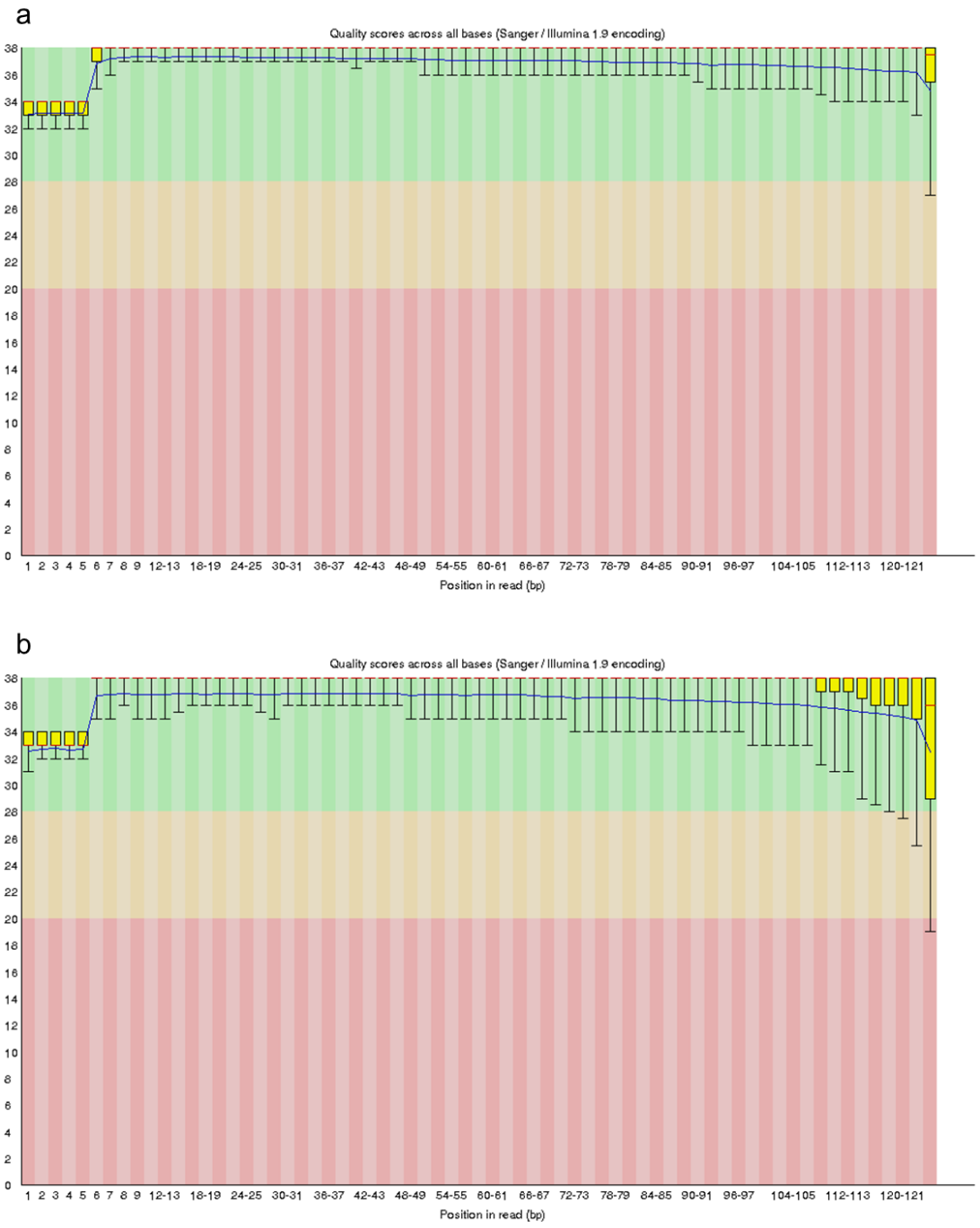


Figure 4.14 – Per base sequence quality from the FastQC report for raw RNA-seq reads.

Results for one representative sample, showing high quality per base sequencing across the entire read for the **(a)** forward (read 1) and **(b)** reverse (read 2) reads from the same sample. A box and whisker plot is drawn for each position, with the central line corresponding to the median value, the yellow box representing the inter-quartile range (25-75%), the upper and lower whiskers representing the 10% and 90% points, and the blue line representing the mean quality. The background of the graph divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). The quality of calls on most platforms will degrade as the run progresses, so it is common to see base calls falling into the orange area towards the end of a read.

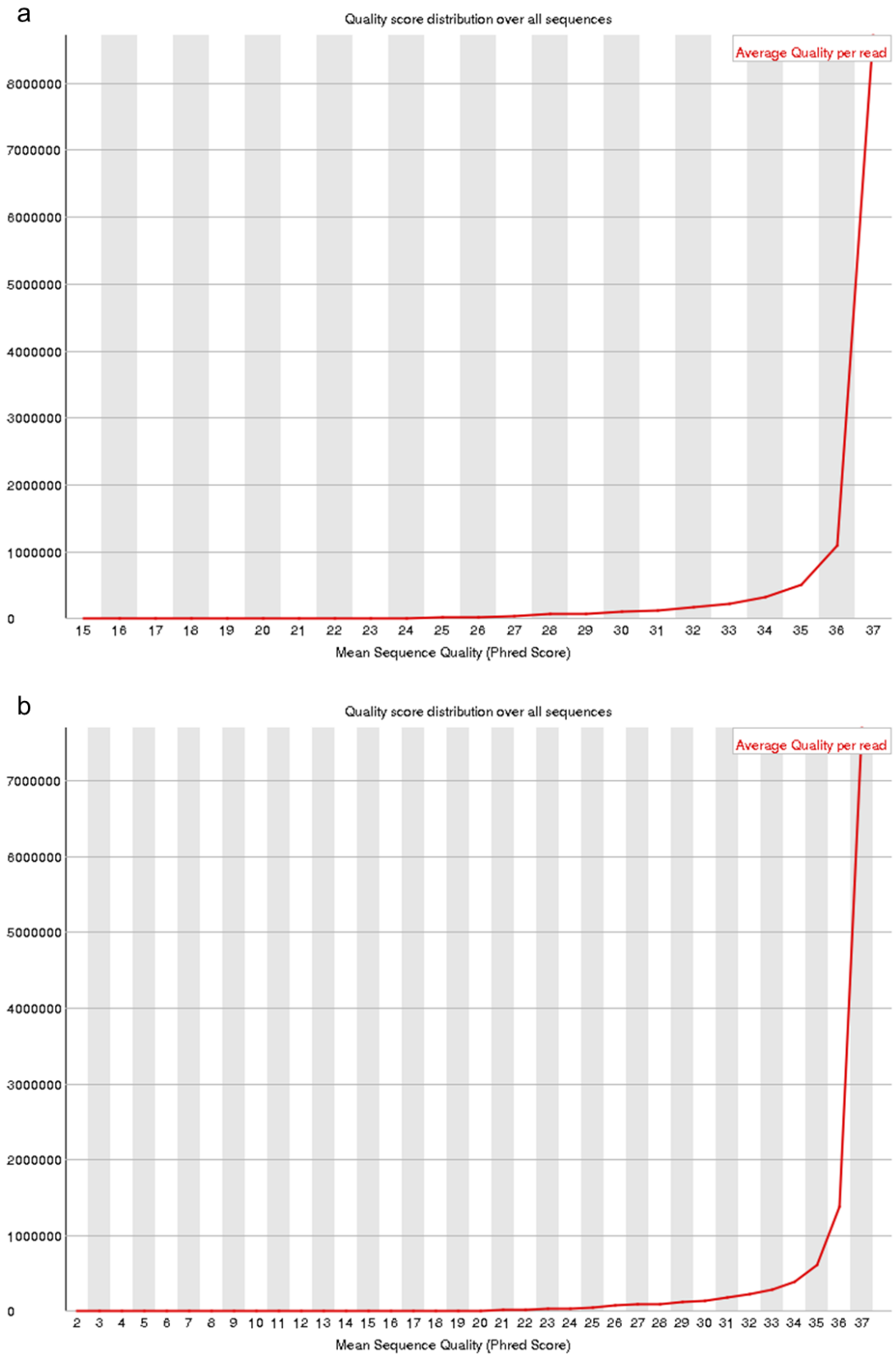


Figure 4.15 – Per sequencing quality scores from the FastQC report for raw RNA-seq reads.
Results for the **(a)** forward (read 1) and **(b)** reverse (read 2) from one representative sample, showing high quality score distribution over all sequences measured by Phred (Q) quality score.

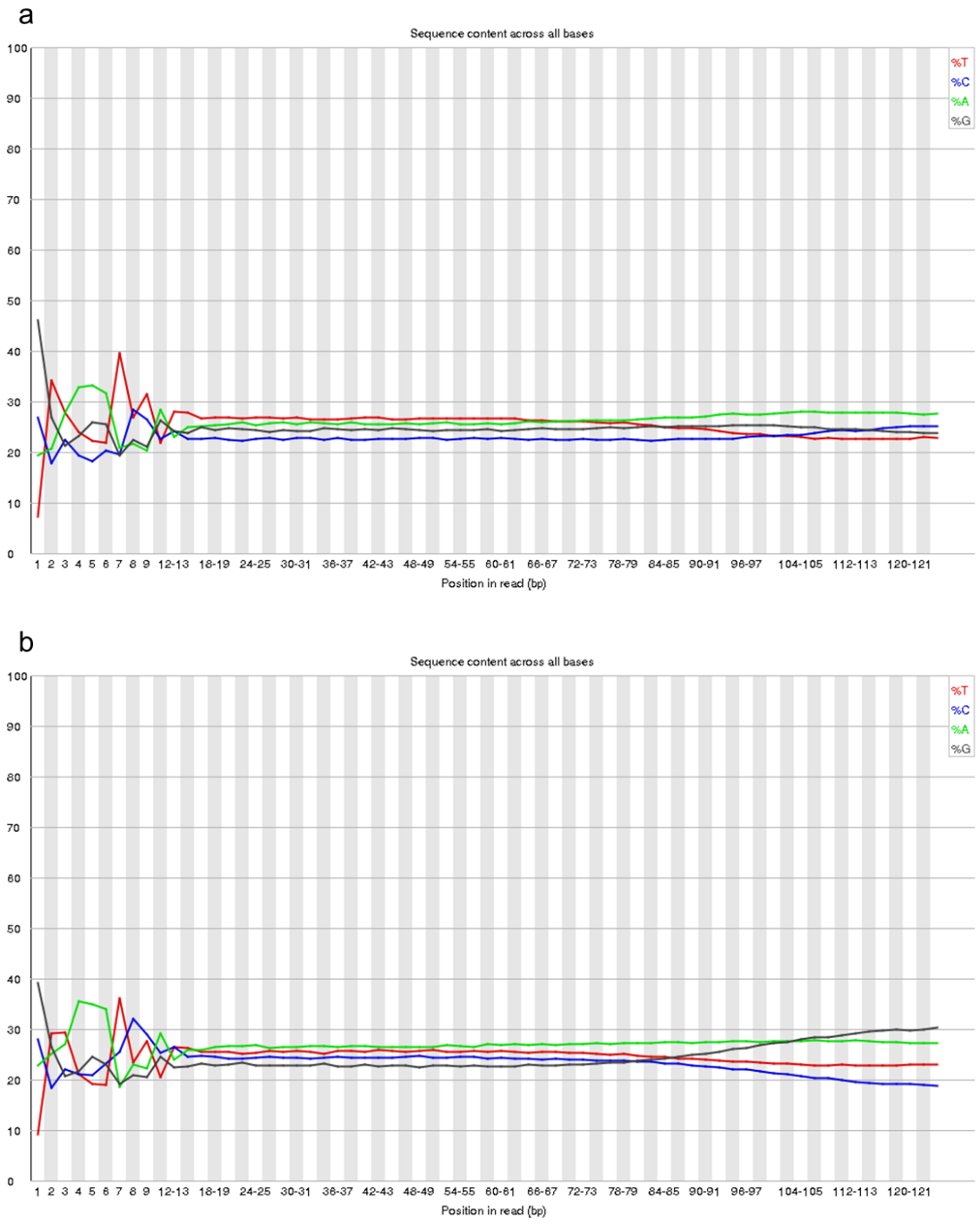


Figure 4.16 – Per base sequence content from the FastQC report for raw RNA-seq reads.

Results for the **(a)** forward (read 1) and **(b)** reverse (read 2) from one representative sample, showing the RNA-seq typical base composition bias at the start of each read.

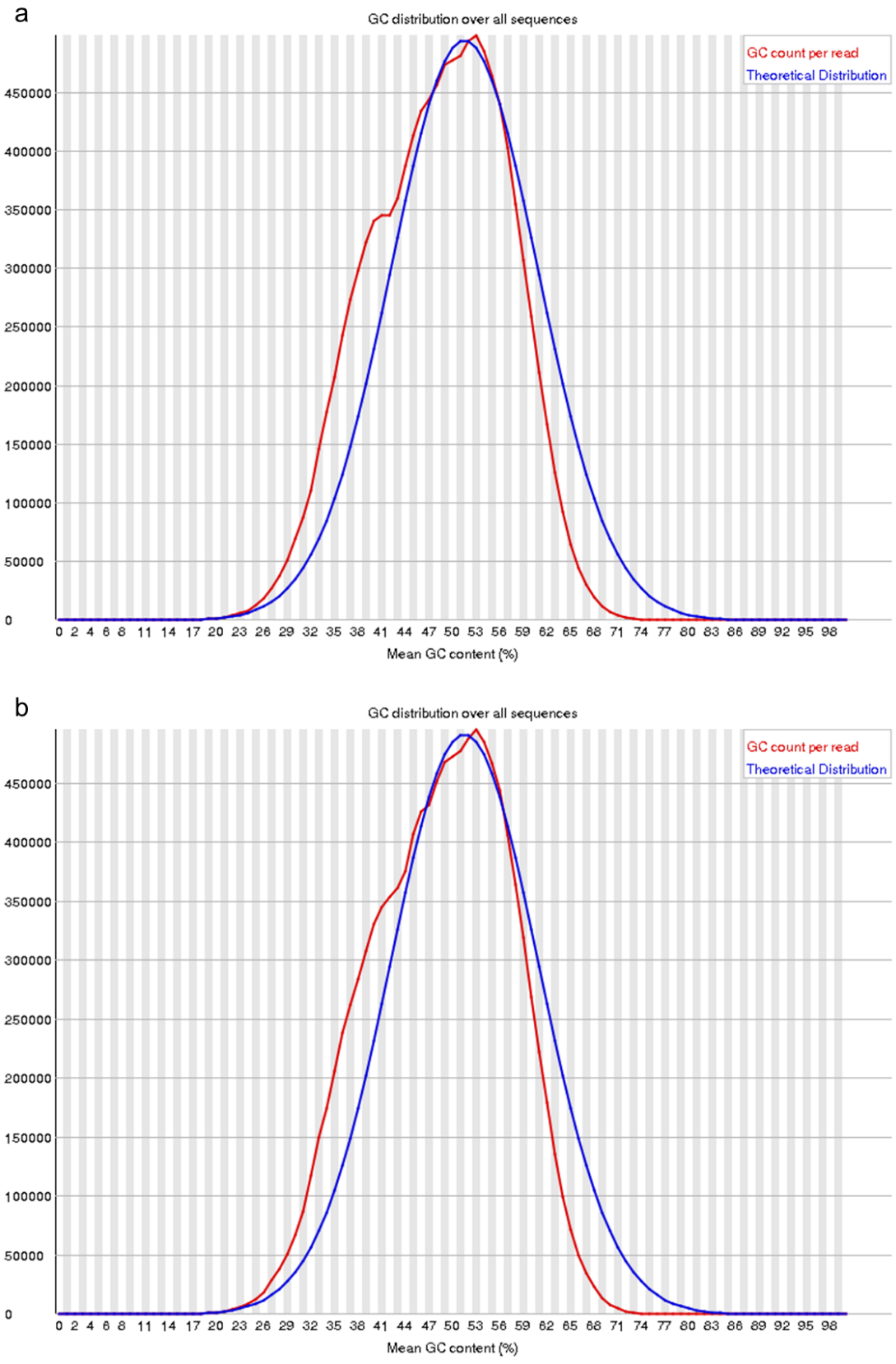


Figure 4.17 – Per sequence GC content from the FastQC report for raw RNA-seq reads.
Results for the **(a)** forward (read 1) and **(b)** reverse (read 2) from one representative sample.

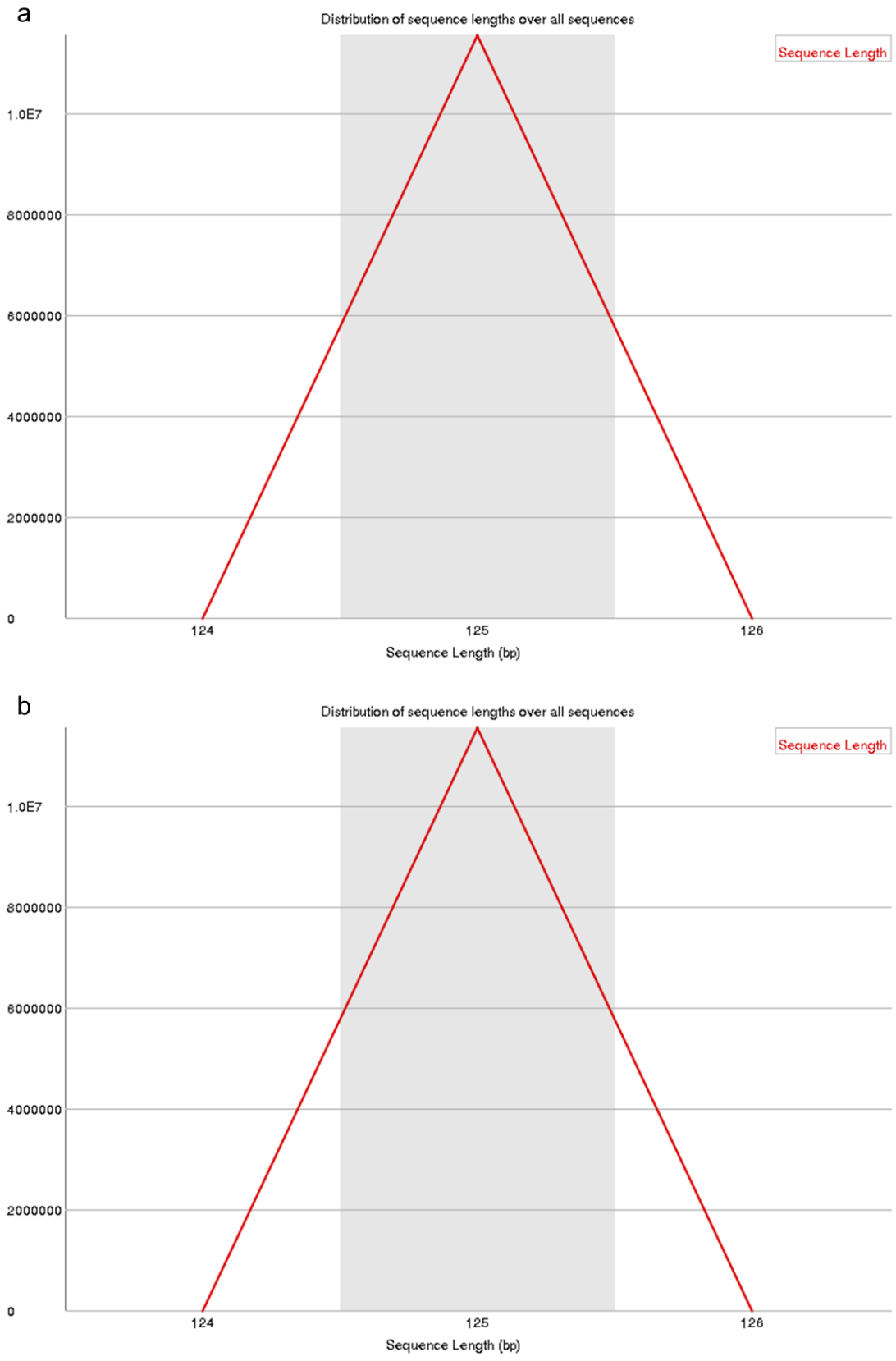


Figure 4.18 – Sequence length distribution from the FastQC report for raw RNA-seq reads.
Results for the **(a)** forward (read 1) and **(b)** reverse (read 2) from one representative sample.

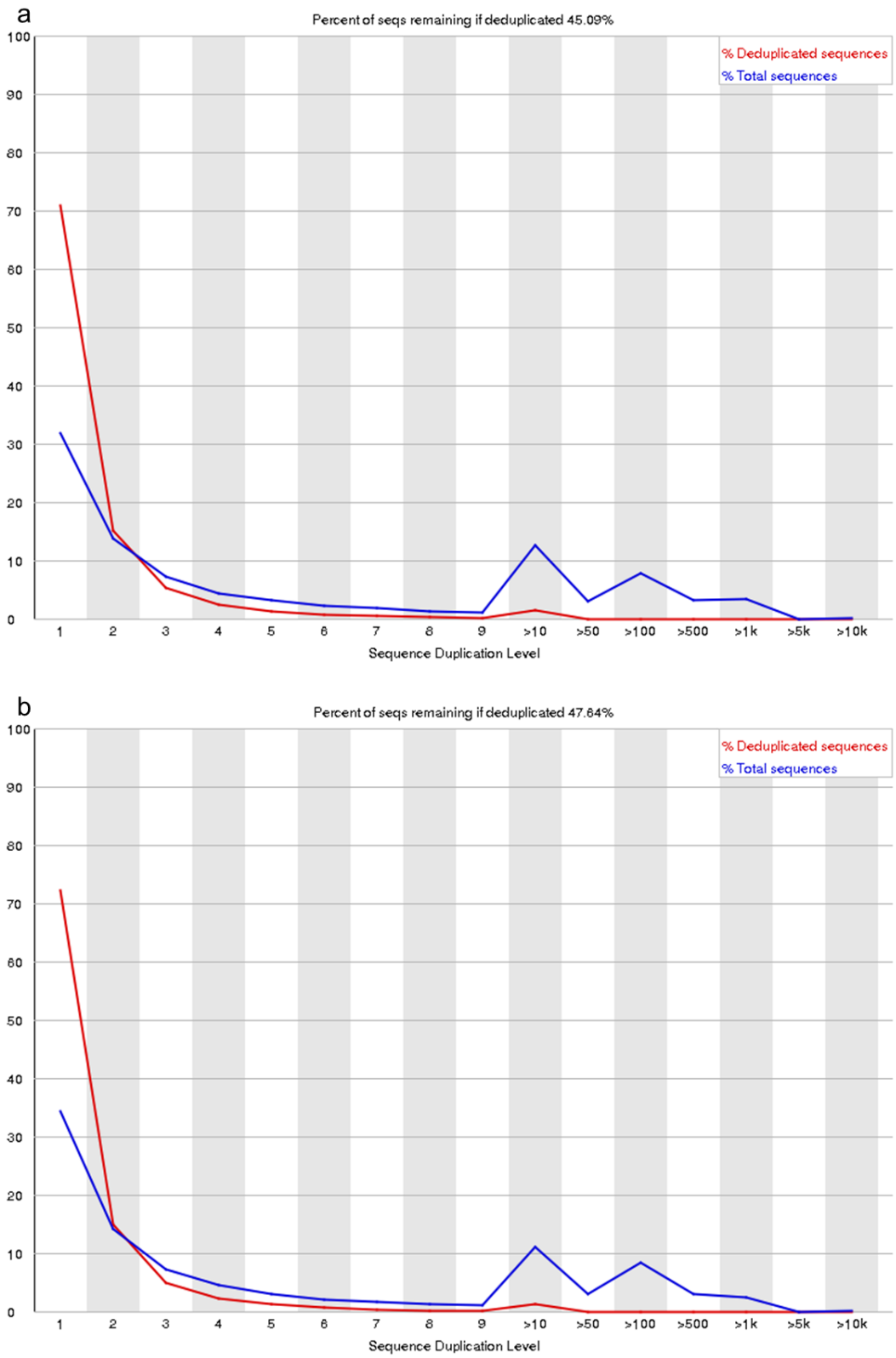


Figure 4.19 – Sequence duplication levels from the FastQC report for raw RNA-seq reads.

Results for the **(a)** forward (read 1) and **(b)** reverse (read 2) from one representative sample. The blue line corresponds to the full sequence set and shows how its duplication levels are distributed. The red line corresponds to the sequences that are de-duplicated and shows the proportions of the deduplicated set which come from different duplication levels in the original data.

4.2.4.3. Trimming and quality filtering

After inspection of the quality-control plots, additional filtering steps were necessary, including removal of reads containing ambiguous bases, removal or trimming Illumina's adapter sequences, removal of ERCC RNA spike-in controls, and removal of PCR duplicates. Other labels, such as multiplexing identifiers and primers, also needed to be removed. All of these are best identified when using paired-end reads, which was the case. Of note, currently there is no consensus about the optimal quality thresholds for trimming (which allows the removal of low-quality ends of reads) and filtering (which removes the entire read) in the context of RNA-seq; despite trimming low quality bases may improve the alignment of reads to a reference, it also reduces size and number of reads – choosing a quality threshold is therefore a trade-off between the two (Aronesty, 2011). Considering this, trimming and filtering (ribosomal sequences removal, quality threshold 20, minimum sequence length 35) was performed with *fastqmcf* (Aronesty, 2011) (version 1.0). This software scans the *FASTQ* file for adapters, and, based on a log-scaled threshold, determines a set of clipping parameters, performing the following steps: 1) detection and removal of sequencing adapters and primers; 2) detection of limited skewing at the ends of reads; 3) detection and removal of poor quality at the ends of reads; 4) detection and removal of Ns from the ends of reads; 5) discarding sequences that are too short after previous steps (Aronesty, 2011). In addition, before running *fastqmcf*, ribosomal sequences were removed using *bowtie2* (Langmead et al., 2009) (version 2.3.0).

The number of RNA-seq reads after filtering for all samples is shown in **Figure 4.20** and **Figure 4.21**. A mean output of 18.18 (SD = 3.33) million filtered sequencing reads per sample was obtained for the rTg4510 dataset, and a mean output of 21.27 (SD = 2.93) million filtered sequencing reads per sample was obtained for the J20 dataset. A representative example of the quality measures reported by *FastQC* for the same sample shown above is provided in **Figure 4.22** to **Figure 4.27**. Following confirmation of genotype and QC, 7 samples were excluded from subsequent analysis leaving a final number of 121 high-quality RNA-seq datasets (representing 6-8 animals per group).

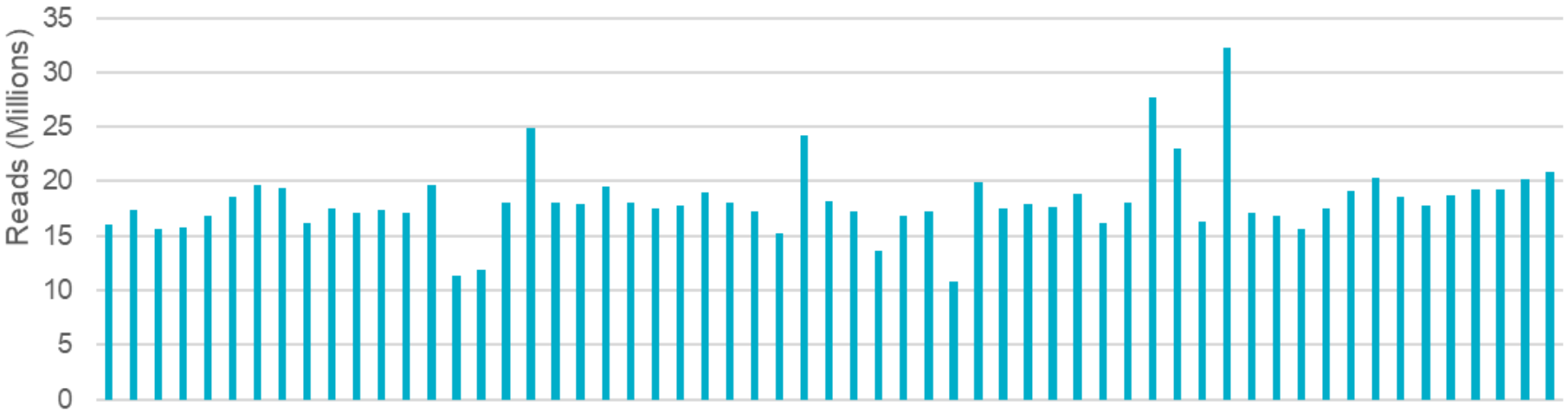


Figure 4.20 – Number of filtered RNA-seq reads for each sample in the rTg4510 dataset.

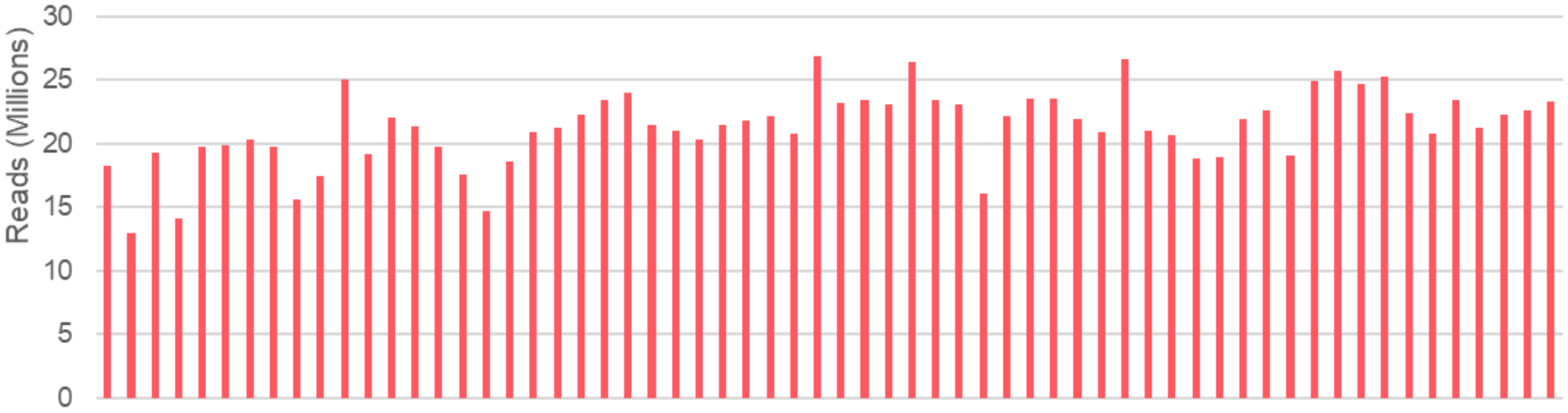


Figure 4.21 – Number of filtered RNA-seq reads for each sample in the J20 dataset.

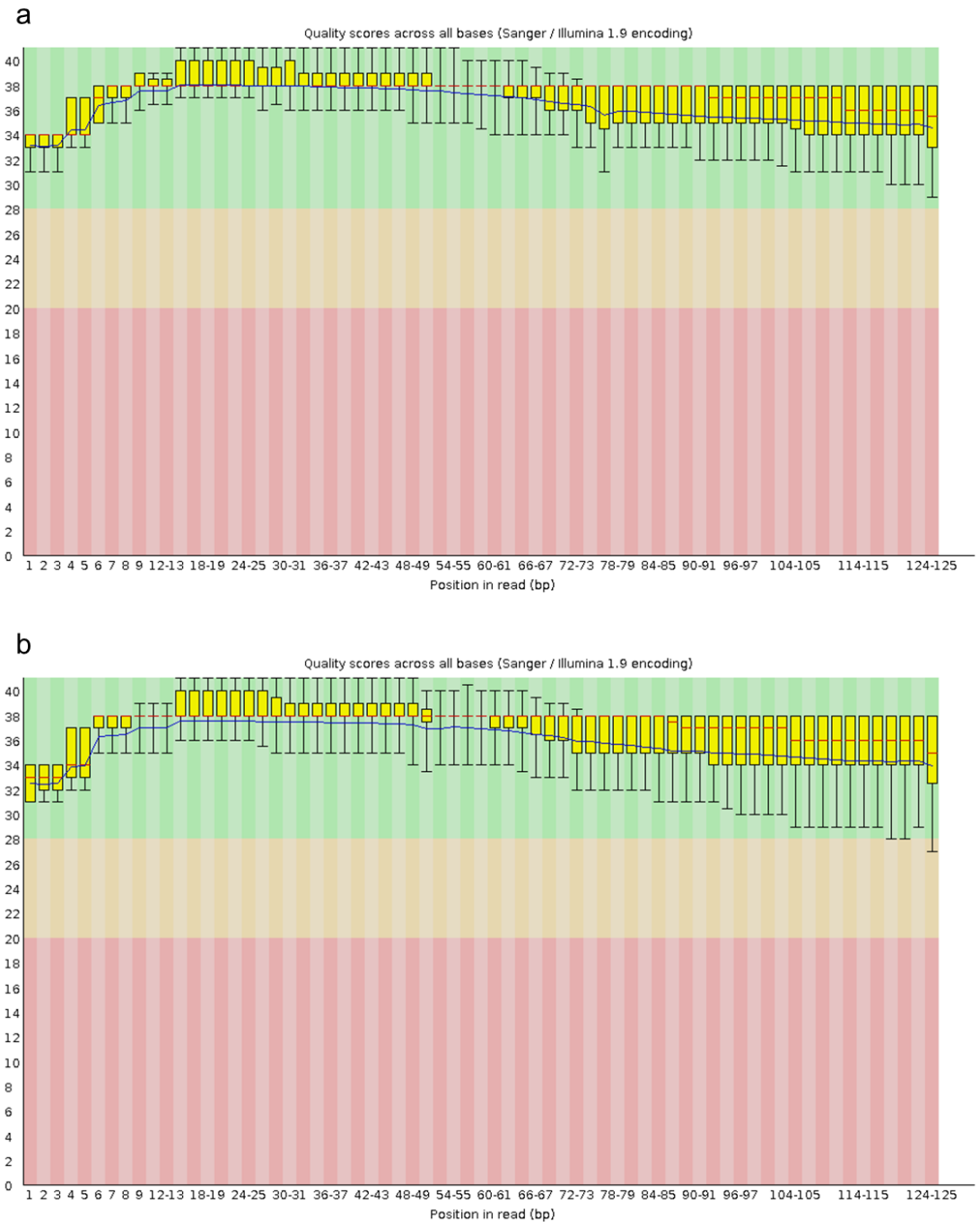


Figure 4.22 – Per base sequence quality from the FastQC report for filtered RNA-seq reads.

Results for one representative sample, showing high quality per base sequencing across the entire read for the **(a)** forward (read 1) and **(b)** reverse (read 2) reads from the same sample.

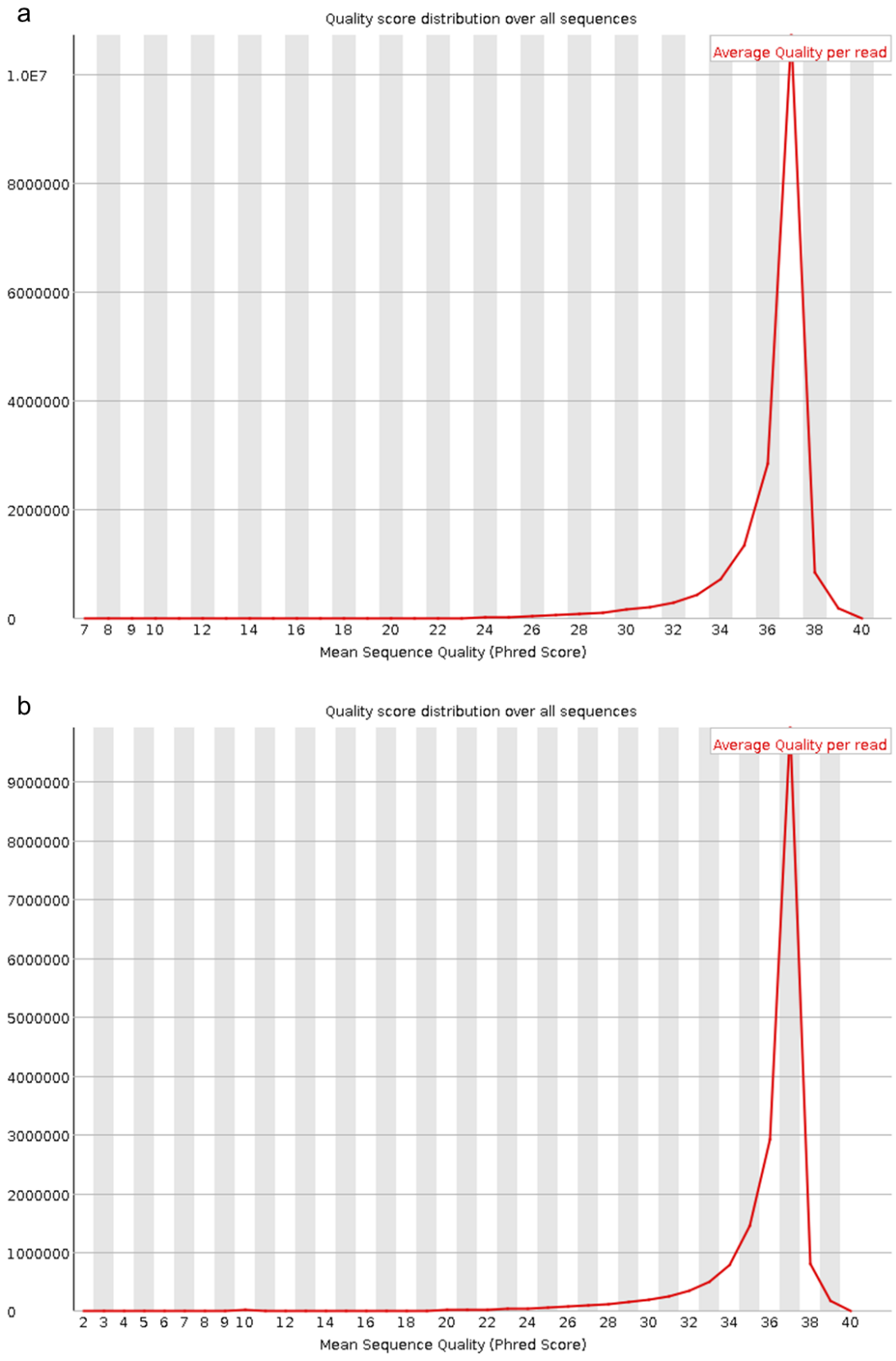


Figure 4.23 – Per sequencing quality scores from the FastQC report for filtered RNA-seq reads.
Results for the (a) forward (read 1) and (b) reverse (read 2) from one representative sample, showing high quality score distribution over all sequences measured by Phred (Q) quality score.

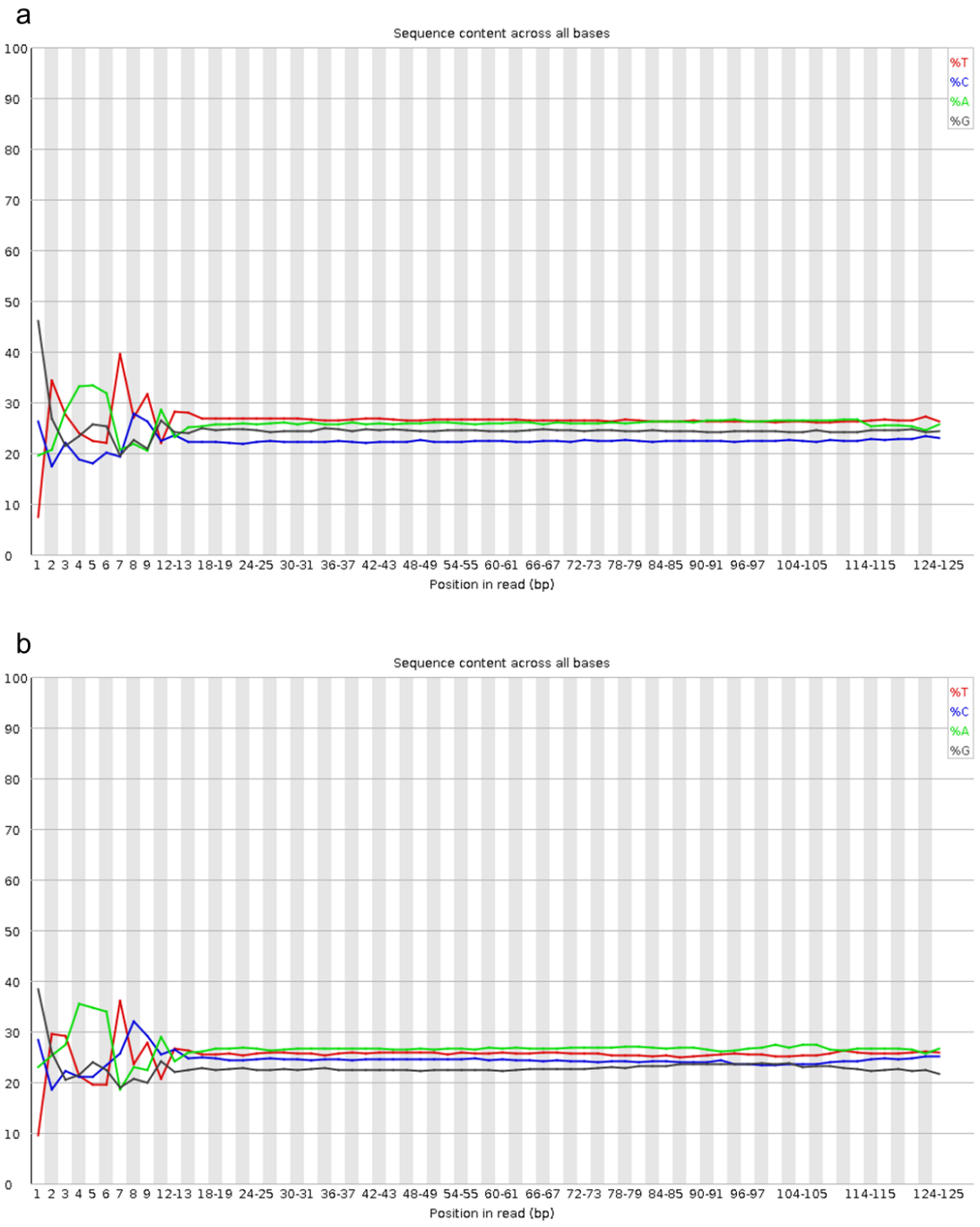


Figure 4.24 – Per base sequence content from the FastQC report for filtered RNA-seq reads.
Results for the **(a)** forward (read 1) and **(b)** reverse (read 2) from one representative sample, showing the RNA-seq typical base composition bias at the start of each read.

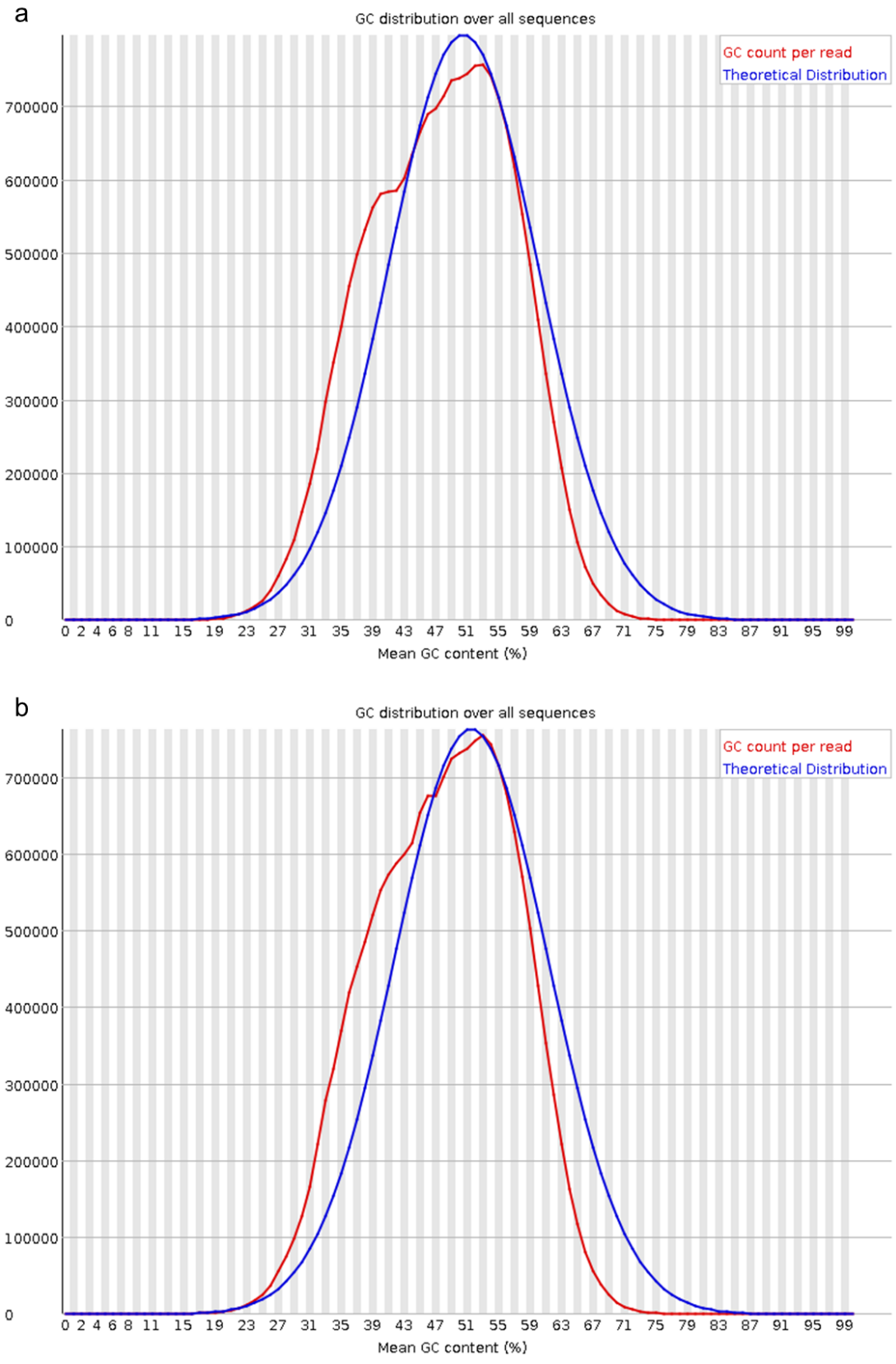


Figure 4.25 – Per sequence GC content from the FastQC report for filtered RNA-seq reads.
Results for the **(a)** forward (read 1) and **(b)** reverse (read 2) from one representative sample.

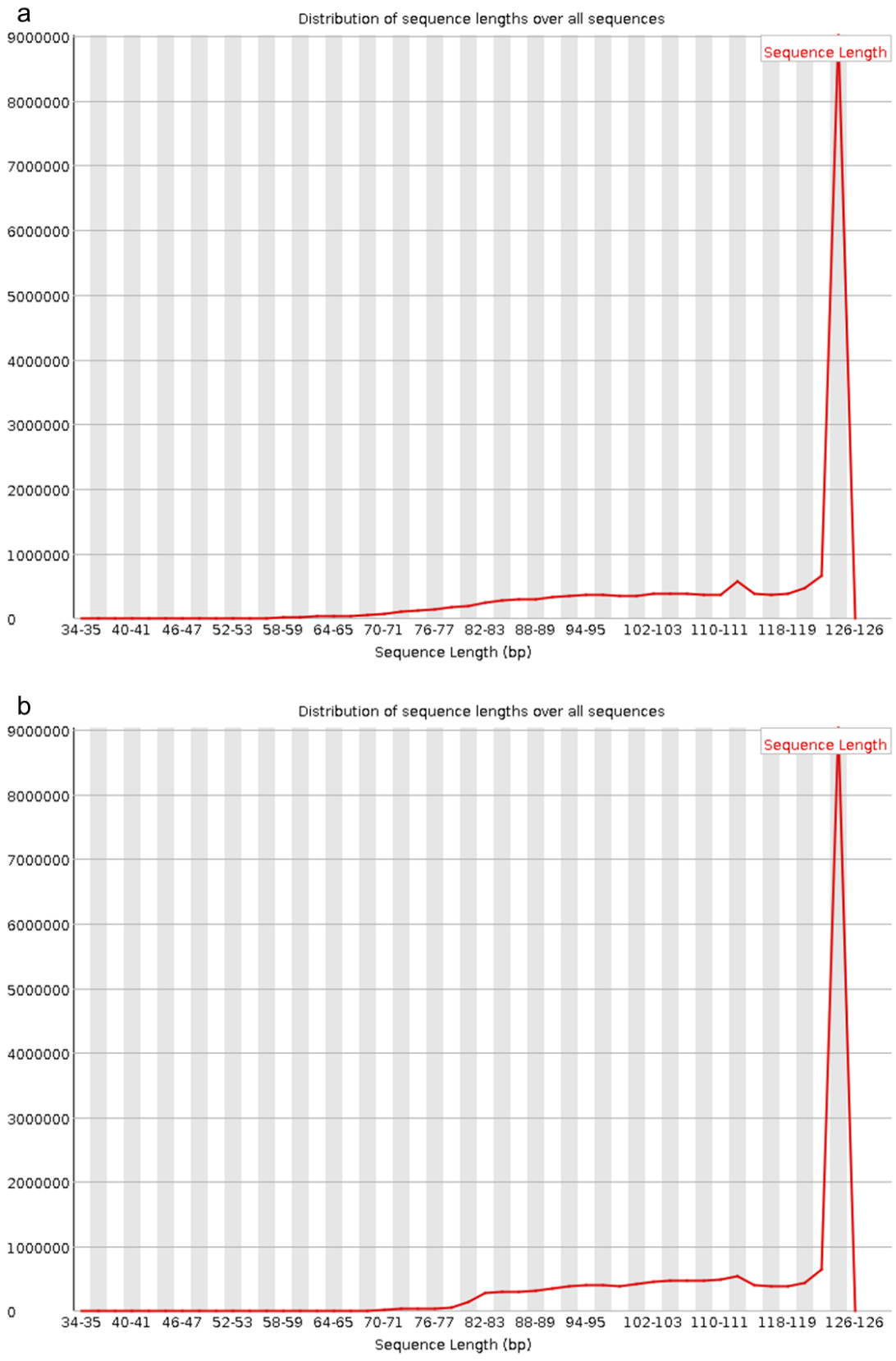


Figure 4.26 – Sequencelength distribution from the FastQC report for filtered RNA-seq reads.
Results for the **(a)** forward (read 1) and **(b)** reverse (read 2) from one representative sample.

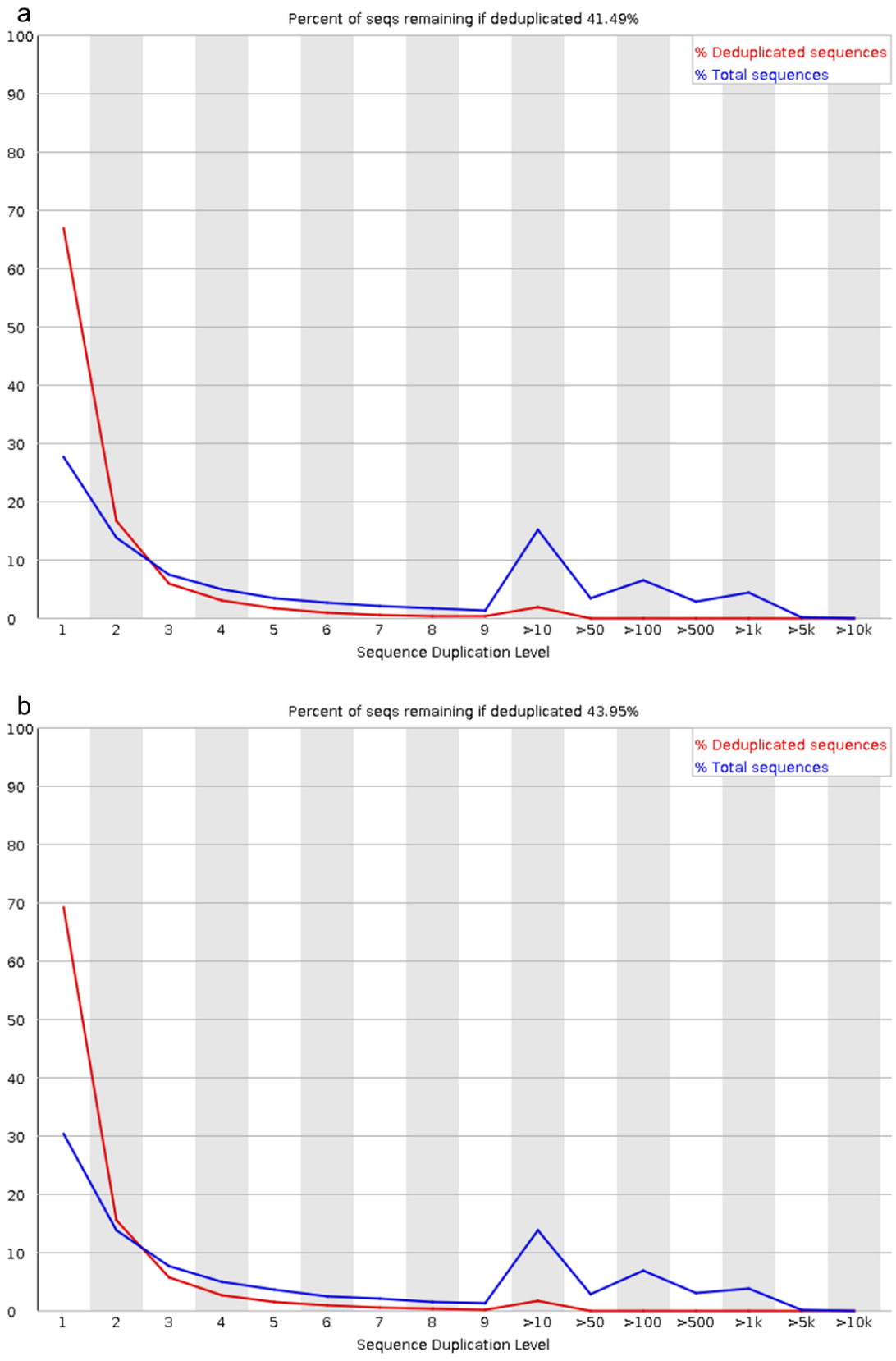


Figure 4.27 – Sequence duplication levels from the FastQC report for filtered RNA-seq reads.
Results for the **(a)** forward (read 1) and **(b)** reverse (read 2) from one representative sample.

4.2.4.4. Alignment to the mouse reference genome and mapping quality control

Aligning or mapping sequencing reads refers to the lining up of the sequences to a reference, allowing us to estimate where the read originated from. Mapping of RNA-seq reads is usually achieved by mapping them to a reference genome, or a reference transcriptome (**Figure 4.28**). Mapping reads to a reference genome provides genomic location information that can be used for the discovery of new genes and/or transcripts, in addition to quantifying gene expression. If a reference genome is not available, or if we wish to quantify known transcripts only, reads can be mapped to the transcriptome instead (Korpelainen et al., 2014). Mapping can be quite challenging for a myriad of reasons, namely: 1) sequencing reads are relatively short and there are millions of them; 2) genomes can be large (e.g. mammalian genomes) and contain non-unique sequences (e.g. repeated sequences and pseudogenes), which increases mapping complexity; 3) mismatches and indels (defined as an insertion or deletion of bases in the genome of an organism) can be present, caused by genomic variation and sequencing errors; and 4) RNA-seq reads align to the genome non-contiguously (given that introns are removed from RNA during splicing events), meaning that aligners have to place spliced reads across introns and determine exon-intron boundaries correctly (Korpelainen et al., 2014).

The list of available software for each step of the RNA-seq analysis is constantly expanding (e.g.: more than ninety aligners are currently available (Fonseca et al., 2012)). Examples of typical workflows and software available for each step are shown in **Figure 4.28**.

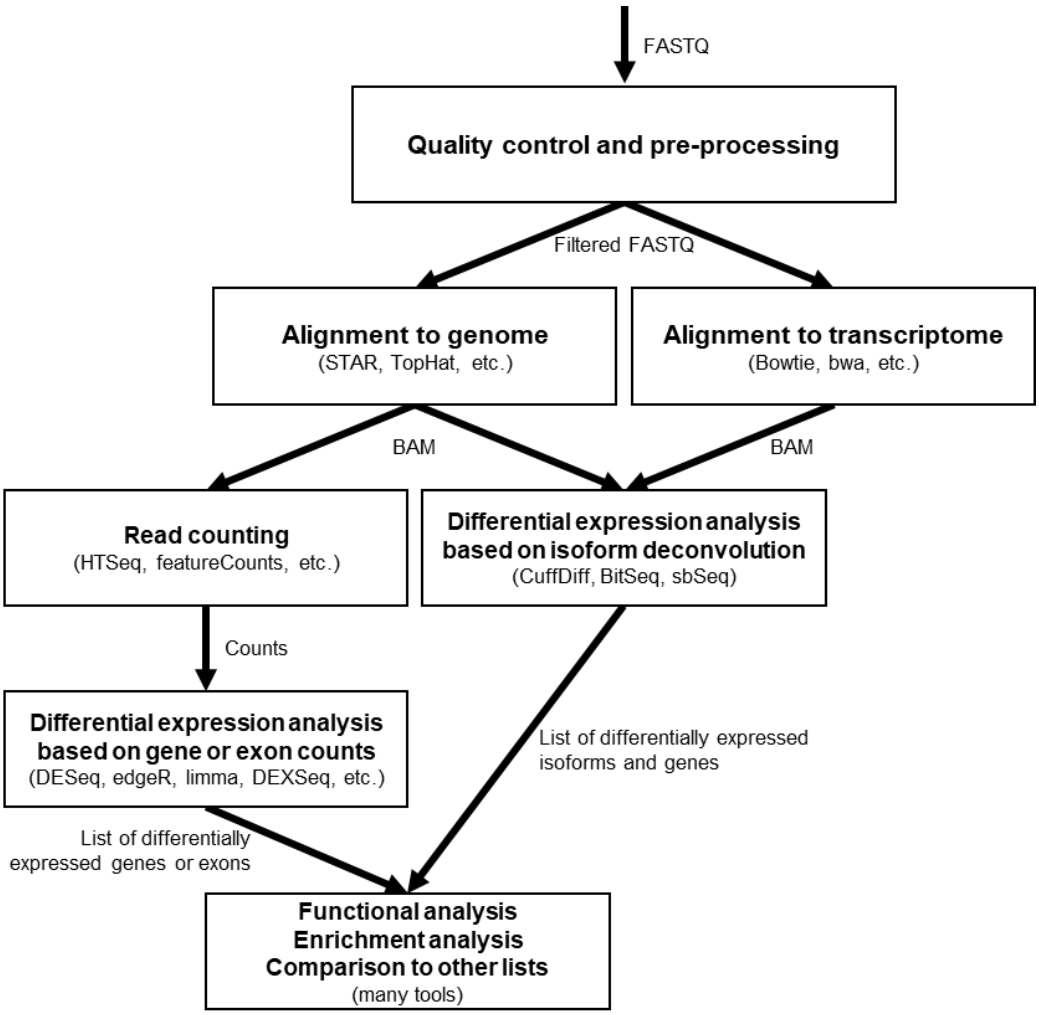


Figure 4.28 – Typical workflows used in RNA-seq data analysis, showing the variety of software available for each step.
Figure adapted from Korpelainen et al. (2014).

As briefly mentioned in the beginning of **Section 4.2**, I tested several approaches based on the analysis pipelines that have been suggested and widely described as the best practices for RNA-seq data analysis (Conesa et al., 2016, Korpelainen et al., 2014). The two main analytical protocols (or "pipelines") that I have explored in higher depth are shown in **Figure 4.29**. First, I used the "Tuxedo" protocol, where trimmed samples were aligned to the mm10 (GRCm38.p4) reference mouse genome using *Tophat* (Trapnell et al., 2009) (version 2.1.1), a program that is built on the short read mapping program Bowtie (Langmead et al., 2009). This pipeline turned out to be suboptimal for our analysis for several reasons (explained in more detail in **Section 4.2.5**), and so I decided to explore extensively other options. Based on the workflow recommended by Bioconductor (www.bioconductor.org) (Love et al., 2016b), additional recommendations by others (Datta and Nettleton, 2014), and my own research about each method, I

elected the “alternative” protocol described on the right side of **Figure 4.29** as the best approach for my RNA-seq data analysis, for which the results described in **Section 4.3** are based on. Within this protocol, I used *STAR* (Dobin et al., 2013) (version 2.5.3a) to align the trimmed *FASTQ* samples to the mm10 (GRCm38.p4) reference mouse genome, which performed a much faster mapping of the reads compared to *Tophat2*, and resulted on an average mapping ≥ 85% (**Figure 4.30**, **Figure 4.31**, **Supplementary Table 1** and **Supplementary Table 2**). Both *TopHat2* and *STAR* are spliced aligners, meaning that both aligners can map RNA-seq reads to genomes that contain introns (Korpelainen et al., 2014).

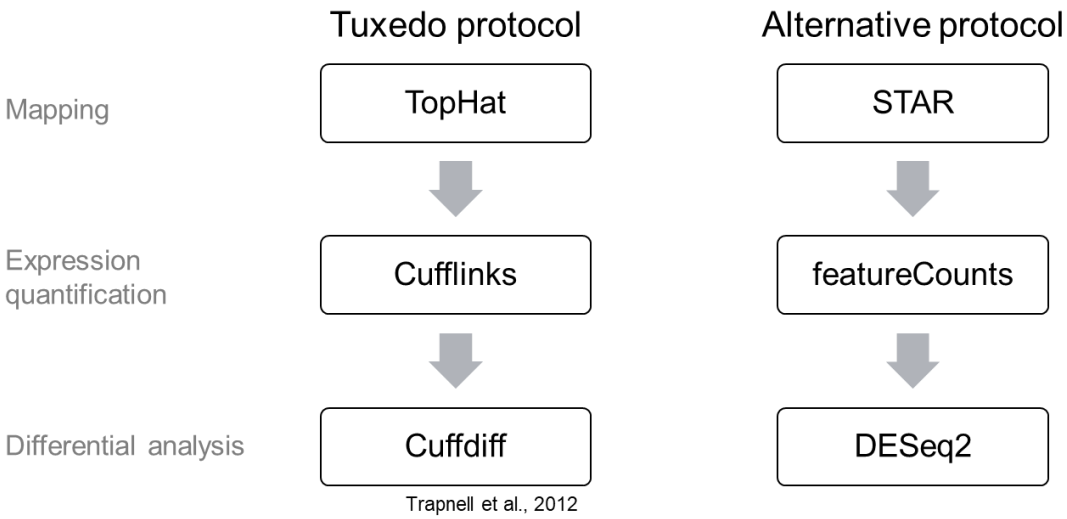


Figure 4.29 – The two protocols (or pipelines) tested for analysis of my RNA-seq data.

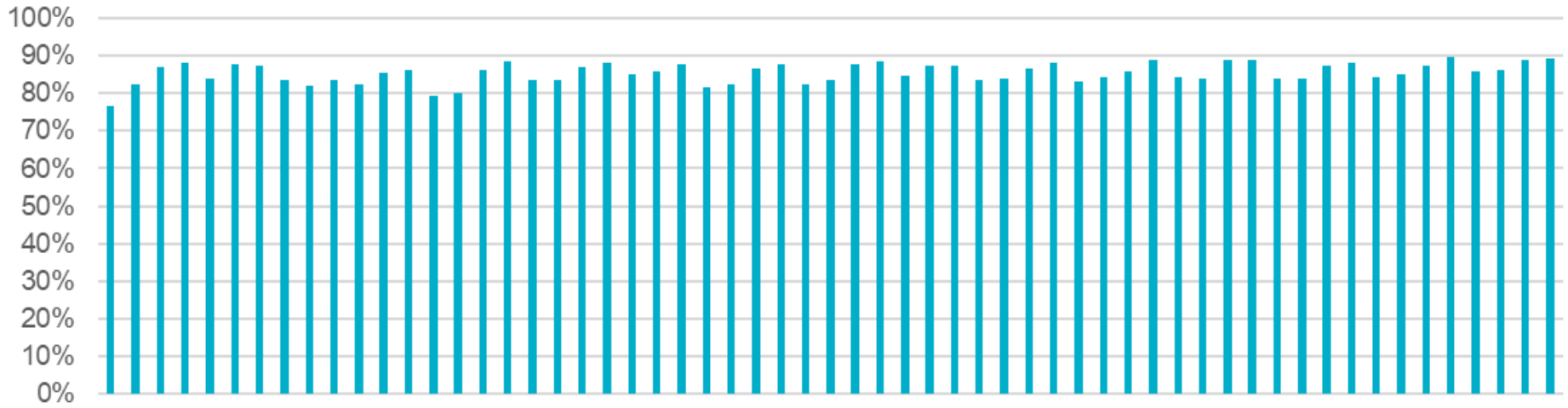


Figure 4.30 – Ratio of filtered sequences that mapped to the mouse reference genome for each sample in the rTg4510 RNA-seq dataset.

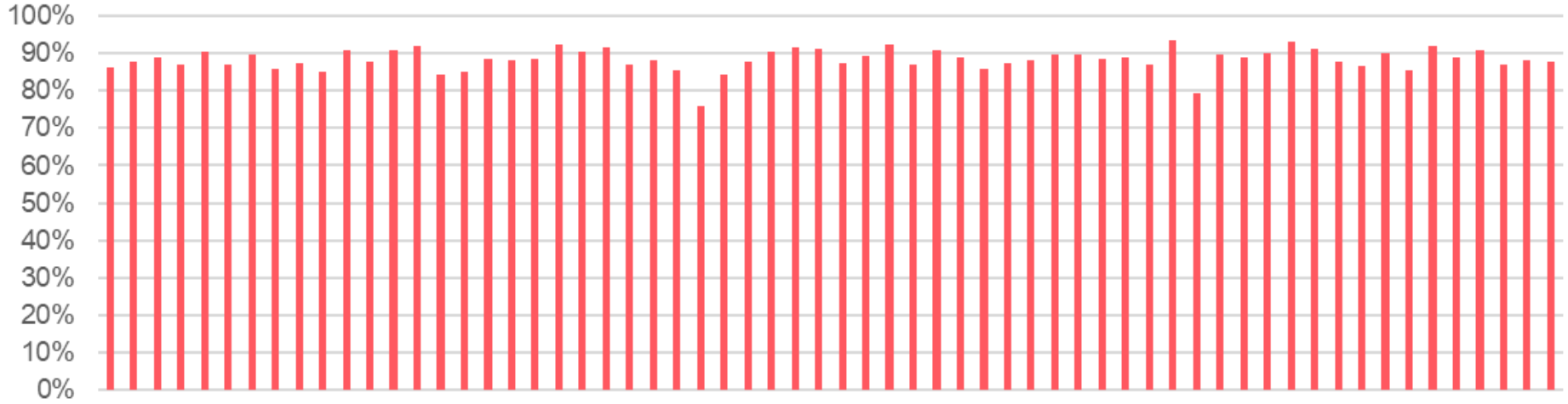


Figure 4.31 – Ratio of filtered sequences that mapped to the mouse reference genome for each sample in the J20 RNA-seq dataset.

4.2.5. Differential expression

4.2.5.1. Gene expression analysis using the "Tuxedo" pipeline

As explained in **Section 4.2.4.4** and **Figure 4.29**, the "Tuxedo" protocol was the first workflow that I tested, where the "classic" workflow includes read mapping with *TopHat*, assembly with *Cufflinks*, and visualization and exploration of results with *CummeRbund* (**Figure 4.32**) (Trapnell et al., 2012). *Tophat*, used for alignment of the sequencing reads, was described in **Section 4.2.4.4**. *Cufflinks* (Trapnell et al., 2012, Trapnell et al., 2010), a tool for transcriptome assembly and isoform quantification from RNA-seq reads, combines the deconvolution of expression data into isoforms with differential expression testing (Korpelainen et al., 2014); it is both the name of a suite of tools and a program within that suite., where the program assembles transcriptomes from RNA-Seq data and quantifies their expression. *CummeRbund*, originally designed for microarray data, is an R visualisation package designed to simplify the analysis of RNA-seq data output from *Cufflinks*.

Despite allowing me to perform a preliminary analysis of my RNA-seq data relatively easily, this workflow turned out to have very little flexibility, with very limited opportunity to perform more complex statistical analyses. For example, it only allowed me to compare samples by genotype (WT versus TG) at each time point, without offering me the possibility of implementing a statistical model containing an interaction term between genotype and age. P-values below 10^{-5} were also not possible to obtain with the latest version of the workflow, which meant that I ended up with many genes with a p-value of 1×10^{-5}. Furthermore, despite normalising the data for transcript length, this workflow does not include filtering of outliers, resulting in the presence of a few clear false positives within my results. For all these reasons I concluded that the Tuxedo protocol was inadequate for the implementation of my final analysis and I decided to explore an alternative approach.

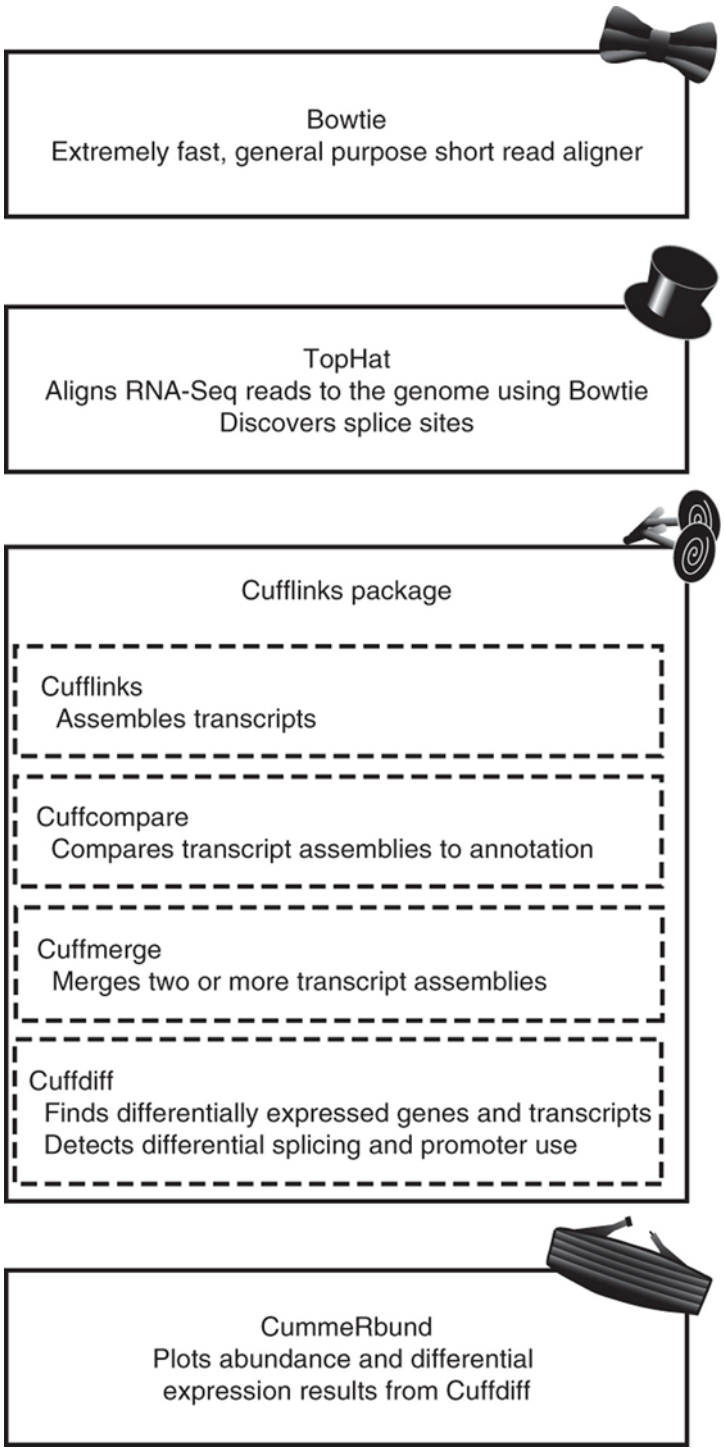


Figure 4.32 – General overview of the Tuxedo protocol and its software components.
Bowtie forms the algorithmic core of the TopHat, aligner. TopHat's read alignments are assembled by Cufflinks and its associated utility program to produce a transcriptome annotation of the genome. Cuffdiff quantifies this transcriptome across multiple conditions using the TopHat read alignments. CummeRbund facilitates exploration and visualisation of the data produced by Cuffdiff, including differentially expressed genes and transcripts. Figure taken, and legend adapted from Trapnell et al. (2012).

4.2.5.2. Gene expression analysis using featureCounts and DESeq2

Gene expression quantification (i.e. the quantification of fragments or templates, hereby referred as read counts) was achieved using *featureCounts* (Liao et al., 2014) (version 1.5.2), a read summarization program that counts mapped reads for genomic features such as genes, exons, promoters, gene bodies, genomic bins and chromosomal locations.

Read counts were analysed for differential expression using the R package *DESeq2* (Love et al., 2014) (version 1.16.1) downloaded from Bioconductor (www.bioconductor.org). *DESeq2* uses the raw read counts, applies an internal normalisation method, and does estimation of library size, estimation of dispersion, and negative binomial generalised linear model fitting. The internal normalisation consists of calculating the geometric mean for each gene across all samples, with the counts for a gene in each sample being divided by this mean. The median of these ratios in a sample is the size factor for that sample. This procedure corrects for library size and RNA composition bias, which can arise for example when only a small number of genes are very highly expressed in one experiment condition but not in the other. Furthermore, *DESeq2* uses shrinkage estimation for dispersions and fold changes, with a dispersion value estimated for each gene through a model fit procedure, only possible if an appropriate number of replicates is included in the experiment. For every gene and for every sample, *DESeq* calculates Cook's distance, a diagnostic test for outliers, automatically flagging genes which contain a Cook's distance above a cutoff for samples which have three or more replicates. The p-values for these genes are set to "NA". When there are seven or more replicates, *DESeq* automatically replaces counts with large Cook's distance with the trimmed mean over all samples, scaled up by the size factor or normalisation factor for that sample, i.e., it replaces the outlier value with the value predicted by the null hypothesis (Love et al., 2016a).

Datasets were filtered for non-expressed and lowly expressed genes (minimum of 6 counts across all samples), and similarity in the genome-wide expression profile between samples was visualised in a heatmap clustered by Euclidean distance (**Figure 4.34** and **Figure 4.35**) and a principal component analysis (PCA) plot of the first two principal components (**Figure 4.36** and **Figure 4.37**). I was interested in detecting both genotype effects and progressive changes

across age between the transgenic and wild type samples (**Figure 4.33**). For that purpose, I used the following statistical model, including main effects for both Genotype and Age (both coded as categorical variables) and an interaction between these two terms:

$$Gene\ expression = Genotype + Age + Genotype * Age$$

To identify significant Genotype effects a Wald test was used, and to identify significant effects of age and interaction effects (i.e. Genotype*Age) we used the likelihood-ratio test, both applied with the *DESeq* function from the *DESeq2* package (Love et al., 2014). *P* values were adjusted for multiple testing, using the false discovery rate (FDR) method (also known as Benjamini and Hochberg correction) (Benjamini and Hochberg, 1995) implemented with the R function *p.adjust*; FDR-adjusted *P* values < 0.05 were defined as significant. Potential differences in proportions of genes with increased versus decreased expression as well as overlapping fold changes in both models were interrogated using the binomial test. Functional annotation and gene ontology analyses were done with *GOseq* (Young et al., 2010*)* (1.30.0), based on genes with FDR < 0.05. Goseq is an approach for gene ontology analysis of RNA-seq data, which takes gene length bias into account (Oshlack and Wakefield, 2009). All analyses were performed in R (version 3.4.3).

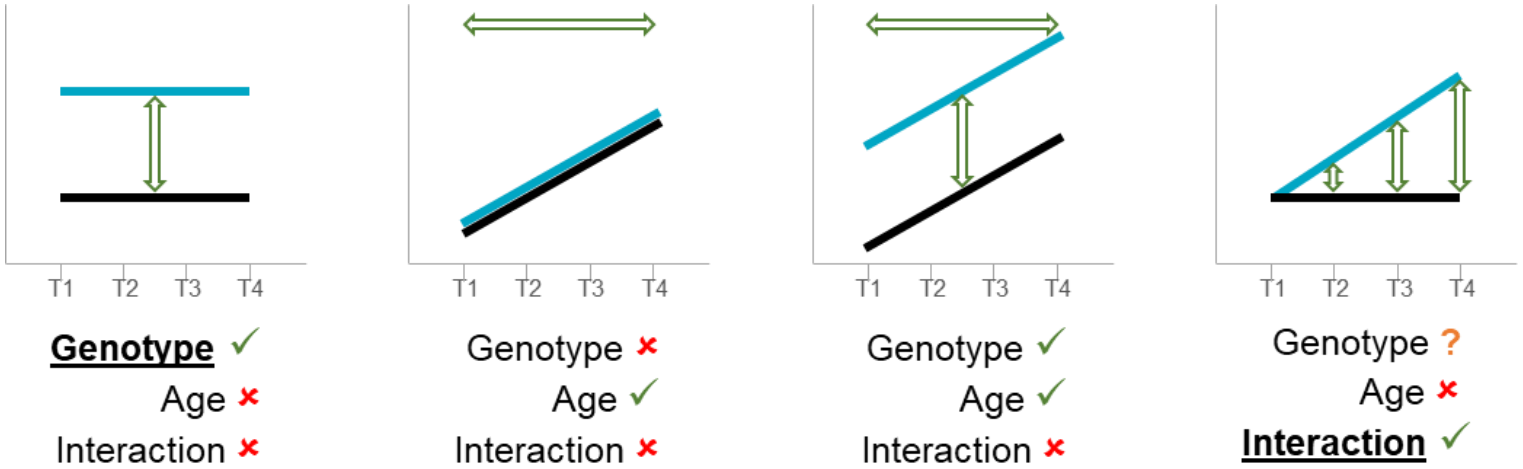


Figure 4.33 – Strategy for the statistical approach used.

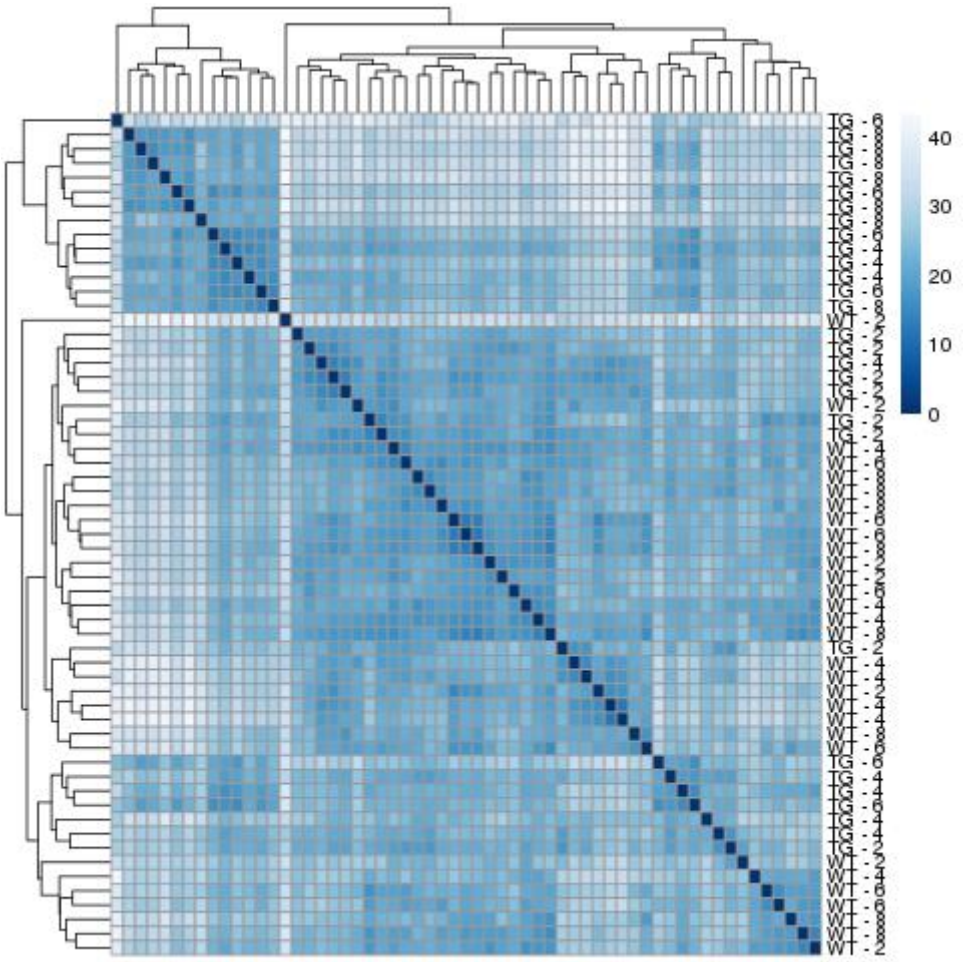


Figure 4.34 – Euclidean distance between samples in rTg4510 mice.
Heatmap of sample-to-sample distances using rlog-transformed values.

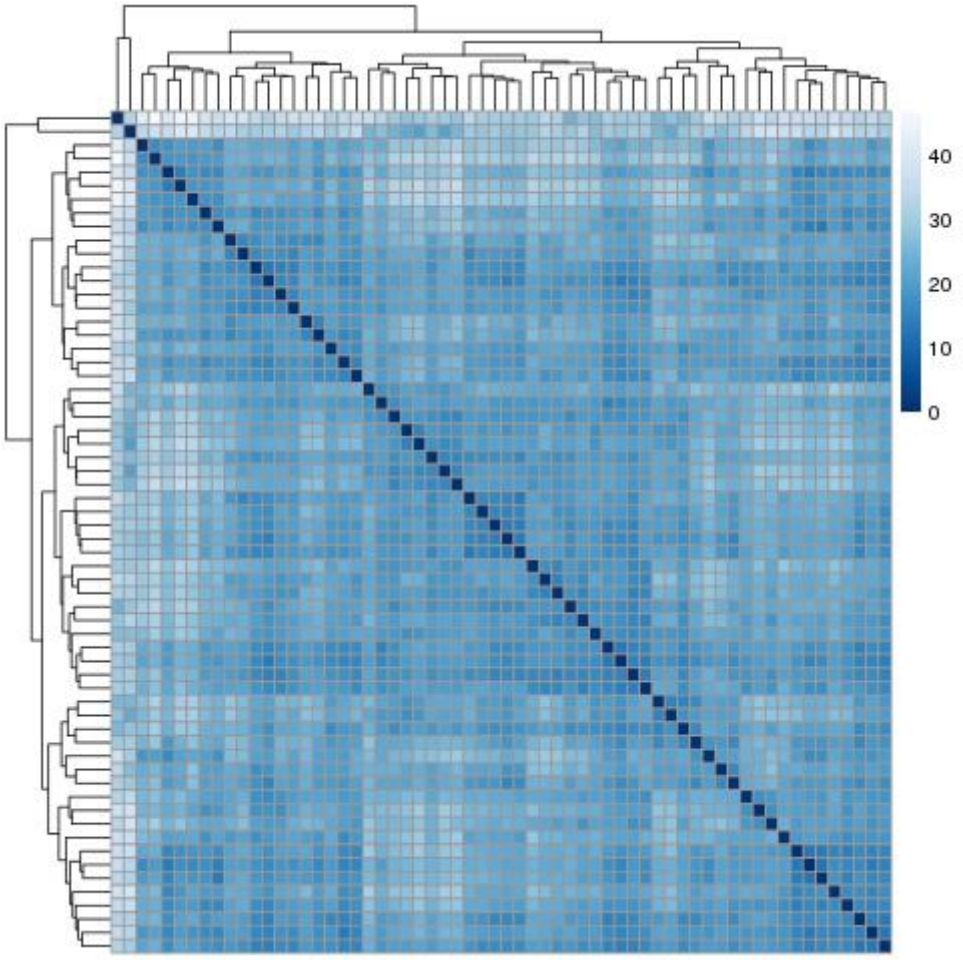


Figure 4.35 – Euclidean distance between samples in J20 mice.
Heatmap of sample-to-sample distances using rlog-transformed values.

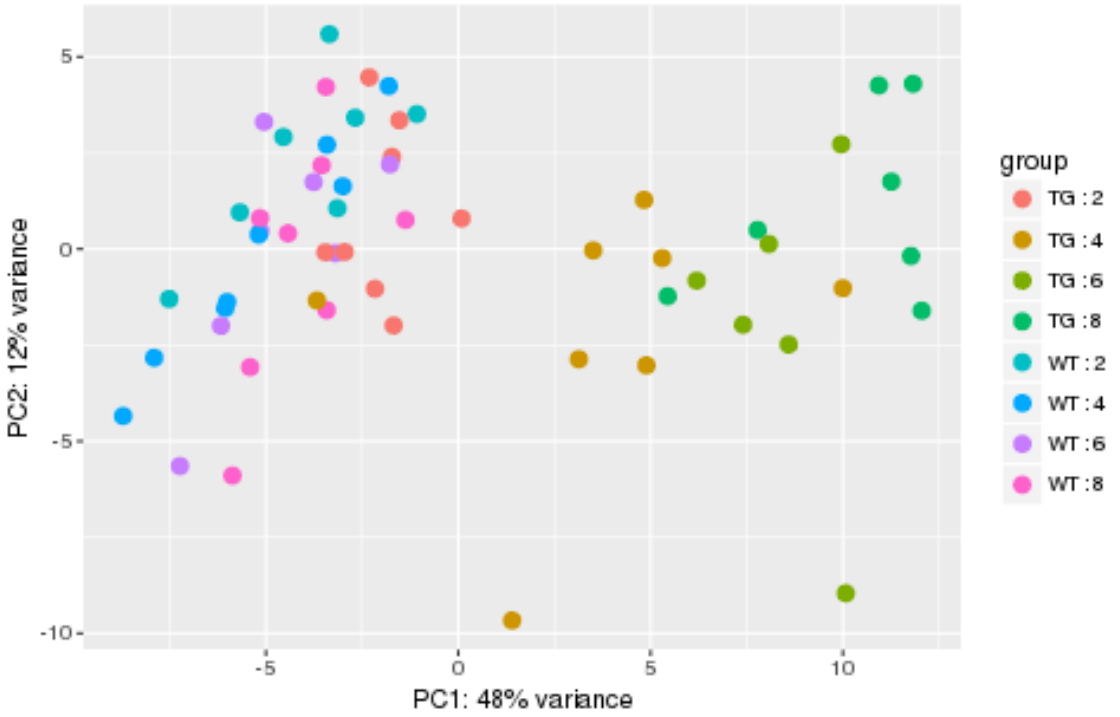


Figure 4.36 – Principal component analysis (PCA) plot of the first two principal components in rTg4510 samples.

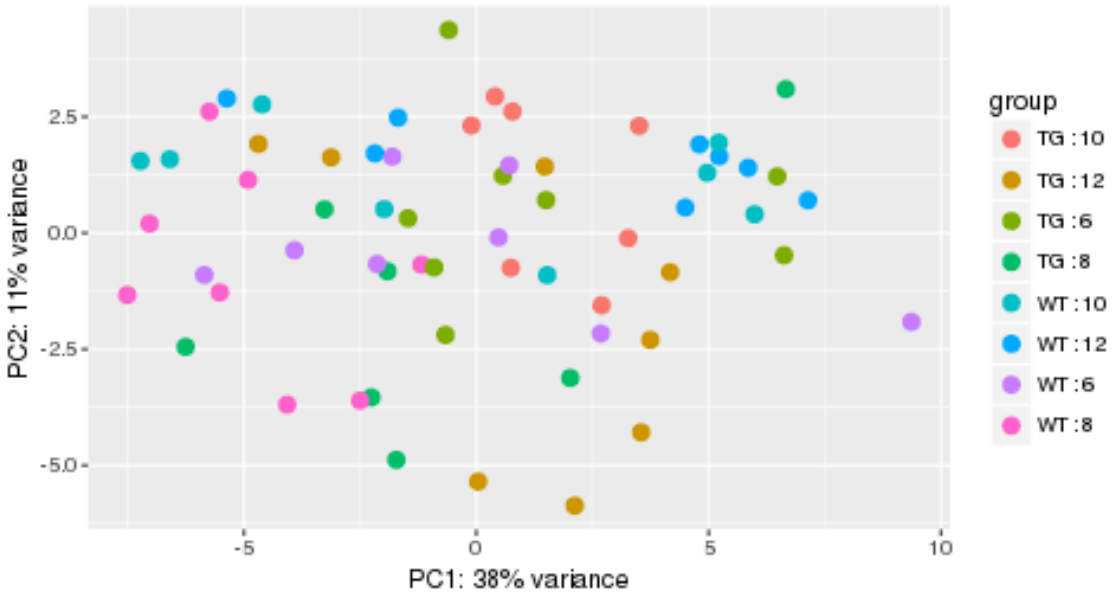


Figure 4.37 – Principal component analysis (PCA) plot of the first two principal components in J20 samples.

4.2.5.3 Quantifying human transgene expression

Mouse and human *App*/*APP* and *Mapt*/*MAPT* sequences were compared using BLAT (Kent, 2002) for divergent transcript sequences representing specific mouse and human gene sequences. Two 200bp regions spanning 4 exons were chosen as representative of mouse-specific *App*. Similar regions consisting of two 200bp exonic regions were also chosen for human *APP*. Mouse-specific *Mapt* and human-specific *MAPT* sequences were chosen from a 2kb region present in the 3'UTR. Using *Bowtie2* (Langmead et al., 2009) (version 2.3.4.3), indices based on these sequences were then built, and alignments were performed using the *FASTQ* read 1 sequences. Counts of read alignments for mouse and human specific indices were then plotted as a ratio of unique (mouse or human) reads relative to the total number of input reads.

4.2.5.4. Comparison with RNA-seq data from the Mouseac database

RNA-seq data (transcripts per million, TPM) from two mouse models (Matarin et al., 2015, Salih et al., 2018) (TAU (Camk2a-MAPTP301L) and TAS10 (SwAPP, K670N/M671L)) were downloaded from the Mouseac online database (www.mouseac.org), with corresponding detailed phenotypic data downloaded from GEO (Barrett et al., 2013, Edgar et al., 2002) (accession number GSE64398). Only genes identified as differentially expressed (FDR < 0.05) in our analysis (for rTg4510 and J20 mice) were kept for further statistical analysis. TPM was log transformed (log2(x+1)) and the same linear regression model described above ($Gene\ expression = Genotype + Age + Genotype * Age$), using ANOVA to test for significant differences associated with either the Genotype, Age, or Genotype*Age terms, was used. *P* values were corrected for the number of genes compared across datasets using Bonferroni correction. Potential differences in proportions of genes with increased versus decreased expression as well as overlapping fold changes in both models were interrogated using the binomial test.

4.2.6. Co-expression network analysis

Gene co-expression changes were detected using weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008, Zhang and Horvath, 2005), a systems biology method for evaluating correlation expression patterns among genes across samples. It can be used 1) to find clusters (modules) of highly correlated genes; 2) to summarise those clusters using the module eigengene or an intramodular hub gene; 3) to relate modules to other modules; 4) to relate modules to external sample traits (using eigengene network methodology); and 5) to calculate module membership measures.

Using the *WGCNA* R package (Langfelder and Horvath, 2008) (version 1.63), I constructed a signed co-expression network for each mouse model using log transformed counts from all samples. Logarithmic transformation of raw counts was achieved using the rlog function from *DESeq2*, which minimises variability in genes with low counts (Love et al., 2014). I checked the data for missing values and outliers, and removed one sample from the analysis for the rTg4510 dataset (flagged as an outlier) before building the networks. Signed WGCNA co-expression networks were built using the lowest power for which the scale-free topology fit index curve flattened out after reaching 0.90 resulting in a soft-threshold power of 10 and 9 for rTg4510 and J20 datasets, respectively, and a minimum module size of 30. For each module of highly interconnected genes, colour-labelled according to the WGCNA conventions, I calculated the module eigengenes (MEs) as the first principal component of the expression matrix, which provide a representative expression profile for each module (Langfelder and Horvath, 2008). In order to identify modules significantly associated with pathology burden, I calculated correlation coefficients between these MEs with the available pathology data and explored the most significant associations. In addition we used the same linear regression model as described for the gene-level analysis ($ME = Genotype + Age + Genotype * Age$) using ANOVA to test for significant differences due to either the Genotype, Age, or Genotype*Age terms. *P* values were corrected for multiple comparisons using the Bonferroni correction, to correct for 18 modules in the rTg4510 (statistical threshold was adjusted to 0.05/18 = 0.0028) and 21 modules in the J20 (statistical threshold was adjusted to 0.05/21 = 0.0024) mice. I used the *GOseq* R package (version 1.30.0) to perform functional annotation and gene ontology (GO) analyses for

each module, where significant pathways were selected using an FDR threshold of 0.05 as previously described (Young et al., 2010). Finally, I used *Cytoscape* (Shannon et al., 2003) (version 3.7.0) for network visualisation using the topological overlap matrix for the log transformed expression data.

4.2.6.1. Comparison with human co-expression networks

Significant rTg4510 co-expression modules were compared to AD-associated co-expression modules, reported in a recent human post-mortem RNA-seq meta-analysis (Logsdon et al., 2019), which was done in collaboration with Emma Laing from Eli Lilly. The six rTg4510-associated co-expression modules identified in this study ("salmon", "turquoise", "purple", "yellow", "light-cyan", and "red", see **Section 4.3.6**), and AD-associated human co-expression modules in the dorsolateral prefrontal cortex (DLPFC) and temporal cortex (TCX) from Logsdon et al. were reduced to contain only mouse-human homologs as defined by *Ensembl* (Zerbino et al., 2018) (accessed on 14/11/2018). The level of overlap between gene members of each pair of modules was assessed via a hypergeometric test using the R (version 3.5.1) function *phyper*. *P* values were corrected for multiple comparisons using Bonferroni correction in a tissue-specific manner, where only the set of raw *P* values related to DLPFC (statistical threshold was adjusted to 0.05/24 = 0.0021) or TCX (statistical threshold was adjusted to 0.05/30 = 0.0017) modules' overlap were considered. Using *GOseq* (version 1.30.0), we performed functional annotation and GO analyses for the common genes in each overlapping pair of modules.

4.2.7. Iba1 immunohistochemistry

Given the enrichments for immune-related biological pathways amongst the genes characterised progressive gene expression changes identified in the rTg4510 mice (see **Section 4.3.4**), I quantified Iba1, a microglia/macrophage-specific calcium-binding protein (Ohsawa et al., 2004), in matched tissue sections from the right brain hemisphere (n = 7-10 animals per group, total n = 70). This was performed as described in **Chapter 3** (**Section 3.2**) for PG-5, except that I used rabbit polyclonal anti-Iba1 (Wako, catalog number 019-19741, lot number LKG5732) as the primary antibody (diluted 1:6000), and biotinylated goat anti-

rabbit IgG (Vector labs, catalog number BA-1000, lot number ZB1007) as the secondary antibody (diluted 1:200).

4.3. Results

4.3.1. Sequencing metrics

Raw RNA-seq data are available for download from the Gene Expression Omnibus (GEO) database (Barrett et al., 2013, Edgar et al., 2002) (accession number GSE125957). No differences in read-depth were identified between TG and WT controls in both the rTg4510 (n = 59 animals, two-tailed unpaired t-test, t(57)= 1.35, *P* = 0.18) and J20 (n = 62 animals, two-tailed unpaired t-test, t(60) = 0.41, P = 0.18) mouse models (**Figure 4.38**). In addition, no differences in the number of sequences mapped to the mouse genome were identified when comparing TG to WT samples in both rTg4510 (n = 59 animals, two-tailed unpaired t-test, t(57) = -1.36, P = 0.089) and J20 (n = 62 animals, two-tailed unpaired t-test, t(60) = 0.79, P = 0.22).

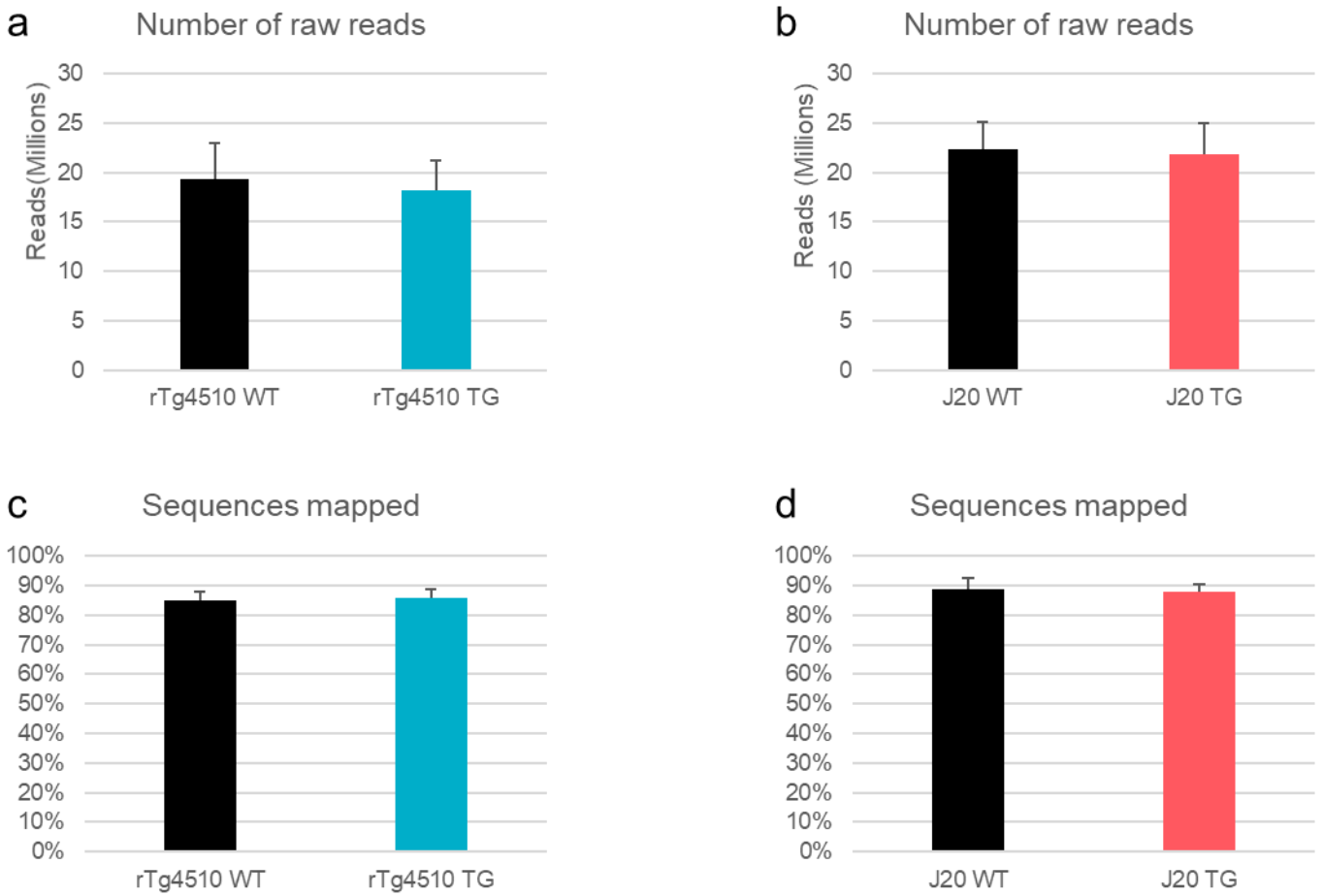


Figure 4.38 – Comparison of RNA-seq reads in transgenic and wild type animals for both models.
The number of raw RNA-seq reads did not differ between WT and TG mice for comparisons of either the **(a)** rTg4510 or **(b)** J20 models. The proportion of RNA-seq reads uniquely mapped to the mouse genome did not differ between WT and TG mice for comparisons of either the **(c)** rTg4510 or **(d)** J20 models.

4.3.2. The rTg4510 model of tau pathology is characterised by widespread transcriptional differences in the entorhinal cortex

Across all samples, striking differences in gene expression were identified in rTg4510 TG animals relative to WT control mice (n = 29 TG, n = 30 WT); gene expression results for all 18,822 detected transcripts are available to download from our online database (www.epigenomicslab.com/ADmice). In total, we identified 154 differentially-expressed transcripts at false discovery rate (FDR) < 0.05 (**Figure 4.39** and **Supplementary Table 3**). Among these, there was a significant (exact binomial test, n = 154 transcripts, *P* = 0.00014) enrichment of transcripts with decreased expression (n = 101 (66%) transcripts with reduced expression in TG compared to n = 53 (34%) transcripts with elevated expression in TG). Of note, differences for five of these transcripts are likely to reflect known deletions of the transgene integration sites for the Camk2a-tTA (encompassing *Wdr60*, *Esyt2*, *Ncapg2*, and *Ptprn2*) and *MAPT* (encompassing *Fgf14*) transgenes (Goodwin et al., 2017). Given the high homology between transcribed regions of the human and mouse tau gene, I also found highly elevated levels of *Mapt* (Wald test, Wald statistic = 11.11, log2 fold change = 0.50, FDR = 7.08E-25) (**Figure 4.40a**) confirming stable activation of the *MAPT* transgene in TG mice; of note, human-specific *MAPT* sequence domains were only detected in TG RNA-seq datasets (**Figure 4.40b-c**). Furthermore, because the rTg4510 transgene is inserted into the context of two untranslated exons of the mouse prion protein gene (*Prnp*), as expected we observed elevated expression of these *Prnp* domains in TG mice (Wald statistic = 25.40, log2 fold change = 1.54, FDR = 4.88E-138).

Beyond these expected direct transgene-induced changes, we observed evidence for widespread transcriptional consequences of the rTg4510 genotype. The most significant rTg4510-associated differentially-expressed transcript is *Car4*, which encodes carbonic anhydrase 4 (**Table 4.2** and **Figure 4.41a**; increased expression in TG mice). Other differentially expressed genes in mice carrying the rTg4510 transgene include *Gpr17*, which encodes the G protein-coupled receptor 17 that is involved in regulating oligodendrocyte differentiation and maturation (Chen et al., 2009) (**Table 4.2** and **Figure 4.41b**; decreased expression in TG mice;); *Blnk*, which encodes a cytoplasmic linker protein that

plays a critical role in B cell development and is involved in the TREM2 activation pathway (Zajkowicz et al., 2018) (**Table 4.2** and **Figure 4.41c**; increased expression in TG mice); and *Hspa5* (also known as *Bip* or *Grp78*), which encodes a member of the heat shock protein 70 (HSP70) family that is localised in the lumen of the endoplasmic reticulum (ER) and involved in the folding and assembly of proteins, and has been previously implicated in neuroprotection and AD (Casas, 2017, Hoozemans et al., 2005) (**Table 4.2** and **Figure 4.41d**; decreased expression in TG mice).

Hierarchical clustering of individual mice based on expression levels for genotype-associated transcripts robustly discriminates between rTg4510 and WT groups (**Figure 4.39**). Within the rTg4510 TG group, samples also cluster by time-point, suggesting, importantly, that there are progressive changes in gene expression within the mutant mice, and highlighting the value of performing longitudinal analyses.

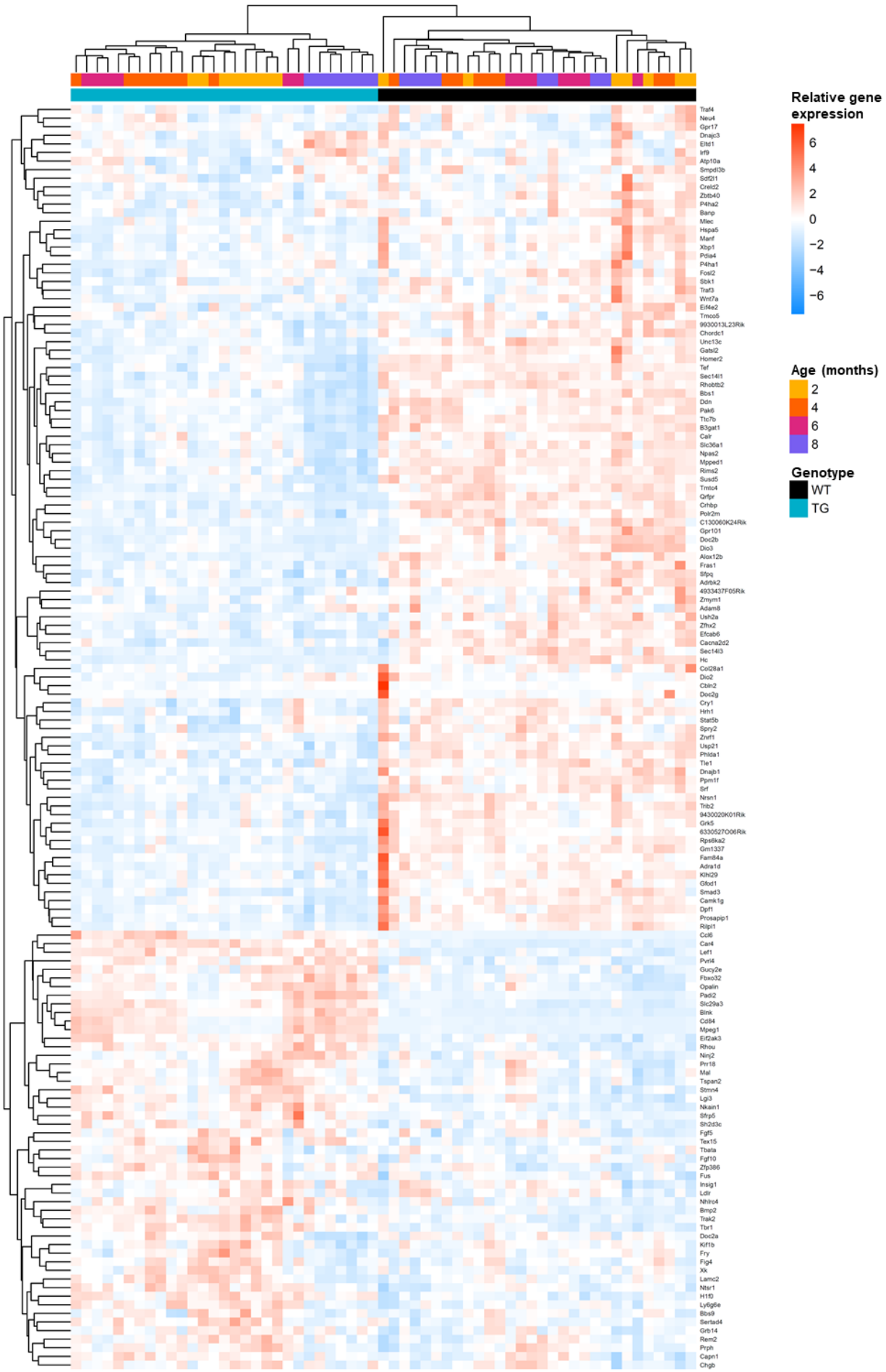

Relative gene expression
6
4
2
0
−2
−4
−6
Age (months)
2
4
6
8
Genotype
WT
TG
Traf4
Neu4
Gpr17
Dnajc3
Eltd1
Irf9
Atp10a
Smpdl3b
Sdf2l1
Creld2
Zbtb40
P4ha2
Banp
Mlec
Hspa5
Manf
Xbp1
Pdia4
P4ha1
Fosl2
Sbk1
Traf3
Wnt7a
Eif4e2
Tmco5
9930013L23Rik
Chordc1
Unc13c
Gatsl2
Homer2
Tef
Sec14l1
Rhobtb2
Bbs1
Ddn
Pak6
Ttc7b
B3galt1
Calr
Slc36a1
Npas2
Mpped1
Rims2
Susd5
Tmtc4
Qrfpr
Crhbp
Polr2m
C130060K24Rik
Gpr101
Doc2b
Dio3
Alox12b
Fras1
Sfpq
Adrbk2
4933437F05Rik
Zmym1
Adam8
Ush2a
Zfhx2
Efcab6
Cacna2d2
Sec14l3
Hc
Col28a1
Dio2
Cbln2
Doc2g
Cry1
Heh1
Stat5b
Spry2
Znrf1
Usp21
Phlda1
Tle1
Dnajb1
Ppm1f
Srf
Nrsn1
Trib2
9430020K01Rik
Grk5
6330527O06Rik
Rps6ka2
Gm1337
Fam84a
Adra1d
Klhl29
Gtsd1
Smad3
Camk1g
Dpf1
Prosapip1
Rilpl1
Ccl6
Car4
Lef1
Pvrl4
Gucy2e
Fbxo32
Opalin
Pasl2
Slc29a3
Blnk
Cd84
Mpeg1
Eif2ak3
Rhou
Nin2
Prr18
Mal
Tspan2
Stmn4
Lgi3
Nkain1
Sfrp5
Sh2d3c
Fgf5
Tex15
Tbata
Fgf10
Zfp386
Fus
Insig1
Ldlr
Nhlrc4
Bmp2
Trak2
Tbr1
Doc2a
Kif1b
Fry
Fig4
Xk
Lamc2
Ntsr1
H1f0
Ly6g6e
Bbs9
Serta04
Grb14
Rem2
Prph
Capn1
Chgb

Figure 4.39 – Genotype-associated transcriptional variation robustly discriminates between rTg4510 transgenic and wild type mice.
Hierarchical clustering of each individual mouse based on expression levels for differentially expressed genes (DEGs) associated with rTg4510 genotype (n = 59 mice (29 TG, 30 WT), 147 transcripts). Direction of normalised DESeq2 read counts, relative to mean levels of expression across all individual mice ("relative gene expression"), is represented in the heatmap (scaled) from high (red) to low (blue).

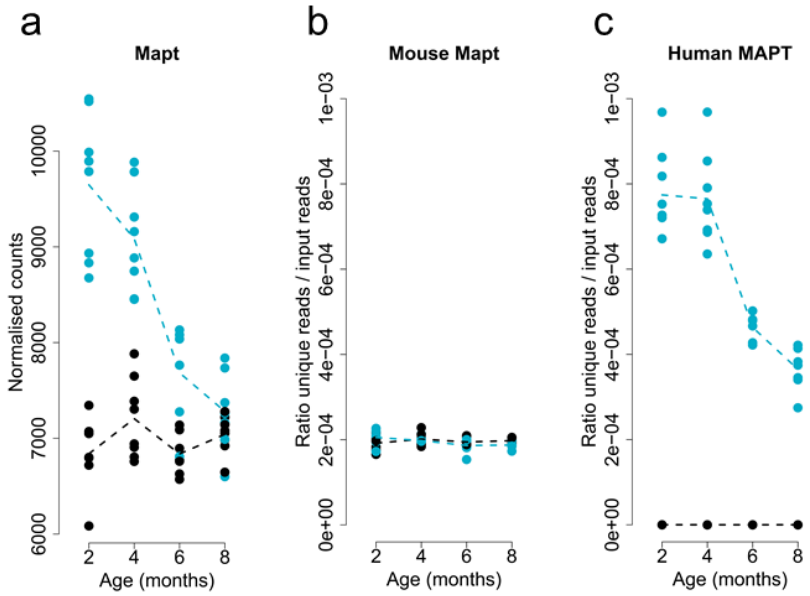


Figure 4.40 – Apparent up-regulation of Mapt in rTg4510 mice results from human-specific sequences only present in transgenic mice.
Apparent up-regulation of Mapt in rTg4510 transgenic (TG, blue) female mice compared to wild type (WT, black) littermate control mice (n = 59 animals). **(a)** *Mapt* gene expression from DESeq2 highlights elevated expression of *Mapt* in TG mice. **(b)** Ratio of unique reads that mapped to mouse-specific sections of *Mapt* or **(c)** human-specific sections of *MAPT*. Normalised counts were obtained using DESeq2. Dashed lines represent mean paths across age groups.

Table 4.2 – Top-ranked differentially expressed transcripts associated with rTg4510 genotype.

Transcript	Statistical test	Wald statistic	log2 fold change	FDR	N
Car4	Wald test	8.36	1.11	2.41E-13	59
Gpr17		-6.73	-0.62	5.11E-08	
Blnk		6.48	0.80	2.12E-07	
Hspa5		-6.16	-0.58	1.37E-06	

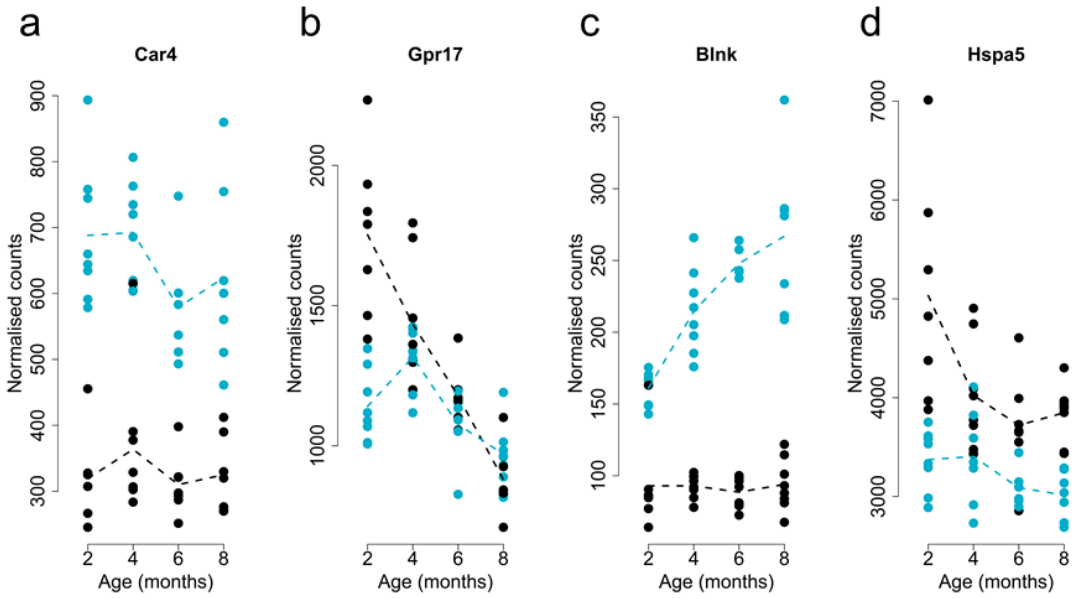


Figure 4.41 – Top-ranked differentially expressed transcripts associated with rTg4510 genotype.
Shown are individual plots for: **(a)** *Car4*, **(b)** *Gpr17*, **(c)** *Blnk*, and **(d)** *Hspa5.* Total n = 59 animals (29 TG, 30 WT). Normalised counts were obtained using DESeq2. Dashed lines represent mean paths across age groups. rTg4510 transgenic (TG, blue) female mice compared to wild type (WT, black) littermate control mice.

4.3.3. The J20 model of amyloid pathology is characterised by differential expression of *Ccdc80*, *Abca8a*, *Htr1a* and *Hspa5*

Relative to the widespread transcriptional signatures associated with the rTg4510 model, fewer significant expression differences were identified in J20 TG mice compared to WT control mice (n = 30 TG, n = 32 WT); gene expression results for all 18,745 expressed transcripts are available to download from our online database (www.epigenomicslab.com/ADmice). As expected, there was an apparent increase in expression of *App* (Wald test, Wald statistic = 8.55, log2 fold change = 0.66, FDR = 2.37E-13) (**Figure 4.42a**), reflecting the high sequence homology with the human *APP* transgene, and confirming stable activation of the mutant transgene in TG mice; of note, we mapped our RNA-seq reads to human-specific *APP* sequence domains and only observed signal in TG animals (**Figure 4.42b-c**). In total we identified four additional differentially-expressed transcripts at a stringent false discovery rate (FDR) < 0.05 (**Figure 4.43** and **Supplementary Table 4**): *Ccdc80*, encoding a protein involved in cell adhesion and matrix assembly (O'Leary et al., 2013) (increased expression in TG samples); *Abca8a*, encoding a member of the A-subclass of ATP-binding cassette (ABC) transporter family which regulates brain lipid homeostasis and has been implicated in AD (Piehler et al., 2012) (decreased expression in TG samples); *Htr1a*, encoding a major G-protein-coupled serotonin receptor, the 5-HT_{1A} receptor, that is widely expressed in the central nervous system (decreased in TG samples); and *Hspa5* (decreased expression in TG samples) (**Table 4.3** and **Figure 4.44**). Overall, expression of these genotype-associated transcripts discriminates between J20 and WT groups (**Figure 4.43**), although, in contrast to the rTg5410 differentially expressed transcripts, there are no clear age effects in the J20 TG mice. Even though the transcriptional changes associated with the rTg4510 and J20 genotypes are generally distinct – there is no robust correlation of effect sizes (TG vs WT) between models for differentially expressed transcripts identified in either the rTg4510 (Pearson correlation, r = 0.15, *P* = 0.063, **Figure 4.45a**) or J20 (r = 0.66, *P* = 0.23, **Figure 4.45b**) models – it is noteworthy that the expression of *Hspa5* is significantly decreased (FDR < 0.05) in the same direction in both models (**Figure 4.41d** and **Figure 4.44d**), implicating a role for ER stress in both mouse models.

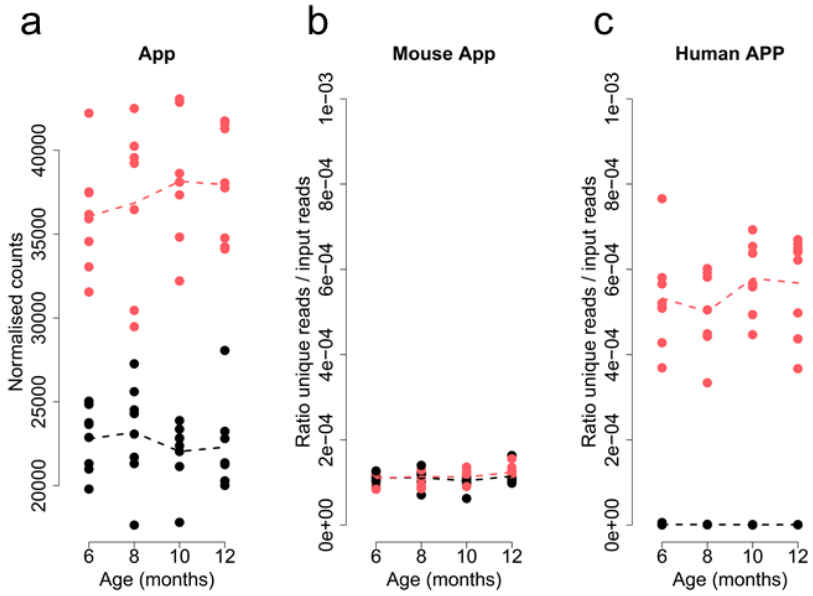


Figure 4.42 – Apparent up-regulation of App in J20 mice results from human-specific sequences only present in transgenic mice.
Apparent up-regulation of *App* in J20 transgenic (TG, red) female mice compared to wild type (WT, black) littermate control mice (n = 62 animals). **(a)** *App* gene expression from DESeq2 highlights elevated expression of *App* in TG mice. **(b)** Ratio of unique reads that mapped to mouse-specific sections of *App* or **(c)** human-specific sections of *APP*. Normalised counts were obtained using DESeq2. Dashed lines represent mean paths across age groups.

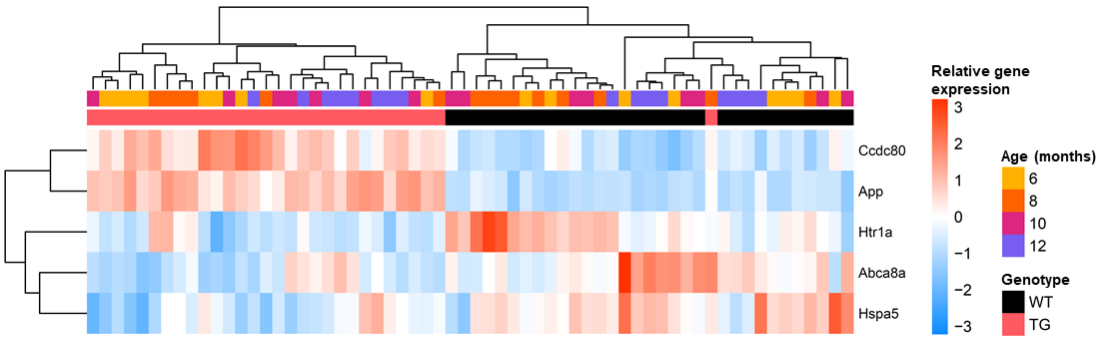


Figure 4.43 – Genotype-associated transcriptional variation robustly discriminates between J20 transgenic and wild type mice.
Hierarchical clustering of each individual mouse based on expression levels for differentially expressed genes (DEGs) associated with J20 genotype (n = 62 mice (30 TG, 32 WT), 5 transcripts). Direction of normalised DESeq2 read counts, relative to mean levels of expression across all individual mice ("relative gene expression"), is represented in the heatmap (scaled) from high (red) to low (blue).

Table 4.3– Top-ranked differentially expressed transcripts associated with J20 genotype.

Transcript	Statistical test	Wald statistic	log2 fold change	FDR	N
Ccdc80	Wald test	6.37	0.81	1.74E-06	62
Abca8a		-4.67	-0.81	0.020	
Htr1a		-4.48	0.51	0.035	
Hspa5		-4.36	-0.28	0.049	

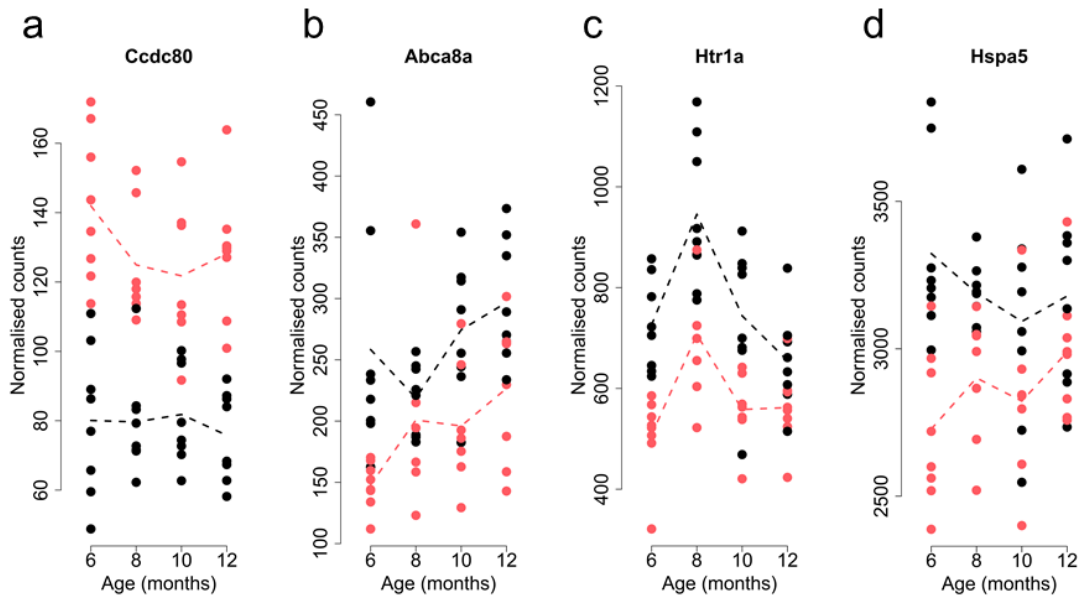


Figure 4.44 – Top-ranked differentially expressed transcripts associated with J20 genotype.
Shown are individual plots for: **(a)** *Ccdc80*, **(b)** *Abca8a*, **(c)** *Htr1a*, and **(d)** *Hspa5*. Total n = 62 animals (30 TG, 32 WT). Normalised counts were obtained using DESeq2. Dashed lines represent mean paths across age groups. J20 transgenic (TG, red) female mice compared to wild type (WT, black) littermate control mice.

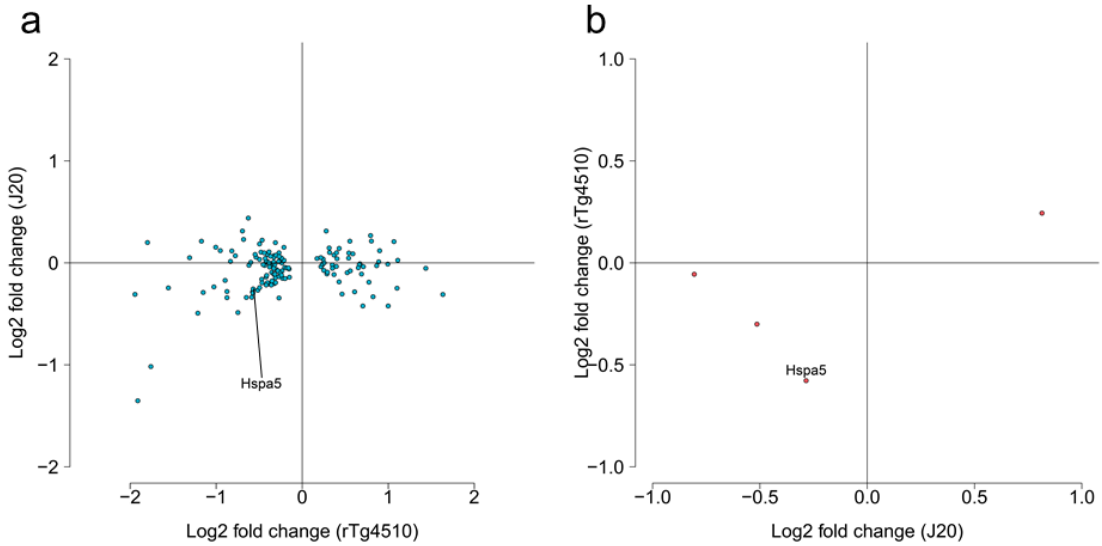


Figure 4.45 – There is no overall consistency between mouse models in effect sizes for genotype-associated differentially-expressed transcripts identified in each individual mouse model.
(a) Negligible correlation for effect size (Log2 fold change) across models for differentially expressed transcripts (FDR < 0.05) associated with rTg4510 genotype (Pearson correlation, r = 0.18, *P* = 0.029; exact binomial test, n = 147 transcripts, *P* = 0.69). **(b)** No correlation for effect size (Log2 fold change) across models for differentially expressed transcripts (FDR < 0.05) associated with J20 genotype (Pearson correlation, r = 0.66, *P* = 0.23; exact binomial test, n = 4 transcripts, *P* = 0.13). Despite the overall lack of consistency between models, the expression of one transcript (*Hspa5*) is differentially decreased (FDR < 0.05) in the same direction in both models.

4.3.4. Progressive changes in gene expression in the entorhinal cortex mirror the development of tau neuropathology in rTg4510 mice

Given the progressive accumulation of brain neuropathology in TG mice (see **Section 3.3.1** in **Chapter 3**), I next explored temporal changes in gene expression associated with genotype to identify transcriptional signatures paralleling the increases in tau and amyloid pathology in TG mice over time (**Figure 4.33**). I initially focused on the rTg4510 mice given the clear temporal clustering of samples amongst genotype-associated differentially-expressed transcripts identified in this model (**Figure 4.39**). Using an approach designed to identify interactions between genotype (TG vs WT) and age group, we identified 1,762 transcripts (FDR < 0.05) whose expression significantly changed with the progression of tau pathology in rTg4510 mice (**Supplementary Table 5**).

Expression differences at these transcripts were found to progressively increase with age relative to baseline (age 2 months) (**Figure 4.46** and **Table 4.4**), paralleling the accumulation of tau pathology in these same individual animals. The top tau-associated differentially-expressed gene in rTg4510 TG mice was *Gfap*, encoding glial fibrillary acidic protein (GFAP), a gene predominantly expressed in both mouse and human astrocytes (Raff et al., 1979, Zhang et al., 2016), and known to be upregulated in reactive astrocytes associated with brain pathology (Ben Haim et al., 2015). Expression of *Gfap* was dramatically amplified with progressive tau pathology (**Table 4.5** and **Figure 4.47a**), similar to results from another study reporting age-dependent (12-18 months) increase in expression of hippocampal *Gfap* in tau (Camk2a-MAPT P301L) and amyloid (*APP*/*PSEN1*) mouse models (Matarin et al., 2015), and paralleling the astrogliosis observed in human AD brain (Liddelow et al., 2017, Panter et al., 1985). Other top-ranked genes progressively altered in rTg4510 mice were notably enriched for microglial markers previously shown to have their expression augmented in AD (Hopperton et al., 2018, Keren-Shaul et al., 2017), including *Cd68* (**Table 4.5** and **Figure 4.47b**), *Itgax* (or *Cd11c*) (**Table 4.5** and **Figure 4.47c**), and *Clec7a* (**Table 4.5** and **Figure 4.47d**). Expression of all these genes has been previously reported to be increased in hippocampal tissue from 6-month old rTg4510 female mice (Wes et al., 2014), in isolated microglia from rTg4510 mice (Wang et al., 2018), in the cortex of amyloid mice at late stages of pathology

(Rothman et al., 2018), and in the neocortex, hippocampus and microglia of mice with amyloid and tau pathology (Keren-Shaul et al., 2017, Landel et al., 2014). Furthermore, recent transcriptional studies in human brain have shown that microglial gene networks exhibit an increase in expression in response to AD neuropathology (Felsky et al., 2019).

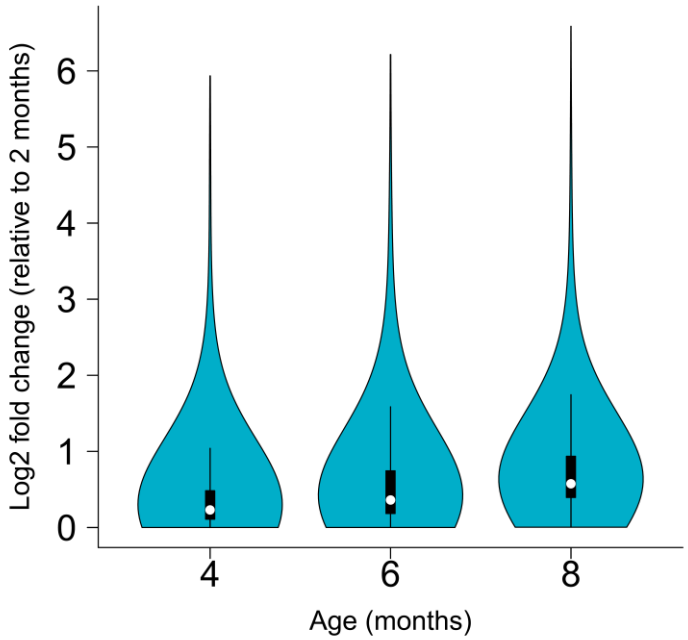


Figure 4.46 – Transcriptional trajectories associated with the accumulation of tau pathology in rTg4510 mice.
Violin plot showing increasing absolute effect size (Log2 fold change) across age groups for transcripts characterised by significant (FDR < 0.05, n = 1,762 transcripts) temporal changes in gene expression associated with genotype. White dots represent the median absolute fold-change.

Table 4.4 – Significant progressive increase with age of expression differences associated with the accumulation of tau pathology.

Age group (relative to 2 months)	**Absolute mean difference (log2 fold change)**	**Statistical test**	**W**	**P-value**
4 months	0.42	Mann-Whitney U test	W (4-6 months) = 1182700	P (4-6 months) < 2.20e-16
6 months	0.60		W (6-8 months) = 1031000	P (6-8 months) < 2.20e-16
8 months	0.78		W (4-8 months) = 679520	P (4-8 months) < 2.20e-16

Table 4.5 – Top-ranked differentially expressed transcripts associated with progression of tau pathology in rTg4510 mice.

Transcript	Statistical test	LRT statistic	log2 fold change (2-8 months)	FDR	N
Gfap	Likelihood-ratio test	106.32	2.75	1.28E-18	59
Cd68		103.77	1.85	2.26E-18	
Itgax		86.85	4.42	6.54E-15	
Clec7a		83.20	5.37	2.97E-14	

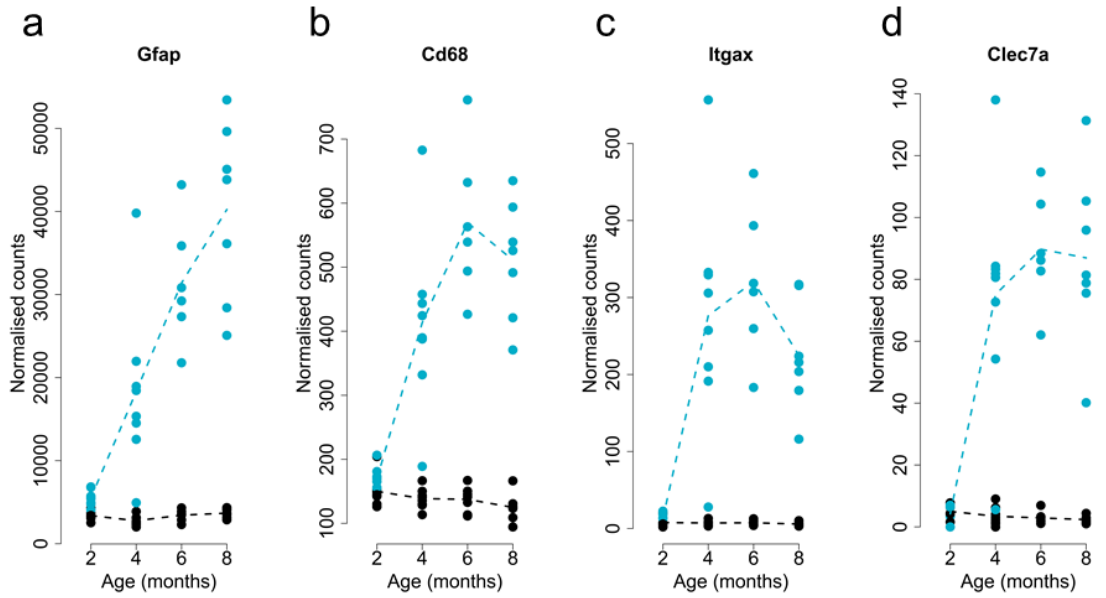


Figure 4.47 – Top-ranked differentially expressed transcripts associated with progression of tau pathology in rTg4510.
Shown are individual plots for: **(a)** *Gfap*, **(b)** *Cd68*, **(c)** *Itgax*, and **(d)** *Clec7a*. Normalised RNA-seq read counts were obtained using DESeq2. Dashed lines represent mean paths for each time point. rTg4510 transgenic (TG, blue) female mice compared to wild type (WT, black) littermate control mice. Total n = 59 animals (6-8 animals per group).

I used *GOseq* (see **Section 4.2.5.2**) to identify ontological enrichments amongst genes characterised by progressively-altered gene expression in rTg4510 mice, finding highly-significant enrichments for immune-related biological pathways including "immune system process" (FDR = 1.03E-25), “defence response” (FDR = 2.98E-24) and “immune response” (FDR = 4.79E-24) (**Supplementary Table 6**). Given these findings, I decided to next quantify Iba1 (see **Section 4.2.7**), a microglia/macrophage-specific calcium-binding protein (Ohsawa et al., 2004), in matched tissue sections from the right brain hemisphere (n = 7-10 animals per group, total n = 70). I observed a significant increase in Iba1 in all brain regions assessed (hippocampus: factorial ANOVA, F(3,62) = 12.60, *P* = 1.56E-06; cortex: factorial ANOVA, F(3,62) = 18.13, *P* = 1.47E-08; thalamus: factorial ANOVA, F(3,62) = 18.85, *P* = 8.37E-09) (**Figure 4.48**).

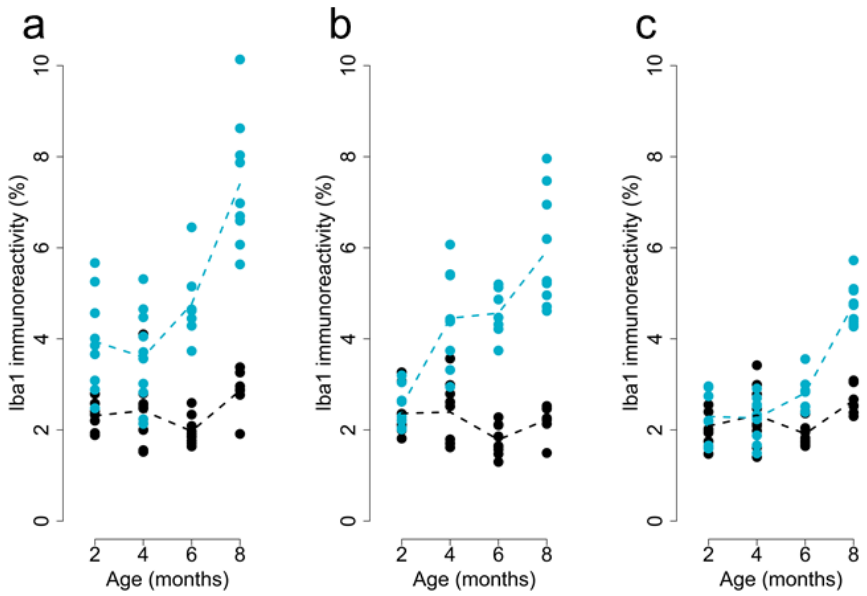


Figure 4.48 – rTg4510 mice show a progressive increase in Iba1 immunoreactivity. **(a-c)** Quantification of Iba1 immunoreactivity in each tested region: **(a)** hippocampus (total n = 70 animals, 7-10 animals per group), **(b)** cortex (total n = 70 animals, 7-10 animals per group), **(c)** thalamus (total n = 70 animals, 7-10 animals per group). rTg4510 transgenic (TG, blue) female mice compared to wild type (WT, black) littermate control mice.

4.3.4.1. Progressively changed transcripts include genes nominated from genetic studies of AD

The list of transcripts progressively altered in rTg4510 mice includes genes robustly associated with familial AD from genetic studies of human patients, including *App* (**Figure 4.49a**; Likelihood-ratio test, LRT statistic = 13.88, log2 fold change (2-8 months) = -0.35, FDR = 0.037), a key driver of amyloid pathology. It also includes genes annotated to both common and rare variants identified in GWAS and exome-sequencing studies of late-onset sporadic AD (LOAD), including *Trem2* (Likelihood-ratio test, LRT statistic = 43.82, log2 fold change (2-8 months) = 1.46, FDR = 3.73E-07), *Pld3* (Likelihood-ratio test, LRT statistic = 36.80, log2 fold change (2-8 months) = -0.58, FDR = 5.80E-06), *Frmd4a* (Likelihood-ratio test, LRT statistic = 27.81, log2 fold change (2-8 months) = 0.36, FDR = 0.00022), *Clu* (Likelihood-ratio test, LRT statistic = 27.73, log2 fold change (2-8 months) = 0.80, FDR = 0.00023), *Apoe* (Likelihood-ratio test, LRT statistic = 22.99, log2 fold change (2-8 months) = 0.86, FDR = 0.0014), *Picalm* (Likelihood-ratio test, LRT statistic = 21.37, log2 fold change (2-8 months) = 0.26, FDR = 0.0025), *Cd33* (Likelihood-ratio test, LRT statistic = 27.32, log2 fold change (2-8 months) = 1.14, FDR = 0.00026), and *Abi3* (Likelihood-ratio test, LRT statistic = 17.10, log2 fold change (2-8 months) = 0.70, FDR = 0.012) (**Figure 4.49b-i**).

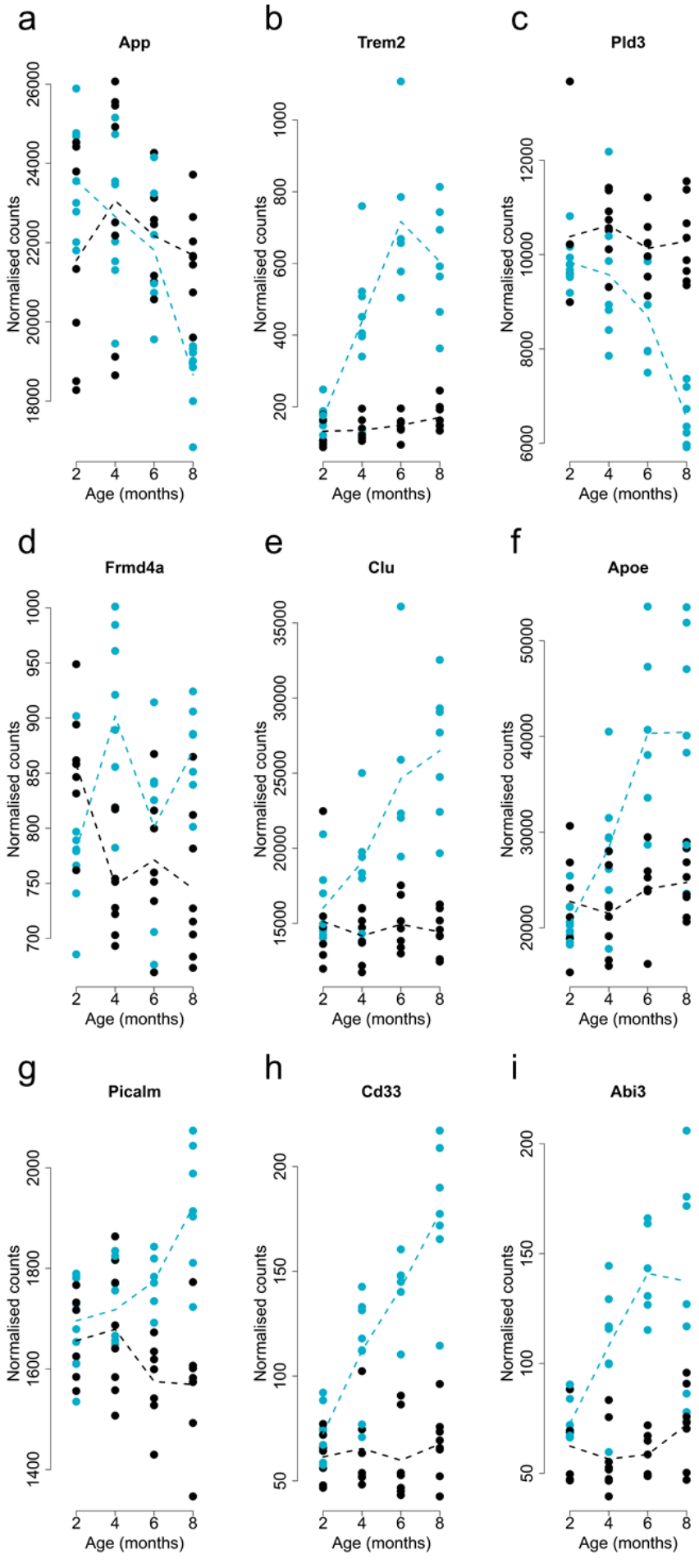

a
App
Normalised counts
18000
20000
22000
24000
26000
2
4
6
8
Age (months)
b
Trem2
Normalised counts
200
400
600
800
1000
2
4
6
8
Age (months)
c
Pld3
Normalised counts
6000
8000
10000
12000
2
4
6
8
Age (months)
d
Frmd4a
Normalised counts
700
750
800
850
900
950
1000
2
4
6
8
Age (months)
e
Clu
Normalised counts
15000
20000
25000
30000
35000
2
4
6
8
Age (months)
f
Apoe
Normalised counts
20000
30000
40000
50000
2
4
6
8
Age (months)
g
Picalm
Normalised counts
1400
1600
1800
2000
2
4
6
8
Age (months)
h
Cd33
Normalised counts
50
100
150
200
2
4
6
8
Age (months)
i
Abi3
Normalised counts
50
100
150
200
2
4
6
8
Age (months)

Figure 4.49 – Transcripts progressively altered in rTg4510 mice include many genes implicated in familial and sporadic AD from genetic association studies.
Genes showing temporal shifts in expression in rTg4510 mice included: **(a)** App; **(b)** Trem2; **(c)** *Pld3*; **(d)** *Frmd4a*; **(e)** Clu; **(f)** *Apoe*; **(g)** *Picalm*; **(h)** *Cd33*; and **(i)** *Abi3*. Total n = 59 animals (29 TG, 30 WT). Normalised counts were obtained using DESeq2. Dashed lines represent mean paths across age groups. rTg4510 transgenic (TG, blue) mice compared to wild type (WT, black) littermate controls.

4.3.5. Progressive changes in gene expression in the entorhinal cortex mirror the development of amyloid neuropathology in J20 mice

In contrast to the dramatic and progressive changes in gene expression identified in rTg4510 mice, fewer significant temporal transcriptional differences associated with the progression of amyloid pathology were identified in the J20 mice; in total we identified five transcripts (*Cst7*, *Wdfy1*, *Grxcr2*, *Itgax*, and *Ifitm1*) whose expression profile significantly changed (FDR < 0.05) with the progression of amyloid pathology (**Table 4.6** and **Supplementary Table 7**). The relatively low number of significantly-altered genes in J20 mice potentially reflects the slower and later accumulation of pathology in these mice compared to rTg4510 mice. Previous work has shown relatively limited amyloid pathology in J20 entorhinal cortex even at 14 months of age (Harris et al., 2010), with neuronal cell loss varying by brain region (Wright et al., 2013). Nevertheless, we found that effect sizes for the 1,762 transcripts identified as being progressively dysregulated in rTg4510 mice were significantly correlated across both models, suggesting some common molecular signals associated with both tau and amyloid pathology (**Figure 4.50**; Pearson correlation, r = 0.46, *P* = 1.50E-92; exact binomial test, n = 1762 transcripts, *P* = 1.97e-05). Interestingly, two genes, *Cst7* and *Itgax,* identified as being associated with progressive tau pathology in rTg4510 mice (**Figure 4.52a** and **Figure 4.47c**) were also significantly associated with amyloid pathology in J20 mice (**Figure 4.51**, **Figure 4.52b**, and **Figure 4.53**). *Cst7* encodes the protein cystatin F, a member of the cystatin superfamily of naturally occurring cysteine protease inhibitors (Hamilton et al., 2008, Ni et al., 1998). Cysteine proteases are involved in protein disruption during programmed cell death, and the regulation of their activity is essential for cell homeostasis, with other cysteine protease inhibitors (e.g. cystatin C) having been shown to play a

protective role in AD and other neurodegenerative diseases (Hasanbasic et al., 2016). According to Hamilton et al. (2008), the aminopeptidase cathepsin C is a major target of cystatin F in different immune cell types, which regulates the activation of effector serine proteases in T cells, natural killer cells, neutrophils and mast cells, suggesting an important role for cystatin F in the regulation of various immune cell effector functions. *Cst7* is expressed selectively in immune cells, and like *Itgax* is a marker for activated microglia (Hansen et al., 2018). Of note, the expression of *Cst7* has been reported to be increased in the presence of AD pathology in other studies (Keren-Shaul et al., 2017, Ofengeim et al., 2017), and was found as the top gene showing an increase in expression in cortex samples from 12 month-old APP NL-G-F knock-in mice (Castillo et al., 2017). *Itgax* codes for the integrin CD11c (or integrin alpha X) protein, which combines with integrin β2 (ITGB2/CD18) to form the leukocyte-specific integrin known as inactivated-C3b (iC3b) receptor 4 (CR4) (Bilsland et al., 1994). The resulting complex has a role in the phagocytosis of complement coated particles. *CD11c* is highly expressed in most dendritic cells, but also on monocytes, macrophages, neutrophils, and some B cells (Bilsland et al., 1994, Ugarova and Yakubenko, 2001). Only a limited number of cells in the mouse brain express *CD11c* under physiological conditions; however, in pathological conditions, including AD, changes in the number, morphology, and distribution of CD11c-expressing microglia occur, and CD11c+ microglia have been observed to be spatially associated with amyloid plaques in an AD mouse model (Kamphuis et al., 2016). Additionally, CD11c has been reported to be increased in microglia surrounding Aβ plaques in human AD brains in immunohistochemistry analyses (Akiyama and McGeer, 1990). Together these results suggest common mechanisms of microglia activation in response to tau pathology in rTg4510 and amyloid pathology in J20 mice.

Table 4.6 – Differentially expressed transcripts associated with progression of amyloid pathology in J20 mice.

Transcript	Statistical test	LRT statistic	log2 fold change (2-8 months)	FDR	N
Cst7	Likelihood-ratio test	37.37	2.42	0.00072	62
Wdfy1		29.34	0.65	0.018	
Grxcr2		27.61	1.36	0.022	
Itgax		27.49	1.44	0.022	
Ifitm1		26.42	1.55	0.029	

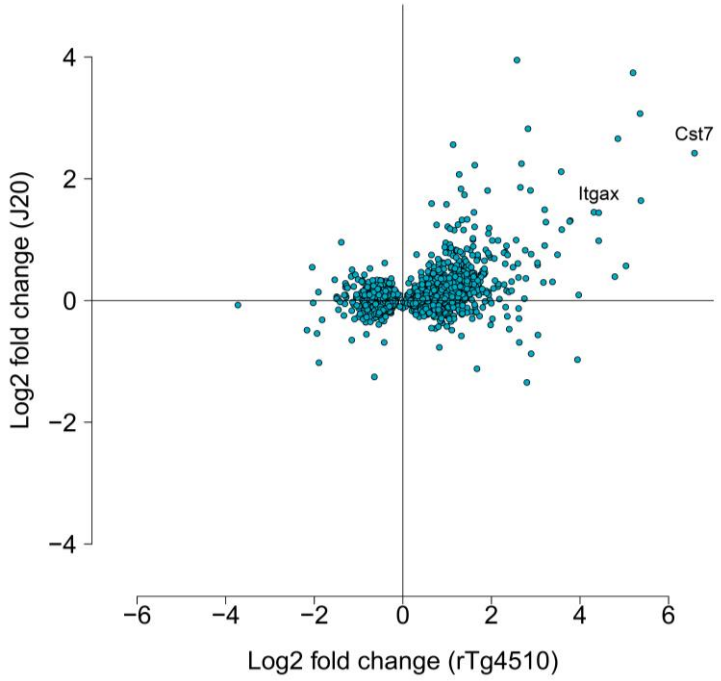


Figure 4.50 – Effect sizes at differentially expressed transcripts associated with the progression of tau in rTg4510 mice are correlated with those associated with the progression of amyloid in J20 mice.
Positive correlation for effect size (Log2 fold change from latest time point compared to baseline) for significant transcripts in rTg4510 mice (Pearson correlation, r = 0.46, *P* = 1.50E-92; exact binomial test, n = 1762 transcripts, *P* = 1.97e-05).

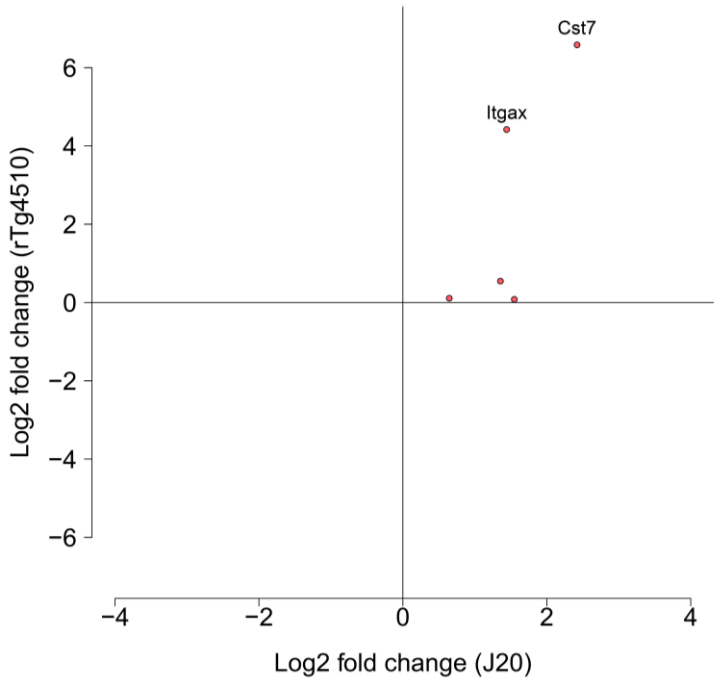


Figure 4.51 – Effect sizes at differentially expressed transcripts associated with the progression of amyloid in J20 mice are correlated with those associated with the progression of tau in rTg4510 mice.
Two transcripts (*Cst7* and *Itgax*) were significantly associated with the progression of both tau (Tg4510) and amyloid (J20) pathology (Pearson correlation, r = 0.77, *P* = 0.13; exact binomial test, n = 5 transcripts, *P* = 0.13).

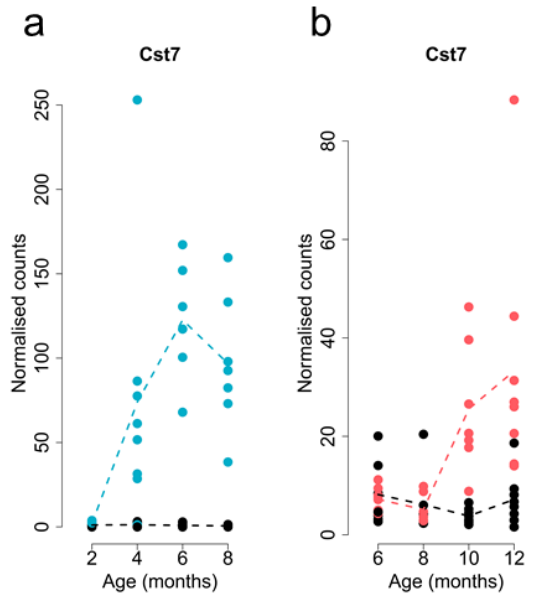


Figure 4.52 – Cst7 was significantly associated with the progression of both tau and amyloid pathology.

(a) *Cst7* gene expression in rTg4510 mice (Total n = 59 animals, 6-8 animals per group, Likelihood-ratio test, LRT statistic = 36.10, log2 fold change (2-8 months) = 6.59, FDR = 7.71E-06). rTg4510 transgenic (TG, blue) female mice compared to wild type (WT, black) littermate control mice. **(b)** *Cst7* gene expression in the J20 mice (Total n = 62 animals, 6-8 animals per group, Likelihood-ratio test, LRT statistic = 37.37, log2 fold change (6-12 months) = 2.42, FDR = 0.00072). J20 transgenic (TG, red) female mice compared to wild type (WT, black) littermate control mice. Normalised counts were obtained using DESeq2. Dashed lines represent mean paths for each time point.

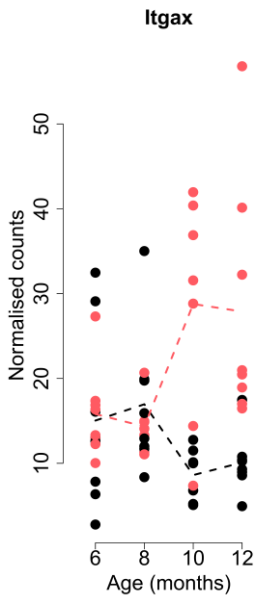


Figure 4.53 – Itgax gene expression increases with age and the progression of amyloid pathology in J20 mice.
The expression of *Itgax* in J20 mice (Total n = 62 animals, 6-8 animals per group, Likelihood-ratio test, LRT statistic = 37.37, log2 fold change (6-12 months) = 2.42, FDR = 0.00072) reflects the changes seen in rTg4510 mice (see **Figure 4.47c**). Normalised counts were obtained using DESeq2. Dashed lines represent mean paths for each time point. J20 transgenic (TG, red) female mice compared to wild type (WT, black) littermate control mice.

4.3.6. Transcriptional changes identified in rTg4510 mice reflect those observed in other models of tau pathology

A number of recent studies have described further evidence for differential gene expression in transgenic models of familial AD gene mutations (Castillo et al., 2017, Keren-Shaul et al., 2017, Landel et al., 2014, Matarin et al., 2015, Rothman et al., 2018, Wang et al., 2018, Wes et al., 2014) (see **Section 1.5.2.1** in **Chapter 1**). We therefore explored hippocampal RNA-seq data from two other transgenic models (TAU (Camk2a-MAPTP301L) and TAS10 (SwAPP, K670N/M671L)) downloaded from the Mouseac database (Matarin et al., 2015, Salih et al., 2018) (www.mouseac.org) to identify consistencies in the transcriptional signatures between different models of tau and amyloid pathology. Effect sizes for transcripts identified as associated with rTg4510 genotype and also present in the Mouseac TAU RNA-seq dataset (n = 138) were significantly correlated between the two models (r = 0.33, *P* = 7.7E-05). Despite this consistency in effect sizes, many of the differentially expressed genes associated with rTg4510 genotype were not statistically replicated in the TAU model (**Table 4.7**, **Figure 4.54** and **Supplementary Table 8**), although this likely reflects the distinct genetic background of the different transgenic lines and the modest power to detect effects given the small number of samples profiled in the Mouseac dataset (n = 49 RNA-seq samples, 1-4 animals per age group, after filtering for samples with complete phenotypic data). Differential expression of *Gpr17* was associated with TAU genotype (Bonferroni-corrected *P* < 0.00035), although the exact profile for this gene differed to that observed in the rTg4510 mice (**Figure 4.55b**). This putative receptor for leukotrienes, uracil nucleotides and/or oxysterols may warrant further investigation as its expression is associated with damage to neural tissue including white matter (Fumagalli et al., 2016). As expected, given the limited evidence for consistency in genotype effects between rTg4510 and J20 mice, there was no correlation between effects observed in rTg4510 and TAS10 mice for the 145 rTg4510-associated genes present in both datasets.

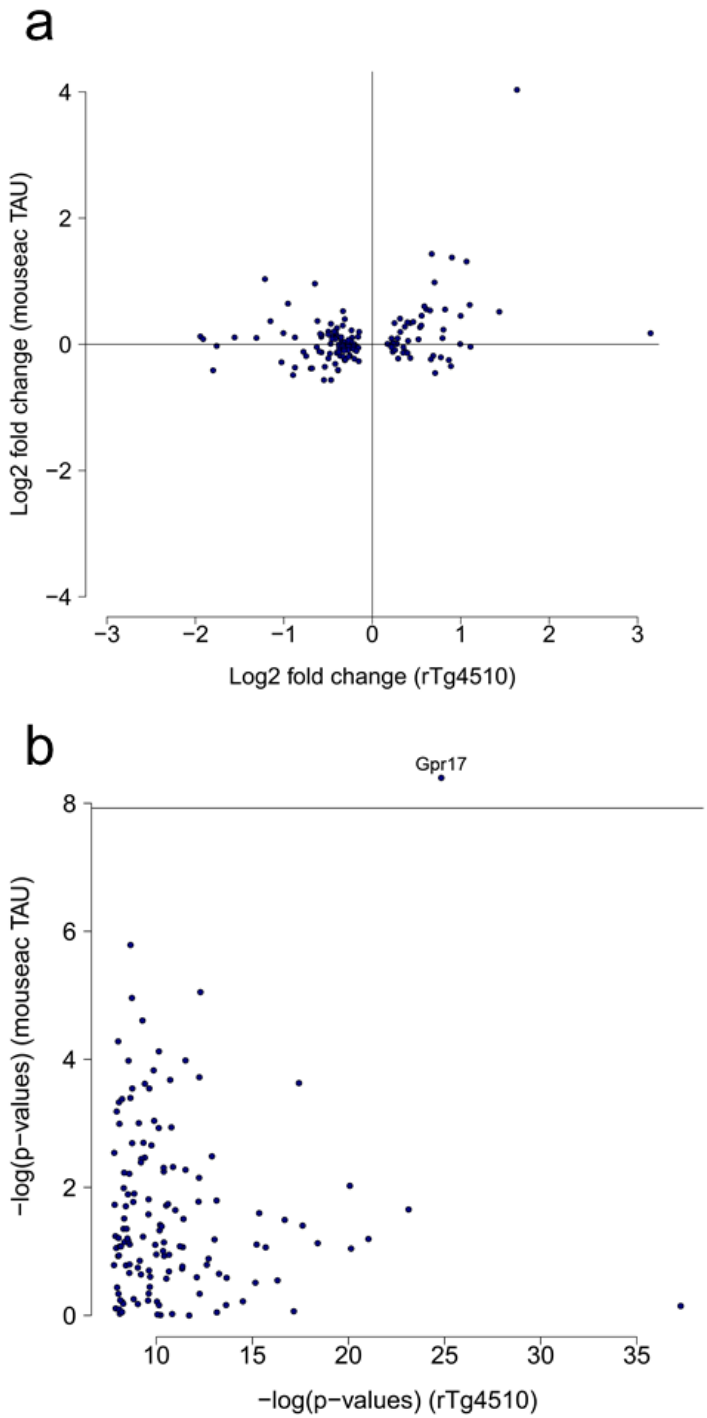


Figure 4.54 – Comparison of rTg4510 genotype-associated transcriptional differences in another transgenic model of tau pathology.
(a) Correlation of effect size (log2 fold change) between rTg4510 genotype-associated genes and the same genes in the TAU mouse model from Mouseac (Pearson correlation, r = 0.33, *P* = 7.73E-5, exact binomial test, n = 138 transcripts, *P* = 0.11). **(b)** Correlation of genotype-associated *P* values (applying the same linear regression model) between rTg4510 and the same genes in the TAU mouse model from Mouseac (Pearson correlation, r = 0.02, *P* = 0.82).

Table 4.7 – The majority of the top-ranked differentially expressed transcripts associated with rTg4510 genotype were not statistically replicated in the TAU mouse model from Mouseac.
Bonferroni-corrected P < 0.00035.

Transcript	Statistical test	t-value	β	P-value	N
Car4	Linear regression	t(20) = -0.18	-1.31	0.87	25
Gpr17		t(20) = 4.49	3.13	0.00023	
Blnk		t(20) = 1.35	1.11	0.19	
Hspa5		t(20) = 1.06	20.72	0.30	

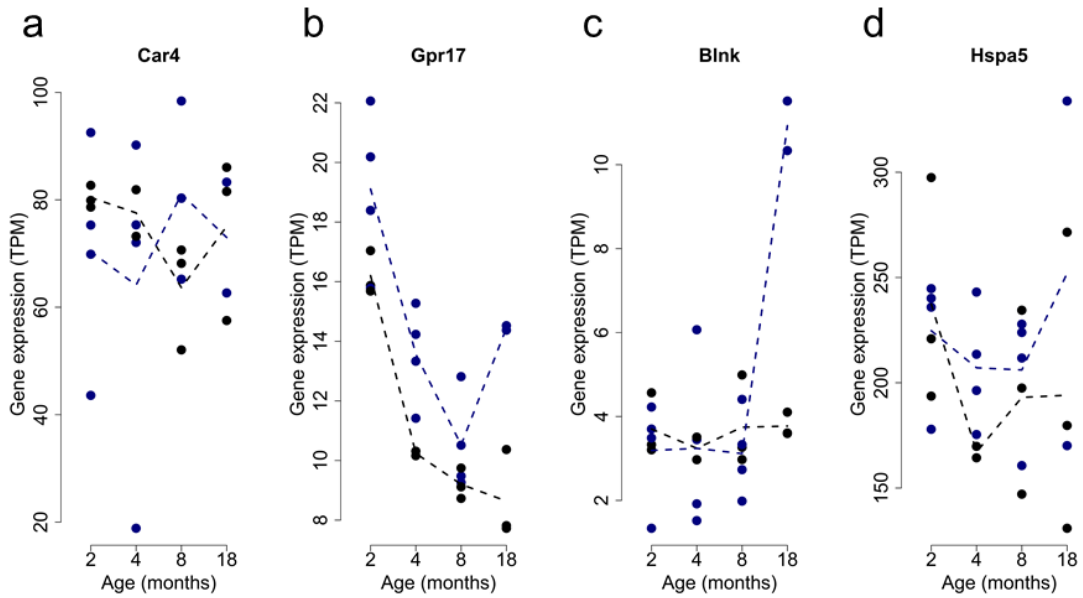


Figure 4.55 – Most of top-ranked differentially expressed transcripts associated with rTg4510 genotype were replicated in another mouse model of tau pathology.
Shown are individual plots for: **(a)** *Car4*. **(b)** *Gpr17*. **(c)** *Blnk*. **(d)** *Hspa5*. Total n = 25 animals (2-4 animals per group). TAU transgenic transcriptomic data obtained from Mouseac (dark blue) compared to wild type (black) controls. Dashed lines represent mean paths across age groups.

In contrast, association statistics for the 1640 transcripts identified as being progressively altered with age in rTg4510 mice and also present in the Mouseac datasets (**Supplementary Table 9** and **Supplementary Table 10**) were significantly correlated with those for the same genes in both TAU (r = 0.46, *P* = 1.2e-86) and TAS10 (r = 0.23, *P* = 3.9E-21) transgenic mice (**Figure 4.56**). Given the small number of progressive alterations observed in J20 mice, it was not possible to systematically explore overlaps between differentially regulated genes in this model and the two Mouseac models. Of note, however, the two genes identified as being temporally-altered in both rTg4510 and J20 mice – *Cst7* and *Itgax* – were both similarly altered in both the TAU and TAS10 models (**Table 4.8** and **Figure 4.57**).

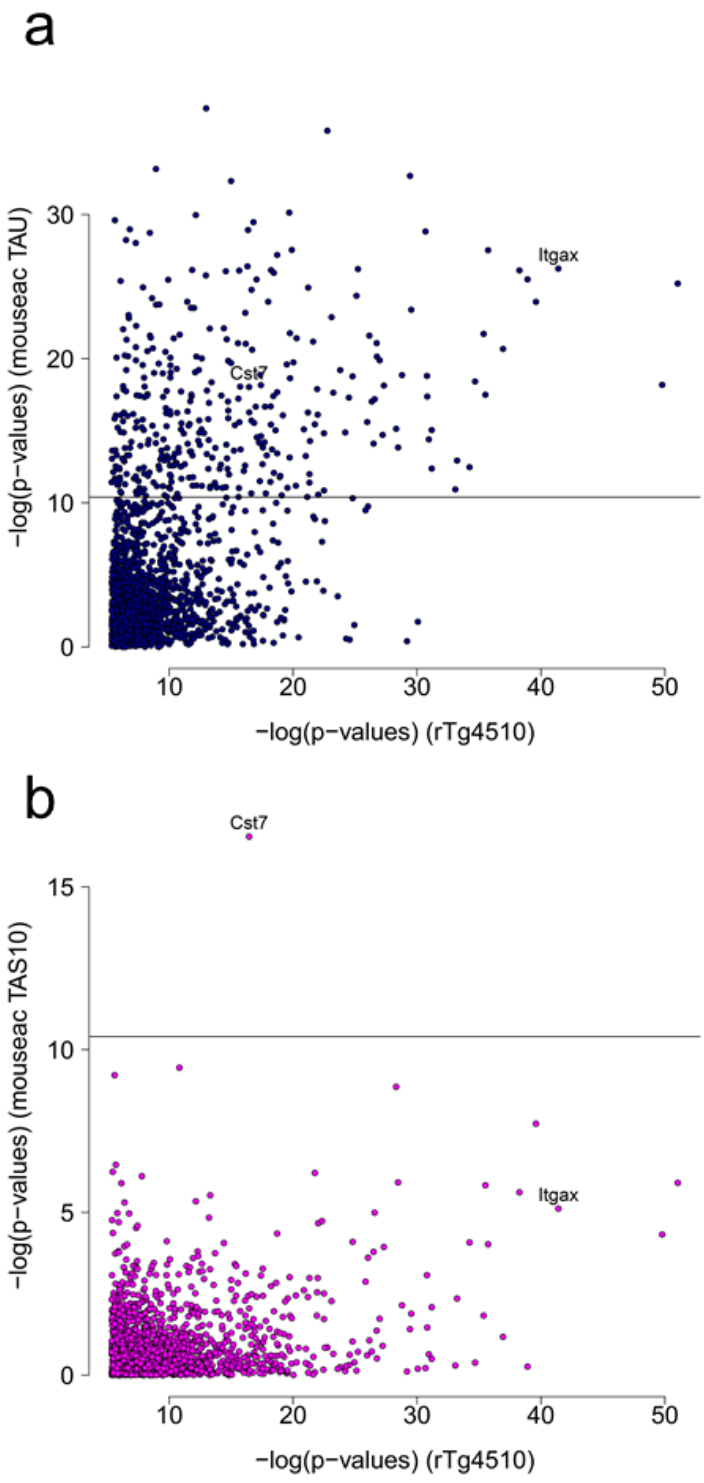


Figure 4.56 – Analysis of genes showing progressive transcriptional differences in rTg4510 mice in other transgenic models.
(a) Correlation of *P* values between the rTg4510 and TAU mouse models (using transcriptional data from Mouseac) for an analysis of interactions between and genotype and age for genes identified as significant in rTg4510 mice (Pearson correlation, r = 0.46, *P* = 1.22E-86, exact binomial test, n = 1640 transcripts, *P* < 2.2e-16). **(b)** Correlation of *P* values between the rTg4510 and TAS10 mouse models (using transcriptional data from Mouseac) for an analysis of interactions between genotype and age for genes identified as significant in rTg4510 mice (Pearson correlation, r = 0.23, *P* = 3.89E-21, exact binomial test, n = 1640 transcripts, *P* < 2.2e-16).

Table 4.8 – Differential expression of Cst7 and Itgax were statistically replicated in TAU and TAS10 Mouseac mouse models.
Bonferroni-corrected P < 3.05E-5.

Mouse model	Transcript	Statistical test	F-value	β	P-value	N
TAU	*Cst7*	Linear regression	F(3,16) = 48.73	187.00	1.46E-8	25
TAS10			F(3,15) = 43.42	11.58	6.50E-8	24
TAU	*Itgax*		F(3,16) = 137.86	28.59	3.95E-12	25
TAS10			F(3,15) = 6.03	0.93	0.0060	24

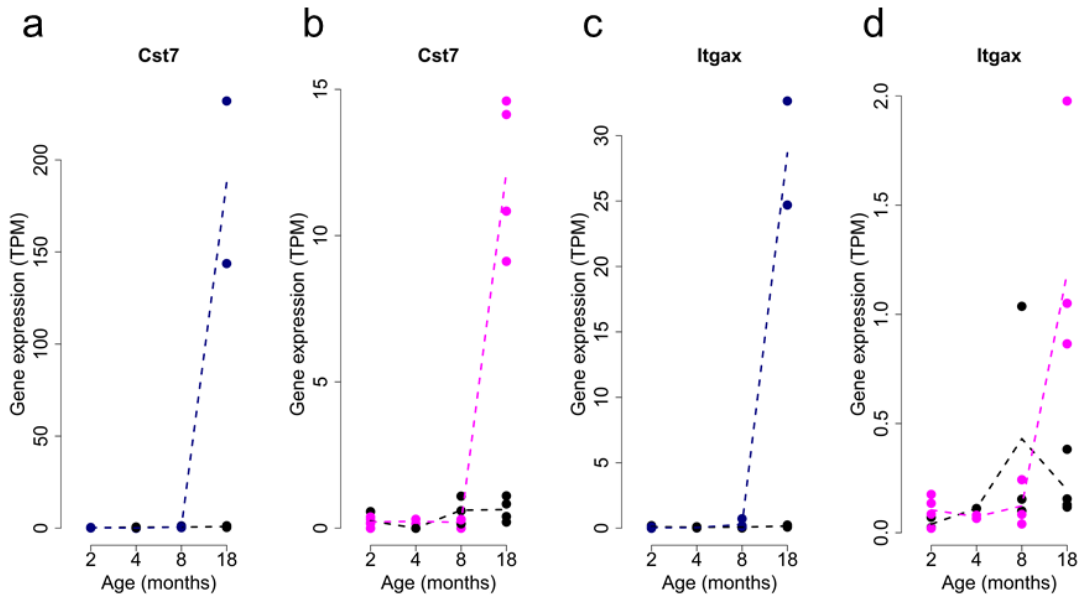


Figure 4.57 – Cst7 and Itgax were similarly altered in both the TAU and TAS10 models.
(a) *Cst7* in TAU mice from Mouseac (Total n = 25 animals, 2-4 animals per group). TAU transgenic mice (dark blue) compared to wild type (black) controls. **(b)** *Cst7* in TAS10 mice from Mouseac (Total n = 24 animals, 1-4 animals per group). TAS10 transgenic mice (pink) compared to wild type (black) controls. **(c)** *Itgax* in TAU mice from Mouseac (Total n = 25 animals, 2-4 animals per group). TAU transgenic mice (dark blue) compared to wild type (black) controls. **(d)** *Itgax* in TAS10 mice from Mouseac (Total n = 24 animals, 1-4 animals per group). TAS10 transgenic mice (pink) compared to wild type (black) controls. Dashed lines represent mean paths across age groups.

4.3.7. Gene co-expression networks associated with the progression of tau pathology are enriched for functional pathways related to AD including synaptic transmission, the immune system, and glial cell activation

Given the dramatic transcriptional changes identified in the rTg4510 mice, I next used WGCNA (see **Section 4.2.6.**) to identify discrete co-expression modules and describe systems-level transcriptional variation associated with rTg4510 genotype and the progression of tau pathology. I constructed co-expression networks using entorhinal cortex RNA-seq data from rTg4510 TG and WT mice (n = 58 mice), identifying 18 discrete co-expression modules (**Figure 4.58**). Next, I used a linear regression model (see **Section 4.2.6.**) and identified six co-expression modules (here named as “salmon”, “turquoise”, “purple”, “yellow”, “light-cyan”, and “red”) that were significantly (Bonferroni corrected, $P < 0.0028$) associated with rTg4510 genotype (**Figure 4.58**, **Table 4.9**, **Figure 4.59** and **Supplementary Table 11**). Strikingly, these tau-associated co-expression modules are highly enriched for molecular functions and biological pathways directly related to AD. The red module, for example, which was down-regulated in TG mice compared to WT mice ($\beta = -0.18$, $P = 1.43E-10$), is highly enriched for functional pathways involved in synaptic transmission (**Supplementary Table 12**). The turquoise module, which was up-regulated in TG mice compared to WT mice ($\beta = 0.18$, $P = 3.04E-10$), is enriched for pathways involved in activation of the immune system (**Supplementary Table 13**). The salmon module, which was consistently up-regulated in TG mice compared to WT mice ($\beta = 0.14$, $P = 3.58E-06$), is enriched for genes involved in myelination and glial cell activation (**Supplementary Table 14**). The purple module, which was down-regulated in TG mice compared to WT mice ($\beta = -0.13$, $P = 0.00012$), is enriched for pathways related to cellular component disassembly (**Supplementary Table 15**). Finally, the yellow module, which was down-regulated in TG mice compared to WT mice ($\beta = -0.10$, $P = 0.0015$), is enriched for pathways related to mitochondria and synaptic processes (**Supplementary Table 16**).

The module eigengenes for three of these co-expression modules (turquoise, yellow and red) were characterised by a significant interaction between genotype and age in rTg4510 mice (**Figure 4.58** and **Supplementary Table 11**), suggesting that they are temporally linked to the development of tau pathology in

TG mice (**Figure 4.60**). The turquoise module becomes increasingly up-regulated with the development of tau pathology (β = 0.28, *P* = 4.23E-06), the red module becomes increasingly down-regulated with the development of tau pathology (β = -0.21, *P* = 0.0022), and the yellow module becomes down-regulated specifically during the later stages of tau pathology (β = -0.29, *P* = 0.0018). Using the matched immunohistochemistry data generated across multiple brain regions for each mouse we were able to explore the relationship between co-expression modules and actual tau pathology in rTg4510 mice, confirming that the turquoise, yellow, and red modules are robustly associated with the accumulation of tau across the brain (**Figure 4.62**). The association with pathology was particularly strong in highly affected brain regions such as the hippocampus (**Figure 4.61**, **Figure 4.63**, **Figure 4.64**, **Figure 4.65** and **Figure 4.66**), in which the module eigengene for the turquoise module is positively correlated with levels of tau in TG mice (r = 0.85, *P* = 1.20E-16), and those for the yellow and red modules are negatively correlated with levels of tau in TG mice (yellow: r = -0.63, *P* = 2.00E-07, red: r = -0.79, *P* = 4.61E-13). Although these co-expression modules were also correlated with measures of tau pathology in the thalamus, the magnitude of effects was much lower, reflecting the later and less aggressive accumulation of tau in this region of the brain (**Figure 4.67**).

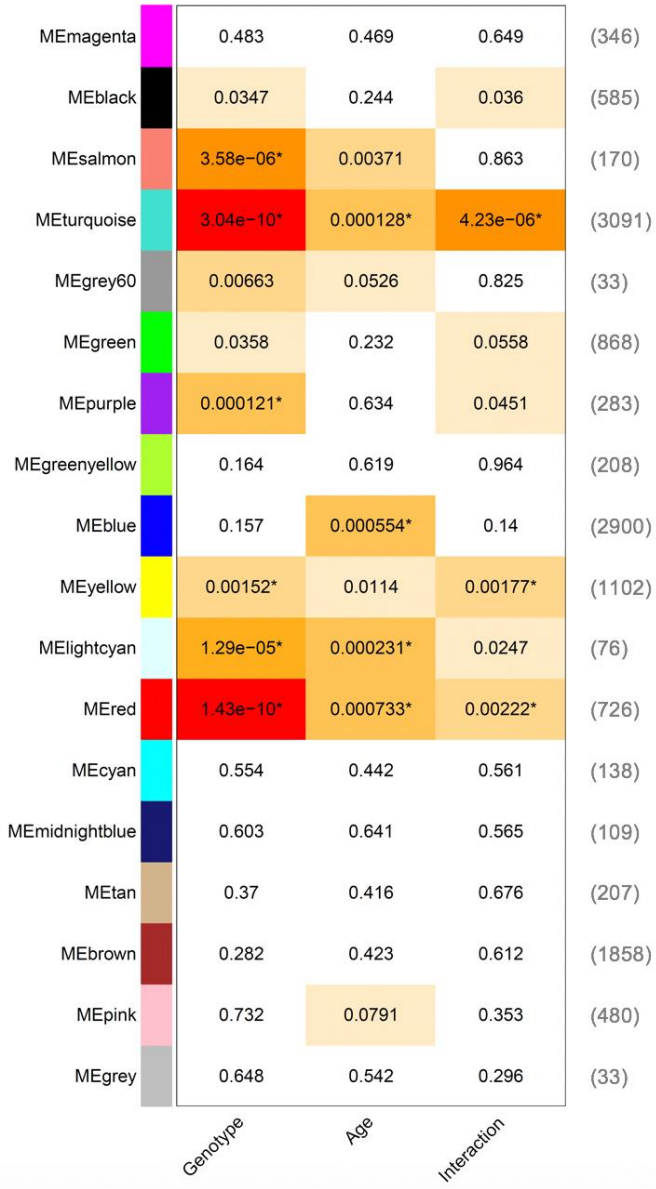


Figure 4.58 – Co-expression modules associated with tau pathology in rTg4510 mice.

In total 18 co-expression modules (each labelled with an arbitrary colour) were identified in the rTg4510 TG and WT mice (n = 58 RNA-seq datasets). Shown is the association of each module with genotype, age and progressive changes in TG mice (i.e. the interaction between genotype and age). Bonferroni corrected significant (α=0.05) *P* values are highlighted with an asterisk. The total number of genes in each module is shown in parentheses.

Table 4.9 – Six co-expression modules were associated with rTg4510 genotype.
Bonferroni corrected P < 0.0028.

WGCNA module	Statistical test	t-value	β	P-value	N
Red	Linear regression	t(53) = -7.93	-0.18	1.43E-10	58
Turquoise		t(53) = 7.73	0.18	3.04E-10	
Salmon		t(53) = 5.17	0.14	3.58E-06	
Light-cyan		t(53) = -4.81	-0.13	1.29E-05	
Purple		t(53) = -4.15	-0.13	0.00012	
Yellow		t(53) = -3.34	-0.10	0.0015	

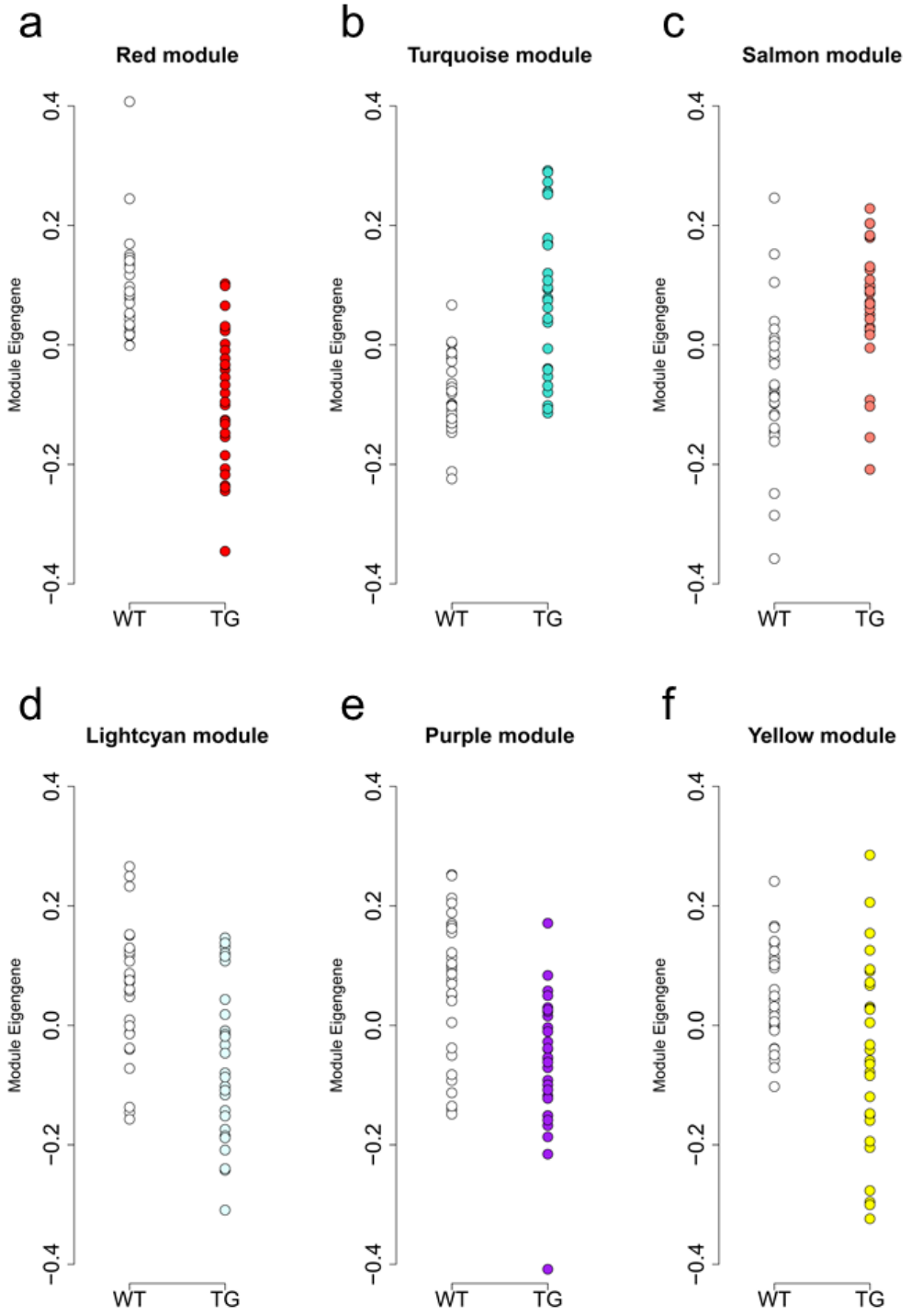


Figure 4.59 – Entorhinal cortex co-expression modules associated with rTg4510 genotype.

Six modules were significantly different (Bonferroni corrected $P < 0.0028$) between TG and WT mice. **(a)** Red module. **(b)** Turquoise module. **(c)** Salmon module. **(d)** Light-cyan module. **(e)** Purple module. **(f)** Yellow module. Shown for each module is the module eigengene value for each individual animal (total n = 58 mice). Coloured circles represent rTg4510 TG mice and white circles represent WT control mice. Each circle represents a single mouse. Total n = 58 animals (6-8 animals per group).

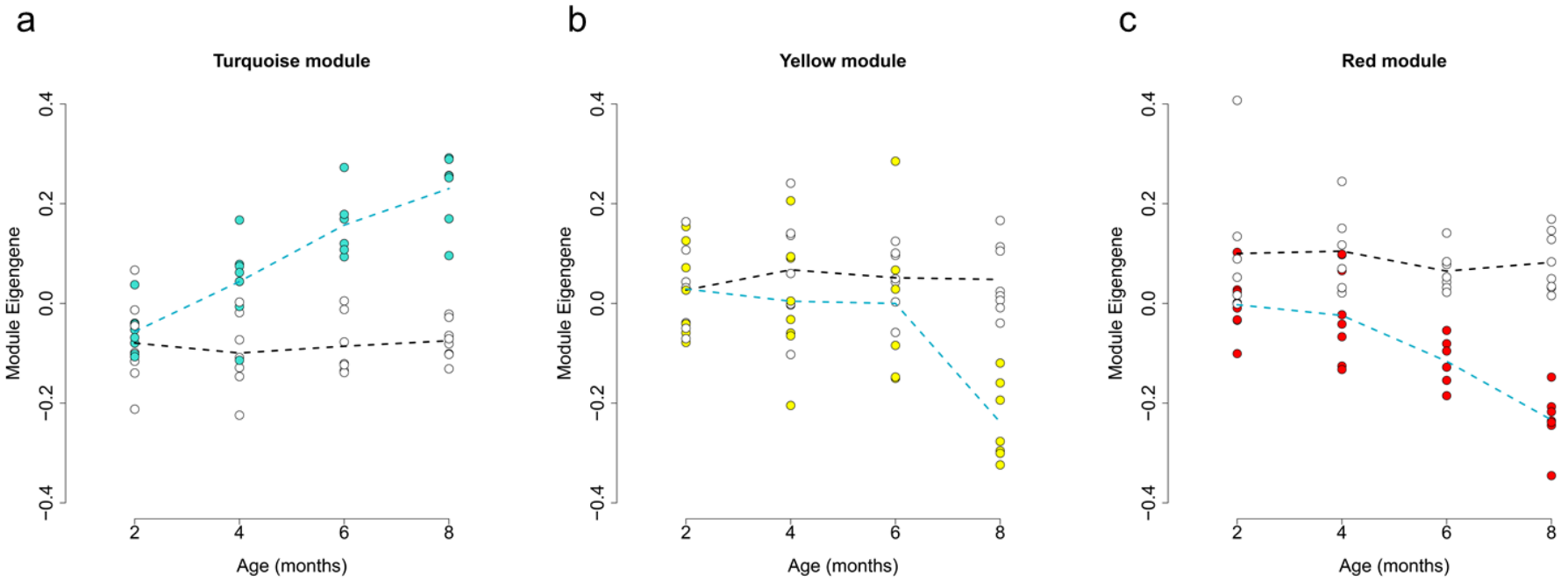


Figure 4.60 – Variation in three entorhinal cortex co-expression modules parallels the accumulation of tau pathology in rTg4510 mice.
Shown are module eigengene values for each individual mouse at four time-points for **(a)** the turquoise module (n = 3091 transcripts, linear regression, F(3,50) = 12.18, β = 0.28, P = 4.23E-06), **(b)** the yellow module (n = 1102 transcripts, linear regression, F(3,50) =5.79, β = -0.29, *P* = 0.0018), and **(c)** the red module (n = 726 transcripts, linear regression, F(3,50) = 5.58, β = -0.21, *P* = 0.0022). Total n = 59 animals (6-8 animals per group). Coloured circles represent rTg4510 TG mice and white circles represent WT control mice. Each circle represents a single individual mouse. Dashed lines represent mean paths for each time point. Total n = 58 animals (6-8 animals per group).

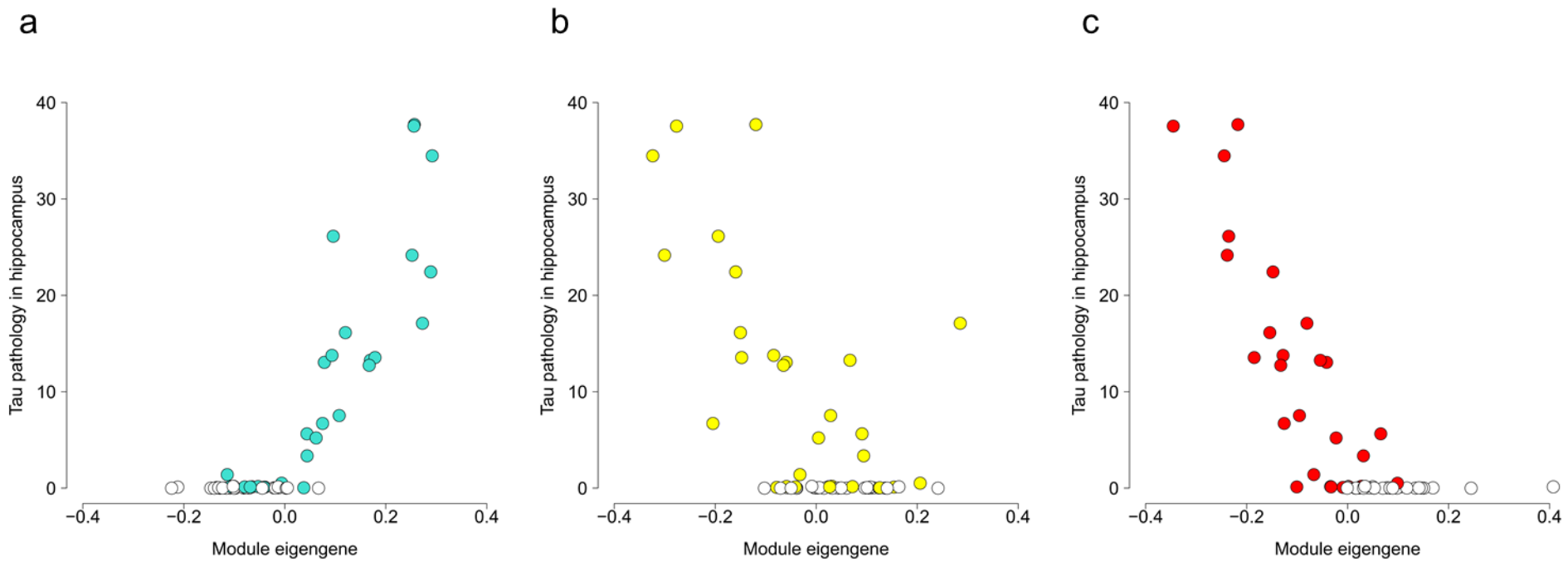


Figure 4.61 – The same three tau-associated co-expression modules are correlated with actual levels of tau pathology quantified using immunohistochemistry in the same mice.

Shown are scatter-plots highlighting the **(a)** positive correlation between module eigengene in the turquoise module and tau pathology in the hippocampus (Pearson correlation, r = 0.85, *P* = 5.00E-17), **(b)** negative correlation between module eigengene in the yellow module and tau pathology in the hippocampus (Pearson correlation, r = -0.63, *P* = 1.00E-07), and **(c)** negative correlation between module eigengene in the red module and tau pathology in the hippocampus (Pearson correlation, r = -0.79, *P* = 2.00E-13). Coloured circles represent rTg4510 TG mice and white circles represent WT control mice. Each circle represents a single individual mouse. Dashed lines represent mean paths for each time point. Total n = 58 animals (6-8 animals per group).

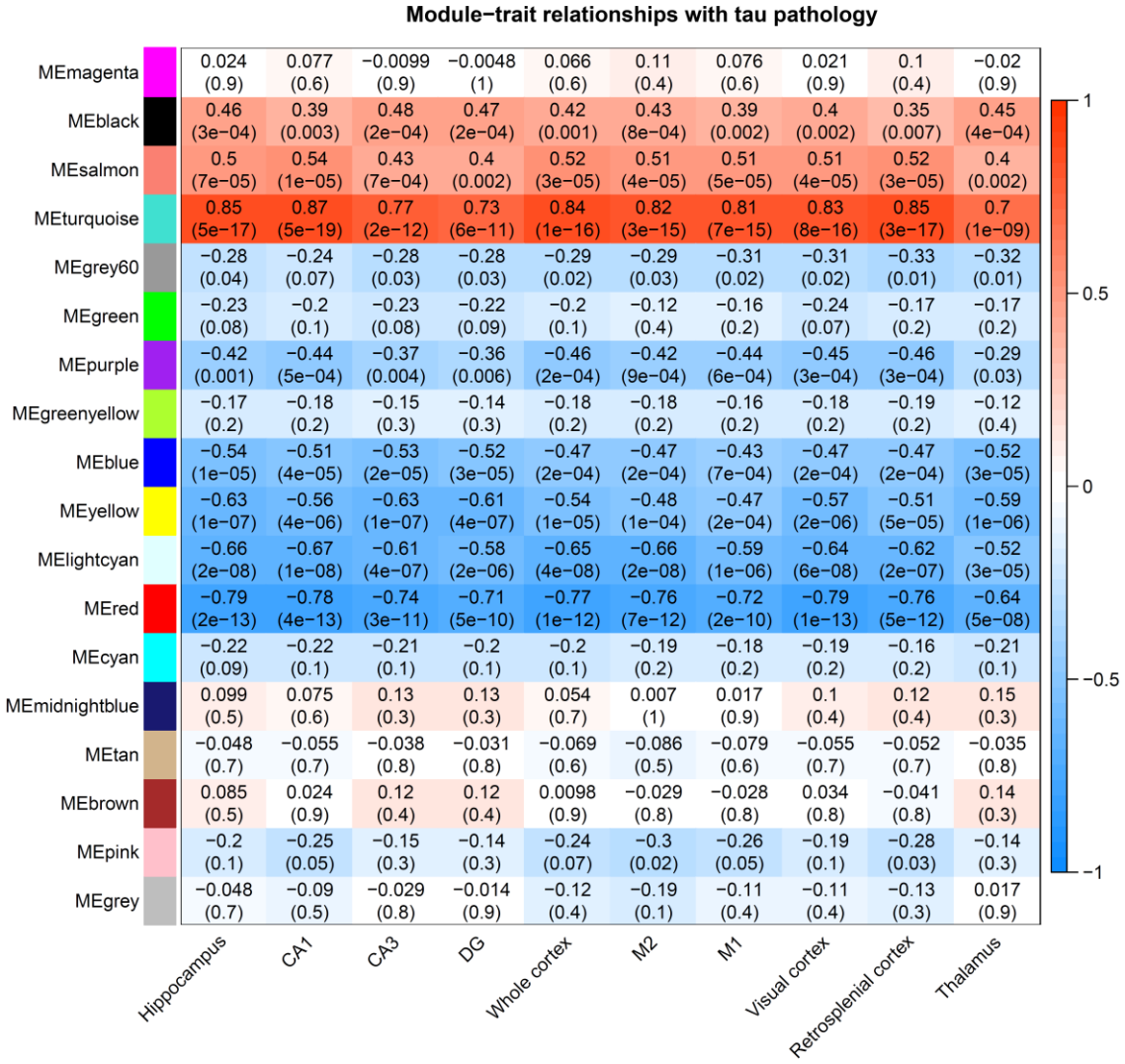


Figure 4.62 – Entorhinal cortex co-expression modules are strongly correlated with tau pathology across multiple brain regions.
Heatmap representing module-trait relationships between module eigengenes and tau pathology measured using immunohistochemistry in each brain region assessed, where red indicates a positive correlation and blue indicates a negative correlation. Values indicate Pearson correlation (r) coefficients and their corresponding *P* values in parentheses. Total n = 58 animals.

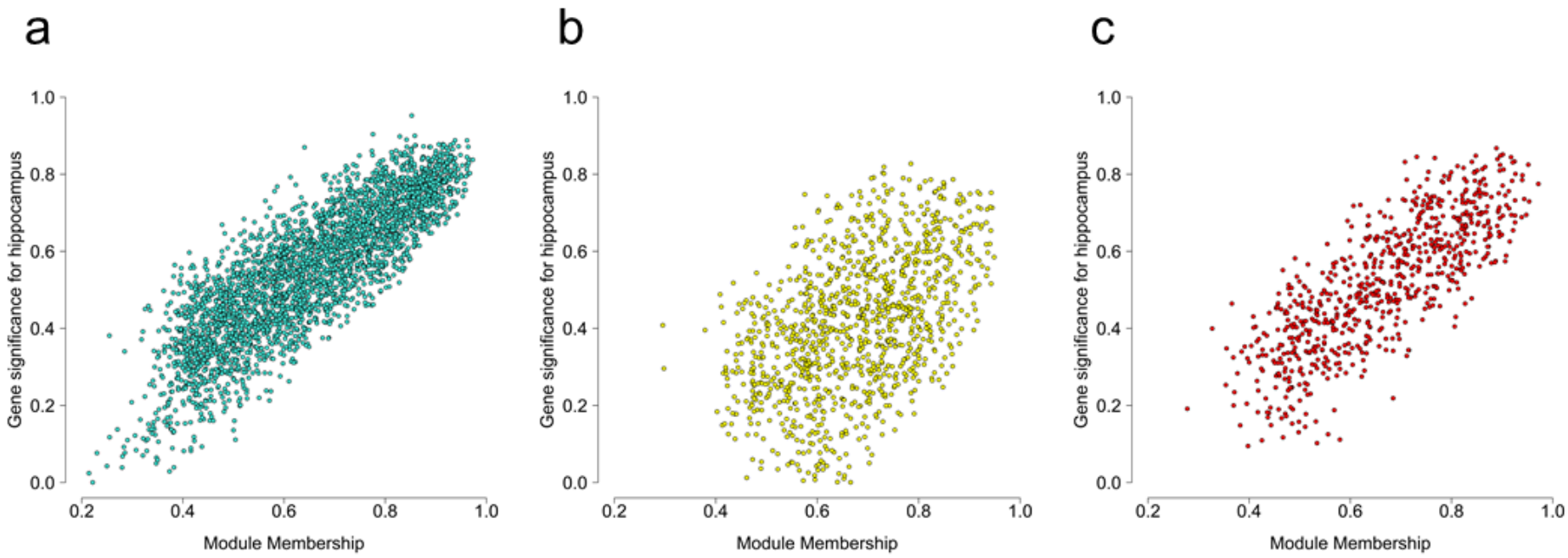


Figure 4.63 – Module membership for each transcript from the three co-expression networks showing progressive changes in rTg4510 mice is strongly correlated with actual levels of pathology in the hippocampus measured using immunohistochemistry.
Shown is the correlation of module membership for each transcript with tau pathology in the hippocampus in **(a)** the turquoise module (n = 3091 transcripts, Pearson correlation, r = 0.85, *P* < 1.00E-200), **(b)** the yellow module (n = 1102 transcripts, Pearson correlation, r = 0.49, *P* = 1.30E-67), and **(c)** the red module (n = 726 transcripts, Pearson correlation, r = 0.80, *P* = 8.88E-163). Each circle represents a single transcript. Total n = 58 animals (6-8 animals per group).

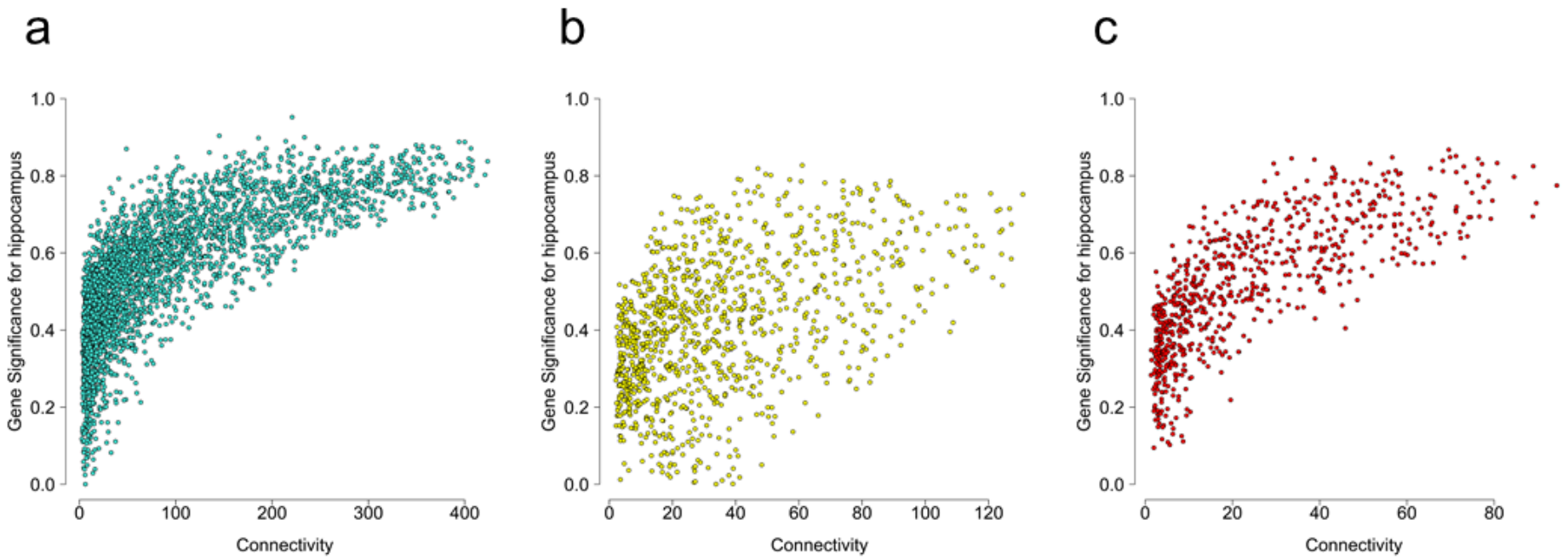


Figure 4.64 – Intramodular connectivity for each transcript from the three co-expression networks showing progressive changes in rTg4510 mice is strongly correlated with actual levels of pathology in the hippocampus measured using immunohistochemistry.
Shown is the correlation of intramodular connectivity for each transcript with tau pathology in the hippocampus in **(a)** the turquoise module (Pearson correlation, r = 0.75, *P* < 1.00E-200), **(b)** the yellow module (Pearson correlation, r = 0.48, *P* = 1.40E-64), and **(c)** the red module (Pearson correlation, r = 0.75, *P* = 4.30E-132). Each circle represents a single transcript. Total n = 58 animals (6-8 animals per group).

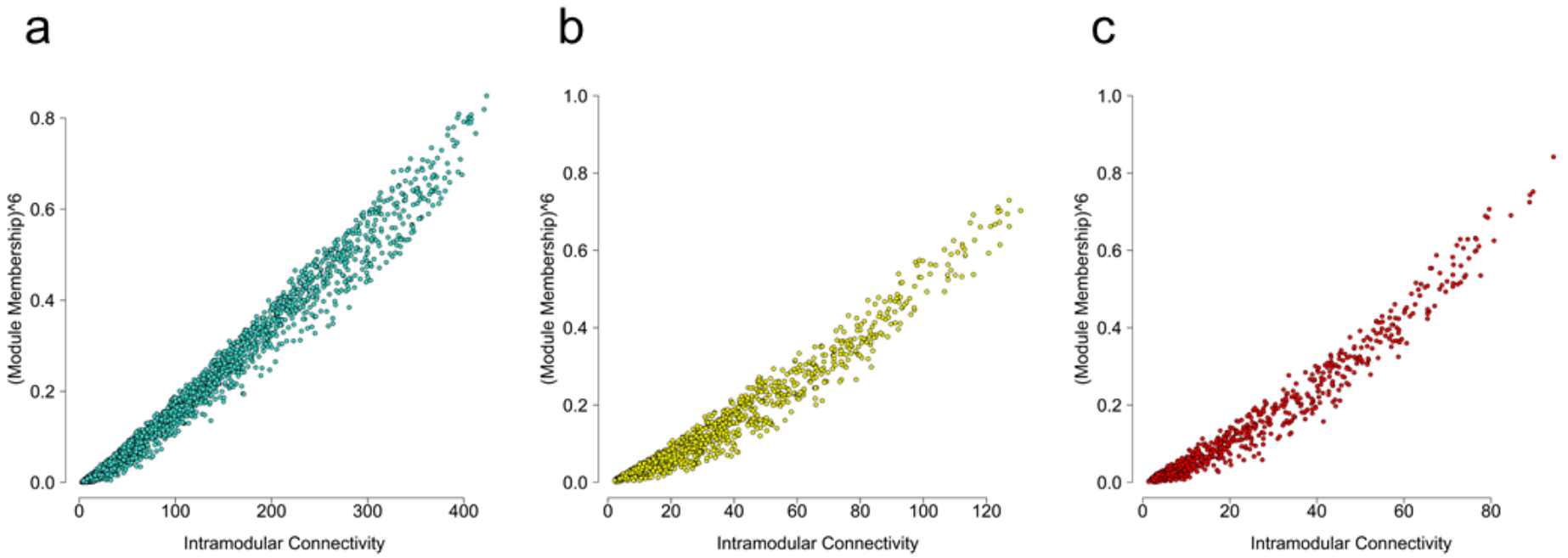


Figure 4.65 – Intramodular connectivity for each transcript from the three co-expression networks showing progressive changes in rTg4510 mice is strongly correlated with its module membership.

Shown is the correlation of module membership raised to power 6 (y-axis) against intramodular connectivity (x-axis) in **(a)** the turquoise module (Pearson correlation, r = 0.99, *P* < 1.00E-200), **(b)** the yellow module (Pearson correlation, r = 0.98, *P* < 1.00E-200), and **(c)** the red module (Pearson correlation, r = 0.98, *P* < 1.00E-200). Each circle represents a single transcript. Total n = 58 animals (6-8 animals per group).

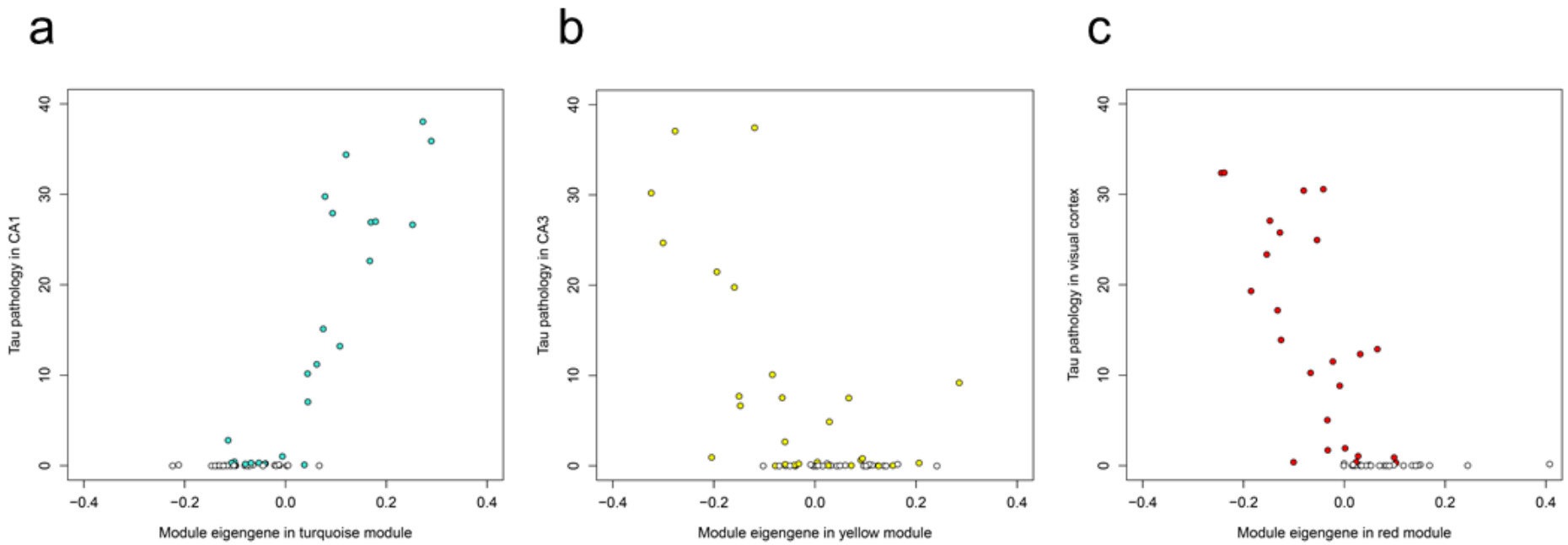


Figure 4.66 – The three tau-associated co-expression modules are strongly correlated with pathological burden measured by immunohistochemistry in regions of the brain affected early in AD.

(a) Correlation of module eigengenes for each sample with tau pathology in the CA1 sub-region of the hippocampus in the turquoise module (Pearson correlation, r = 0.87). **(b)** Correlation of module eigengenes for each sample with tau pathology in the CA3 sub-region of the hippocampus in the yellow module (Pearson correlation, r = -0.63). **(c)** Correlation of module eigengenes for each sample with tau pathology in the visual cortex in the red module (Pearson correlation, r = -0.79). Correlations and corresponding *P* values for all modules and traits can be seen in **Figure 4.62**. Each circle represents a single individual mouse, with coloured circles representing rTg4510 TG mice and white circles representing WT control mice. Total n = 58 animals (6-8 animals per group).

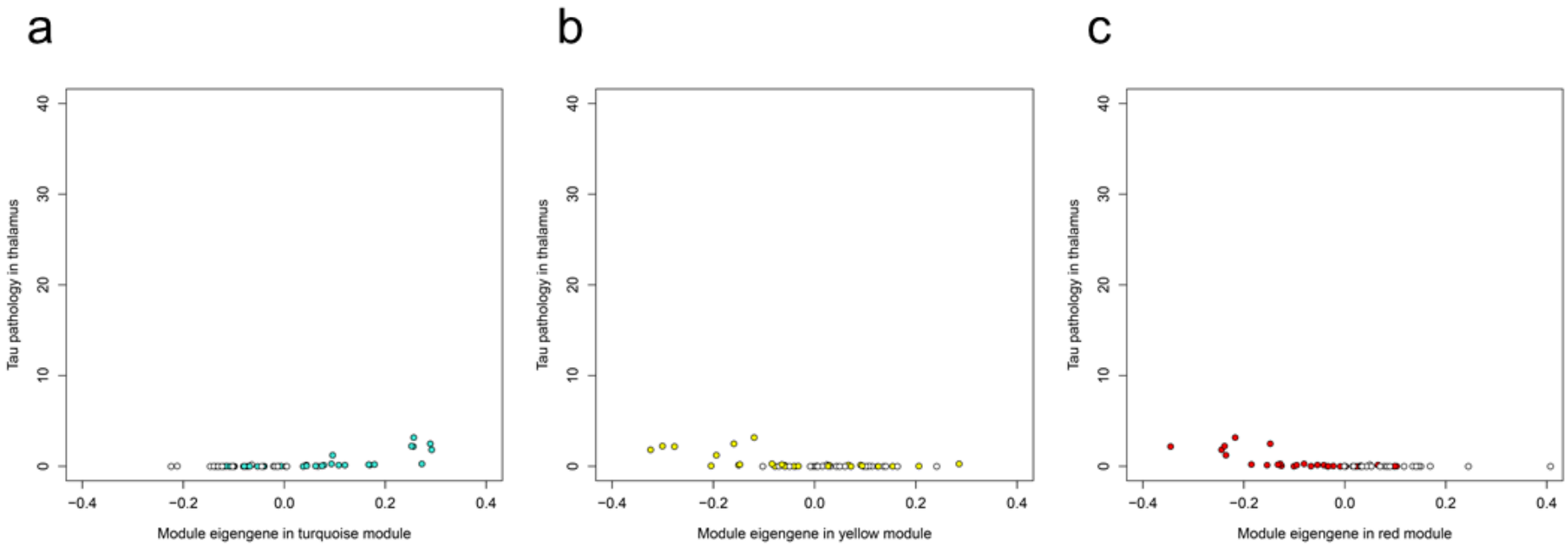


Figure 4.67 – The three tau-associated co-expression modules are less strongly correlated with pathological burden measured by immunohistochemistry in regions of the brain affected in later stages.
(a) Correlation of module eigengenes for each sample with tau pathology in the thalamus in the turquoise module (Pearson correlation, r = 0.70). **(b)** Correlation of module eigengenes for each sample with tau pathology in the thalamus in the yellow module (Pearson correlation, r = -0.59). **(c)** Correlation of module eigengenes for each sample with tau pathology in the thalamus in the red module (Pearson correlation, r = -0.64). Correlations and corresponding *P* values for all modules and traits can be seen in **Figure 4.62**. Each circle represents a single individual mouse, with coloured circles representing rTg4510 TG mice and white circles representing WT control mice. Total n = 58 animals (6-8 animals per group).

Within each of these three modules we ranked transcripts based on their intramodular connectivity to identify “hub” genes within each network, finding many genes known to play a major role in the neuro-immunological and neurodegenerative processes involved in AD. In the turquoise module the four genes with the highest intramodular connectivity (i.e. those with most connections to other genes) were *Cd63*, *Msn*, *Npc2* and *Tnfrsf1a* (**Supplementary Table 17)**, with other highly interconnected transcripts including several genes identified as having a role in LOAD from GWAS (e.g. *Abca1*, *Clu* and *Apoe*) in addition to genes previously implicated in AD pathology (e.g. *Itgax*, *Clec7a* and *Cd68*). Furthermore, genes identified as having the strongest connections (edges) to other genes (nodes) in the turquoise module included *C1qb*, *Mpeg1*, *Tyrobp*, and *Trem2* (**Figure 4.68**). In the yellow module the four genes with the highest intramodular connectivity were *Atp9a*, *Ywhag*, *Rab3a* and *Svop* (**Supplementary Table 18)**, with *App* also being a highly-connected gene in this module. Genes identified as having the strongest connections to other genes in the yellow module included *Atp9a*, *Faim2*, *Ppp2r1a* (**Figure 4.69).** In the red module *Atxn7l3*, *Sept5*, *Cbx6* and *Fbxl16* were the top most connected genes (**Supplementary Table 19**). Genes with the strongest connections to other genes in the red module included *Dlgap3*, *Shank3*, *Epn1*, and *Fbxl16* (**Figure 4.70**).

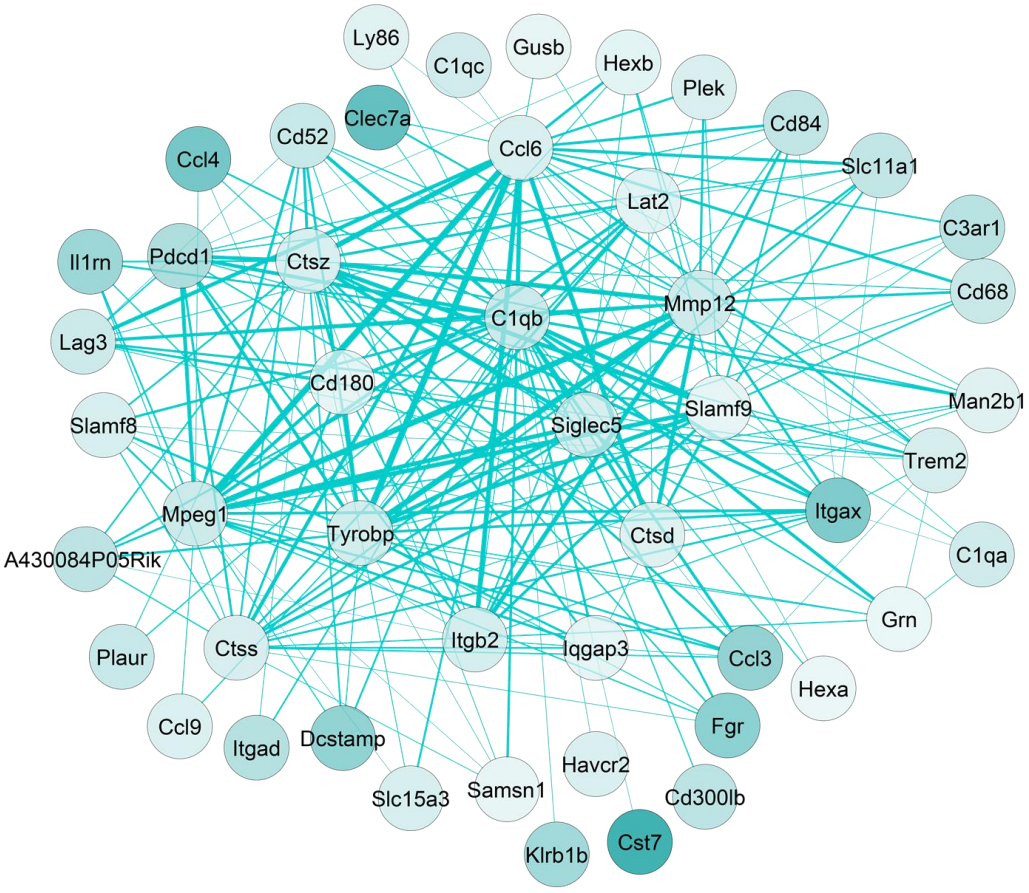


Figure 4.68 – Network plot highlighting core members of the turquoise gene co-expression module associated with the development of tau pathology.
Shown are the top 50 nodes (i.e. genes) with the strongest edges (representing individual connections with other genes) for the turquoise module (all genes with increased expression – log2 fold change for latest time point against baseline). Stronger colours reflect higher absolute log2 fold change (8 months against 2 months). Total n = 58 animals (6-8 animals per group).

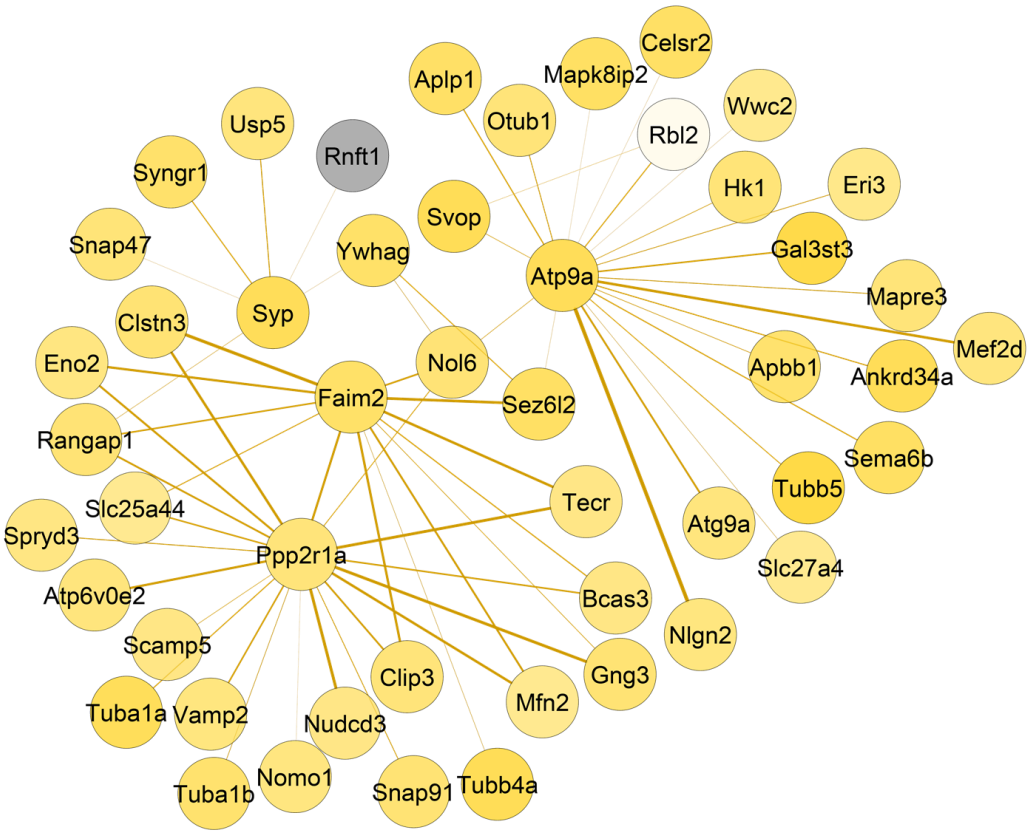


Figure 4.69 – Network plot highlighting core members of the yellow gene co-expression module associated with the development of tau pathology.
Shown are the top 50 nodes (i.e. genes) with the strongest edges (representing individual connections with other genes) for the yellow module (genes with decreased expression are shown in yellow, and genes with increased expression are shown in grey). Stronger colours reflect higher absolute log2 fold change (8 months against 2 months). Total n = 58 animals (6-8 animals per group).

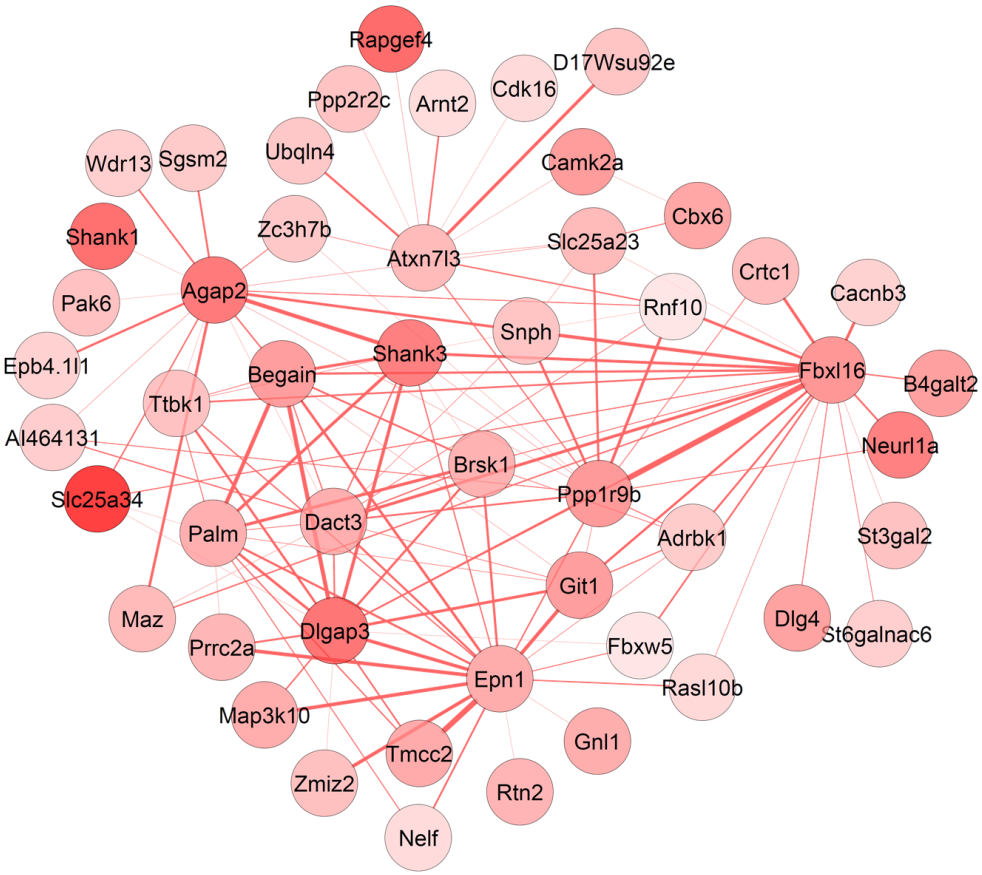


Figure 4.70 – Network plot highlighting core members of the red gene co-expression module associated with the development of tau pathology.
Shown are the top 50 nodes (i.e. genes) with the strongest edges (representing individual connections with other genes) for the red module (all genes with decreased expression). Stronger colours reflect higher absolute log2 fold change (8 months against 2 months). Total n = 58 animals (6-8 animals per group).

4.3.8. Co-expression changes identified in rTg4510 mice overlap with AD-associated co-expression changes from human studies

We next compared the significant rTg4510 co-expression modules to AD-associated co-expression modules reported in a recent human post-mortem RNA-seq meta-analysis, focusing on modules identified in dorsolateral prefrontal cortex (DLPFC) and temporal cortex (TCX) (Logsdon et al., 2019). Briefly, we used a hypergeometric test to identify overlaps between the six rTg4510-associated co-expression modules (“salmon”, “turquoise”, “purple”, “yellow”, “light-cyan”, and “red”) and four DLPFC (**Supplementary Table 20**) and five TCX

(**Supplementary Table 21**) AD-associated human co-expression modules, restricting our analysis to mouse-human homologs (see **Section 4.2.6.1**). After controlling for the number of comparisons performed for each of the human brain regions (DLPFC: Bonferroni-corrected $P < 0.0021$; TCX: Bonferroni-corrected $P < 0.0017$), each of the rTg4510-associated modules was found to significantly overlap with at least one AD-associated module in both human cortical regions. For example, genes in the turquoise rTg4510 module (enriched for pathways involved in activation of the immune system (**Supplementary Table 13**)), were found to overlap significantly with two human DLPFC modules ("DLPFC-blue" and "DLPFC-brown" from Logsdon et al.) and three TCX modules ("TCX-blue", "TCX-turquoise" and "TCX-yellow" from Logsdon et al.) associated with AD; for this module the largest proportion of overlaps in genes were found with the "DLPFC-blue" module (n = 658 genes, 40.82% of the human module gene list, $P < 2.2E-16$) and the "TCX-turquoise" module (n = 389 genes, 39.1% of the human module gene list, $P < 2.2E-16$). Interestingly, GOseq analysis (see **Section 4.2.6**) highlighted a strong enrichment for immune response processes amongst the rTg4510 turquoise module genes overlapping with those in both the "DLPFC-blue" module (**Supplementary Table 22**) and the "TCX-turquoise" module (**Supplementary Table 23**). Reflecting the similarities between these two human cortex modules, the list of overlapping genes includes many of the core hub transcripts identified in the turquoise rTg4510 turquoise module for both the "DLPFC-blue" module (e.g. *CD63*, *ABCA1*, *CLU*, *APOE*, *ITGAX*, *CLEC7A*, *C1QB*, *TYROBP*, and *TREM2*) and "TCX-turquoise" module (e.g. *CD63*, *ITGAX*, *CLEC7A*, *C1QB*, *TYROBP*, and *TREM2*). Together, these results indicate that the transcriptional networks associated with tau pathology in rTg4510 mice overlap considerably with those identified in human AD cortex and are involved in driving common molecular pathways.

4.4. Discussion

4.4.1 Overview of results

In this study, I quantified transcript read counts and profiled gene expression in TG and WT animals in a model of tau (rTg4510) and a model of amyloid (J20) pathology. To my knowledge, this represents the most extensive gene expression dataset generated on rodent models of AD pathology, providing excellent power to identify transcriptional variation associated with mutations in *MAPT* and *APP*, and the progressive changes in gene expression accompanying the development of AD pathology in TG mice. I identified transcriptional changes in the entorhinal cortex associated with the progression of AD-associated pathology in the two transgenic models. I found robust genotype-associated differences in entorhinal cortex gene expression in both models and identified widespread changes in gene expression paralleling the development of tau pathology in rTg4510 mice and reflecting alterations observed in other models of tau pathology. Of note, the list of transcripts progressively altered in rTg4510 mice includes genes robustly associated with familial AD from genetic studies of human patients, including *App* which is a key driver of amyloid pathology. It also includes genes annotated to both common and rare variants identified in GWAS and exome-sequencing studies of late-onset sporadic AD. Systems-level analyses identified discrete co-expression networks associated with the progressive accumulation of tau, with these also being enriched for genes and pathways previously implicated in neuroimmune and neurodegenerative processes driving AD pathology. Further support for upregulation of immune system genes in response to tau pathology comes from our finding of increased expression of complement pathway genes including C1qa, C1qb, and C1qc. Finally, we compared these tau-associated networks to those identified in human post-mortem tissue from AD individuals, finding considerable overlap with disease-associated co-expression modules.

To our knowledge, our study represents the most systematic analysis of transcriptional variation in mouse models of tau and amyloid pathology, and is the first to focus specifically on changes in the entorhinal cortex, a key region of the brain implicated early in the pathogenesis of AD (Braak and Braak, 1991). Compared to previous studies of transcriptional variation in transgenic mouse models of AD we profiled a relatively large number of samples spanning multiple

time-points selected to encompass the development of pathology; my study was therefore well powered to identify gene expression differences associated with both genotype and the progression of AD pathology. Furthermore, we implemented a statistical approach that enabled us to detect progressive changes in gene expression across age between the TG and WT samples, not only identifying stable differences induced by the transgene at each time-point, but also assessing temporal transcriptional changes relative to baseline *within* mutant mice. Our detailed immunohistochemical analyses (described in **Chapter 3**) also allowed me to directly compare transcriptional variation with measures of tau and amyloid pathology measured in the same individual mice.

Together my results reflect the dramatic upregulation of microglial genes observed in studies of other AD rodent models (Kamphuis et al., 2016, Kan et al., 2015, Keren-Shaul et al., 2017, Landel et al., 2014, Matarin et al., 2015, Rothman et al., 2018), and also support a role – either causal or consequential – for dysregulation of the central nervous system (CNS) immune system in the development of AD pathology. In line with other studies using murine models of AD, mostly of amyloid pathology, *Cst7* and *Itgax* (*Cd11c*) showed increased expression both in rTg4510 and J20 entorhinal cortex as a result of accumulation of neuropathology. These two genes are markers for activated microglia; therefore, this observation is most probably microglia-associated, as shown in the single-cell gene expression studies discussed previously (**Chapter 1**, **Section 1.5.2**), suggesting an important role for microglia in the presence of pathological tau and amyloid.

In summary, this study provides compelling evidence for widespread transcriptional changes in the entorhinal cortex paralleling the progression of AD pathology. Our data suggest that the altered expression of multiple genes, including several known AD risk genes is robustly associated with the accumulation of tau, with tau-associated co-expression networks overlapping those altered in human AD cortex. Our data provide further support for an immune-response component in the accumulation of tau, and reveal novel molecular pathways associated with the progression of AD neuropathology.

4.4.2. Limitations

The work presented in this chapter has a number of important limitations that should be considered when interpreting my results.

First, to minimise the heterogeneity in our analysis we only profiled female mice. However, a number of sex differences have been previously reported for these models, with females demonstrating elevated and more progressive pathology than males (Blackmore et al., 2017, Yue et al., 2011). Future work should focus on examining the extent to which the transcriptional profiles identified here are consistent between male and female mice. Second, our analysis was performed on bulk entorhinal cortex tissue, comprising a mix of different neural cell-types; consequently, changes in the fractional contribution of any given cell type to the total cellular population will contribute to the observed outcomes at each time-point. Given the compelling evidence in our data for an enrichment of microglial markers, previously shown to be upregulated in AD (Hopperton et al., 2018, Keren-Shaul et al., 2017), as well as increased expression of canonical markers of astrocytes, future work should focus on identifying changes that occur within these and other brain cell-types. Of note, immunocytochemistry analyses of tissue sections from the left-brain hemisphere of these mice revealed a progressive increase in the microglia/macrophage marker Iba1, indicating that our bulk-tissue RNA-seq measurements reflect real underlying cellular changes. In rTg4510 mice it is also interesting to consider neuron-specific genes of which expression is not decreased, in what is a falling total neuronal population; these might represent transcripts which expression is increased in response to neuropathology in neuronal cells. Third, compared to the rTg4510 model, relatively few transcriptional changes were observed in J20 mice, potentially reflecting the slower and later accumulation of pathology (Harris et al., 2010), as well as the potential absence of neurodegeneration, in the entorhinal cortex in this model; future work should focus on the analysis of other brain regions more directly affected in the early stages of amyloid pathology. Interestingly, however, we found that effect sizes for the transcripts identified as being progressively dysregulated in rTg4510 mice were significantly correlated across both models, suggesting some common transcriptional mechanisms are involved in both tau and amyloid pathology.

Chapter 5. Global DNA modifications profiling in the brain of tau and amyloid mouse models

5.1. Introduction

Despite global changes in DNA modifications being robustly reported in cancer, only limited work has been done in AD, with inconclusive results (see **Section 1.4.1.1** in **Chapter 1)**. Even less studies have been done in rodent models of AD (see **Section 1.5.2.2** in **Chapter 1**), looking at a very low number of samples, with no studies conducted in mouse models of tau pathology to date. One study in particular included brain samples from J20 mice and reported an age-related (4, 8, 16, and 34 months) decrease in 5mC in the dentate gyrus (DG) hippocampal sub-region in J20 transgenic mice compared to controls, as well as a negative correlation between plaque load and 5mC in the DG (Lardenoije et al., 2018). A decrease in 5mC/5hmC ratio in J20 transgenic mice compared to controls in the DG and CA3 hippocampal sub-regions was also reported in this study.

This chapter describes my analysis of global DNA modification levels in brain tissue from rTg4510 and J20 mice, using two independent approaches. First, I quantified global DNA methylation in the hippocampus (see **Section 2.4.1** and **Section 2.5.2** in **Chapter 2** for details about brain dissections and isolation of DNA) from rTg4510 and J20 transgenic (TG) and wild-type (WT) littermate control mice (described in **Chapter 2**) using a modified version of the Luminometric Methylation Assay (LUMA). Second, using immunohistochemistry, I directly quantified levels of both 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) in multiple brain regions (**Figure 5.11**) in rTg4510 and J20 mice, using right hemisphere tissue sections from transgenic (TG) and wild-type (WT) littermate control mice (see **Section 2.4.2** in **Chapter** 2 for details about brain collection).

LUMA involves the use of differential methylation sensitive restriction enzymes followed by polymerase extension assay by pyrosequencing to produce an estimation (percentage) of global methylation (Karimi et al., 2006, Karimi et al., 2011, Luttropp et al., 2015).The LUMA experiments described here were performed in conjunction with an undergraduate student (Hedley Baulf) who I was directly supervising during his Professional Training Year (PTY) placement in our lab.

As explained in **Chapter 3**, immunohistochemistry relies on the selective identification of antigens within tissue sections by means of specific antibodies. Antigen-antibody binding is then visualised using a coloured histochemical reaction, visible by light microscopy (Ramos-Vara and Miller, 2014). All immunohistochemistry experiments for this chapter were performed at Eli Lilly & Co. Ltd. (Windlesham, United Kingdom), where I spent about 3.5 months during my PhD (from May 2017 to August 2017). I prepared the sections with the help of Mark Ward (Eli Lilly) and performed the immunostaining procedures. Josh Harvey (Eli Lilly) helped me with the visualisation and quantification.

5.2. Methods

In this section the experimental procedures used for i) the evaluation of global DNA modification levels in hippocampus samples dissected from rTg4510 and J20 mice (described in **Chapter 2**, **Section 2.4** and **Section 2.5**) using LUMA, and ii) the evaluation of DNA methylation and DNA hydroxymethylation levels in brain tissue sections (described in **Chapter 2**, **Section 2.4**, and in **Chapter 3**) using immunohistochemistry from rTg4510 and J20 mice (**Chapter 2**, **Section 2.1**) are described in detail.

5.2.1. Assessment of DNA quality and integrity

DNA samples used for LUMA were quantified and assessed for purity using a NanoDrop 8000 spectrophotometer (Thermo Scientific), as described in **Section 2.5.3.1**. One rTg4510 sample had a DNA concentration below the required minimum and was therefore excluded, leaving a total of 75 rTg4510 samples (initial mean concentration = 40.29 ng/µL (SD = 11.58 ng/µL)), and 77 J20 samples (initial mean concentration = 53.40 ng/µL (SD = 13.94 ng/µL)).

5.2.2. Sample dilutions for LUMA

To ensure consistency in DNA concentrations across all samples, original DNA samples were diluted to 30 ng/µL in Elution Buffer (Qiagen). 5 µL of each dilution (150 ng) was then transferred into a 96-well plate to be used for LUMA.

5.2.3. LUMA

We used a modified version of LUMA (introduced in **Section 5.1**), using HpaII (Thermo Scientific) and MspI (Thermo Scientific) as the differential methylation sensitive restriction enzymes as originally described (Karimi et al., 2006, Karimi et al., 2011) and MunI (Thermo Scientific) as internal reference (Luttropp et al., 2015). Both HpaII and MspI cleave the sequence 5'-CCGG-3' resulting in 5'-CG overhangs; however, HpaII is methylation sensitive and therefore only cleaves at unmethylated sites, whilst MspI is methylation insensitive and cleaves at both methylated and unmethylated sites (Karimi et al., 2006, Karimi et al., 2011). MunI is unaffected by DNA methylation status and recognises the sequence 5'-GAATTC-3', producing 5'-AATT overhangs which can be distinguished from the 5'-CG overhangs generated by from HpaII and MspI and therefore used for normalisation. We used MunI as an alternative to EcoRI – the enzyme used in the original protocol (Karimi et al., 2006, Karimi et al., 2011) – because of its similar action and lack of altered specificity at suboptimal conditions ("STAR activity") which is a characteristic of EcoRI (Lisanti et al., 2013).

An overview of the modified version of LUMA, including the four-step pyrosequencing reaction, is given in **Figure 5.1**. Samples from each mouse

model were processed separately and individual samples were randomised to ensure that each group was equally represented in each processing batch. Moreover, samples were run in duplicates and each duplicate was run on different days to control for batch effects. A fully methylated DNA control sample (Thermo Scientific) was included in each assay and used as a positive control. All experiments were performed in a blinded and randomised fashion.

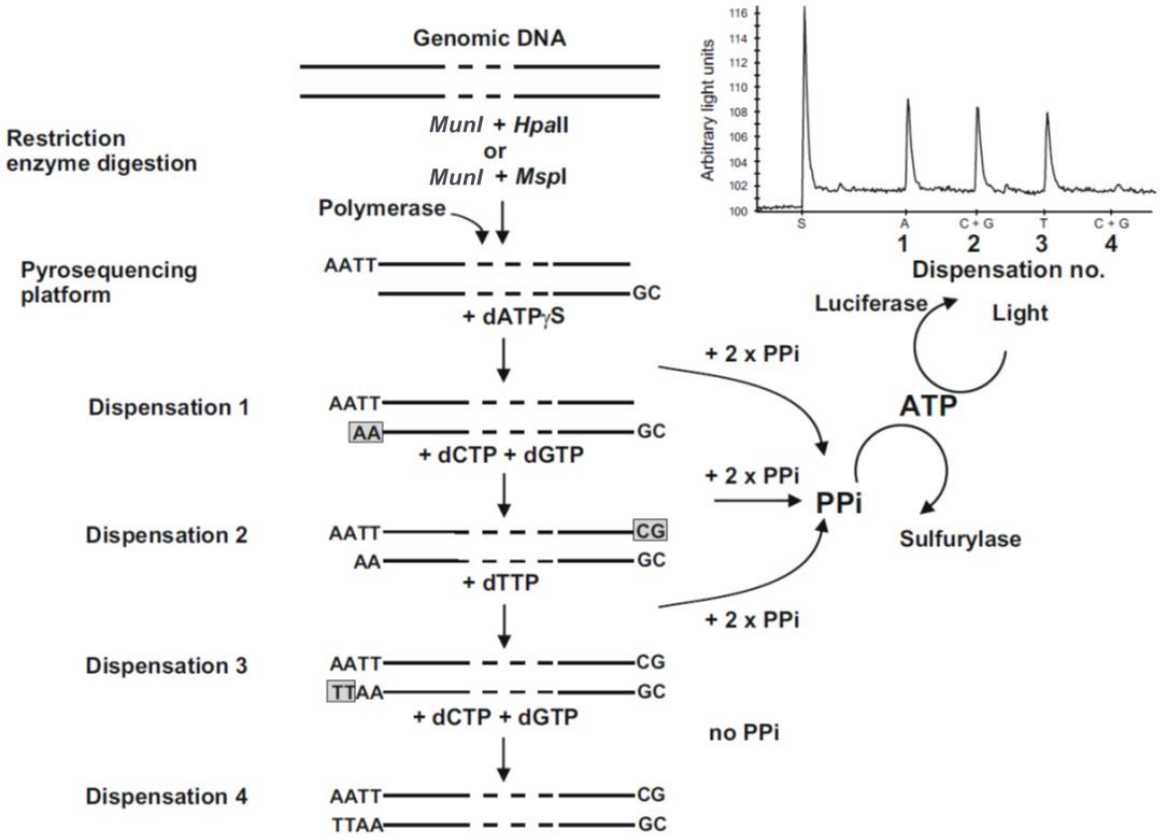


Figure 5.1 – An overview of LUMA.
Genomic DNA from each sample is digested with two combinations of restriction enzymes (MunI + HpaII or MunI + MspI), followed by a polymerase extension assay based on a four-step pyrosequencing reaction. In the pyrosequencer, after each nucleotide dispensation, inorganic pyrophosphate (PPi) is released and converted to ATP by ATP-sulfurylase. The resulting ATP is then used by luciferase to activate luciferin. The amount of luciferin activation produces a proportional amount of visible light, which is detected by a charge coupled device camera. The amount of light is directly proportional to the number of overhangs produced by the respective restriction enzymes with each step in the pyrosequencing reaction correlating to a peak in the resulting pyrogram via light capture by the camera: 1) the A and T peaks correspond to pyrosequencing Step 1 and Step 3, reflecting the MunI cleavage and should be equal; 2) the C + G peak resulting from pyrosequencing Step 2 illustrates HpaII or MspI cleavage; 3) the second C + G peak originating from Step 4 is an internal control that should be close to zero. The height of the light peaks is used for quantification. Figure and legend adapted from Karimi et al. (2006) and Luttropp et al. (2015).

5.2.3.1. Restriction enzymatic digestion of genomic DNA

Digestion reactions containing MunI + HpaII (Mix A) or MunI + MspI (Mix B) were run in parallel on the same DNA sample (**Figure 5.1**). Preparation of the master mixes (Mix A and Mix B) for restriction enzymatic digestion per 5µL of diluted DNA Sample (30 ng/µl) were prepared for each restriction enzyme digestion as follows:

MunI + HpaII (Mix A)	**MunI + MspI (Mix B)**
12 µL DNAse-Free Water	12 µL DNAse-Free Water
2 µL 10 x Tango Buffer	2 µL 10 x Tango Buffer
0.5 µL MunI (10U/µL)	0.5 µL MunI (10U/µL)
0.5 µL HpaII (10U/µL)	0.5 µL MspI (10U/µL)

15 µL of Mix A (MunI + HpaII) or Mix B (MunI + MspI) were added to 5 µL of diluted DNA samples (30 ng/µl) and mixed by pipetting. The plates were then sealed and incubated at 37ºC overnight in a thermocycler. Following the incubation period, 20 µL of Pyrosequencing Annealing Buffer (Qiagen) was added to each well of the pyrosequencing plate and mixed by pipetting. 40 µL of the annealing buffer/restriction digest solution was transferred to a pyrosequencing plate and placed within the pyrosequencer.

5.2.3.2. Evaluation of enzymatic reactions by agarose gel electrophoresis

To evaluate the success of the LUMA enzymatic digestions, two ‘test’ digested DNA samples were run on a 0.8% agarose gel. Agarose gel electrophoresis is explained in detail in **Chapter 2** (**Section 2.5.3.2**). For this experiment we used the fluorescent nucleic acid stain SYTO 60 (Thermo Scientific) for visualisation (10 µL of SYTO 60 in 100 mL of agarose powder dissolved in 1X TBE).

5 µL of the digested DNA sample, as well as the original diluted DNA test samples (30 ng/µL), were mixed with 2 µL of loading buffer (Orange G, Sigma-Aldrich) in a sterile 96-well PCR plate. 5 µL of each sample-loading buffer solution were loaded into an individual well in the agarose gel, and 5 µL of a molecular-weight size marker (1kb DNA ladder, New England BioLabs) were loaded into the first and last wells of the gel. The gel was left to run at 120 v for 40 mins. The LI-COR

Odyssey CLx scanner and LI-COR Image Studio Lite software were used to visualise the presence, size and intensity of each DNA fragment. The resulting gel highlighting digestion of genomic DNA is shown in **Figure 5.2**.

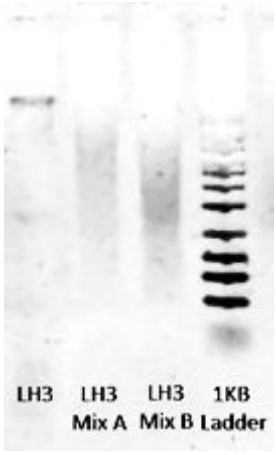


Figure 5.2 – Agarose gel obtained for the evaluation of enzymatic reactions in test samples.
LH3: original sample, showing one band corresponding to the genomic DNA; Mix A: genomic DNA treated with MunI + Hpall; Mix B: genomic DNA treated with MunI + MspI; Ladder: molecular weight marker. Smears indicate enzymatically digested DNA.

5.2.3.3. Pyrosequencing

The ratio of 5'-CG overhangs resulting from the two enzymatic reactions was quantified via the integration of nucleotides (dNTPs) into restriction sites using pyrosequencing. The pyrosequencing reaction involves the addition of dNTPs in four consecutive steps in a predetermined order, discriminating the 5'-CG (HpaII/MspI) from the 5'-AATT (MunI) overhangs. The incorporation of the nucleotides with DNA polymerase results in release of inorganic pyrophosphate (PPi), followed by its conversion into ATP by ATP-sulfurylase and adenosine-5'-phosphosulfate. The newly formed ATP is then used for the conversion of luciferin into oxyluciferin by luciferase, which produces visible light (**Figure 5.1**). The light, proportional to dNTP incorporation in the original 5'-overhang produced by the restriction enzyme, is then detected by a charge coupled device camera and displayed as peaks in the pyrosequencer software. The resulting ratio of overhangs produced from each enzyme digestion reaction can then be used to calculate the percentage of global DNA methylation in each sample.

Pyrosequencing assays were created using the PyroMark Q24 2.0.6 software (Qiagen) using the 'AQ Assay' (SNP) setup. The pyrosequencing cartridge

(**Figure 5.3**) was loaded using Pyromark Gold Q96 Reagents (Qiagen) as described in **Table 5.1** and the dispensation order of the nucleotides was "ACTCGA".

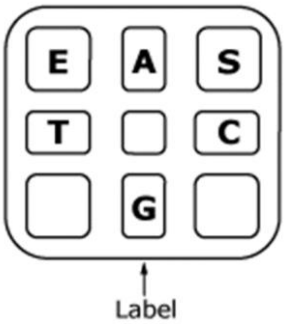


Figure 5.3 Schematic representation of the PyroMark Q24 cartridge.
Image taken from the PyroMark Q24 Cartridge Product Sheet (Qiagen, 2013).

Table 5.1 – Pyrosequencing cartridge loading details for a 24-well pyrosequencing plate.

Position in cartridge	Reagent	Volume
E	Pyrosequencing Enzyme	90 µL
S	Pyrosequencing Substrate	90 µL
A	dATP + dH_2O	50 µL + 50 µL
C	dCTP + dGTP	50 µL + 50 µL
G	dH_2O	100 µL
T	dTTP + dH_2O	50 µL + 50 µL

dH_20 = nuclease-free water.

5.2.3.4. Data analysis

Pyrograms produced from each assay were analysed using the PyroMark Q24 2.0.6 software (Qiagen) as shown in **Figure 5.4**.

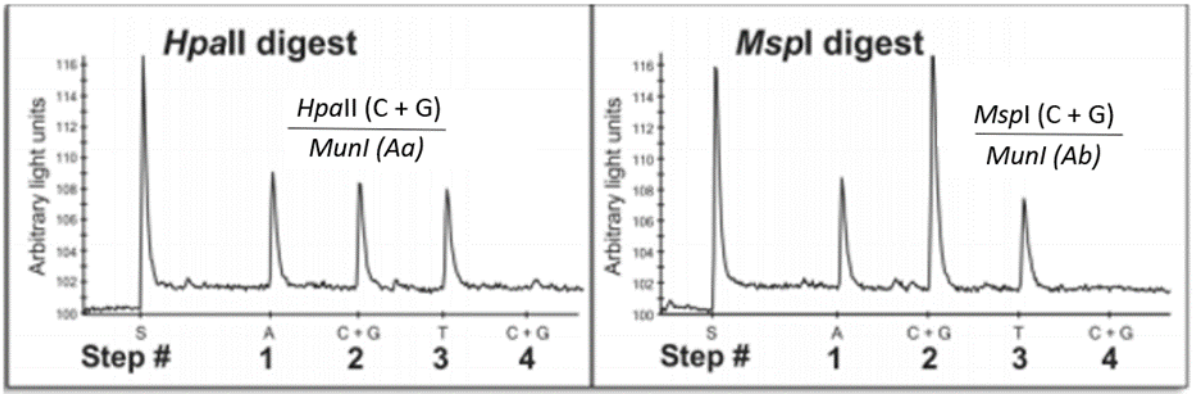


Figure 5.4 – Typical pyrograms for LUMA for each enzyme restriction reaction (HpaII or MspI) from the same DNA sample.
Numbers located on the x-axis correspond to the dispensation steps in **Figure 5.1**.
Image and legend adapted from Karimi et al. (2006).

Peak heights were extracted and global DNA methylation levels (%) were determined using the following equation:

$$Global\ DNA\ methylation\ (\%) = \left(1 - \left(\frac{\frac{Hpa\mathrm{II(C+G)}}{Mun\mathrm{I(Aa)}}}{\frac{Msp\mathrm{I(C+G)}}{Mun\mathrm{I(Ab)}}}\right)\right) X100$$

Where:
HpaII (C+G) = Mean peak height for dispensation 2 (Mix A)
MunI (Aa) = Mean peak height for dispensation 1 (Mix A)
MspI (C+G) = Mean peak height for dispensation 2 (Mix B)
MunI (Ab) = Mean peak height for dispensation 1 (Mix B)

All samples were run in duplicate (**Section 5.2.3**), and an average of the two global methylation values from both duplicates was calculated for each sample.

5.2.4. Tissue preparation for immunohistochemistry

A general overview of the immunohistochemistry procedures for the quantification of 5mC and 5hmC, from tissue preparation to visualisation and quantification, is given in **Figure 5.5**.

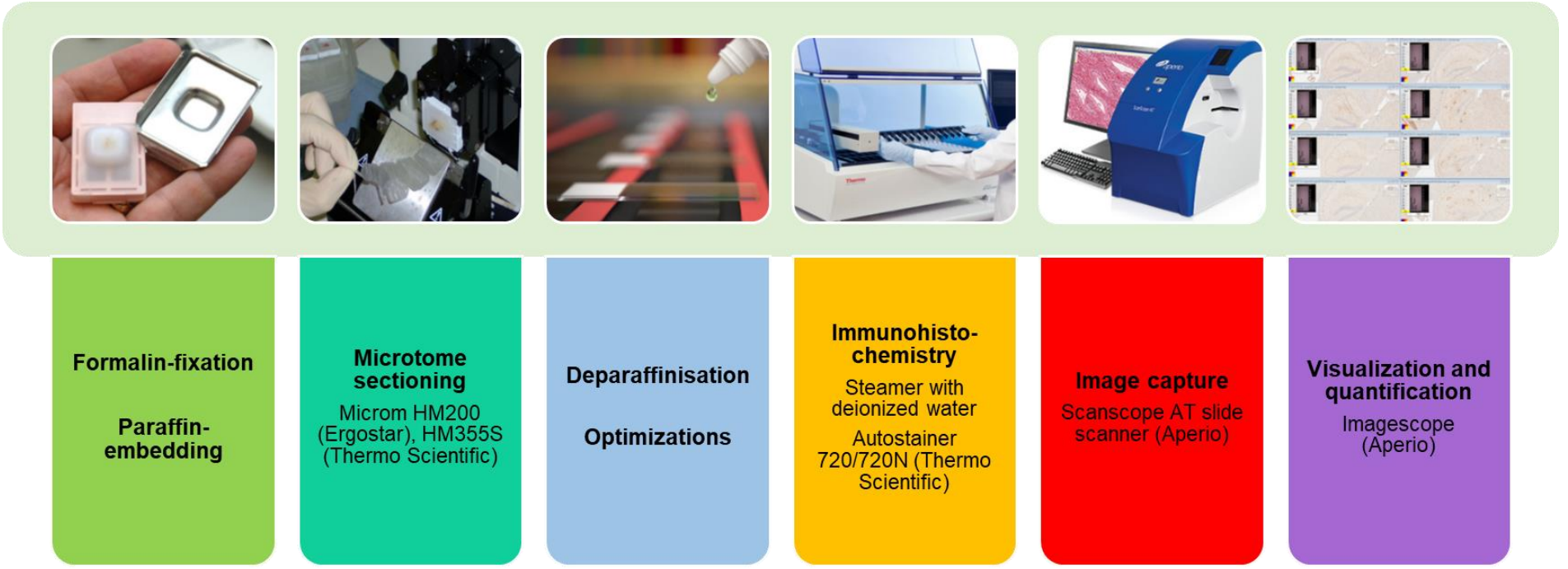


Figure 5.5 – An overview of the steps used for the quantification of 5mC and 5hmC in brain sections from rTg4510 and J20 mice by immunohistochemistry.

As described in **Chapter 3**, right hemisphere brain tissue from each individual mouse was processed at Eli Lilly (in the Molecular Pathology department) using the Tissue TEK® VIP processor (GMI Inc), embedded in paraffin wax and stored for sectioning. During my placement at Eli Lilly, I prepared 6 µm serial sagittal sections (from bregma 0.84 to 1.08) using rotary microtomes (HM 200 from Ergostar and HM 355S from Thermo Scientific), which were mounted on glass slides (two sections per slide) for immunohistochemistry assessments.

5.2.5. Deparaffinisation of the tissue for immunohistochemistry

As described previously in **Chapter 3** (see **Section 3.2.2**), deparaffinisation of the tissue was achieved by washing with xylene (Fisher Scientific), followed by 70% ethanol (industrial methylated spirit, Fisher Scientific) and deionised water for rehydration of the sections.

5.2.6. Immunohistochemistry

Because we have not previously assessed 5mC and 5hmC levels in the mouse brain using immunohistochemistry in our lab, the first step of this study was to optimise the immunostaining experimental procedures. Before starting, I performed a comprehensive search for potential adequate antibodies; despite not finding many options available, according to information from the manufacturer, the antibodies used in this chapter looked promising for use in paraffin-embedded mouse brain tissue sections. Optimisation steps, which gave rise to satisfactory results to move forward with the stains for my rTg4510 and J20 samples, and the final protocol are described below.

All experiments were performed in a blinded and randomised fashion. One slide containing no primary antibody and one slide containing no secondary antibody were included as negative controls.

5.2.6.1. Antigen retrieval and blocking

During the optimisation steps I initially performed heat-induced epitope retrieval in a PT Module (Thermo Scientific) containing citrate buffer (dilution 1:100), as described previously ('Standard protocol', described in section **3.2.3.2** in **Chapter 3**) (**Figure 5.7**, **Figure 5.8**, **Figure 5.9** and **Figure 5.10**). However, we noticed considerable background staining and decided to test if performing heat-induced antigen retrieval in an electric steamer containing deionised water as an alternative ('Alternative protocol') to citrate buffer would give more optimal results. Indeed, this slight modification dramatically improved the background staining (**Figure 5.6**), and therefore was the method that I used for antigen retrieval in the brain sections from rTg4510 and J20 mice used in this study. Samples were blocked using normal goat serum (Vector labs, catalogue number S-1000).

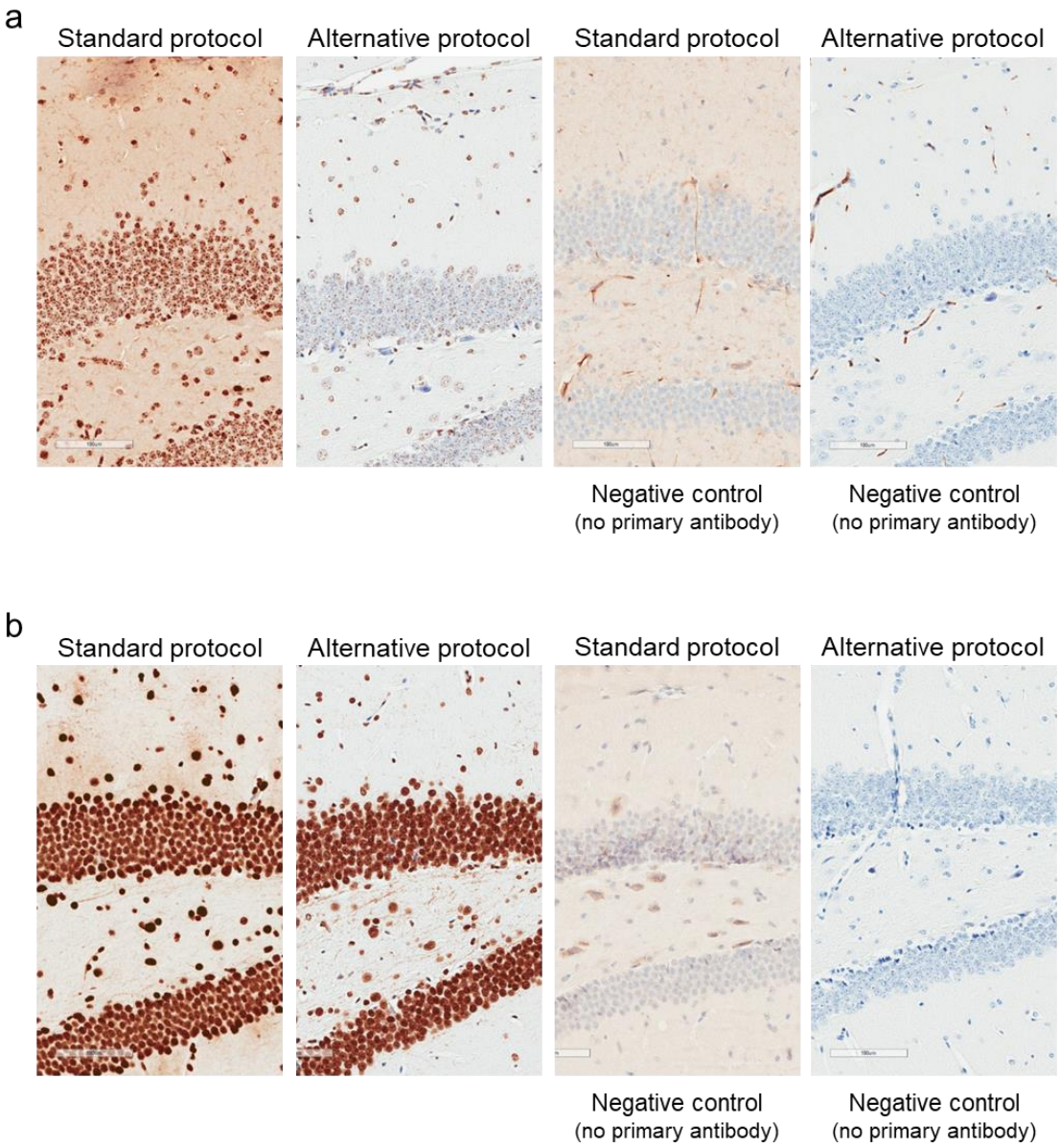


Figure 5.6 – Two methods for antigen retrieval were tested.
(a) Immunostain for the quantification of 5mC (primary antibody: anti-5mC (1:4000); secondary antibody: biotinylated goat anti-mouse (1:200)) showing the two methods tested for antigen retrieval. **(b)** Immunostain for the quantification of 5hmC. (primary antibody: anti-5hmC (1:4000); secondary antibody: biotinylated goat anti-rabbit (1:200)) showing the two methods tested for antigen retrieval. Standard protocol: PT Module with citrate buffer. Alternative protocol: steamer with deionised water.

5.2.6.2. Antigen-antibody labelling

For 5mC, we used mouse monoclonal anti-5-hydroxymethylcytosine (5hmC) antibody [RM236] (Abcam, catalogue number ab10805, lot number GR302463-15) as the primary antibody, which discriminates between the modified base 5mC and unmethylated cytosine, and biotinylated goat anti-mouse IgG (Vector labs, catalogue number BA-9200, lot number ZB0304) as the secondary antibody (1:200).

For 5hmC, we used rabbit monoclonal anti-5-methylcytosine (5mC) antibody [33D3] (Abcam, catalogue number ab214728, lot number GR284756-2) as the primary antibody, which reacts with 5-hydroxymethylcytosine in both single-stranded and double-stranded DNA with no cross reactivity with non-methylated cytosine and methylcytosine in DNA, and biotinylated goat anti-rabbit IgG (Vector labs, catalogue number BA-1000, lot number ZB1007) as the secondary antibody (1:200).

Considering the dilutions recommended by the manufacturer as a reference, I tested a range of dilutions (anti-5mC: 1:1000, 1:2000, 1:4000, 1:8000; anti-5hmC: 1:500, 1:1000, 1:2000, 1:4000) to determine the optimal concentration for the primary antibodies for our experiments (**Figure 5.7**, **Figure 5.8**, **Figure 5.9** and **Figure 5.10**). The 1:4000 dilution looked optimal for both stains, therefore we decided to proceed with a 1:4000 dilution for both anti-5mC and anti-5hmC quantifications in our rTg4510 and J20 brain tissue samples.

Optimisation experiments were performed manually; all samples for each of the final 5mC or 5hmC experiments were immunostained together in an autostainer (Autostainer 720 for rTg4510 samples and 720N for J20 samples, Thermo Scientific).

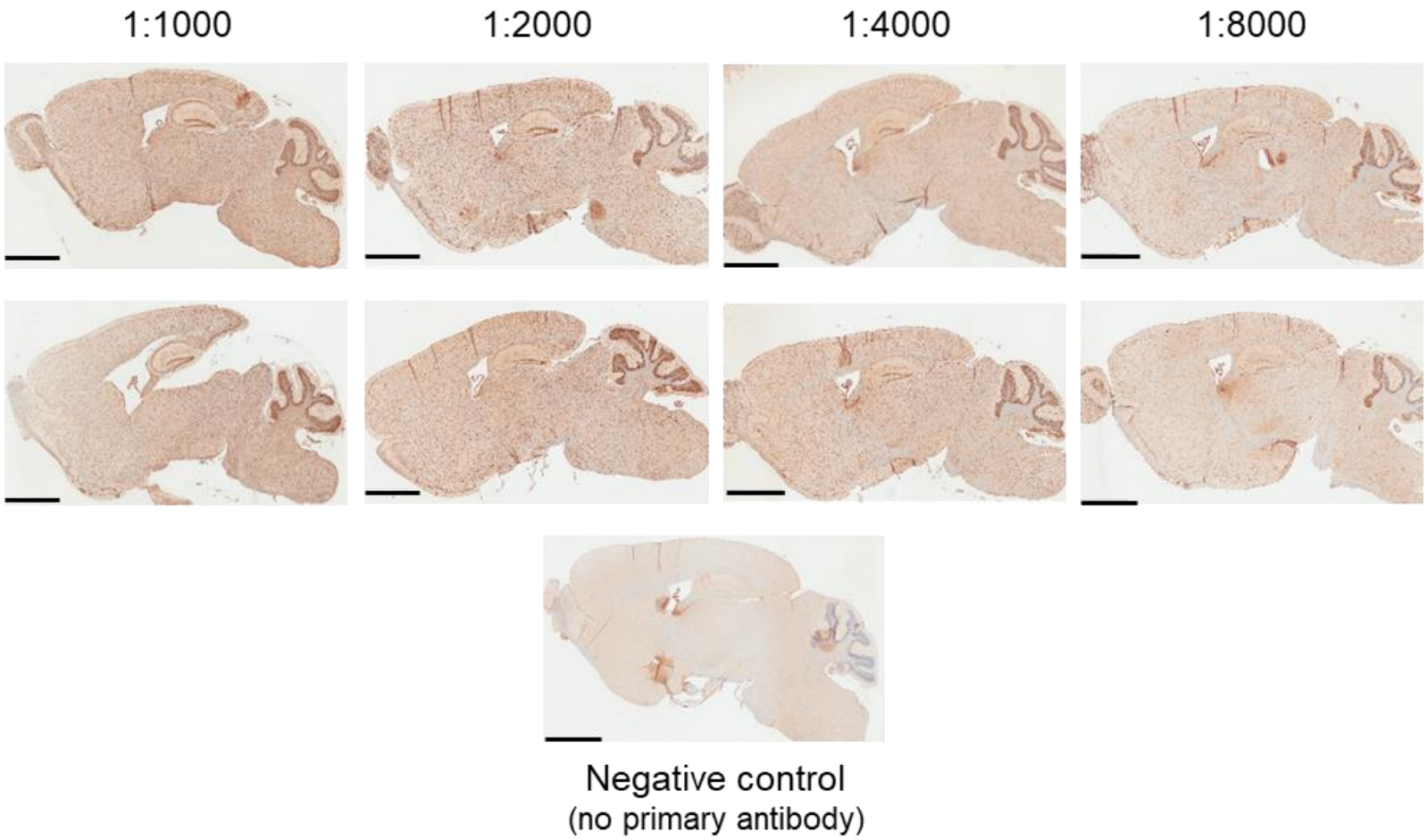


Figure 5.7 – Representative immunohistochemistry images for the different concentrations of anti-5mC primary antibody that were tested.
Sagittal mouse brain sections showing a general view of the stained tissue with the different concentrations (1:1000, 1:2000, 1:4000, 1:8000) of anti-5mC primary antibody (serial dilutions) tested. Primary antibody: anti-5mC (serial dilutions); secondary antibody: biotinylated goat anti-mouse (1:200). Black bar = 2 mm.

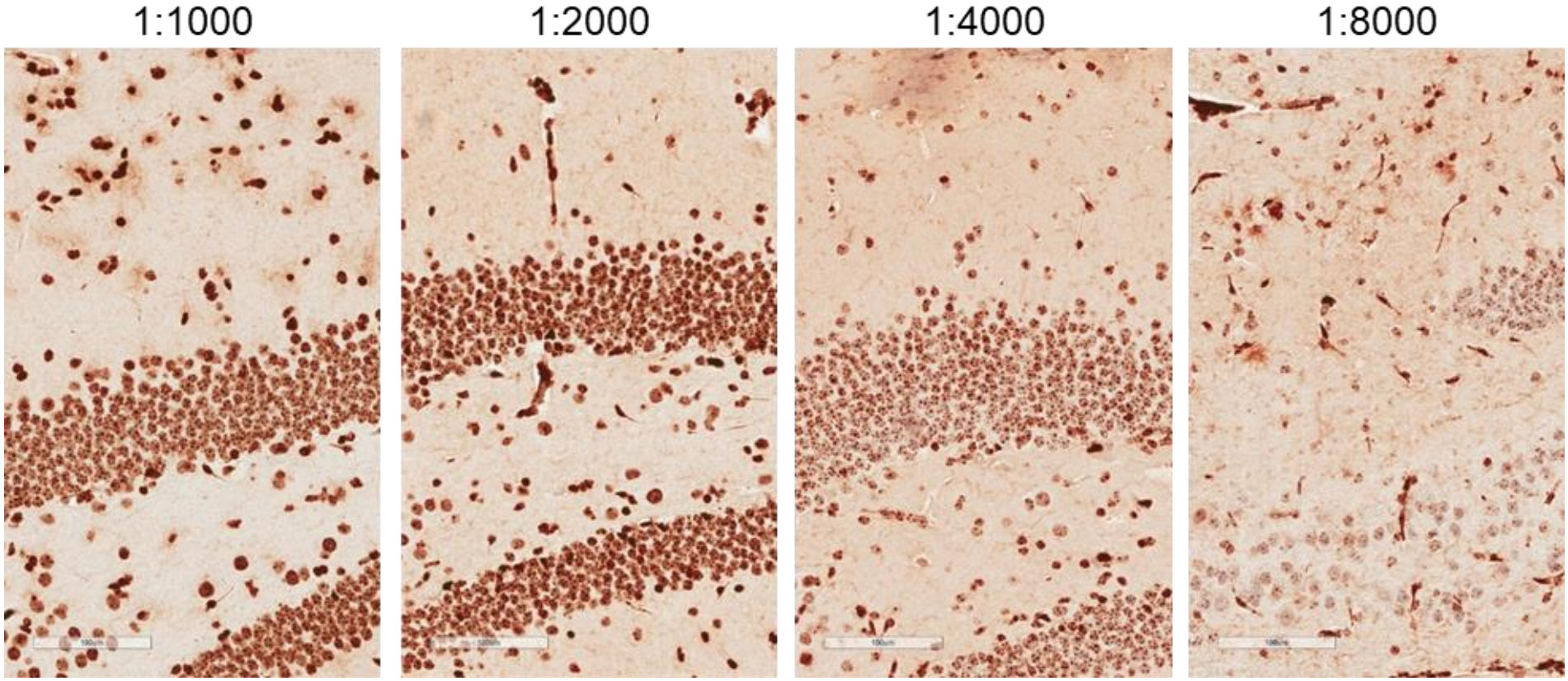


Figure 5.8 – Representative hippocampi immunohistochemistry images for the different concentrations of anti-5mC primary antibody that were tested.

The DG sub-region of the hippocampus is showed in detail. Primary antibody: anti-5mC (serial dilutions); secondary antibody: biotinylated goat anti-mouse (1:200). Bar = 100 µm.

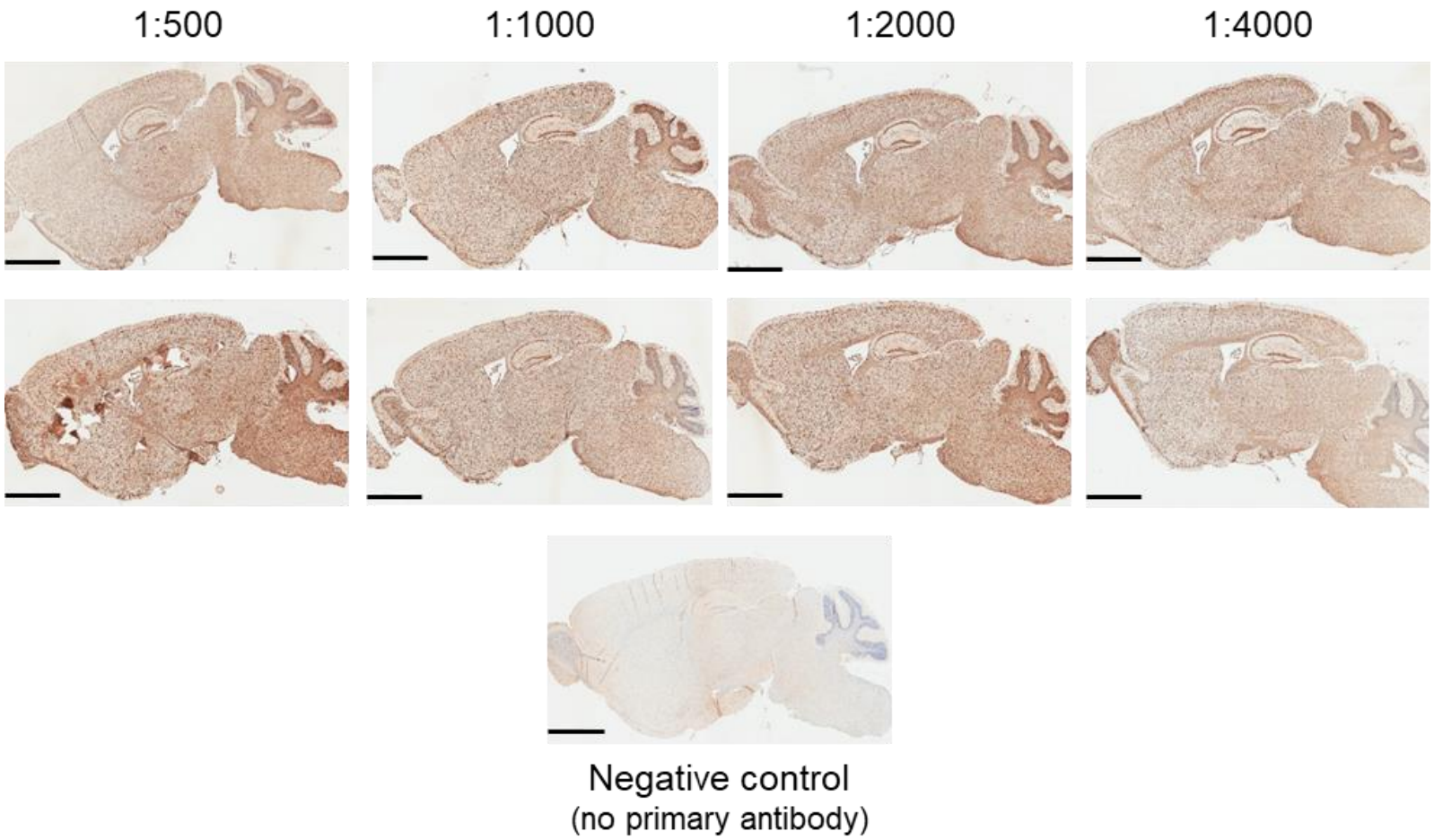


Figure 5.9 – Representative immunohistochemistry images for the different concentrations of anti-5mhC primary antibody that were tested. Sagittal mouse brain sections showing a general view of the stained tissue with the different concentrations (1:500, 1:1000, 1:2000, 1:4000) of anti-5hmC primary antibody (serial dilutions) tested. Primary antibody: anti-5hmC (serial dilutions); secondary antibody: biotinylated goat anti-rabbit (1:200). Black bar = 2 mm.

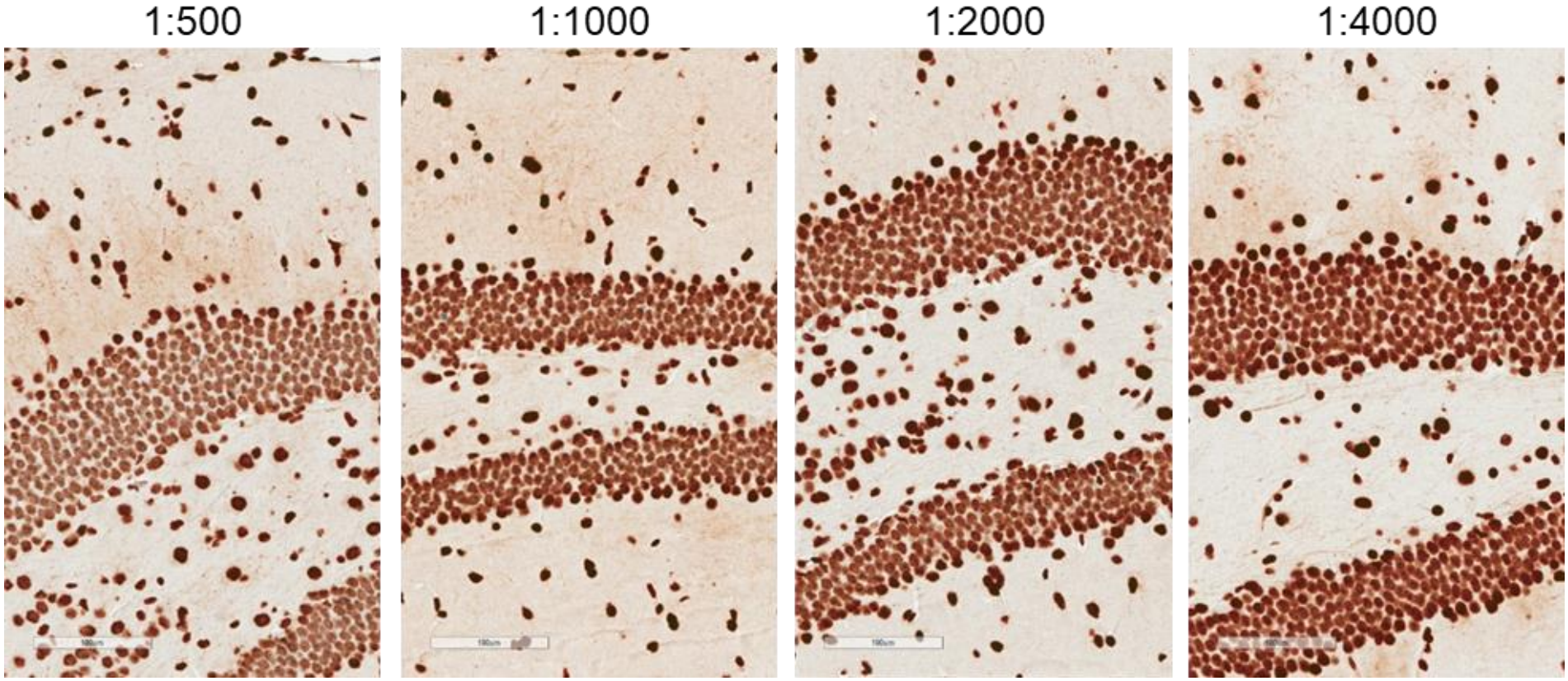


Figure 5.10 – Representative hippocampi immunohistochemistry images for the different concentrations of anti-5hmC primary antibody that were tested.

The DG sub-region of the hippocampus is showed in detail. Primary antibody: anti-5hmC (serial dilutions); secondary antibody: biotinylated goat anti-rabbit (1:200). Bar = 100 µm.

5.2.6.3. Detection

For detection we undertook enzymatic labelling using peroxidase (Vectastain Elite ABC HRP Reagent, Vector Laboratories) and DAB substrate (Vector Laboratories).

5.2.6.4. Counterstain

Tissues were counterstained with haematoxylin to provide contrast and dehydrated with ethanol and xylene to prepare for mounting.

5.2.7. Mounting and image capture

Immunohistochemistry samples were mounted in an automated coverslipper (ClearVue Coverslipper from Thermo Scientific) using ClearVue mountant. Images were digitised with Scanscope AT slide scanner (Aperio) at 20x magnification, and saved in *svs* (ScanScope Virtual Slide) format on an Eli Lilly network server. Images were subsequently accessed using the eSlide Manager system (Aperio).

5.2.8. Visualisation and quantification

Visualisation of the digitised tissue sections stored in the network server and delineation of a representative area of the anatomical brain regions of interest (ROI) was achieved using the software *Imagescope* (version 12.2.1.5005; Aperio). ROI were selected to reflect the known progression of amyloid and tau pathology in the two models (see **Chapter 3**) and to overlap with previous and ongoing studies at Eli Lilly in the same mouse models.

For both mouse models, the ROI defined and quantified were the hippocampi sub-regions CA1, CA3 and dentate gyrus (DG); rostral cortex and caudal cortex; and thalamus (**Figure 5.11**), which exhibits very low levels of tau pathology in rTg4510 mice and no amyloid pathology in J20 mice (see **Chapter 3**). For each sample, ROI showing experimentally-related damaged tissue or staining artefacts were excluded from the analysis.

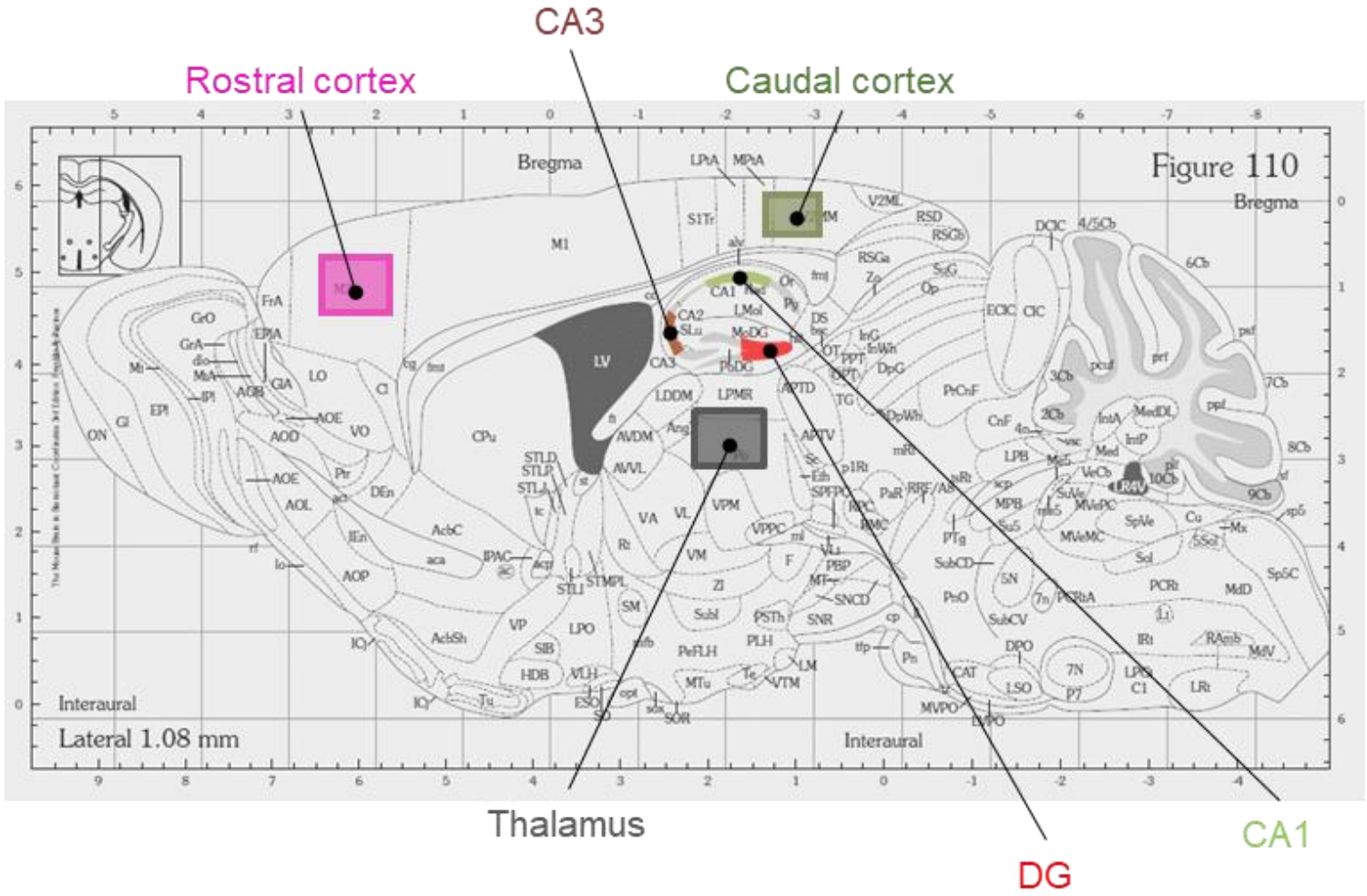


Figure 5.11 – Anatomical regions of the brain analysed.
A representative area was delineated, as shown in the figure. Figure adapted from Paxinos and Franklin (2001).

Positivity was quantified automatically using the Aperio Positive Pixel Count Algorithm (Aperio), which was applied to the images using *ImageScope* (Aperio) as described in detail in **Chapter 3** (**Section 3.2.5**). The values for the algorithm input parameters (**Table 5.2** and **Table 5.3**, see **Section 3.2.5** for details), calibrated to ignore non-specific staining, were based on optimisations from other experiments using similar (nuclear) stains from previous and ongoing studies conducted by my collaborators at Eli Lilly. Representative images depicting how the algorithm distinguished 5mC and 5hmC positivity in my samples are shown in **Figure 5.12**. The quantification of 5mC or 5hmC positivity was expressed as number of positive cells divided by the total number of cells (as a percentage).

Table 5.2 – Algorithm inputs for anti-5mC

Version	9.1
View Width (Pixels)	1000
View Height (Pixels)	1000
Overlap Size	0
Image Zoom	1
Classifier	None
Classifier Neighbourhood	0
Pixel Area (mm^2)	2.46E-07
Hue Value (Centre)	0.1
Hue Width	0.5
Colour Saturation Threshold	0.1
Intensity Threshold WEAK (Upper Limit)	255
Intensity Threshold WEAK (Lower Limit)	145
Intensity Threshold MEDIUM (Upper Limit)	145
Intensity Threshold MEDIUM (Lower Limit)	100
Intensity Threshold STRONG (Upper Limit)	100
Intensity Threshold STRONG (Lower Limit)	0
Intensity Threshold Negative Pixels	-1

Table 5.3 – Algorithm inputs for anti-5hmC

Version	9.1
View Width (Pixels)	1000
View Height (Pixels)	1000
Overlap Size	0
Image Zoom	1
Classifier	None
Classifier Neighbourhood	0
Pixel Area (mm^2)	2.46E-07
Hue Value (Centre)	0.1
Hue Width	0.5
Colour Saturation Threshold	0.1
Intensity Threshold WEAK (Upper Limit)	255
Intensity Threshold WEAK (Lower Limit)	145
Intensity Threshold MEDIUM (Upper Limit)	145
Intensity Threshold MEDIUM (Lower Limit)	100
Intensity Threshold STRONG (Upper Limit)	100
Intensity Threshold STRONG (Lower Limit)	0
Intensity Threshold Negative Pixels	-1

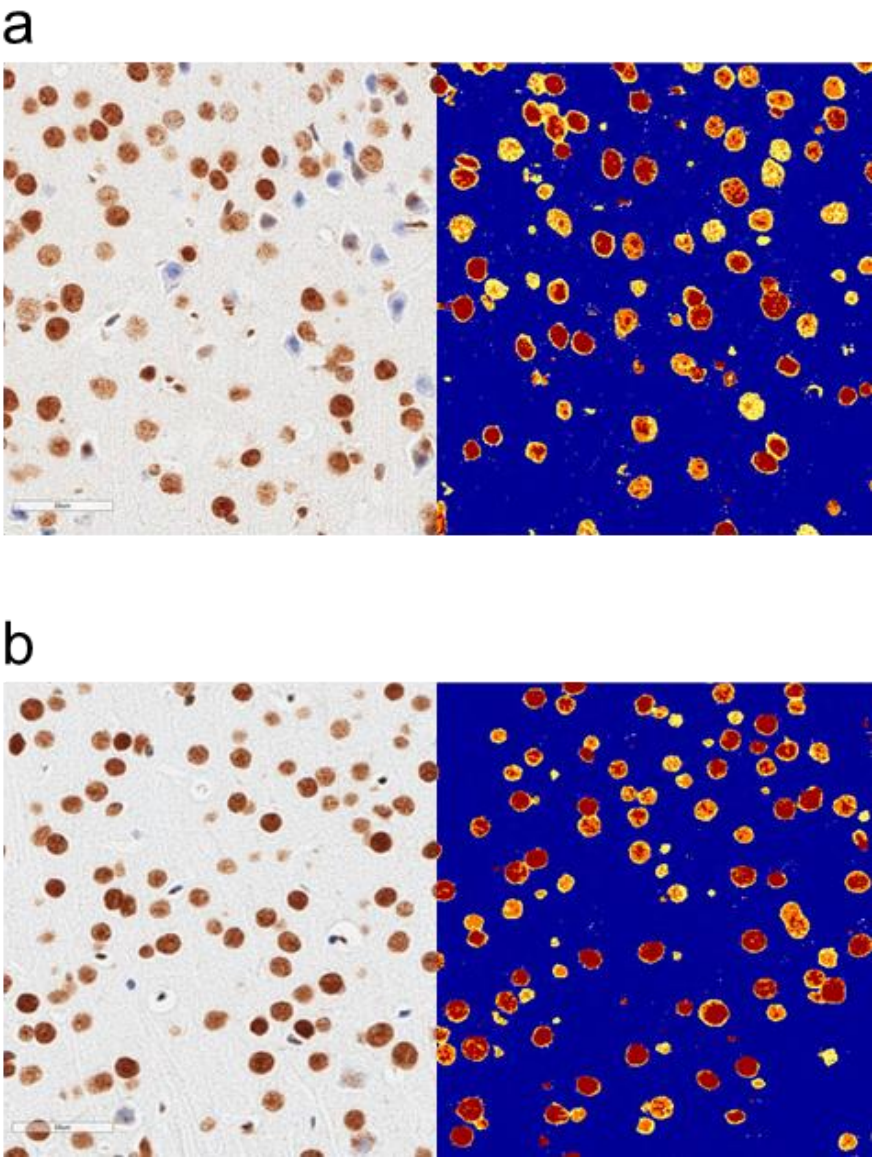


Figure 5.12 – Representative immunohistochemistry images showing how the positive pixel count algorithm recognised the positive stain for quantification of 5mC and 5hmC pathology.
(a) Anti-5mC stain. **(b)** Anti-5hmC stain. Both images are from caudal cortex of a WT mouse. Bar = 50 µm.

5.2.9. Statistical analysis

R (version 3.4.3) was used for all statistical analysis, as well as to plot the data. Two-way factorial ANOVA was used, implementing the following statistical model:

$$DNA\ methylation = Genotype + Age + Genotype * Age$$

5.3. Results

5.3.1. LUMA

No differences in global DNA methylation assessed by LUMA were identified in the hippocampus of rTg4510 (**Figure 5.13** and **Table 5.4**) or J20 (**Figure 5.15** and **Table 5.4**) when comparing TG to WT littermate mice. Furthermore, there was no significant correlation between global DNA methylation and levels of pathology in the hippocampus in the same mice for both rTg4510 (**Figure 5.14**) and J20 (**Figure 5.16**) mouse models.

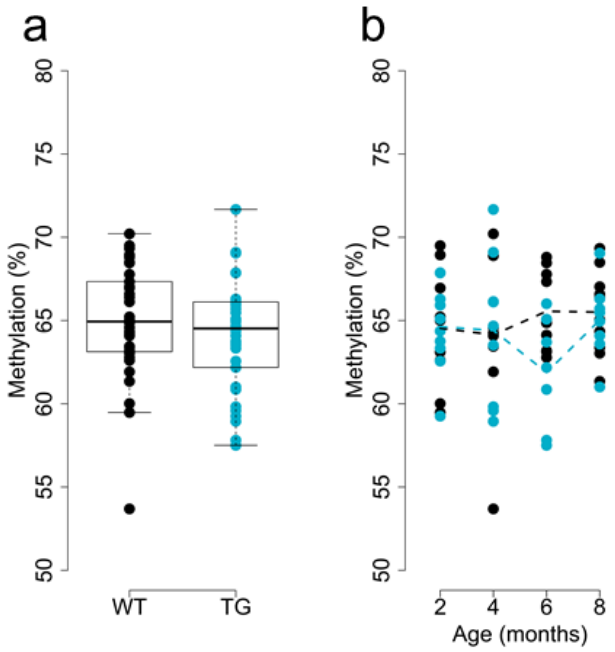


Figure 5.13 – Global DNA methylation in hippocampus from rTg4510 female mice quantified by LUMA.
No significant differences were identified between rTg4510 transgenic (TG, blue) mice compared to wild-type (WT, black) littermate controls, associated with **(a)** genotype or **(b)** interaction between genotype and age (see **Table 5.4**). Total n = 72 animals, 7-10 animals per group. Dashed lines represent mean paths of pathological burden across the four age groups.

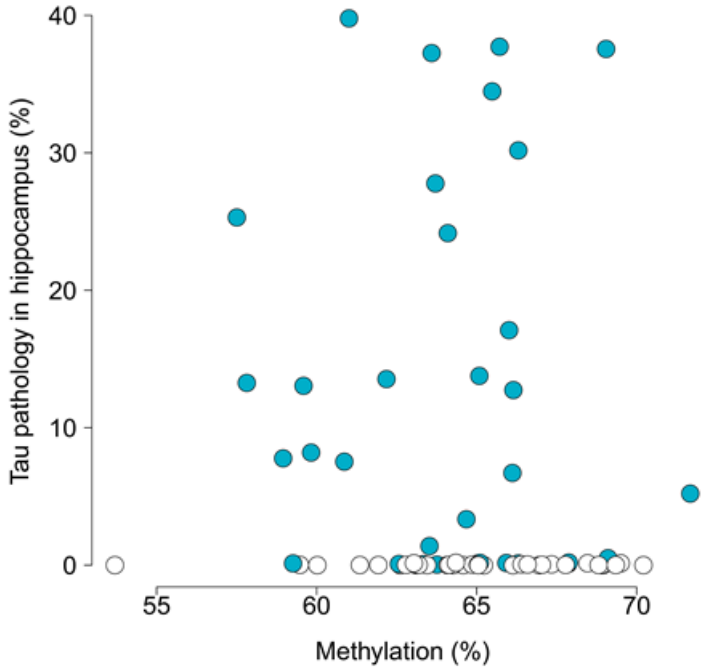


Figure 5.14 – No correlation was identified between global DNA methylation measured by LUMA and tau pathology in the hippocampus in rTg4510 mice.
Pearson correlation r = 0.10, *P* = 0.41. Blue circles represent rTg4510 transgenic (TG) mice, and white circles represent control (WT) mice.

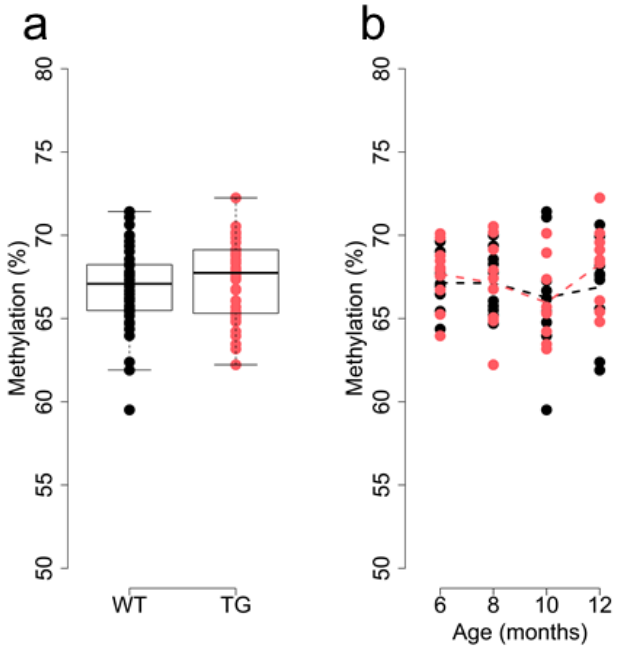


Figure 5.15 – Global DNA methylation in hippocampus from J20 female mice quantified by LUMA.
No differences were identified between J20 transgenic (TG, red) mice compared to wild-type (WT, black) littermate controls, associated with **(a)** genotype or **(b)** interaction between genotype and age (see **Table 5.4**). Total n = 77 animals, 9-10 animals per group. Dashed lines represent mean paths of pathological burden across the four age groups.

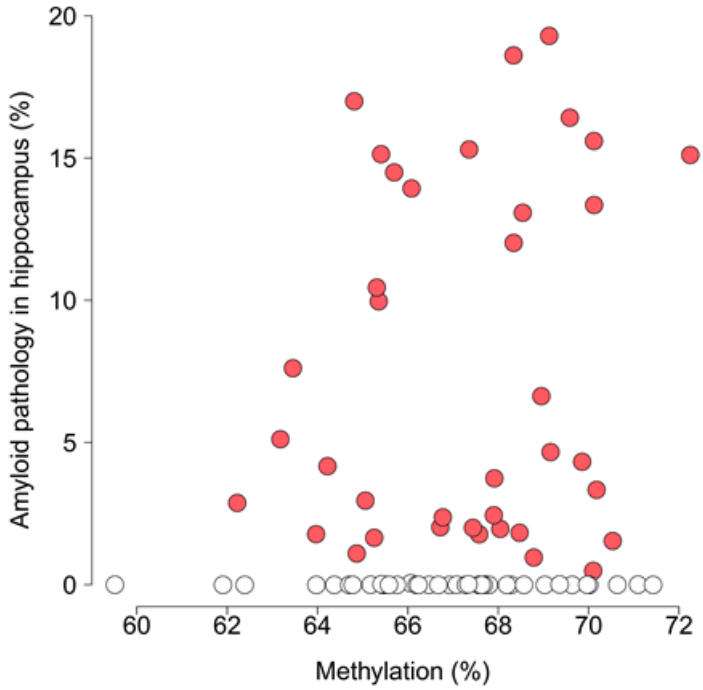


Figure 5.16 – No correlation was identified between global DNA methylation measured by LUMA and amyloid pathology in the hippocampus in J20 mice.
Pearson correlation r = 0.14, *P* = 0.22. Red circles represent J20 transgenic (TG) mice, and white circles represent control (WT) mice.

Table 5.4 – Results for LUMA experiments in hippocampal tissue from rTg4510 and J20 mice.
Statistical results (two-way ANOVA) are shown.

Mouse model	Genotype		Age		Genotype*Age		N
	F-value	P-value	F-value	P-value	F-value	P-value	
rTg4510	F(1,64) = 1.22	0.27	F(3,64) = 0.48	0.70	F(3,64) = 1.30	0.28	72
J20	F(1,69) = 0.63	0.43	F(3,69) = 1.33	0.27	F(3,69) = 0.43	0.73	77

5.3.2. Immunohistochemistry

5.3.2.1. rTg4510

My analysis of 5mC and 5hmc levels in tissue sections using immunohistochemistry revealed both genotype-associated and progressive changes (interaction between genotype and age) in rTg4510 TG mice compared to WT littermate controls in several brain regions (5mc: **Figure 5.17** and **Table 5.5**; 5hmC: **Figure 5.21** and **Table 5.6**). Genotype-associated differences were identified for both 5mC and 5hmC in the hippocampal sub-regions CA1 (**Figure 5.17a** and **Figure 5.21a**) and DG (**Figure 5.17c** and **Figure 5.21c**), with rTg4510 TG mice showing significantly ($P < 0.05$) lower levels for both DNA modifications in both brain regions compare to WT. Levels of 5hmC in the caudal cortex were also lower in rTg4510 TG compared to WT mice (**Figure 5.21e**). Changes in 5mC associated with the progression of tau were identified in CA1 (**Figure 5.17a**, with 5mC decreasing with age in TG mice only), CA3 (**Figure 5.17b**, with 5mC decreasing more intensely with age in TG mice compared to WT controls), and caudal cortex (**Figure 5.17e**). Changes in 5hmC associated with progression of tau were identified only in CA1 (**Figure 5.21a**, with 5hmC decreasing with age in TG mice only), with a similar profile to 5mC in the same hippocampal sub-region (**Figure 5.17a**). Representative immunohistochemistry images for the anti-5mC stain are shown in **Figure 5.18** and **Figure 5.19**, and for the anti-5hmC stain in **Figure 5.22** and **Figure 5.23**.

Levels of both 5mC and 5hmc were also negatively correlated with levels of tau pathology in all hippocampal sub-regions (CA1, CA3, DG) (**Figure 5.20a-c** and **Figure 5.24a-c**). No correlation between DNA modifications and pathology was observed in any of the other brain regions assessed (**Figure 5.20d-f** and **Figure 5.24d-f**).

Age-associated changes were also identified in the CA1, CA3, DG, and caudal cortex for 5mC (**Table 5.5**), and CA1, CA3, and rostral cortex for 5hmC (**Table 5.6**).

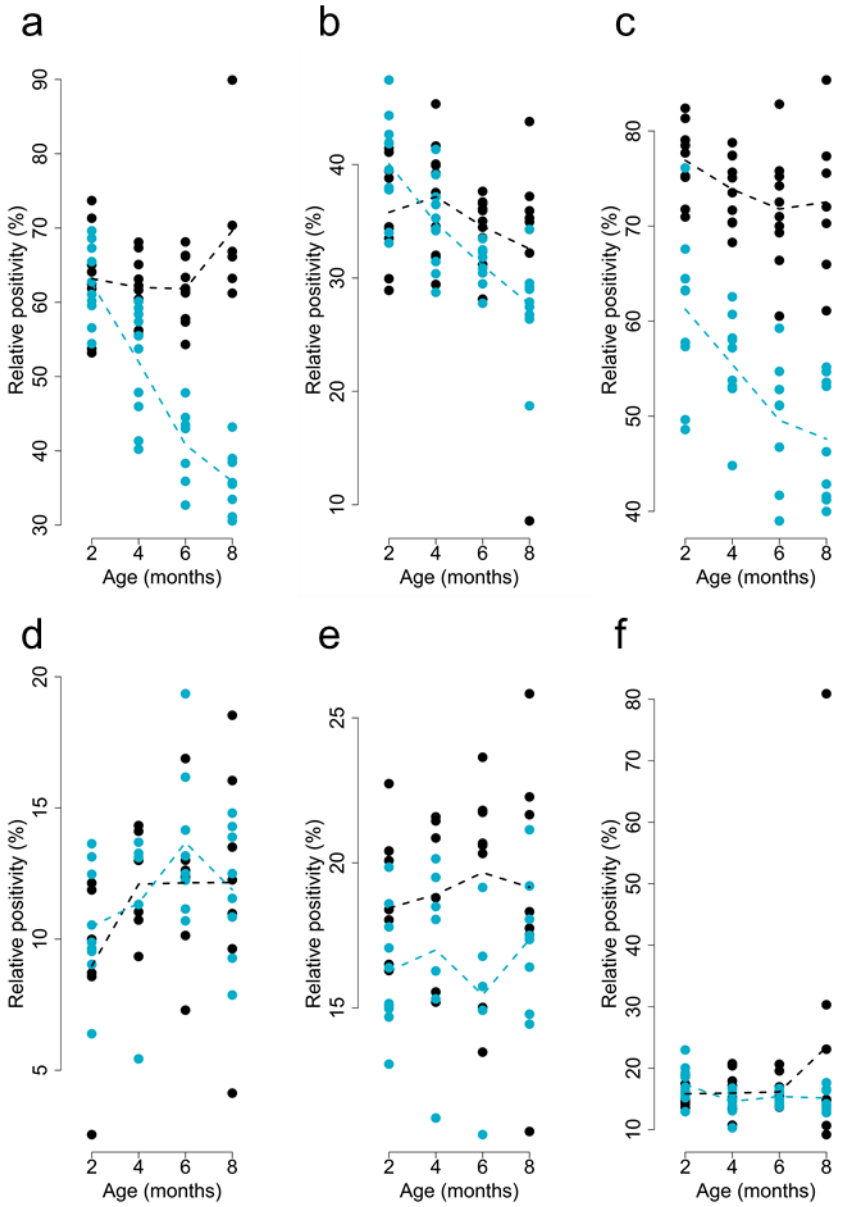


Figure 5.17 – Levels of 5mC in rTg4510 mice.
Quantification of anti-5mC immunoreactivity for each of the tested brain regions: **(a)** CA1, **(b)** CA3, **(c)** DG, **(d)** rostral cortex, **(e)** caudal cortex, **(f)** thalamus. Dashed lines represent mean paths of positivity across the four age groups. Blue circles represent rTg4510 transgenic (TG) mice, and black circles represent control (WT) mice.

Table 5.5 – Results for anti-5mC quantification in brains from rTg4510 mice.
Statistical results (two-way ANOVA) are shown. Significant (α < 0.05) P values are underlined.

Brain region	N	Genotype		Age		Genotype*Age	
		F-value	P-value	F-value	P-value	F-value	P-value
CA1	69	F(1,61) = 96.26	3.68E-14	F(3,61) = 14.68	2.65E-07	F(3,61) = 21.25	1.52E-09
CA3	73	F(1,65) = 1.55	0.22	F(3,65) = 8.51	7.56E-05	F(3,65) = 2.84	0.045
DG	73	F(1,65) = 194.17	3.41E-21	F(3,65) = 8.82	5.48E-05	F(3,65) = 1.95	0.13
Rostral cortex	57	F(1,49) = 0.27	0.60	F(3,49) = 2.80	0.050	F(3,49) = 0.51	0.68
Caudal cortex	73	F(1,65) = 1.55	0.22	F(3,65) = 8.51	7.56E-05	F(3,65) = 2.84	0.045
Thalamus	75	F(1,67) = 1.26	0.27	F(3,67) = 0.91	0.44	F(3,67) = 1.25	0.30

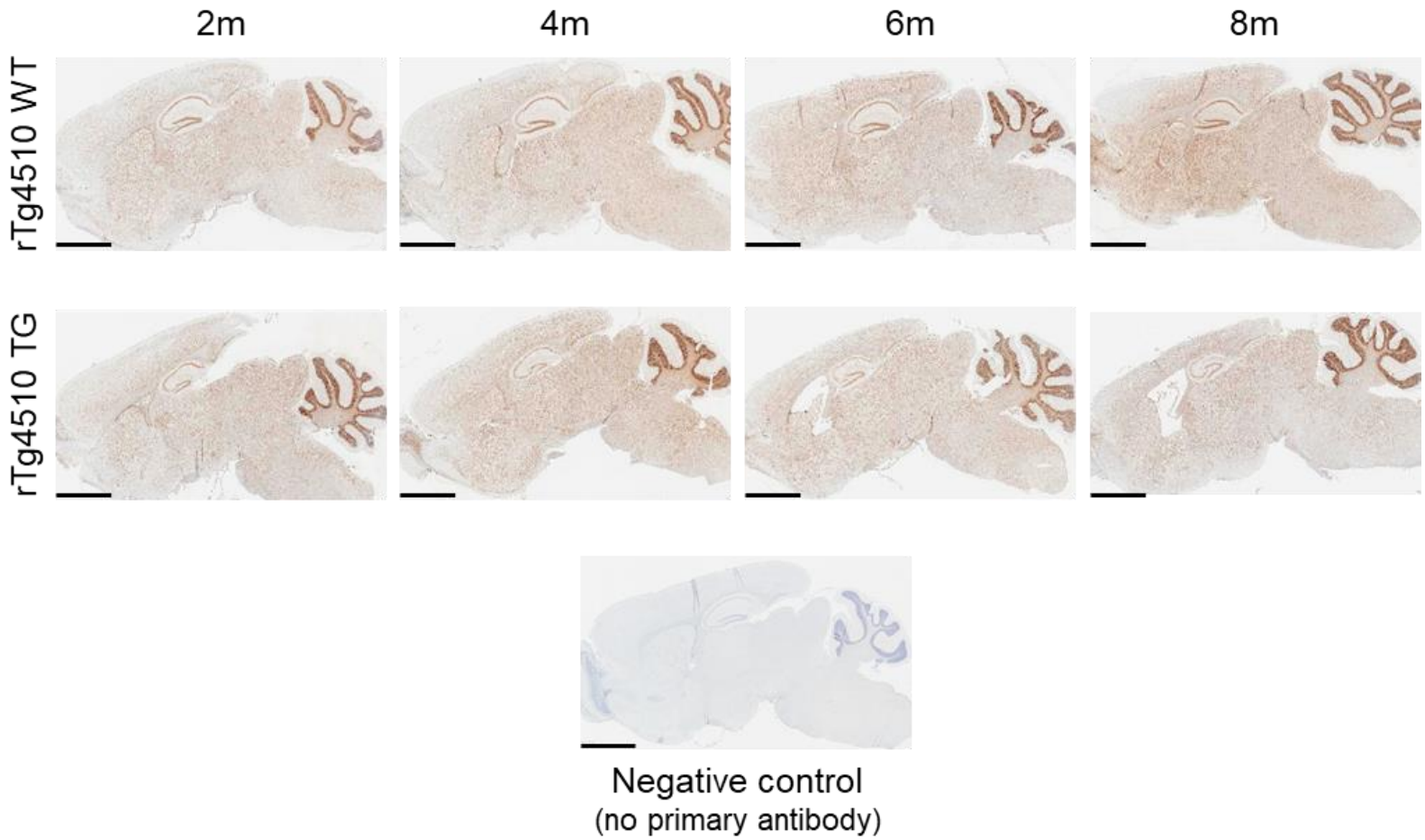


Figure 5.18 – Representative immunohistochemistry images for 5mC staining in rTg4510 mice.
Sagittal sections showing the distribution of anti-5mC in the brains of rTg4510 transgenic (TG) mice (second row) compared to wild-type control (WT) mice (top) at 2, 4, 6 and 8 months of age. Primary antibody: anti-5mC (1:4000); secondary antibody: biotinylated goat anti-mouse (1:200). Black bar = 2 mm.

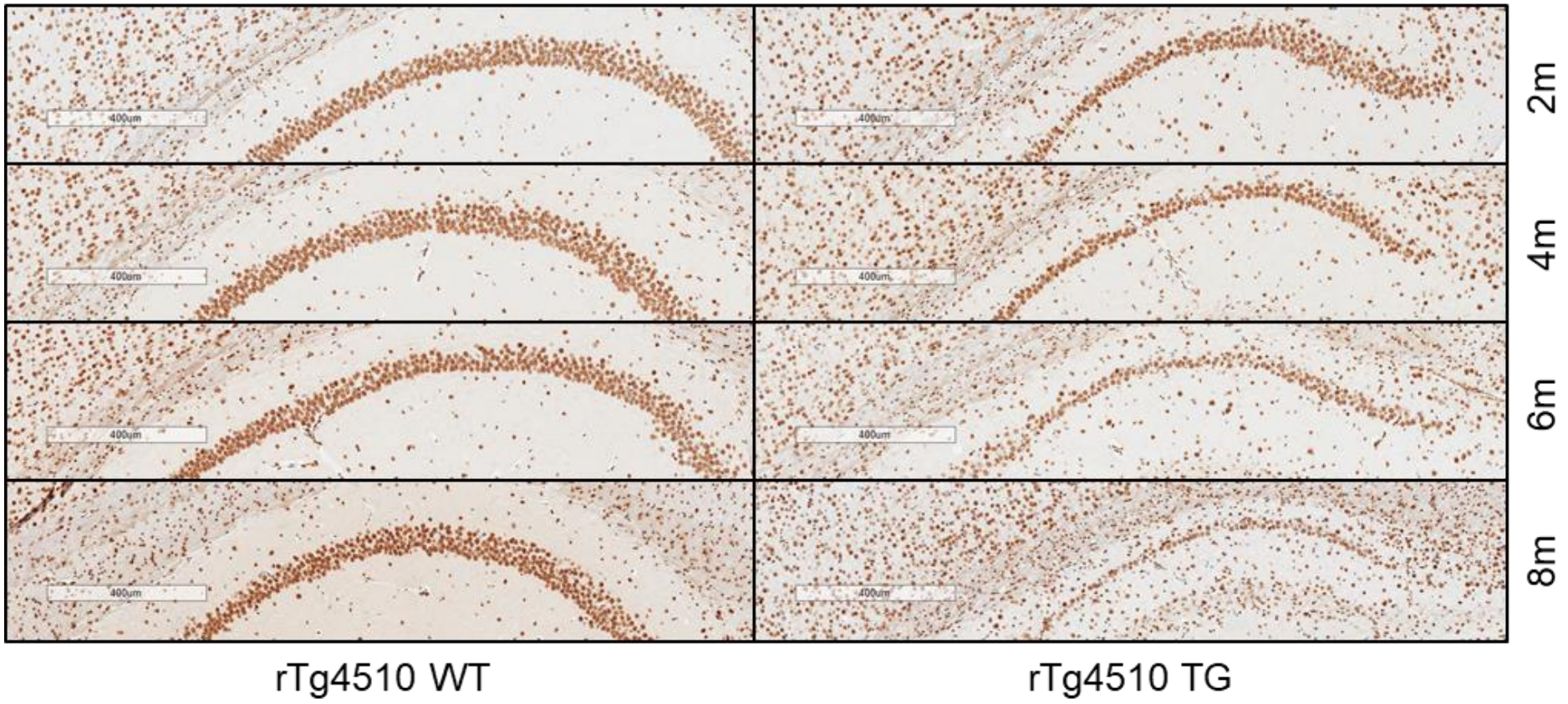


Figure 5.19 – Representative immunohistochemistry images for 5mC staining in rTg4510 mice showing the hippocampal sub-region CA1 in detail.
WT: wild type. TG: transgenic. Primary antibody: anti-5mC (1:4000); secondary antibody: biotinylated goat anti-mouse (1:200). Bar = 400 µm.

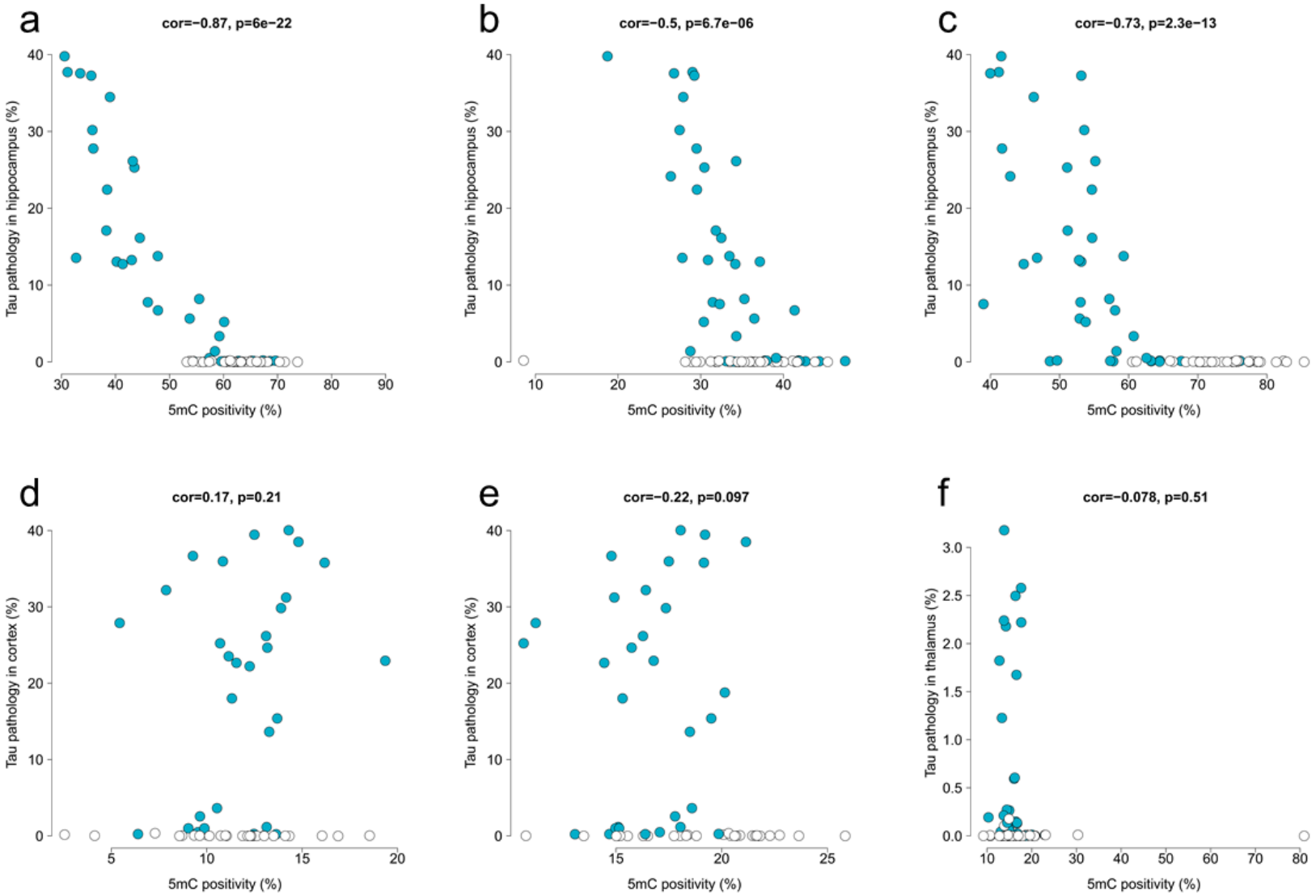


Figure 5.20 – Correlations between levels of 5mC and tau pathology in rTg4510 mice.
(a) CA1, **(b)** CA3, **(c)** DG, **(d)** rostral cortex, **(e)** caudal cortex, **(f)** thalamus. Shown is the Pearson correlation statistic for each comparison. Blue circles represent rTg4510 transgenic (TG) mice, and white circles represent control (WT) mice.

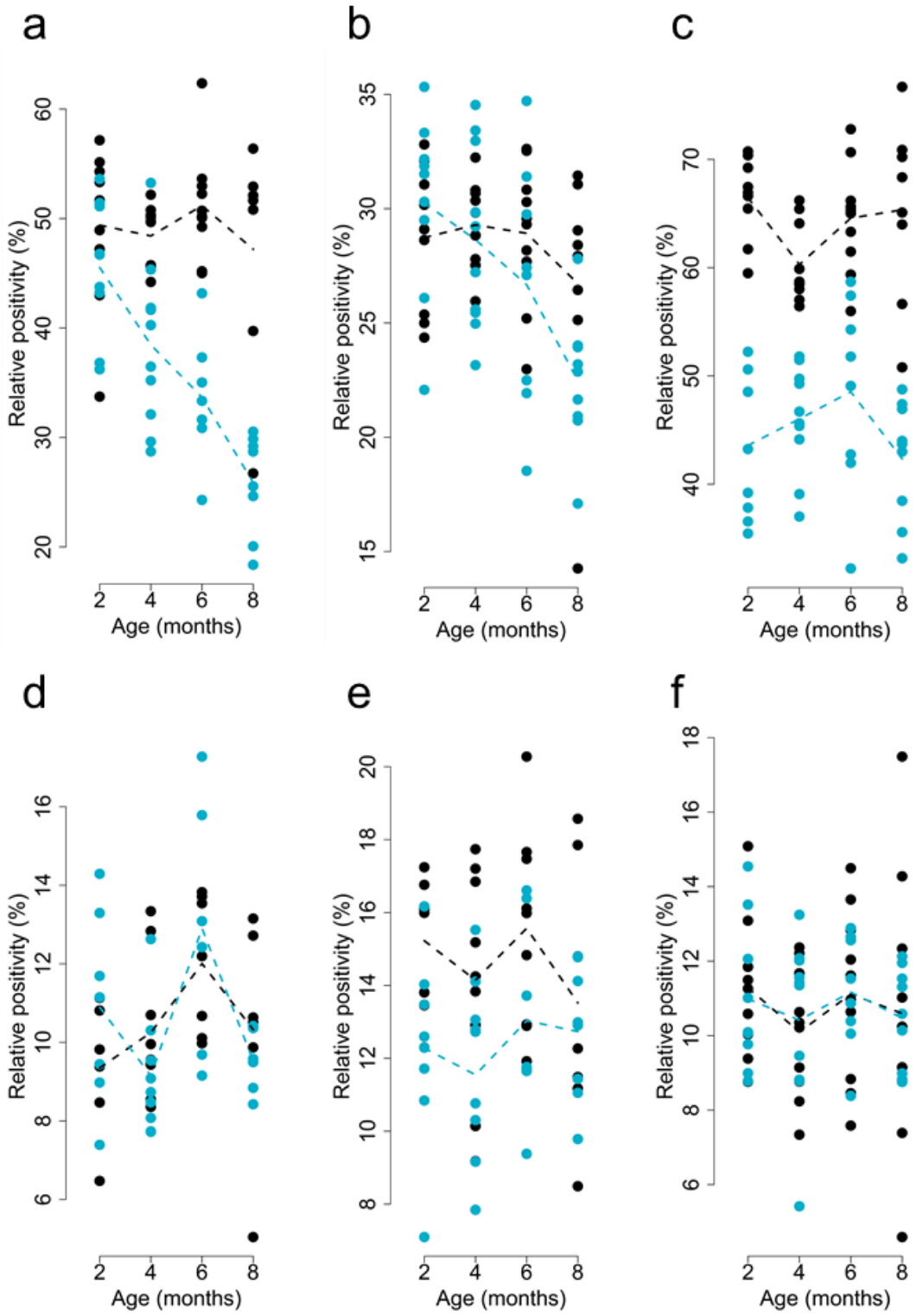


Figure 5.21 – Levels of 5hmC in rTg4510 mice.
Quantification of anti-5hmC immunoreactivity for each of the tested brain regions: **(a)** CA1, **(b)** CA3, **(c)** DG, **(d)** rostral cortex, **(e)** caudal cortex, **(f)** thalamus. Dashed lines represent mean paths of positivity across the four age groups. Blue circles represent rTg4510 transgenic (TG) mice, and black circles represent control (WT) mice.

Table 5.6 – Results for anti-5hmC quantification in brains from rTg4510 mice.
Statistical results (two-way ANOVA) are shown. Significant (α < 0.05) P values are underlined.

Brain region	N	Genotype		Age		Genotype*Age	
		F-value	P-value	F-value	P-value	F-value	P-value
CA1	70	F(1,62) = 69.94	9.30E-12	F(3,62) = 8.48	8.38E-05	F(3,62) = 6.19	0.00095
CA3	73	F(1,65) = 2.59	0.11	F(3,65) = 6.00	0.0011	F(3,65) = 1.79	0.16
DG	73	F(1,65) = 185.70	1.01E-20	F(3,65) = 1.09	0.36	F(3,65) = 2.69	0.053
Rostral cortex	56	F(1,48) = 0.0056	0.94	F(3,48) = 4.94	0.0045	F(3,48) = 1.35	0.27
Caudal cortex	66	F(1,58) = 12.82	0.00070	F(3,58) = 1.023	0.39	F(3,58) = 0.51	0.68
Thalamus	74	F(1,66) = 0.0022	0.96	F(3,66) = 0.74	0.53	F(3,66) = 0.067	0.98

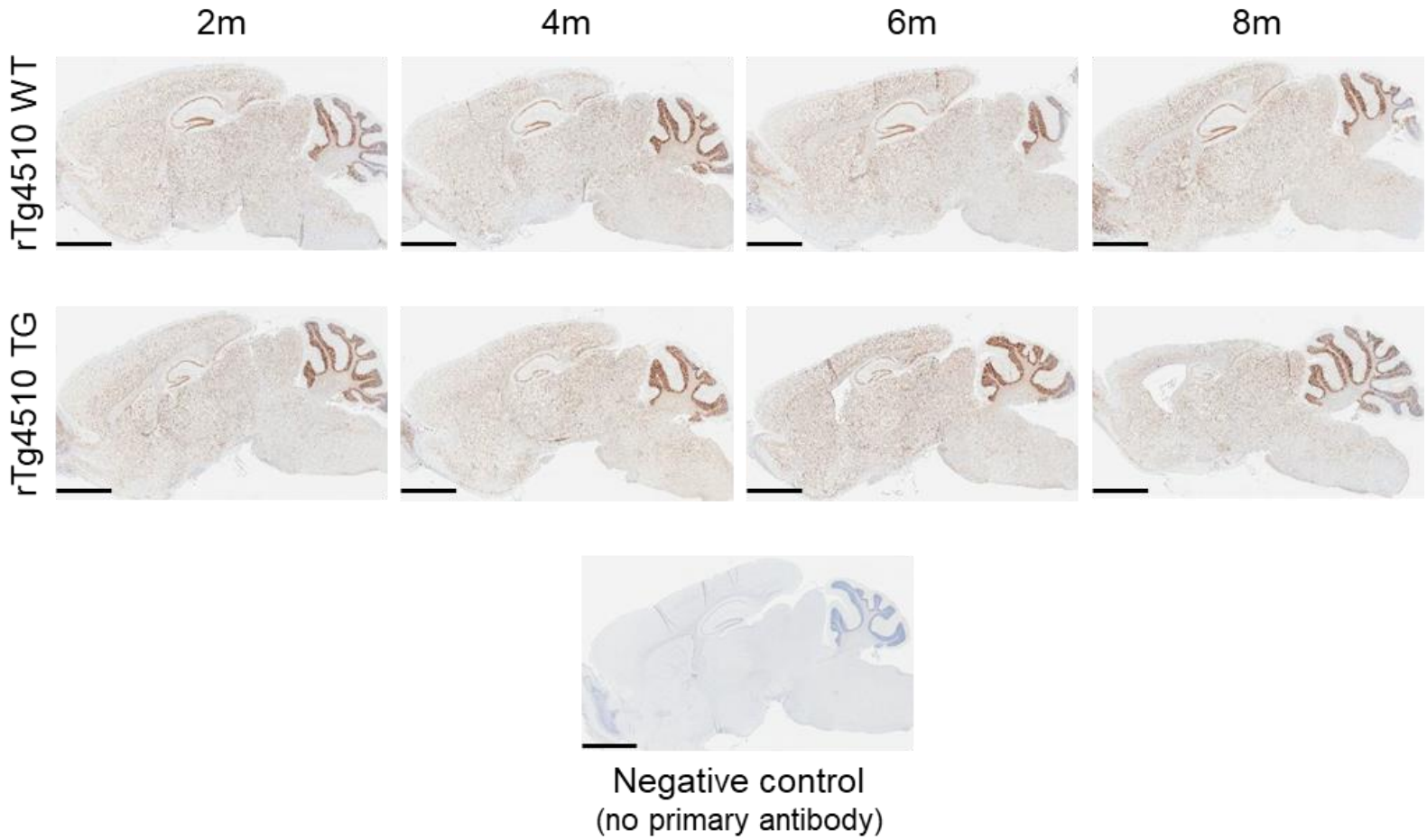


Figure 5.22 – Representative immunohistochemistry images for 5hmC stain in rTg4510 mice.
Sagittal sections showing the distribution of anti-5hmC in the brains of rTg4510 transgenic (TG) mice (second row) compared to wild-type control (WT) mice (top) at 2, 4, 6 and 8 months of age. Primary antibody: anti-5hmC (1:4000); secondary antibody: biotinylated goat anti-rabitt (1:200). Black bar = 2 mm.

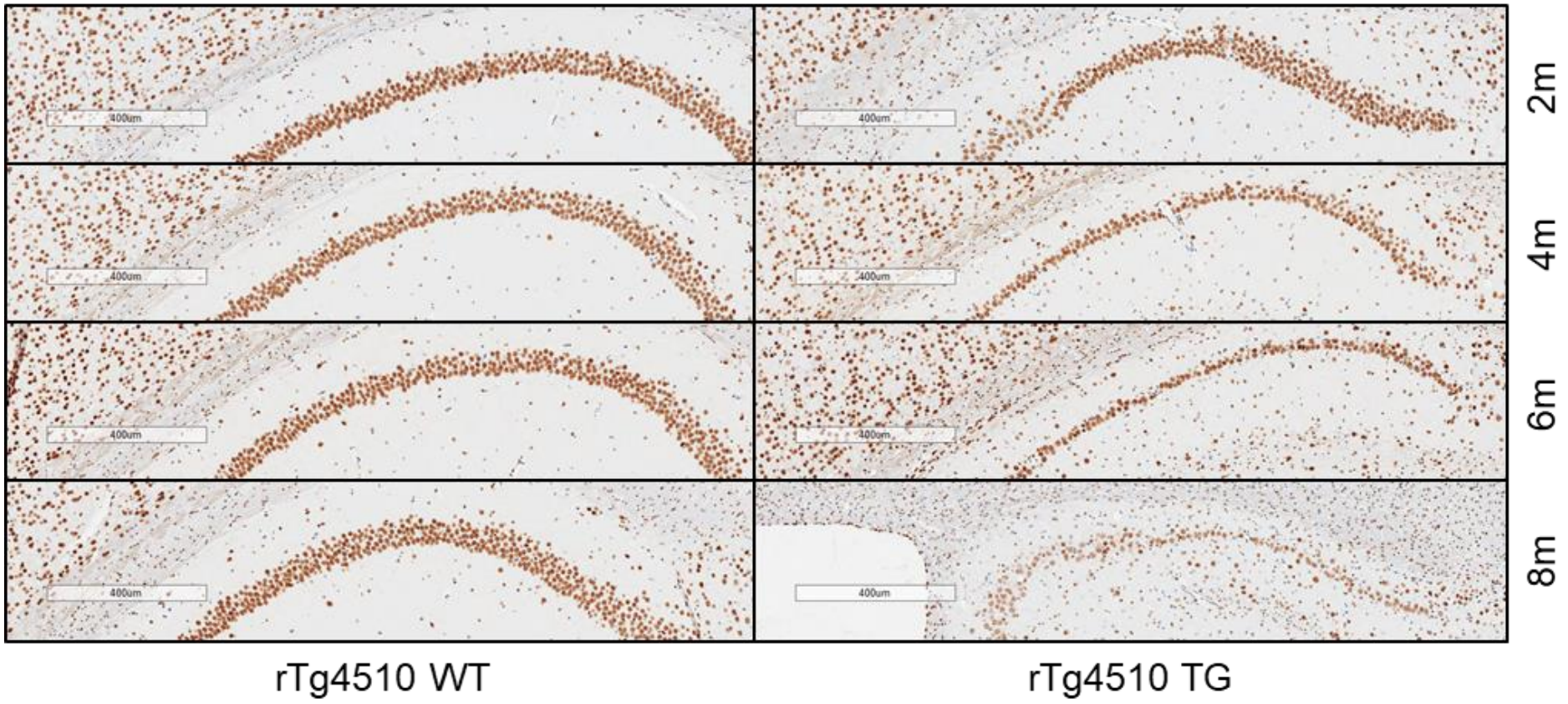


Figure 5.23 – Representative immunohistochemistry images for 5hmC stain in rTg4510 mice showing the hippocampal sub-region CA1 in detail.
WT: wild type. TG: transgenic. Primary antibody: anti-5hmC (1:4000); secondary antibody: biotinylated goat anti-rabitt (1:200). Bar = 400 µm.

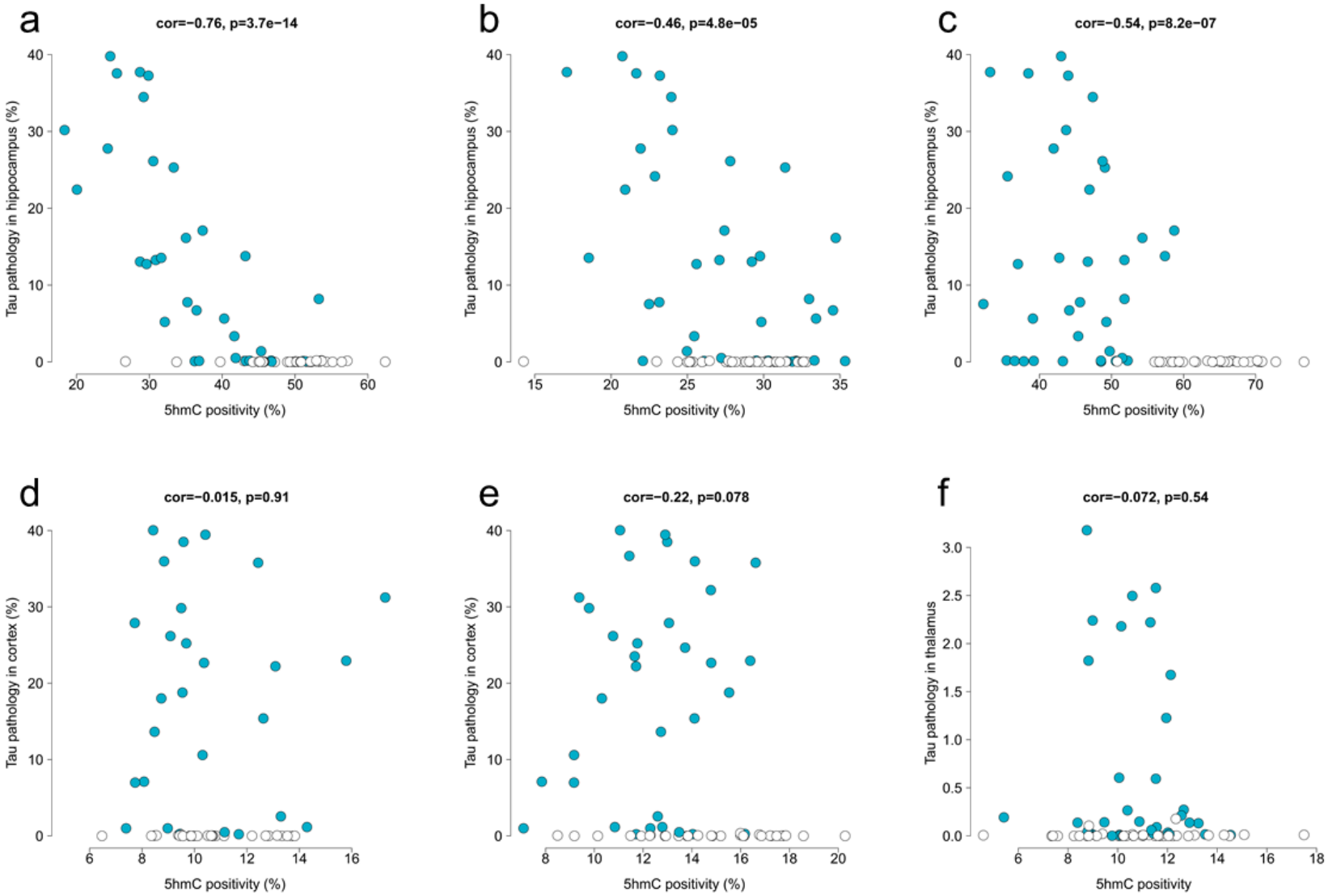


Figure 5.24 – Correlations between levels of 5hmC and tau pathology in rTg4510 mice.
(a) CA1, **(b)** CA3, **(c)** DG, **(d)** rostral cortex, **(e)** caudal cortex, **(f)** thalamus. Shown is the Pearson correlation statistic for each comparison. Blue circles represent rTg4510 transgenic (TG) mice, and white circles represent control (WT) mice.

5.3.2.2. J20

In tissue sections from J20 mice I identified more subtle changes in the levels of the DNA modifications 5mC (**Figure 5.25** and **Table 5.7**) and 5hmC (**Figure 5.29** and **Table 5.8**) compared to those observed in rTg4510 mice. Genotype-associated differences were identified in the hippocampal sub-regions CA1 and CA3 for both 5mC (**Figure 5.25a-b**) and 5hmC (**Figure 5.29 a-b**), with J20 TG mice exhibiting lower levels compared to WT controls for both DNA modifications. No statistically significant ($P < 0.05$) differences for the interaction between genotype and age were detected in any brain region. Representative immunohistochemistry images for the anti-5mC stain are shown in **Figure 5.26** and **Figure 5.27** and for the anti-5hmC stain in **Figure 5.30** and **Figure 5.31**. Interestingly, both levels of 5mC and 5hmc were negatively correlated with levels of amyloid pathology in the hippocampus in the CA1 and CA3 sub-regions (**Figure 5.28a-b** and **Figure 5.32a-b**). Levels of 5mC and 5hmc were also negatively correlated with cortical levels of amyloid pathology for the two cortical (rostral and caudal cortex) sub-regions quantified (**Figure 5.28d-e** and **Figure 5.32d-e**). Age-associated changes were identified in all regions for 5mC (**Table 5.7**), and in all regions except the thalamus for 5hmC (**Table 5.8**).

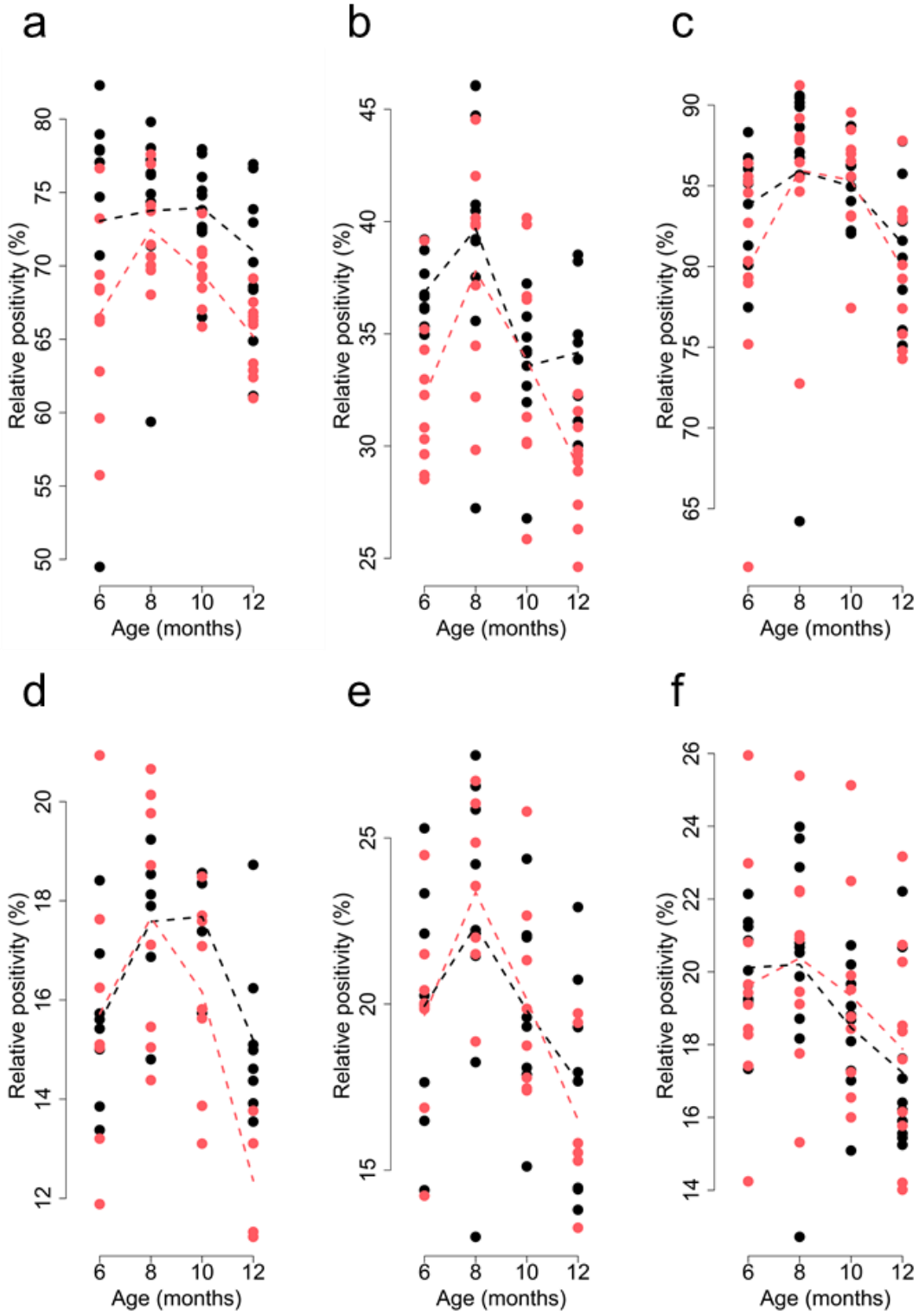


Figure 5.25 – Levels of 5mC in J20 mice.
Quantification of anti-5mC immunoreactivity for each of the tested brain regions: **(a)** CA1, **(b)** CA3, **(c)** DG, **(d)** rostral cortex, **(e)** caudal cortex, **(f)** thalamus. Dashed lines represent mean paths of positivity across the four age groups. Red circles represent J20 transgenic (TG) mice, and black circles represent control (WT) mice.

Table 5.7 – Results for anti-5mC quantification in brains from J20 mice.
Statistical results (two-way ANOVA) are shown. Significant (α < 0.05) P values are underlined.

Brain region	N	Genotype		Age		Genotype*Age	
		F-value	P-value	F-value	P-value	F-value	P-value
CA1	77	F(1,69) = 14.40	0.00031	F(3,69) = 3.17	0.03	F(3,69) = 0.87	0.46
CA3	75	F(1,67) = 11.91	0.00097	F(3,67) = 12.00	2.23E-06	F(3,67) = 2.02	0.12
DG	77	F(1,69) = 1.40	0.24	F(3,69) = 4.63	0.01	F(3,69) = 0.66	0.58
Rostral cortex	54	F(1,46) = 0.44	0.51	F(3,46) = 7.66	0.00030	F(3,46) = 1.58	0.21
Caudal cortex	62	F(1,54) = 0.00059	0.98	F(3,54) = 7.16	0.00039	F(3,54) = 0.28	0.84
Thalamus	77	F(1,69) = 0.25	0.62	F(3,69) = 4.01	0.01	F(3,69) = 0.24	0.87

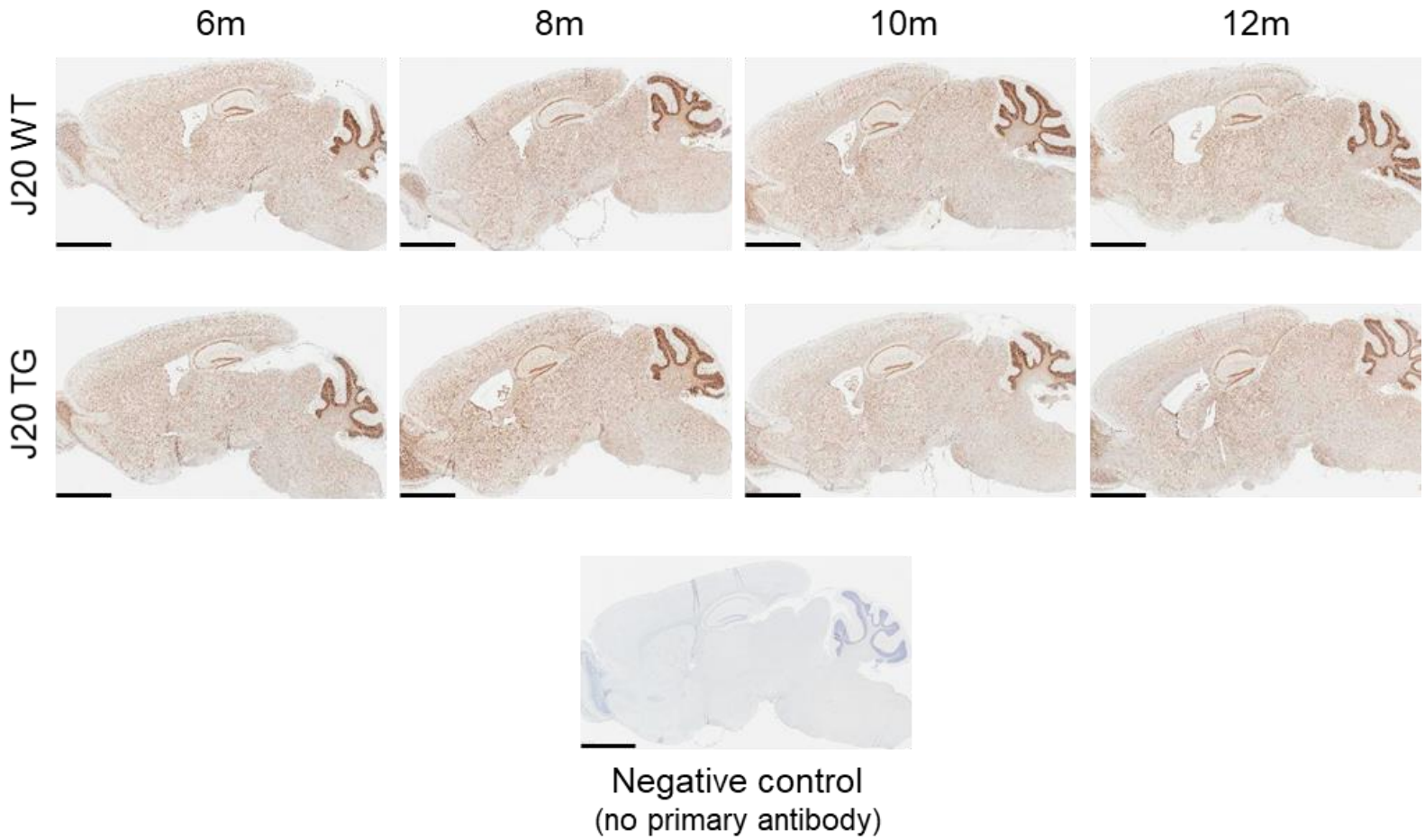


Figure 5.26 – Representative immunohistochemistry images for 5mC stain in J20 mice.
Sagittal sections showing the distribution of anti-5mC in the brains of J20 transgenic (TG) mice (second row) compared to wild-type control (WT) mice (top) at 6, 8, 10 and 12 months of age. Primary antibody: anti-5mC (1:4000); secondary antibody: biotinylated goat anti-mouse (1:200). Black bar = 2 mm.

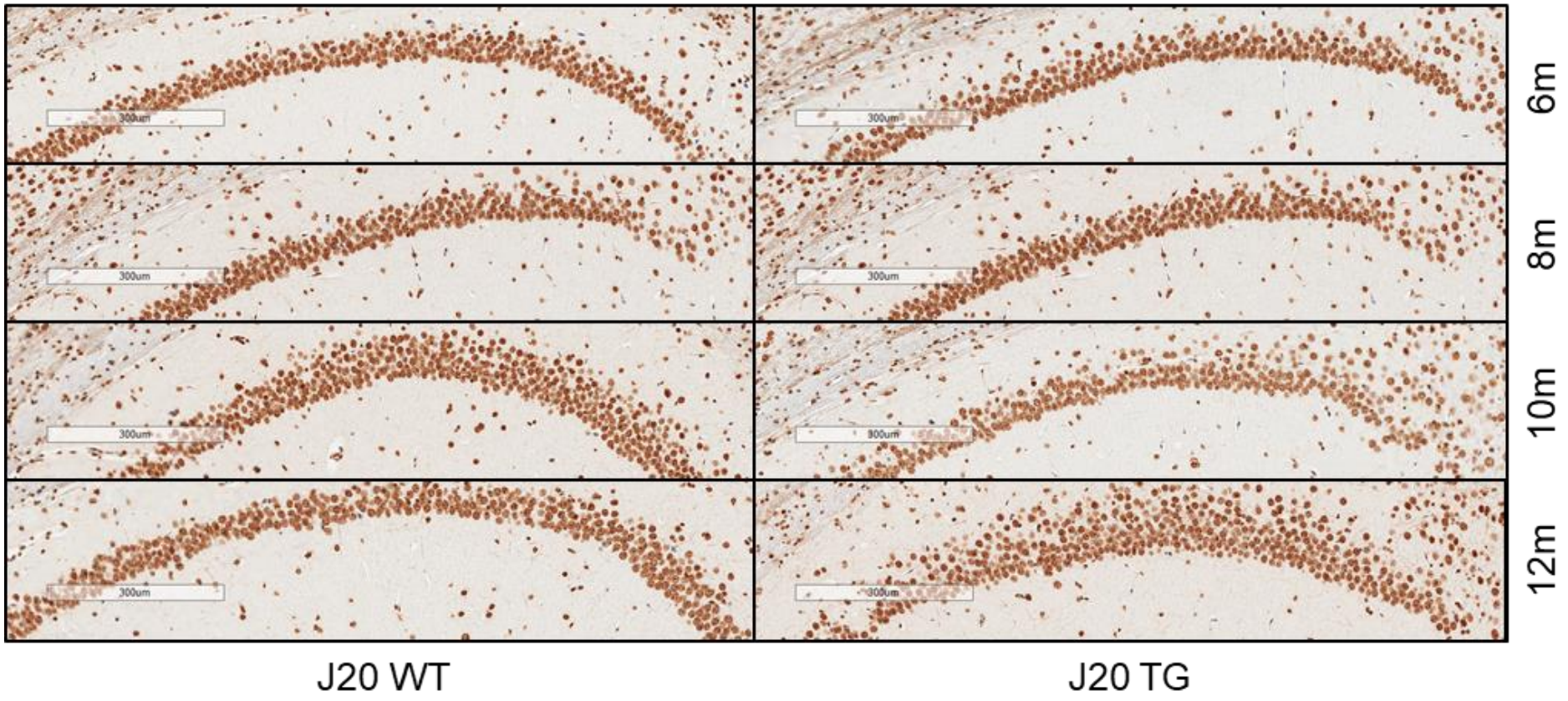


Figure 5.27 – Representative immunohistochemistry images for 5mC staining in J20 mice showing the hippocampal sub-region CA1 in detail.
WT: wild type. TG: transgenic. Primary antibody: anti-5mC (1:4000); secondary antibody: biotinylated goat anti-mouse (1:200). Bar = 300 µm

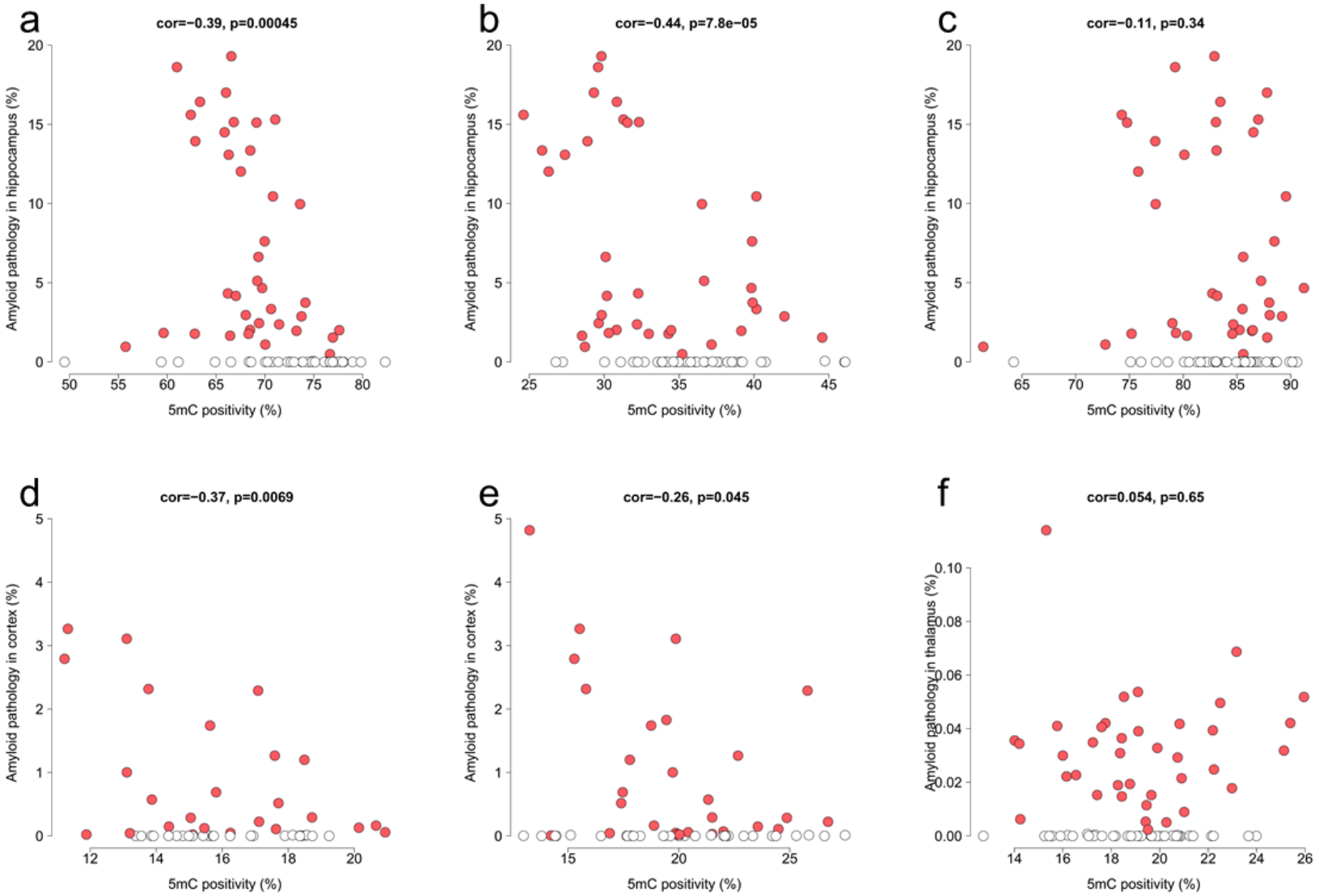


Figure 5.28 – Correlations between levels of 5mC and amyloid pathology in J20 mice.
(a) CA1, **(b)** CA3, **(c)** DG, **(d)** rostral cortex, **(e)** caudal cortex, **(f)** thalamus. Shown is the Pearson correlation statistic for each comparison. Red circles represent J20 transgenic (TG) mice, and white circles represent control (WT) mice.

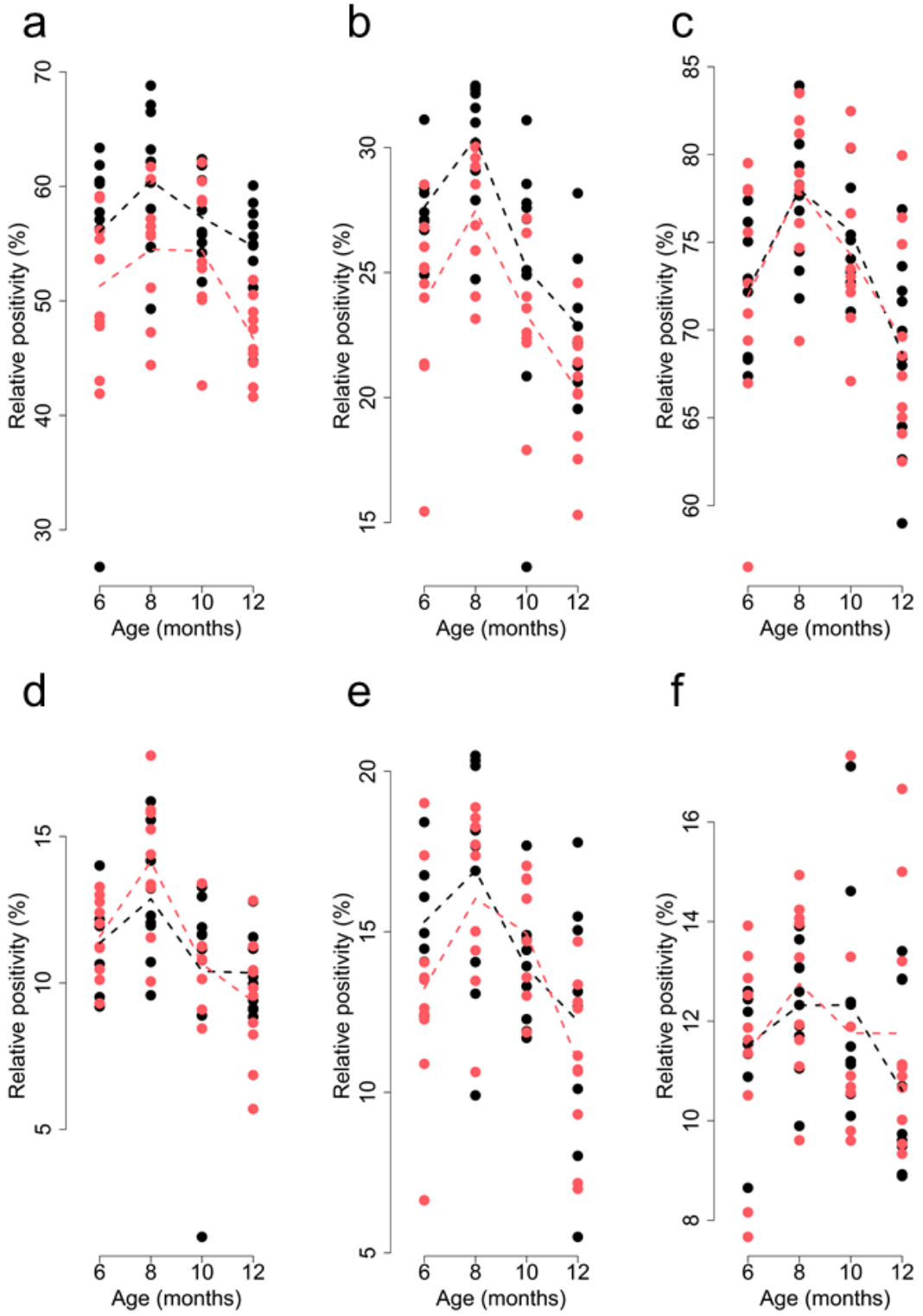


Figure 5.29 – Levels of 5hmC in J20 mice.
Quantification of anti-5mC immunoreactivity for each of the tested brain regions: **(a)** CA1, **(b)** CA3, **(c)** DG, **(d)** rostral cortex, **(e)** caudal cortex, **(f)** thalamus. Dashed lines represent mean paths of positivity across the four age groups. Red circles represent J20 transgenic (TG) mice, and black circles represent control (WT) mice.

Table 5.8 – Results for anti-5hmC quantification in brains from J20 mice.
Statistical results (two-way ANOVA) are shown. Significant (α < 0.05) P values are underlined.

Brain region	N	Genotype		Age		Genotype*Age	
		F-value	P-value	F-value	P-value	F-value	P-value
CA1	76	F(1,68) = 15.11	0.00023	F(3,68) = 4.21	0.0086	F(3,68) = 0.57	0.64
CA3	73	F(1,65) = 15.51	0.00020	F(3,65) = 17.57	1.84E-08	F(3,65) = 0.28	0.84
DG	75	F(1,67) = 0.082	0.78	F(3,67) = 11.49	3.62E-06	F(3,67) = 0.14	0.94
Rostral cortex	72	F(1,64) = 0.16	0.69	F(3,64) = 10.25	1.34E-05	F(3,64) = 0.88	0.45
Caudal cortex	71	F(1,63) = 2.11	0.15	F(3,63) = 8.82	5.75E-05	F(3,63) = 0.89	0.45
Thalamus	75	F(1,67) = 0.23	0.63	F(3,67) = 1.87	0.14	F(3,67) = 0.68	0.57

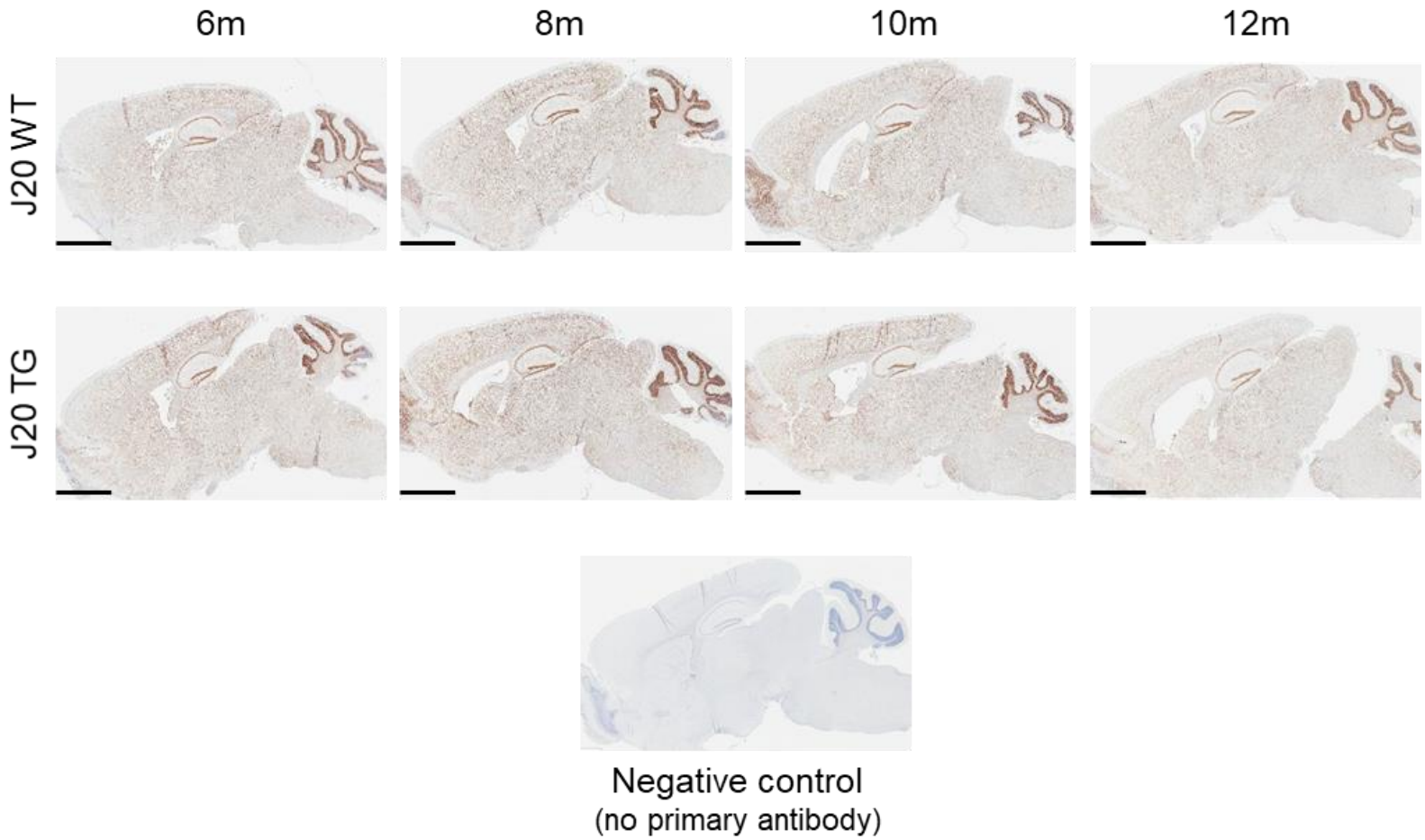


Figure 5.30 – Representative immunohistochemistry images for 5hmC staining in J20 mice.
Sagittal sections showing the distribution of anti-5hmC in the brains of J20 transgenic (TG) mice (second row) compared to wild-type control (WT) mice (top) at 6, 8, 10 and 12 months of age. Primary antibody: anti-5hmC (1:4000); secondary antibody: biotinylated goat anti-rabitt (1:200). Black bar = 2 mm.

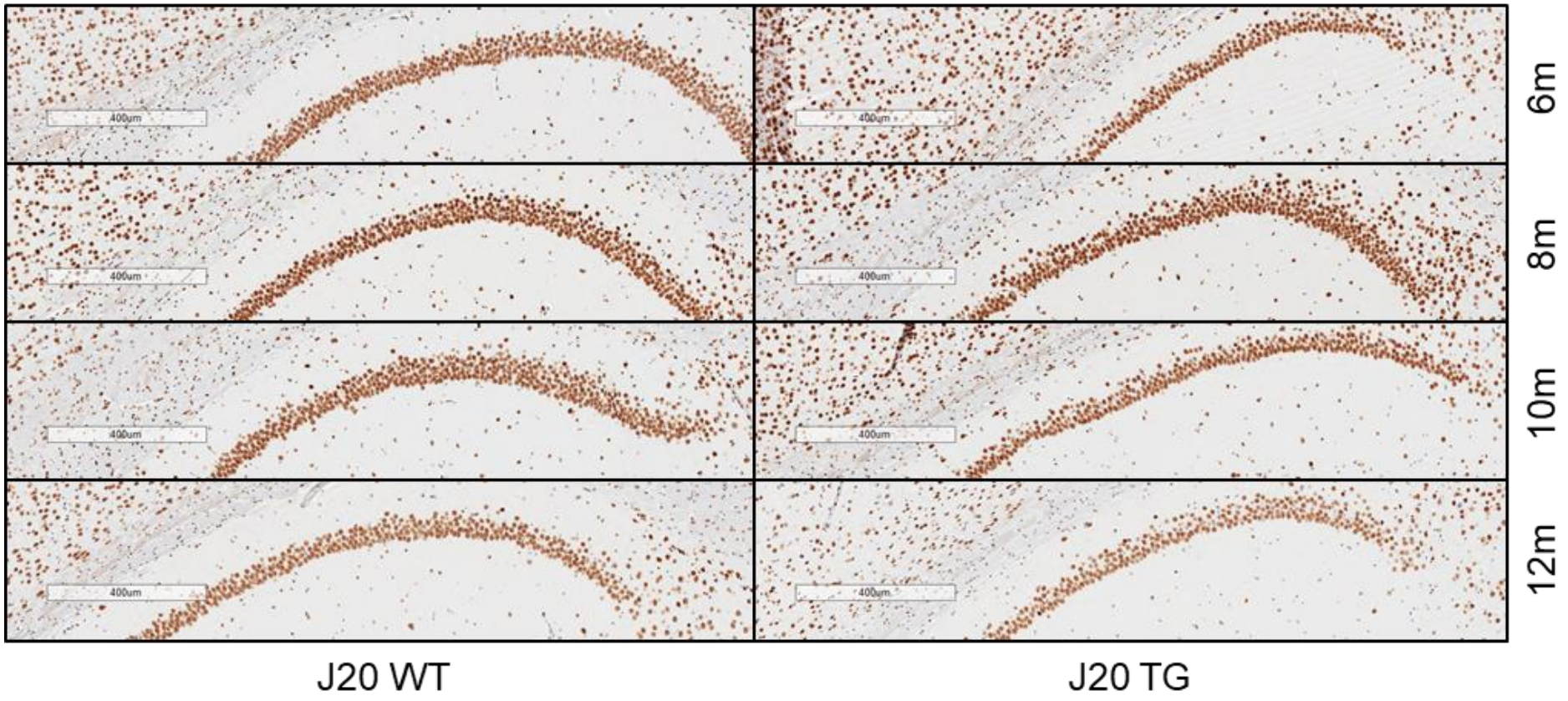


Figure 5.31 – Representative immunohistochemistry images for 5hmC stain in J20 mice showing the hippocampal sub-region CA1 in detail.
WT: wild type. TG: transgenic. Primary antibody: anti-5hmC (1:4000); secondary antibody: biotinylated goat anti-rabitt (1:200). Bar = 400 µm

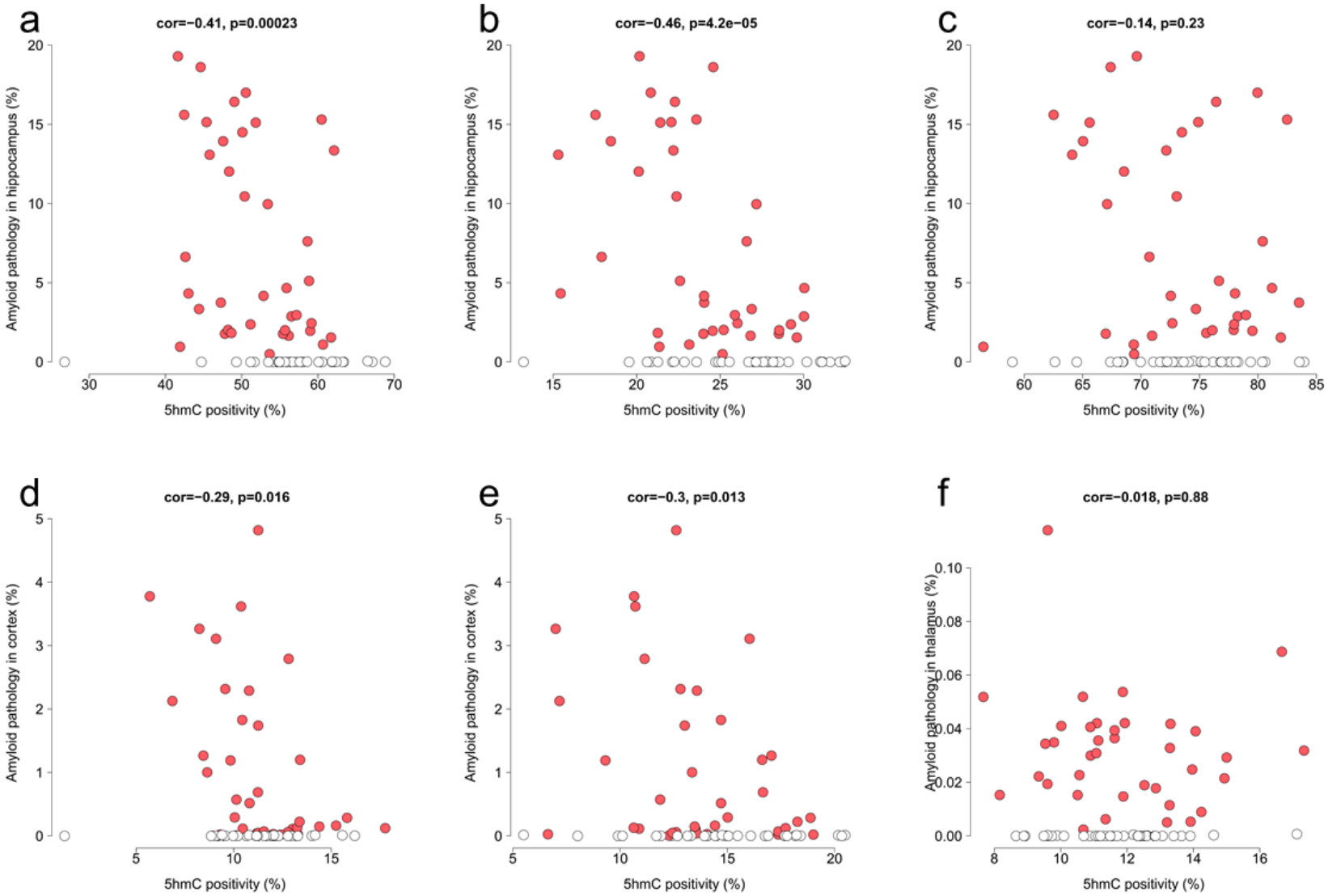


Figure 5.32 – Correlations between levels of 5hmC and amyloid pathology in J20 mice.
(a) CA1, **(b)** CA3, **(c)** DG, **(d)** rostral cortex, **(e)** caudal cortex, **(f)** thalamus. Shown is the Pearson correlation statistic for each comparison. Red circles represent J20 transgenic (TG) mice, and white circles represent control (WT) mice.

5.4. Discussion

In this chapter I identified that decreased levels of two DNA modifications (5mC and 5hmC) are associated with accumulation of tau (in rTg4510 mice) and amyloid (in J20 mice) pathology. Both genotype-associated and progressive changes measured by immunohistochemistry were observed, with these changes notably stronger in rTg4510 TG relatively to J20 TG mice, particularly in the hippocampus. Notably, some of our findings in the J20 mice showed a similar profile to results from a previous study that reported a negative correlation between plaque load and 5mC in the DG hippocampal sub-region in J20 mice as well (Lardenoije et al., 2018). In contrast to my results, an age-related (4, 8, 16, and 34 months) decrease in 5mC in the DG in J20 TG mice compared to controls was also reported in this study, perhaps because it covered a larger aging interval compared to my analysis.

Quantification of 5mC or 5hmC positivity was normalised to cell number in each region quantified (see **Section 5.2.9**). However, given that the rTg4510 mouse model exhibits neuronal cell loss, I cannot exclude the possibility that my observations may, at least partially, reflect the neurodegeneration (and this overall cell number) observed in this mouse model. This might also explain the negative correlation between tau pathology and the levels of both DNA modifications. Interestingly, a negative correlation between amyloid pathology and both DNA modifications was also observed in hippocampal and cortical regions in the J20 mice. The J20 mouse model, previously believed to have absent neurodegeneration, has been shown to exhibit limited age-dependent neuronal cell loss, varying by brain region, with neurodegeneration in CA1 present from 12 weeks of age, but no neuronal cell deficits observed in CA3 at any age (up to 36 weeks of age) (Wright et al., 2013).

In contrast to my immunohistochemistry data, no differences were identified in the LUMA experiments, when comparing rTg4510 TG and J20 TG mice to their respecting controls. This might be due to LUMA not being sensitive enough to detect small differences, or because we looked at DNA from the whole hippocampus; in the immunohistochemistry experiments I quantified specific hippocampal sub-regions. Furthermore, LUMA does not distinguish between

5mC and 5mC, and only quantified DNA modifications across informative CpG sites.

Looking forward, potential complimentary assays for future follow-up studies to evaluate global changes in DNA modifications in AD mouse models could include high performance liquid chromatography (HPLC) and liquid chromatography mass spectrometry (LC-MS/M). However, these have high costs associated to them, and enzyme-linked immunosorbent assay (ELISA)-based methods have also been developed in recent years, in the form of commercialised kits, which should be considered as an additional alternative well.

Chapter 6. Genome-wide methylation profiling of tau and amyloid mice

6.1. Introduction

Building on the dramatic changes in gene expression identified in **Chapter 4**, and the global DNA modification changes identified in **Chapter 5**, this chapter describes my analysis of genome-wide patterns of DNA methylation at single-base resolution in the entorhinal cortex (ECX) from rTg4510 and J20 mice.

As discussed in **Section 1.4.1** (**Chapter 1**), DNA methylation is the most widely studied epigenetic modification in human disease, with several studies robustly linking methylomic variation to AD and other neurodegenerative disorders. A very limited number of studies assessing DNA methylation in AD mouse models have been performed, profiling a very low number of mice (see **Section 1.5.2.2** in **Chapter 1** for details), and no study to date has used sequencing-based technology to achieve this.

Methods to profile DNA methylation have undergone significant improvements in the last decade due to the development of array- and sequencing-based technologies that have allowed the efficient quantification of DNA methylation at individual sites across the genome (Yong et al., 2016). Commonly used methodologies for genome-wide DNA methylation profiling are shown in **Figure 6.1**.

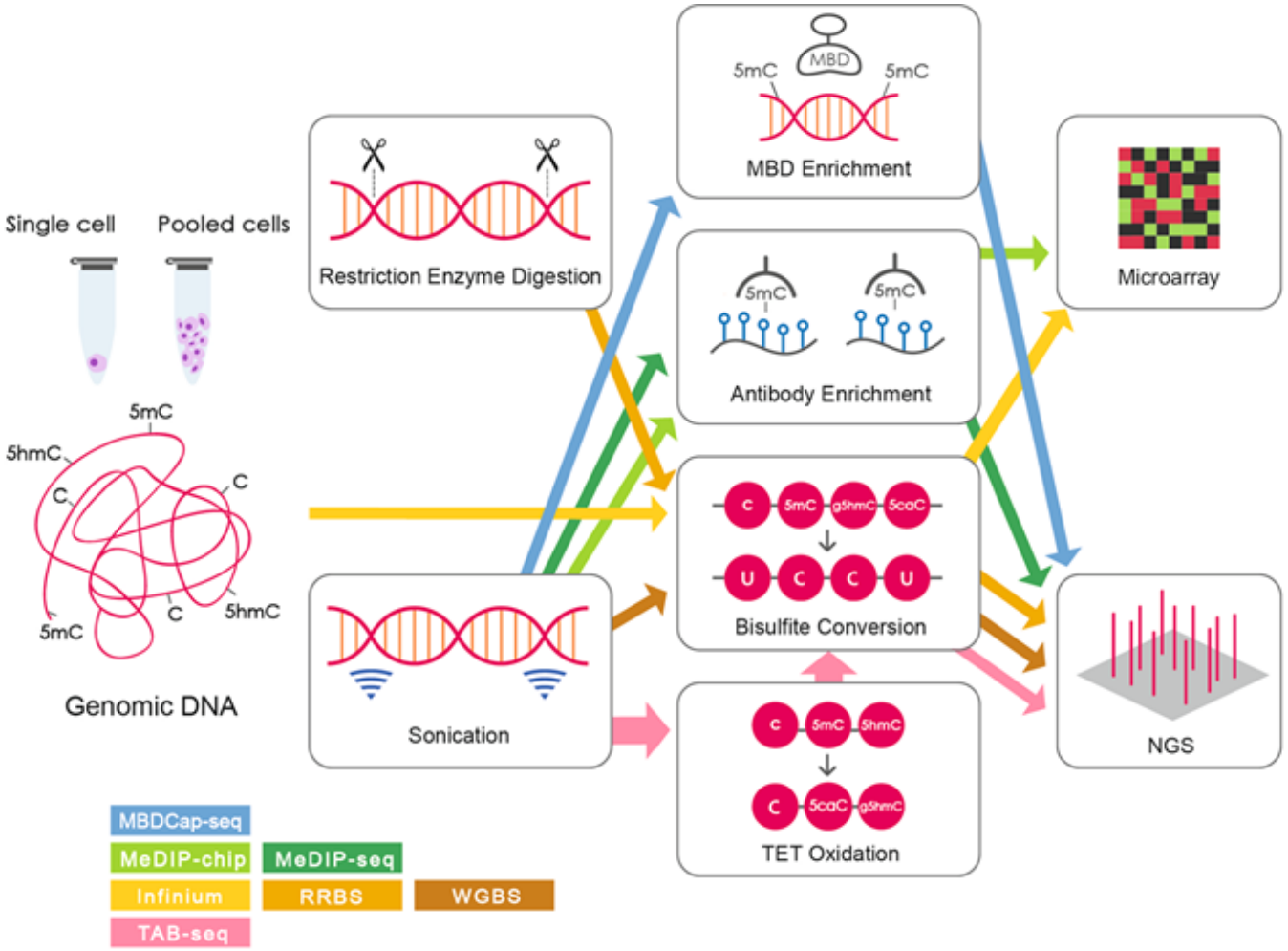


Figure 6.1 – Commonly used methods for genome-wide DNA methylation analysis.
(a) The procedures may involve fragmentation of genomic DNA by restriction enzyme digestion or sonication. The genomic DNA can be subjected to methyl-CpG-binding domain (MBD) enrichment, antibody enrichment, bisulfite conversion or oxidation by ten eleven translocation (TET) enzymes, before being analysed by microarray or next-generation sequencing (NGS) platforms. Figure and legend adapted from Yong et al. (2016).

Bisulfite-based approaches coupled with high-throughput sequencing, such as reduced representation bisulfite sequencing (RRBS) and whole-genome bisulfite sequencing (WGBS), allow for quantitative, genome-wide, site-specific methylation assessments (Doherty and Couldrey, 2014). Bisulfite conversion involves treating genomic DNA with sodium bisulfite, which reacts differently with unmethylated versus methylated cytosine, resulting in a positive display of methylation (Fouse et al., 2010). Sodium bisulfite promotes the conversion of unmethylated cytosine (C) to uracil (U), which is then replaced by thymine (T) by polymerase chain reaction (PCR) amplification. Methylated cytosines are unaffected by the bisulfite conversion, therefore remaining as cytosine after the bisulfite treatment reaction (Fouse et al., 2010, Yong et al., 2016). PCR clones derived from a single genomic DNA fragment can then be sequenced, allowing methylation status to be quantified at single-nucleotide resolution (Fouse et al., 2010, Gu et al., 2011) using second-generation sequencing (discussed in more detail in **Chapter 4**). Bisulfite sequencing comes, however, with its challenges, particularly for studies in mammals: 1) repetitive elements (or repeated sequences), which occur in multiple copies throughout the genome and are usually highly methylated, make the alignment of short-read sequencing reads difficult; 2) the conversion of unmethylated cytosine to uracil/thymidine reduces genomic complexity to three bases in unmethylated regions, further complicating the alignment of sequencing reads; and 3) the costs associated with sequencing large genomes are relatively high, making large studies unfeasible (Fouse et al., 2010).

RRBS, introduced in 2005 by Meissner et al., combines the use of a CpG-specific restriction enzyme and bisulfite sequencing in order to enrich for regions of the genome with high CpG content, i.e., only a small proportion of the bisulfite-treated genome is interrogated (‘reduced representation’), emerging as a cost-efficient alternative to WGBS (Meissner et al., 2005). An overview of the RRBS workflow is given in **Figure 6.2**. In this approach genomic DNA is fragmented using the methylation-insensitive restriction enzyme MspI, which recognises and cleaves 5'-CCGG-3' (Lee et al., 2014). Sequencing libraries are then constructed from size-selected MspI DNA fragments; in total only about 1% of the genome is sequenced, with the ‘reduced genome’ still including the majority of promoters and repeat sequences. This decreases the amount of sequencing required, reducing cost significantly and making this a relatively efficient method for large

studies (Gu et al., 2011). Despite being very cost-effect and characterised by very low input requirements (10-300 ng) (Gu et al., 2011, Yong et al., 2016), RRBS also has some limitations. An obvious disadvantage of RRBS, compared to other approaches such as WGBS, is the fact that only CpG-rich regions of the genome are sequenced, with RRBS failing to capture DNA methylation across all functionally-relevant domains. Furthermore, because the library is size-selected after enzymatic digestion with MspI, but before adapter ligation, any MspI fragments outside of the size-selection range (~40-250 bp) will not be sequenced (Doherty and Couldrey, 2014), i.e., since coverage is limited to regions in proximity to the restriction enzyme's recognition sites, intergenic and distal regulatory elements exhibit low coverage (Yong et al., 2016). Another limitation of any method based on sodium bisulfte conversion, including RRBS, is the inability to distinguish between 5-methycytosine (5mC, 'true methylation') and other DNA modifications such as 5hmC; with RRBS or WGBS we can only determine if a particular cytosine residue contains a DNA modification or not. As mentioned in **Section 1.4.1.4**, 5hmC is particularly enriched in the central nervous system.

Despite the described limitations, their advantages make targeted approaches such as RRBS the methods of choice for sequencing-based genome-wide studies in mammals, particularly in studies with a high number of samples (Fouse et al., 2010, Yong et al., 2016). Because methods for profiling DNA methylation using RRBS are still in development, I fully-optimised both laboratory and analytical approaches as part of my PhD research.

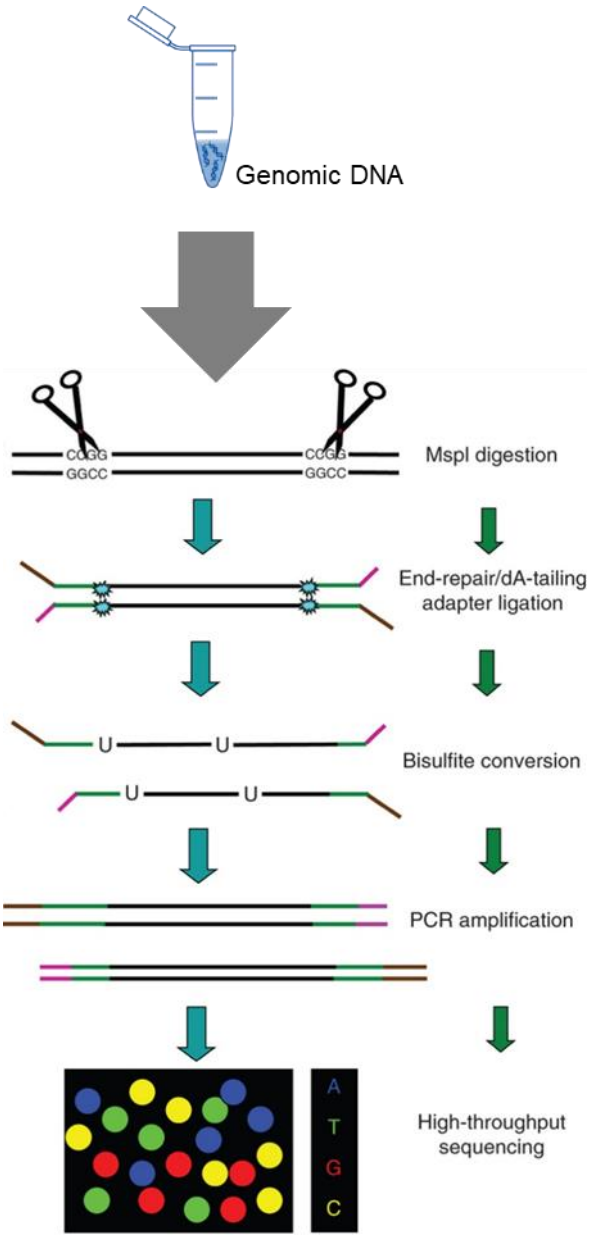


Figure 6.2 – Flowchart of the experimental procedures for RRBS.
Adapted from Guo et al. (2015).

6.2 Methods

In this section the experimental procedures used for the evaluation of genome-wide DNA methylation profiles using RRBS in samples from the ECX (described in **Chapter 2**, **Section 2.4** and **Section 2.5**) dissected from rTg4510 and J20 mice (**Chapter 2**, **Section 2.1**) are described in detail.

When I started this work, the RRBS method had not been optimised to profile DNA methylation in a large number of samples, and the original protocol required certain modifications (see **Sections 6.2.2.6** and **6.2.2.7**). Therefore, my first step in this study was to optimise the method by performing a pilot experiment with eight test/control samples. After carefully checking the quality of the pilot data, I processed the rTg4510 and J20 samples. These were separated into two batches, with the two mouse models processed independently – 64 samples were processed in each batch (64 samples for the rTg4510 model, and 63 samples plus a negative control for the J20 model). All RRBS libraries were sequenced together (total number of samples = 128 samples), and this study represents one of the largest analyses of RRBS data – and the first in transgenic models of AD pathology – yet undertaken.

6.2.1. Assessment of DNA quality and concentration

All DNA samples used for my RRBS experiments were quantified and checked for purity using a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific), as described in **Section 2.5.3.1**.

Before RRBS library preparation, the concentration of each sample was determined using a fluorescence-based assay for double-stranded DNA (dsDNA) – the Qubit high sensitivity assay (Qubit dsDNA HS Assay, Thermo Fisher Scientific) – that utilises target-selective dyes that emit fluorescence when bound to DNA. Briefly, following the manufacturer's instructions, a working reagent (WR) was freshly prepared by mixing HS Reagent and HS Buffer (both provided with the kit) in a 1:200 proportion. 'Standard' samples 1 and 2 (provided with the kit) were prepared by adding 10µL of each to 190 µL of WR in their respective Qubit Assay Tubes (Thermo Fisher Scientific). 'Test' samples were prepared by adding 1 µL of the original DNA sample to 199 µL of WR in individual Qubit Assay Tubes.

All tubes containing a final volume of 200 µL were mixed by vortexing for 2-3 seconds and left to incubate at room temperature for 2 minutes. Following incubation, standards and samples were read in a Qubit 3.0 Fluorometer (Thermo Fisher Scientific) and the concentration of each sample was recorded.

6.2.2. Library preparation

RRBS libraries were prepared using the Premium RRBS kit (Diagenode) with some modifications. An overview of the kit is given in **Figure 6.3**.

All reagents were provided with the kit unless otherwise stated, with all reagents and consumables being sterile and DNase and RNase free. All experimental procedures were performed manually (in a PCR workstation, or PCR hood) as automation has not yet been established for this protocol by the manufacturer.

The full protocol for the 64 samples from each batch (each mouse model processed separately, as described above) took 6 days, with the protocol repeated twice (total number of samples = 128 samples, 64 samples for each mouse model / batch). All volume calculations (e.g. for the digestion master mix) were performed in advance. On the first day, all reagents, except 80% ethanol, were prepared as indicated in the manufacturer's instructions. 80% ethanol was prepared fresh from 100% ethanol (molecular biology grade, Sigma-Aldrich) on each day. On each day, before starting, every surface and all equipment were thoroughly cleaned using 70% ethanol followed by DNAZap (Thermo Fisher Scientific).

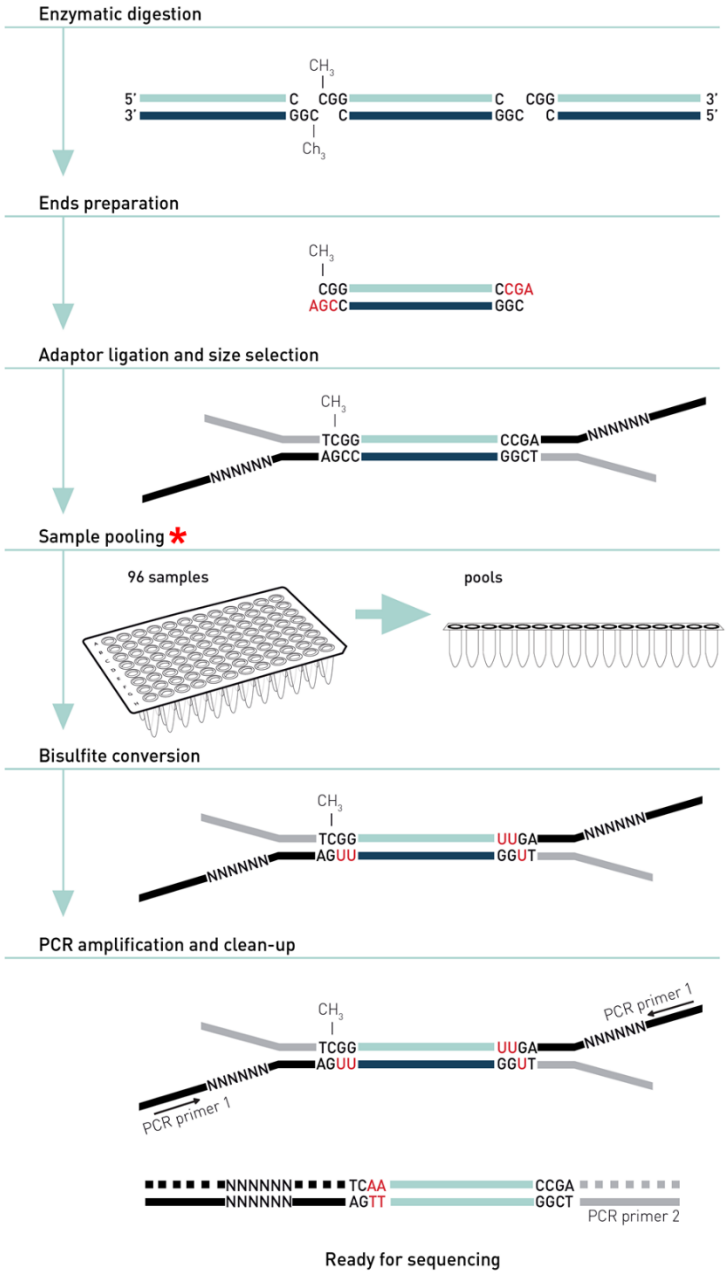


Figure 6.3 – Modified premium RRBS kit (Diagenode) workflow.
The red asterisk indicates the key step modified in our study – an absolute quantification of the libraries was used instead of the relative quantification from the kit. The number of samples pooled together was also different than recommended in the original protocol. Figure adapted from the Premium RRBS kit manual (Diagenode, 2016).

6.2.2.1. Sample dilutions (day 1)

After determining the concentration of each sample using the fluorescence-based Qubit High Sensitivity assay (see section **6.2.1**), samples were diluted (100 ng of DNA from each sample in a total volume of 26 µL of buffer EB (Qiagen, see section **2.5.2.3**)) and aliquoted into a 96-well plate, starting at well A01 and proceeding in columns.

6.2.2.2. Enzymatic digestion (day 2)

The Digestion Mix was prepared in a 1.5 mL tube as described below:

Components	**Volume / reaction (x1)**
Enzyme Buffer	3 µL
Restriction Enzyme	1 µL
Total	4 µL

4 µL of the Digestion Mix was added to each sample in the 96-well plate and mixed by pipetting 10 times (total volume = 30 µL). The plate was placed in a thermocycler and left to run using the following program overnight:

37°C	12 hours
8°C	∞

6.2.2.3. Ends preparation (day 3)

After digestion, the ends of the DNA fragments need to be filled-in with dGTP and methylated dCTP, followed by the addition of an A overhang to enable adaptor ligation (Fouse et al., 2010). The Ends Preparation Mix was prepared in a 1.5 mL tube as described below:

Components	**Volume / reaction (x1)**
Ends Preparation Enzyme	1 µL
dNTP mix	1 µL
Unmethylated spike-in control	1 µL
Methylated spike-in control	1 µL
Total	4 µL

A fully methylated control DNA fragment and an unmethylated control DNA fragment were added to each sample to control for conversion efficiency during sodium bisulfite treatment

4 µL of the Ends Preparation Mix was distributed to each sample in the 96-well plate and mixed by pipetting 10 times (total volume = 34 µL). The plate was placed in a thermocycler and left to run using the following program:

30°C	20 minutes
37°C	20 minutes
75°C	20 minutes
8°C	∞

6.2.2.4. Adaptor ligation (day 3)

5 µL of the adaptors was added to each sample, using a multichannel pipette to pipette 8 adaptors at a time, alternating between the different sets (strips A-F) of adaptors (this is important for sample pooling, as described in **Section 6.2.2.6**). As described in **Chapter 4** (**Section 4.2.2.6**), chemically synthesised adaptors are short oligonucleotides that can be added to the ends of DNA molecules. In the Premium RRBS kit, the adaptors include indexes (or "barcodes", described in **Figure 6.4**) that allow several samples (or libraries) to be pooled and sequenced simultaneously (multiplex sequencing). RRBS adaptors are methylated at cytosine residues to prevent conversion during sodium bisulfite treatment (Fouse et al., 2010).

Adaptor tube strips	Adaptor ID	Index sequences
Strip A	17	GTAGAG
	24	GGTAGC
	26	ATGAGC
	28	CAAAAG
Strip B	29	CAACTA
	30	CACCGG
	31	CACGAT
	32	CACTCA
Strip C	33	CAGGCG
	34	CATGGC
	35	CATTTT
	36	CCAACA

Adaptor tube strips	Adaptor ID	Index sequences
Strip D	37	CGGAAT
	38	CTAGCT
	39	CTATAC
	40	CTCAGA
Strip E	41	GCGCTA
	42	TAATCG
	43	TACAGC
	44	TATAAT
Strip F	45	TCATTC
	46	TCCCGA
	47	TCGAAG
	48	TCGGCA

Figure 6.4 – Distribution of the 24 adaptors in the tube strips and sequences of the 24 corresponding indexes used in my RRBS experiments.
Figure taken from the Premium RRBS kit manual (Diagenode, 2016).

The Ligation Mix was subsequently prepared in a 15 mL tube as described below:

Components	Volume / reaction (x1)
Ligation Buffer	40 µL
Ligase	1 µL
Total	4 µL

41 µL of the Ligation Mix was distributed to each sample in the 96-well plate and mixed by pipetting 10 times (total volume = 80 µL). The plate was placed in a thermocycler and left to run using the following program:

25°C	20 minutes
65°C	10 minutes
8°C	∞

6.2.2.5. Size selection

60 µL of AMPure XP Beads (Beckman Coulter) was added to each well of the 96-well plate using a multichannel pipette, mixed by pipetting at least 10 times, and left to incubate for 15 minutes. The plate was then placed on a 96-well magnet for at least 5 minutes (until the supernatant was clear).

Working column by column, the supernatant was removed using a multichannel pipette (taking care not to touch the beads) and the pellets were washed for 5 seconds with 100 µL of freshly prepared 80% ethanol twice. The pellets from each column were left to air-dry for at least three minutes. The 96-well plate was taken from the magnet and 25 µL of Resuspension Buffer was added to each well using a multichannel pipette. Samples were then vortexed to resuspend the beads and left to incubate for 5 minutes out of the magnet to elute the DNA. The 96-well plate was placed on the magnet again for 5 minutes and the supernatant transferred to a new 96-well plate, discarding the beads. 2 µL of each sample was taken from this new plate and added to another 96-well plate containing 118 µL of NEB dilution buffer (New England Biolabs) 1x (1/60 dilution, optimised as the ideal dilution to bring the libraries' concentrations to the interval of the standards' concentrations and therefore accurately quantify the libraries), for quantification using qPCR (**6.2.2.6**). The remainder of the libraries were stored at -20°C during the quantification by qPCR and until used for sample pooling (**6.2.2.7**).

6.2.2.6. Quantification (day 4)

Previous RRBS experiments in our lab using the Premium RRBS kit (Diagenode) had highlighted considerable variability in the number of sequencing reads between samples within and between pools (**Figure 6.5**). To overcome this issue, I decided to replace the relative quantification step described in the official Premium RRBS kit protocol with an absolute quantification approach (**Figure 6.3**), using the NEBNext Library Quant Kit (New England Biolabs) with modifications to quantify the libraries. My pilot experiment to test this approach revealed notably reduced variation in the number of reads obtained after sequencing (**Figure 6.9**).

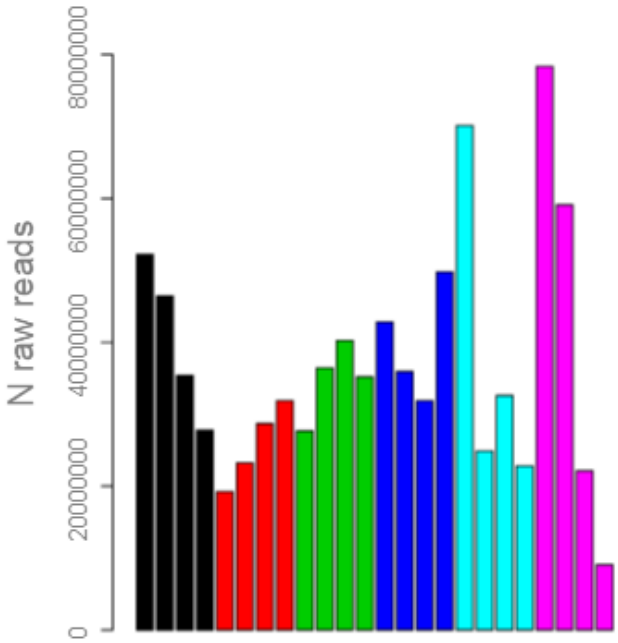


Figure 6.5 – Number of sequencing reads obtained from a previous experiment using the Premium RRBS kit.
The number of raw sequencing reads from previous experiments in our group looking at DNA methylation in the mouse hippocampus exhibited high variability between and within pools. Each column represents a sample and each colour represents a pool.

All reagents for quantification were provided with the NEBNext Library Quant Kit, with the Quantification Mix prepared according to the manufacturers' instructions in advance and stored 'ready to use'. Four standards (NEBNext Library Quant DNA Standards 10 pM, 1 pM, 0.1 pM, and 0.01 pM, provided with the kit) and

one negative control (no template control or NTC, containing dilution buffer only) were included. Diluted samples (1/60 dilution, see **6.2.2.5**), standards and the negative control were prepared in triplicate in 96-well qPCR optical plates as described below, using a multichannel pipette:

Components	Volume / reaction (x1)
NEB Quant Master Mix (containing primers & ROX)	8 µL
Diluted sample, standard, or negative control	2 µL
Total	10 µL

The plate was sealed with a qPCR-specific adhesive plate seal, mixed by vortexing, centrifuged briefly and placed in a qPCR thermocycler using the following program (the second plate was kept at 4°C during the first run):

Cycle step	Temperature	Time	Cycles
Initial Denaturation	95°C	3 minutes	1
Denaturation	95°C	15 seconds	35
Annealing	63°C	45 seconds	
Hold	4°C	∞	

Using the standard curve generated for each plate (**Figure 6.6a**), sample concentrations were calculated with the qPCR instrument software (StepOnePlus Real-Time PCR System, Thermo Fisher Scientific). The concentrations of the diluted libraries were within the range of interval of the standards (**Figure 6.6b**), allowing for an accurate quantification of the RRBS libraries.

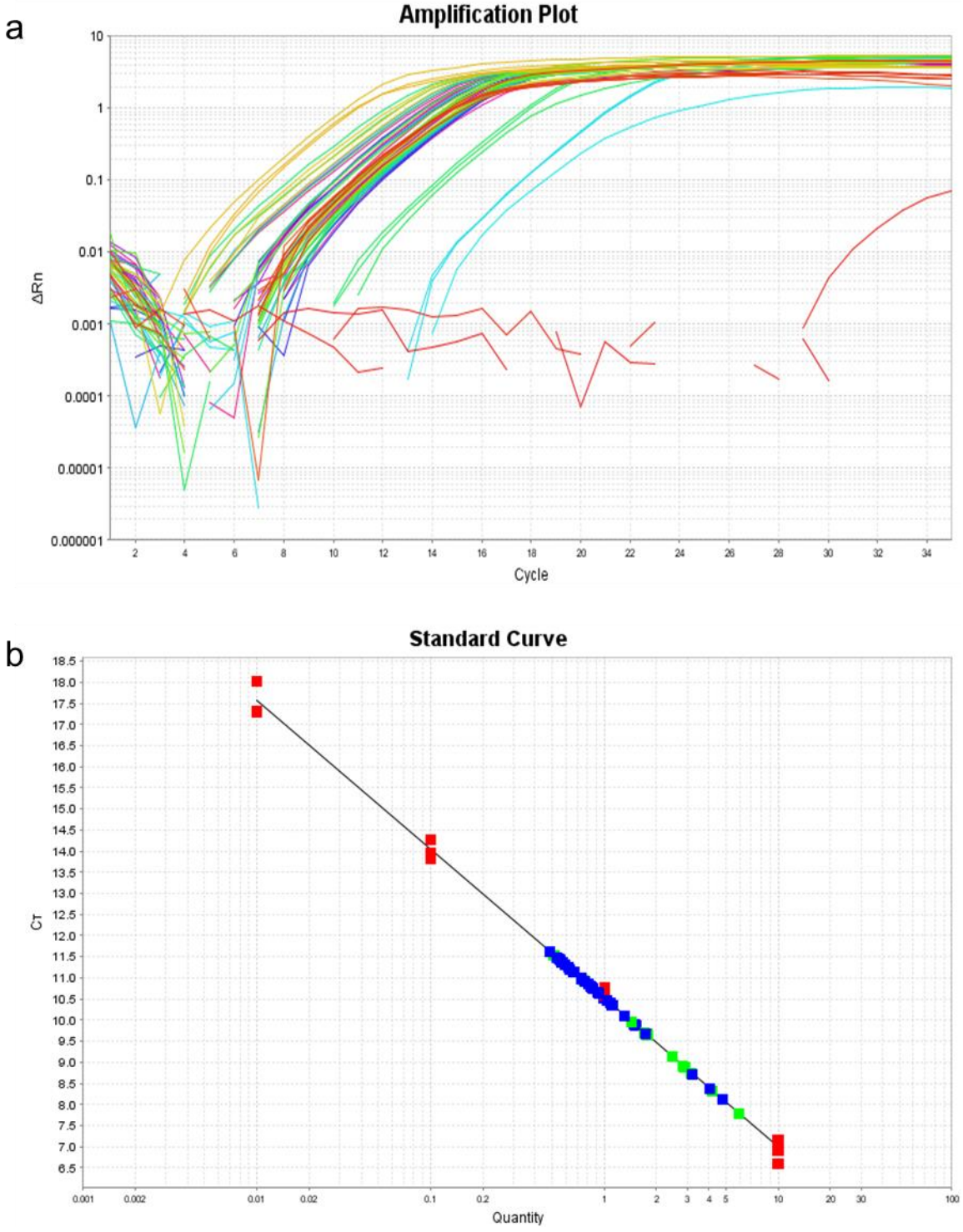


Figure 6.6 – Representative amplification plot and corresponding standard curve from qPCR quantification of an RRBS library.

(a) Amplification plot showing that the diluted samples were within the range of the NEBNext Library Quant DNA standards. **(b)** Standard curve showing data generated from the four (10 pM, 1 pM, 0.1 pM, and 0.01 pM) NEBNext Library Quant DNA standards (red) and samples (blue and green). Samples in green were flagged by the software and therefore repeated in a new qPCR plate. $R^2 = 0.996$.

6.2.2.7. Sample pooling (day 5)

Making sure that there were no duplicate indexes (described in **6.2.2.4.**) in the same pool, pools of 8 samples (pool concentration between 15 nM and 30 nM) were prepared in 1.5 mL tubes. Sample volumes for each pool were calculated using the next-generation sequencing (NGS) Laboratory Information Management System (LIMS) software (Illumina), adding Resuspension Buffer (from **6.2.2.5**) to reach a volume of 120 µL per pool.

Despite the original protocol being optimised for pools of six samples, our previous experiments using the Premium RRBS kit (Diagenode) revealed lower than optimal sequencing depth when running six samples in a single sequencing lane. Sequencing depth is a key parameter when evaluating DNA methylation profiles, with a greater depth providing higher statistical power to identify differentially methylated sites (Yong et al., 2016). Therefore, we decided that sequencing four samples on each Illumina lane (see **Section 6.2.3**) would likely provide more optimal output per sample. To maximise read depth per sample and to avoid batch effects, we pooled eight samples together and distributed each pool of eight samples across two sequencing lanes (**Section 6.2.3**). The concentration for each pool was determined using the Qubit dsDNA HS Assay (Thermo Fisher Scientific) as described in **6.2.1**.

240 µL of AMPure XP Beads (Beckman Coulter) were added to each pool, mixed well by pipetting, and incubated for 15 minutes at room temperature to allow DNA binding to the beads. After incubation, the tubes were placed in a magnetic 1.5 mL tube rack for at least 5 minutes (until the supernatant appeared clear), and the supernatant was subsequently discarded. Pools were washed twice with 500 µL of freshly prepared 80% ethanol for 5 seconds, and the tubes were left opened in the magnet, to enable the beads to air-dry for at least 25 minutes (until beads were dry). Tubes were taken off the magnet and 36 µL of Resuspension Buffer was added to each tube. Each tube was mixed by pipetting 10 times, centrifuged quickly and incubated for 5 minutes out of the magnet to elute DNA. The tubes were then place on the magnet once again for at least 5 minutes (until the supernatant was clear) and the supernatant was transferred to a strip tube.

6.2.2.8. Bisulfite conversion (days 5 and 6)

BS Conversion Reagent was prepared by adding 790 µL of BS Solubilization Buffer and 300 µL of BS Dilution Buffer to a tube of BS Conversion Reagent, vortexing for 10 minutes at room temperature, adding 160 µL of BS Reaction Buffer, and mixing for an additional 1 minute.

33 µL of each library pool was transferred to a new 96-well plate and 117 µL of BS Conversion reagent was added (total volume = 150 µL), followed by mixing and brief centrifuging. The 96-well plate was then left to incubate overnight in a thermocycler, as indicated below:

Cycle step	Temperature	Time	Cycles
Denaturation	95°C	1 minute	20
Conversion	60°C	10 minutes	
Hold	4°C	∞	

On the subsequent morning, BS Spin columns were placed into the provided BS Collection Tubes, and 600 µL of BS Binding Buffer were added into each BS Spin Column. Bisulfite-converted library pools were loaded into the BS Spin Columns containing the BS Binding Buffer, mixed by inverting the column several times, and centrifuged at 12,000 x g for 30 seconds. The flow-through was discarded, and 100 µL of BS Wash Buffer was added to each column, followed by centrifuging at 12,000 x g for 30 seconds. 200 µL of BS Desulphonation Buffer was added to each column and let stand at room temperature (20-30°C) for exactly 30 minutes. After incubation, columns were centrifuged at 12,000 x g for 30 seconds, 200 µL of BS Wash Buffer was added to each column followed by centrifugation at 12,000 x g for 30 seconds, and the flow-through was discarded. Another 200 µL of BS Wash Buffer was added to each column, followed by centrifugation for an additional 30 seconds. Each column was then placed into a new 1.5 mL tube, and 22 µL of BS Elution Buffer were added directly to the centre of the column matrix. After incubation for 2 minutes, columns were centrifuged for 30 seconds at 12,000 x g and the bisulfite-converted DNA was kept on ice during the qPCR step (**6.2.2.9**) to determine the optimal cycle number for the enrichment PCR (**6.2.10**).

6.2.2.9. Determination of the optimal cycle number for the enrichment PCR (day 6)

Bisulfite-converted pools were quantified using the NEBNext Library Quant Kit (New England Biolabs) with modifications, as described in section **6.2.2.6**. A single qPCR per library was performed instead of triplicates, as recommended in the Premium RRBS kit manual (Diagenode, 2016), and the pools were not diluted but quantified directly. The remainder of the pooled samples were kept at 4°C (or on ice) during the quantification as mentioned in **6.2.2.8**. A standard curve was generated for each plate (**Figure 6.7a**), and the concentrations of the libraries were observed to be within the range of intervals for the standards and very close to each other (**Figure 6.7b**). Ct values for each pool were used to determine the optimal cycle number (Ct - 1) for the enrichment PCR (**6.2.2.10**).

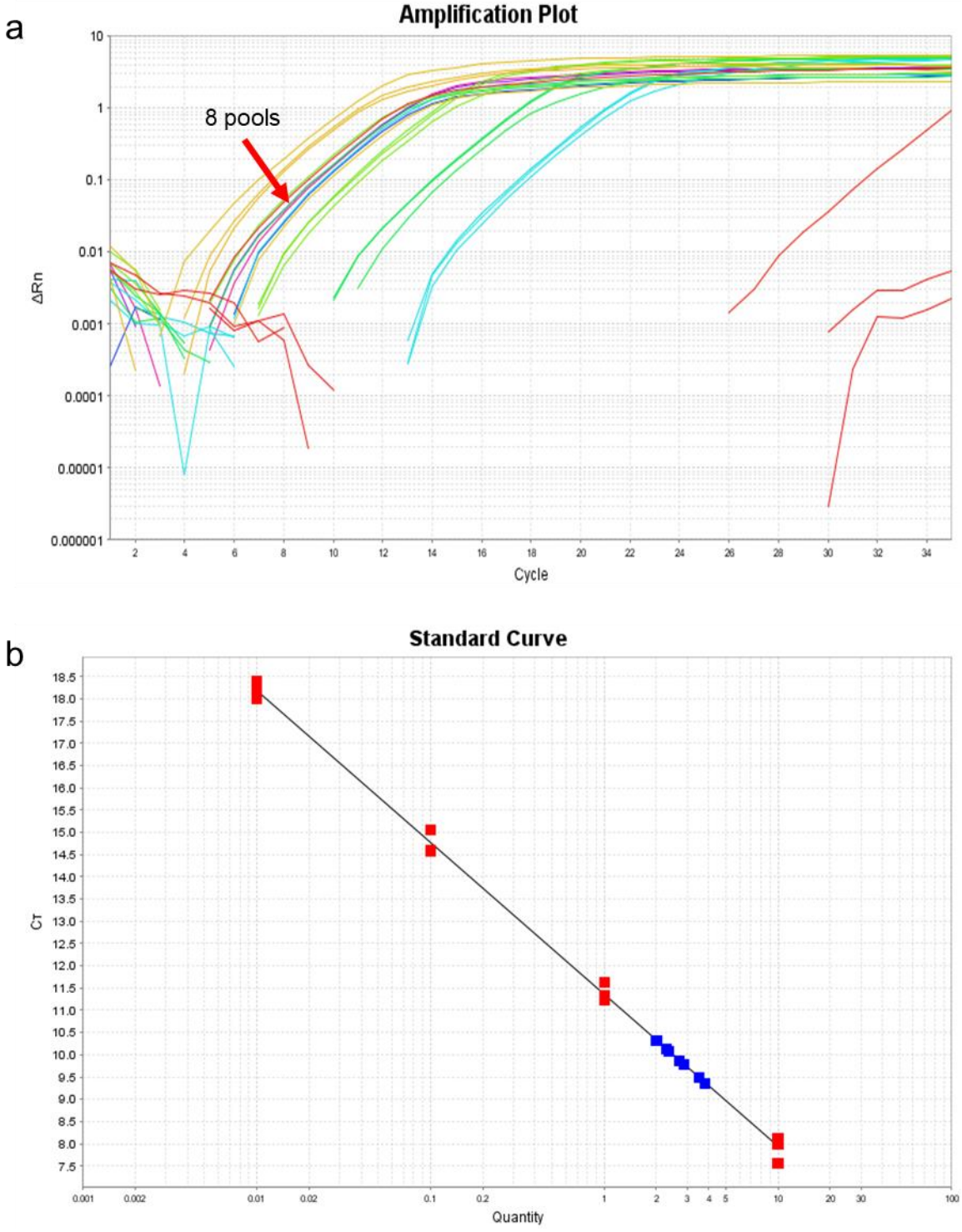


Figure 6.7 –Representative amplification plot and corresponding standard curve from one of the qPCR runs to calculate the Ct value for each RRBS pool.
(a) Amplification plot showing that the RRBS pools (indicated with the red arrow had very similar Ct values). NEBNext Library Quant DNA Standards. **(b)** Standard curve showing the four (10 pM, 1 pM, 0.1 pM, and 0.01 pM) NEBNext Library Quant DNA Standards (red) and RRBS pools (blue). $R^2 = 0.997$.

6.2.2.10. Enrichment PCR (day 6)

The Amplification Mix was prepared in a 1.5 mL tube as described below:

Components	Volume / reaction (x1)
2X MethylTaq Plus Master Mix	25 µL
Primer mix	2.5 µL
Water	3.5 µL
Total	31 µL

In a new 96-well PCR plate, 31 µL of Amplification Mix and 19 µL of each bisulfite converted RRBS pool (final volume = 50 µL) were added to each well and incubated in a thermocycler as indicated below:

Cycle step	Temperature	Time	Cycles
Denaturation	95°C	5 minutes	1
Denaturation	98°C	20 seconds	Ct – 1 (**6.2.2.9**)
Annealing	60°C	15 seconds	
Extension	72°C	45 seconds	
Final extension	72°C	7 minutes	1
Hold	4°C	∞	

6.2.2.11. Clean up (day 6)

50 µL of AMPure XP Beads were added to the samples, mixed by pipetting at least 10 times, and left to incubate for 15 minutes at room temperature to allow DNA binding to the beads. After incubation, the plate was placed on a 96-well magnet for at least 5 minutes (until the supernatant appeared clear), and the supernatant was removed and discarded. Libraries were sequentially washed twice for 5 seconds with 100 µL of freshly prepared 80% ethanol and the beads were left to air-dry for at least 10 minutes (until the beads appeared dry). Beads were then resuspended in 15 µL of Resuspension Buffer and incubated for 5 minutes to elute the DNA from the beads. The 96-well plate was placed on a 96-

well magnet for at least 5 minutes (until the supernatant appeared clear), and the clear supernatant was transferred to a new 0.6 mL tube for each pool.

6.2.2.12. Quality control (day 6)

Libraries were checked using the High Sensitivity D1000 Screentape and 2200 TapeStation System (Agilent Technologies). The TapeStation System is described in detail in **Section 2.5.3.4**, and the protocol for the High Sensitivity D1000 Screentape is described in **Section 4.2.2.8**. When the run was finished, I checked the profile of each RRBS library using the TapeStation Software. An example is given in **Figure 6.8**, where the area observed between the lower and upper peaks represents the RRBS library.

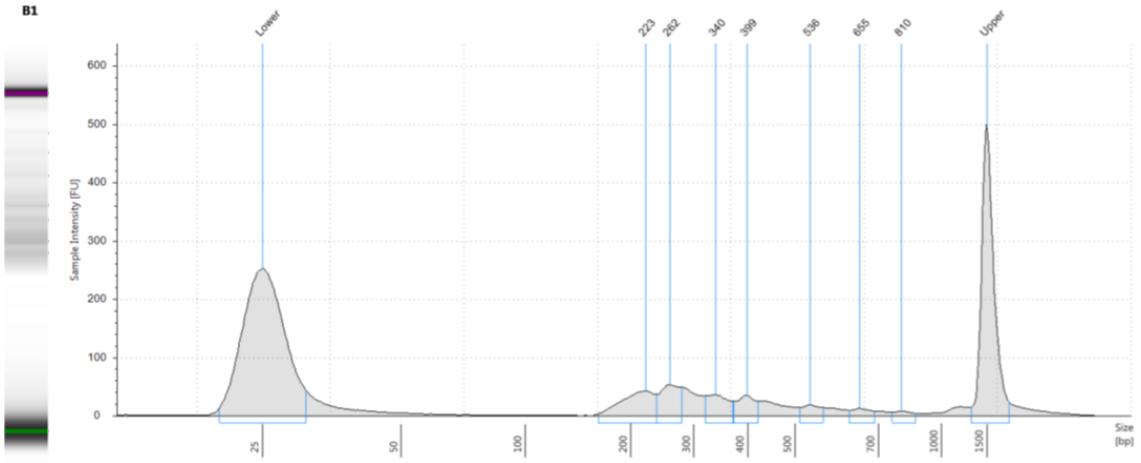


Figure 6.8 – Example of a TapeStation gel electrophoresis and corresponding electropherograms for one representative RRBS library pool.
Each pool was checked using High Sensitivity D1000 Screentapes and 2200 TapeStation System (Agilent Technologies). The area observed between the lower and upper peaks represents the library.

6.2.3. Sequencing

After I prepared the libraries, they were sequenced at the University of Exeter Sequencing Service. The library pool from the pilot experiment was subjected to 150 bp single-end sequencing using a MiSeq sequencer (Illumina). Final library pools (eight pools per mouse model) were distributed across thirty-two HiSeq2500 (Illumina) lanes (16 lanes for samples from each mouse model, with

each pool of eight samples distributed in two lanes) and subjected to 50 bp single-end sequencing, as recommended by others (Gu et al., 2011).

6.2.4. Sequencing data processing

All sequencing data processing was performed on a Unix-based operating system server.

6.2.4.1. Data pre-processing

Initial demultiplexing of the raw sequencing files into *FASTQ* files, contamination checks and methylated spike-in control quality checks were performed in collaboration with Paul O'Neill at the University of Exeter Sequencing Service.

6.2.4.2. Quality control of raw data using FastQC

The randomised *FASTQ* files underwent quality control (QC) assessments using *FastQC* (Andrews, 2010) (version 0.11.7). Read coverage for each sample from the pilot experiment is shown in **Figure 6.9**, and read coverage for all rTg4510 and J20 samples can be seen in **Figure 6.10** and **Figure 6.11**, respectively. A mean coverage of 40.67 (SD = 6.65) million sequencing reads per sample was obtained for the rTg4510 dataset, and a mean coverage of 42.36 (SD = 7.24) million sequencing reads per sample was obtained for the J20 dataset.

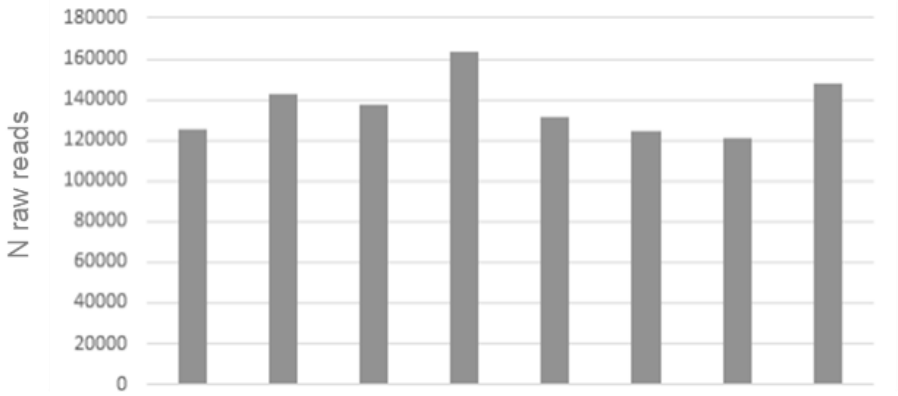


Figure 6.9 – Number of sequencing reads obtained from the pilot experiment using the Premium RRBS kit with modifications.
We observed much less variability between samples in the same pool than before. Each column represents a sample.

A representative example of the quality measures reported graphically by *FastQC* (explained in more detail in **Chapter 4**, **Section 4.2.4.2**) for one sample is shown in **Figure 6.12** to **Figure 6.17**. Quality measures for all samples looked as expected for RRBS data, indicating high quality sequencing data, with all samples passing QC.

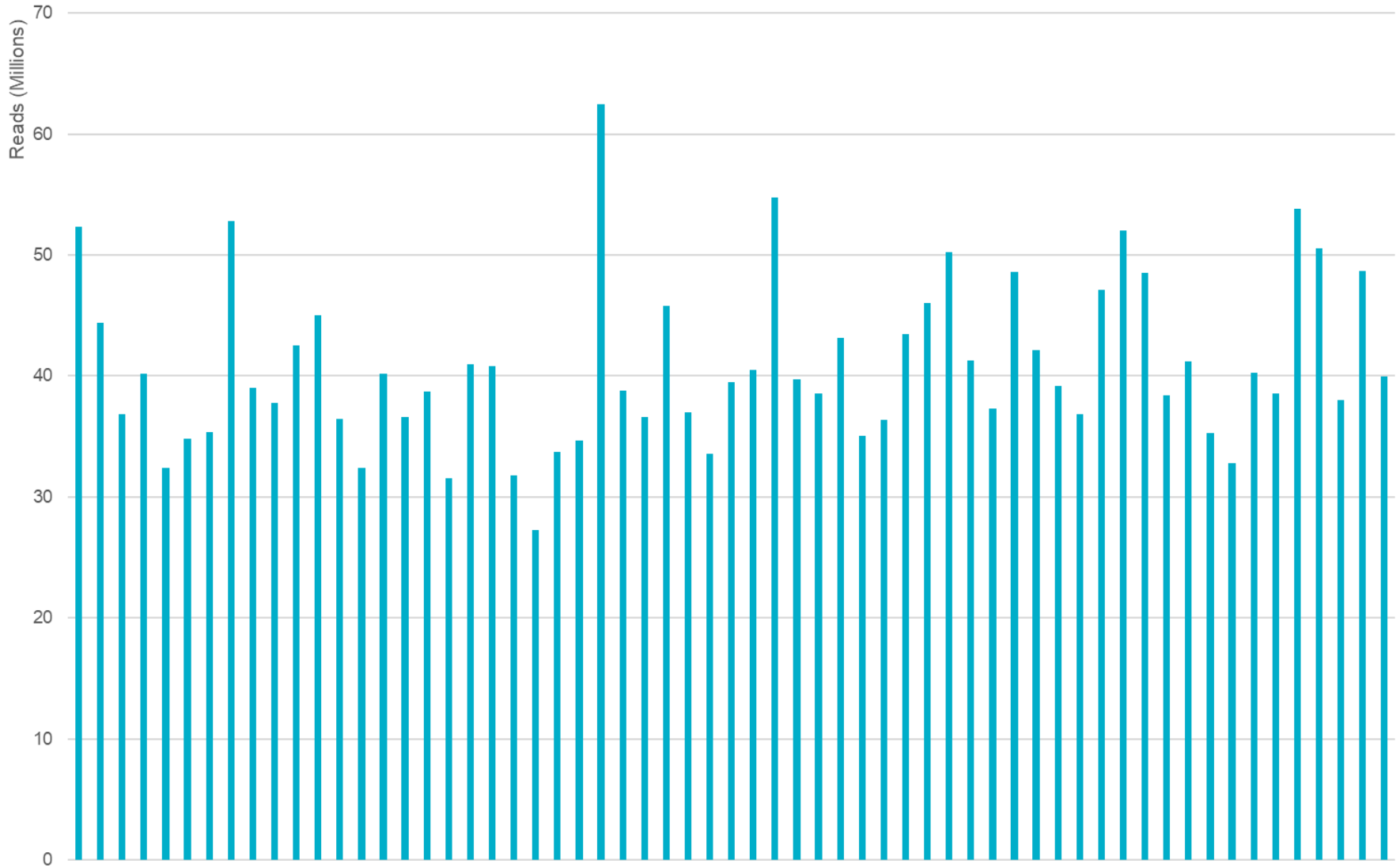


Figure 6.10 – Number of raw RRBS reads for each sample in the rTg4510 dataset.

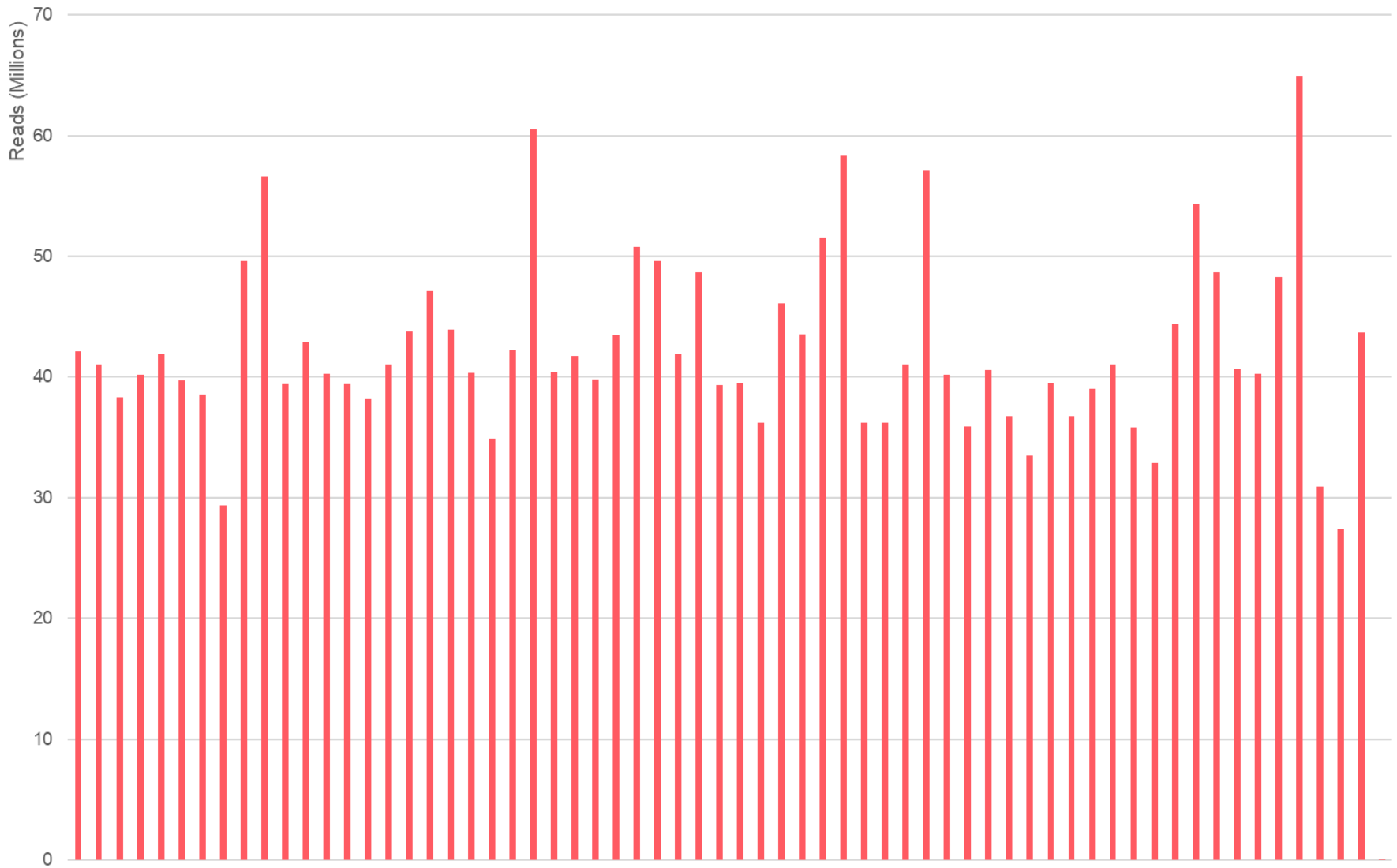


Figure 6.11 – Number of raw RRBS reads for each sample in the J20 dataset.
The last column corresponds to a negative control (containing elution buffer instead of DNA template) included in the experiment, that was removed from further analysist.

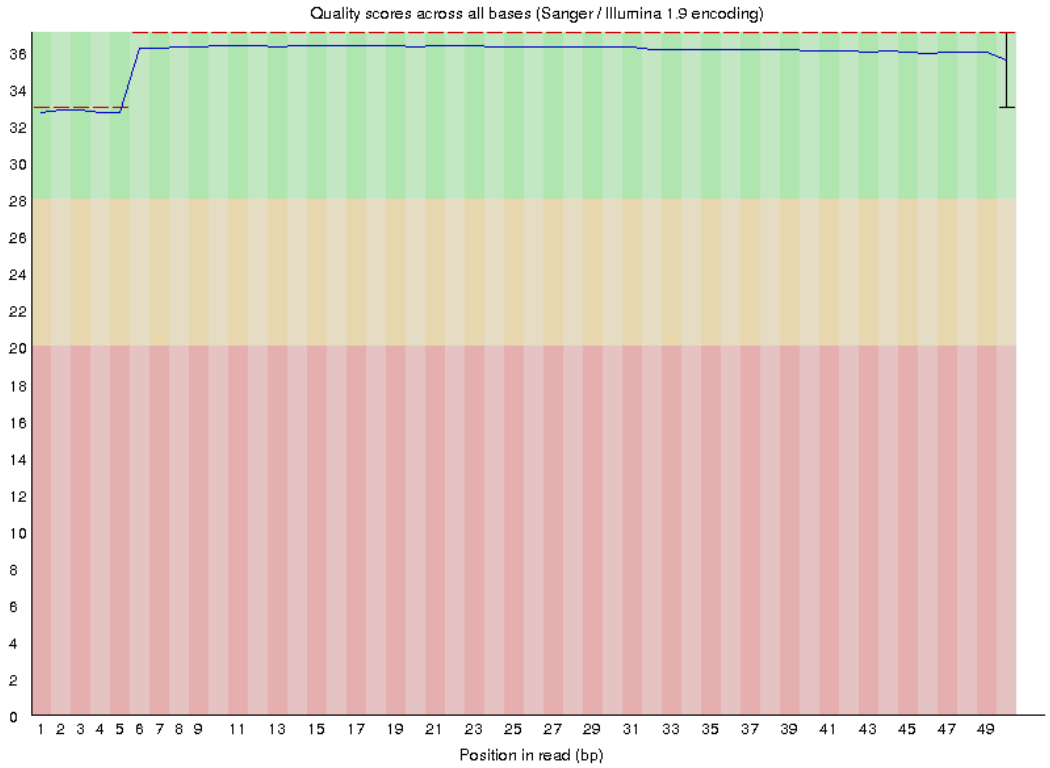


Figure 6.12 – Per base sequence quality from the FastQC report for raw RRBS reads.
Results for one representative sample, showing high quality per base sequencing across the entire read.

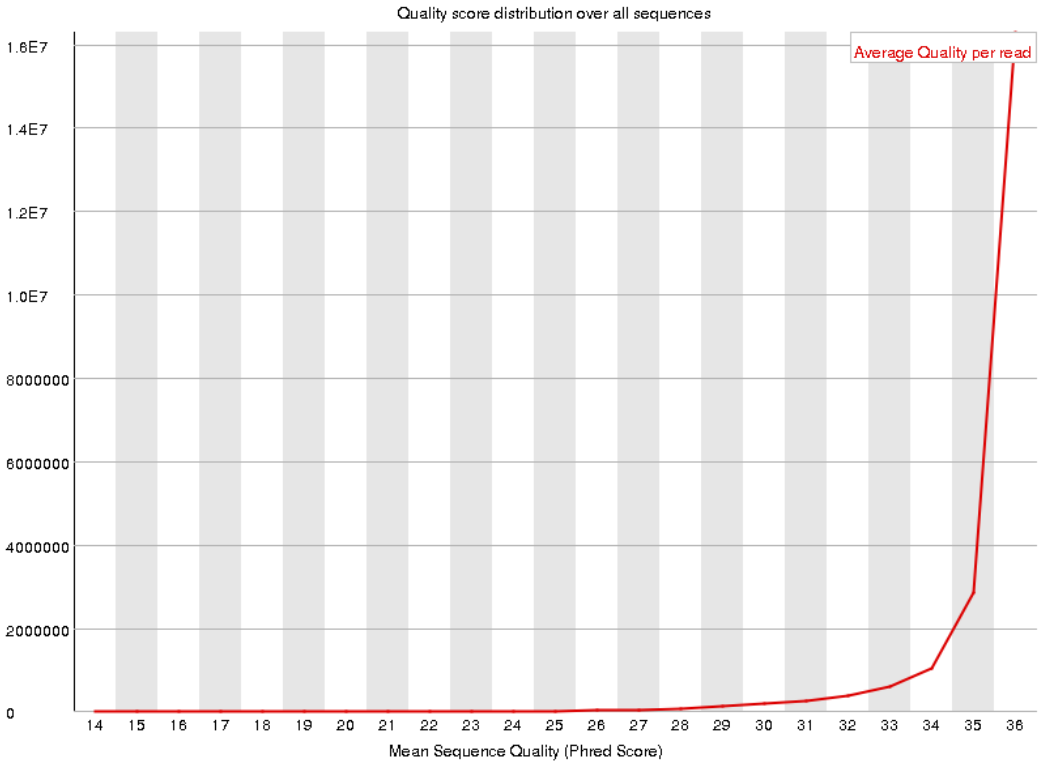


Figure 6.13 – Per sequencing quality scores from the FastQC report for raw RRBS reads.

Results for one representative sample, showing high quality score distribution over all sequences measured by Phred (Q) quality score (Phred score > 30, where a quality score of 30 means that the chances of that base having been called incorrectly are 1 in 1000).

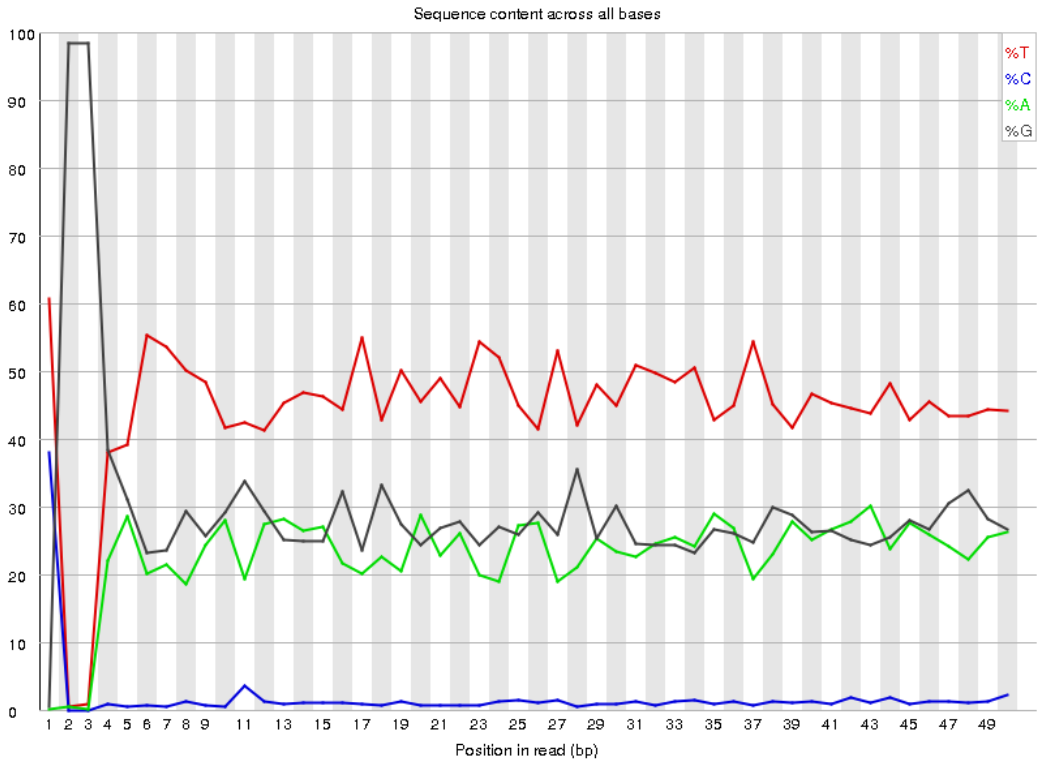


Figure 6.14 – Per base sequence content from the FastQC report for raw RRBS reads.
Results for one representative sample, showing the expected pattern of bases for RRBS as a result of bisulfite conversion.

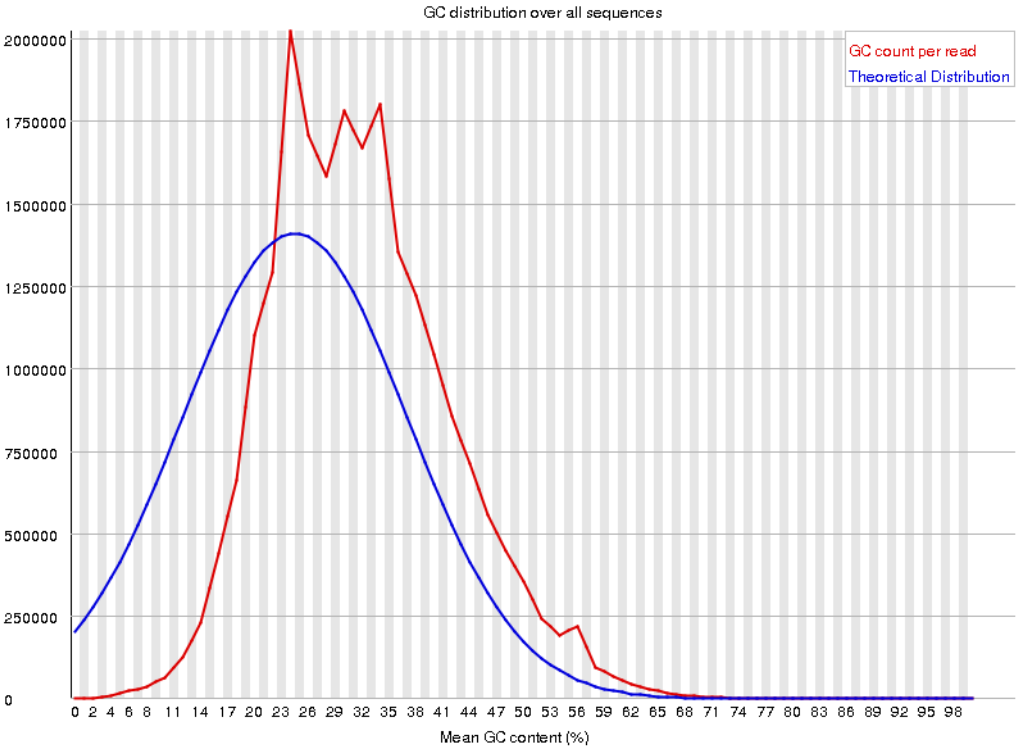


Figure 6.15 – Per sequence GC content from the FastQC report for raw RRBS reads.
Results for one representative sample are shown as an example.

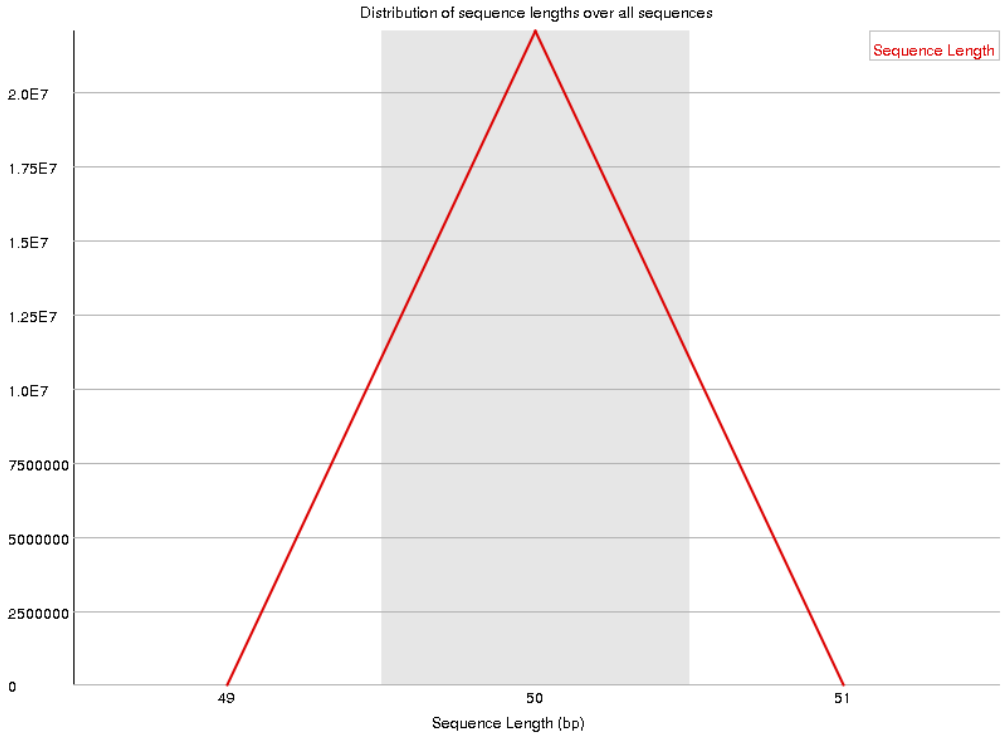


Figure 6.16 – Sequence length distribution from the FastQC report for raw RRBS reads.
Results for one sample are shown as an example.

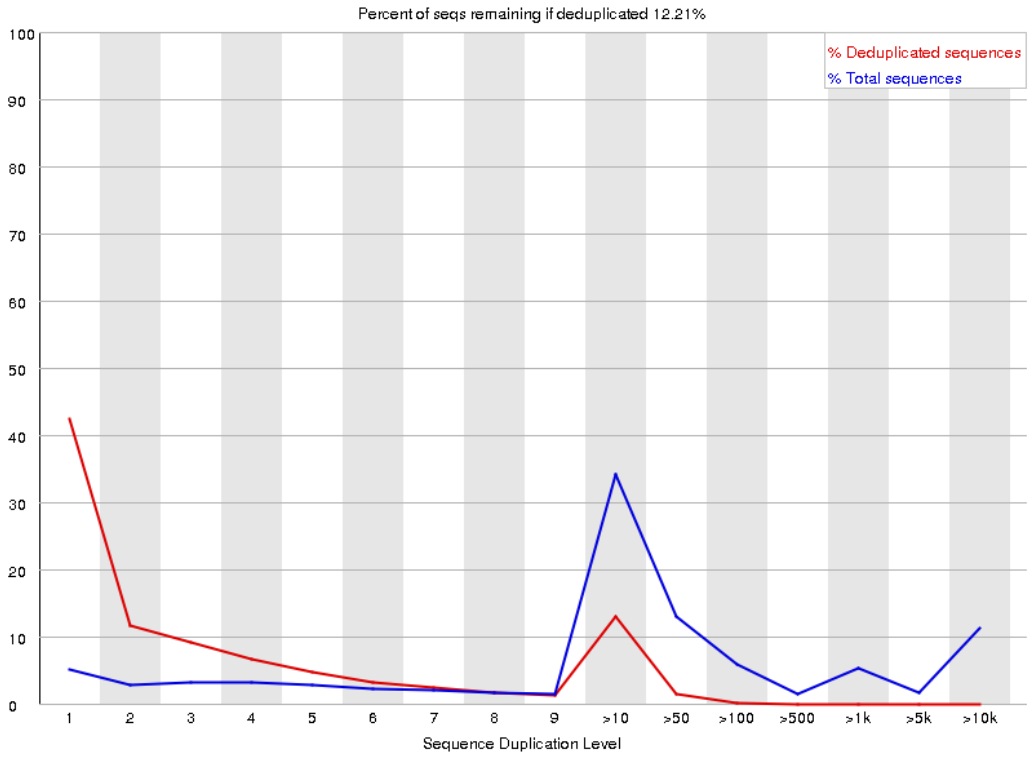


Figure 6.17 – Sequence duplication levels from the FastQC report for raw RRBS reads.
Results for one sample are shown as an example.

6.2.4.3. Trimming and quality filtering

Trimming (quality threshold 20, error rate 0.2) was performed with *Trim Galore* (Krueger, 2017) (version 0.4.5). A Phred score of 20 was used as the quality cut-off value, as previously recommended (Doherty and Couldrey, 2014), which relates to a 1/100 chance of the assigned nucleotide being incorrect. This represents a balance between using high quality data without discarding too many sequencing reads. Trimmed *FASTQ* files underwent QC assessments using *FastQC* (Andrews, 2010) (version 0.11.7) once again, with very little changes in most quality measures for each sample. All samples passed these stringent QC measures. At this stage, however, two samples were removed from the rTg4510 dataset as neuropathology observations using immunohistochemistry (**Chapter 3**) identified these samples to be incorrectly assigned to their experimental groups.

The number of RRBS reads after filtering for all samples is shown in **Figure 6.18** and **Figure 6.19**. A mean output of 40.19 (SD = 6.12) million filtered sequencing reads per sample was obtained for the rTg4510 dataset, and a mean output of 42.20 (SD = 7.23) million filtered sequencing reads per sample was obtained for the J20 dataset. A representative example of the quality measures reported by *FastQC* for the same sample shown above is provided in **Figure 6.20** to **Figure 6.25**.

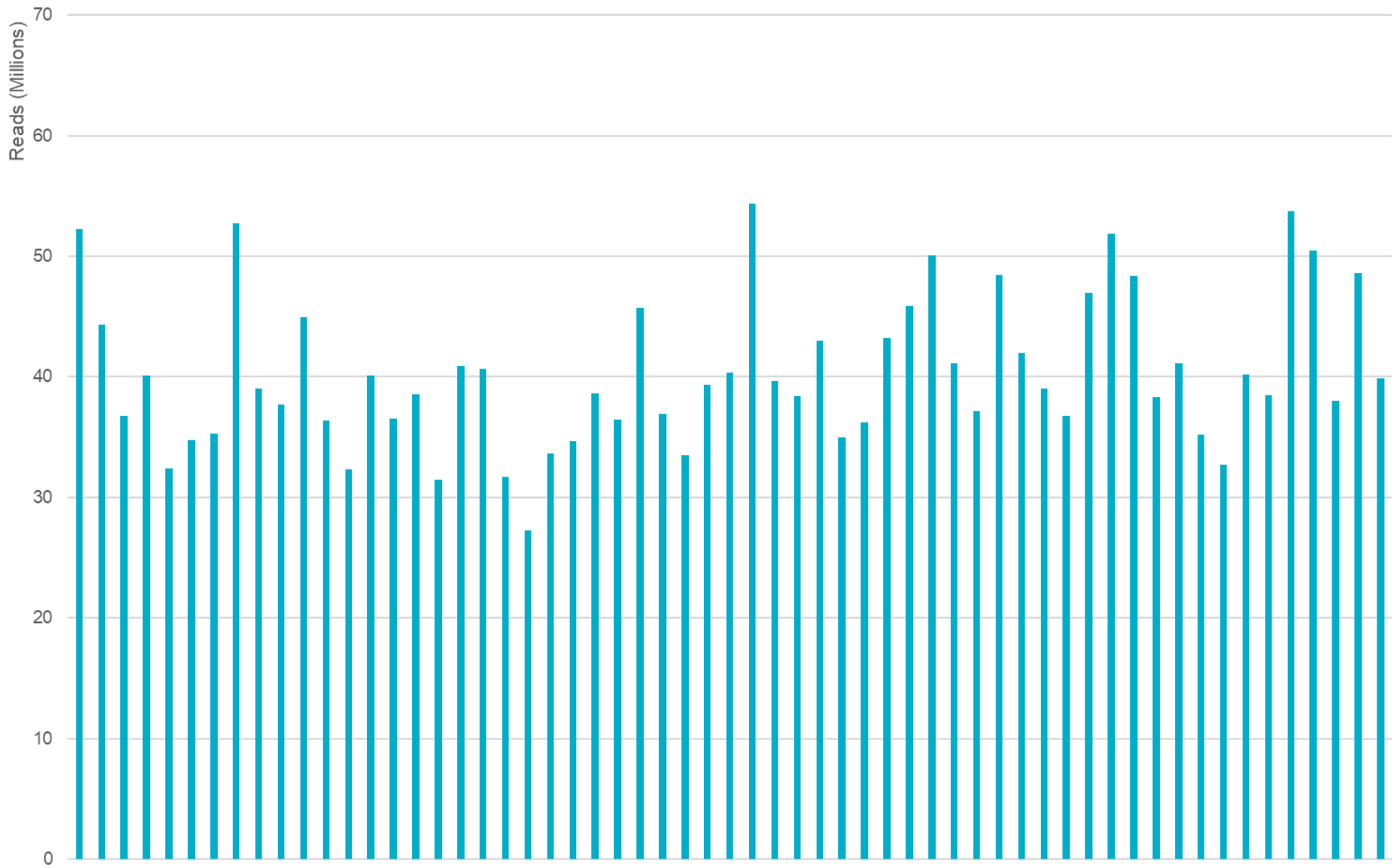


Figure 6.18 – Number of sequencing reads after trimming for each sample in the rTg4510 dataset.

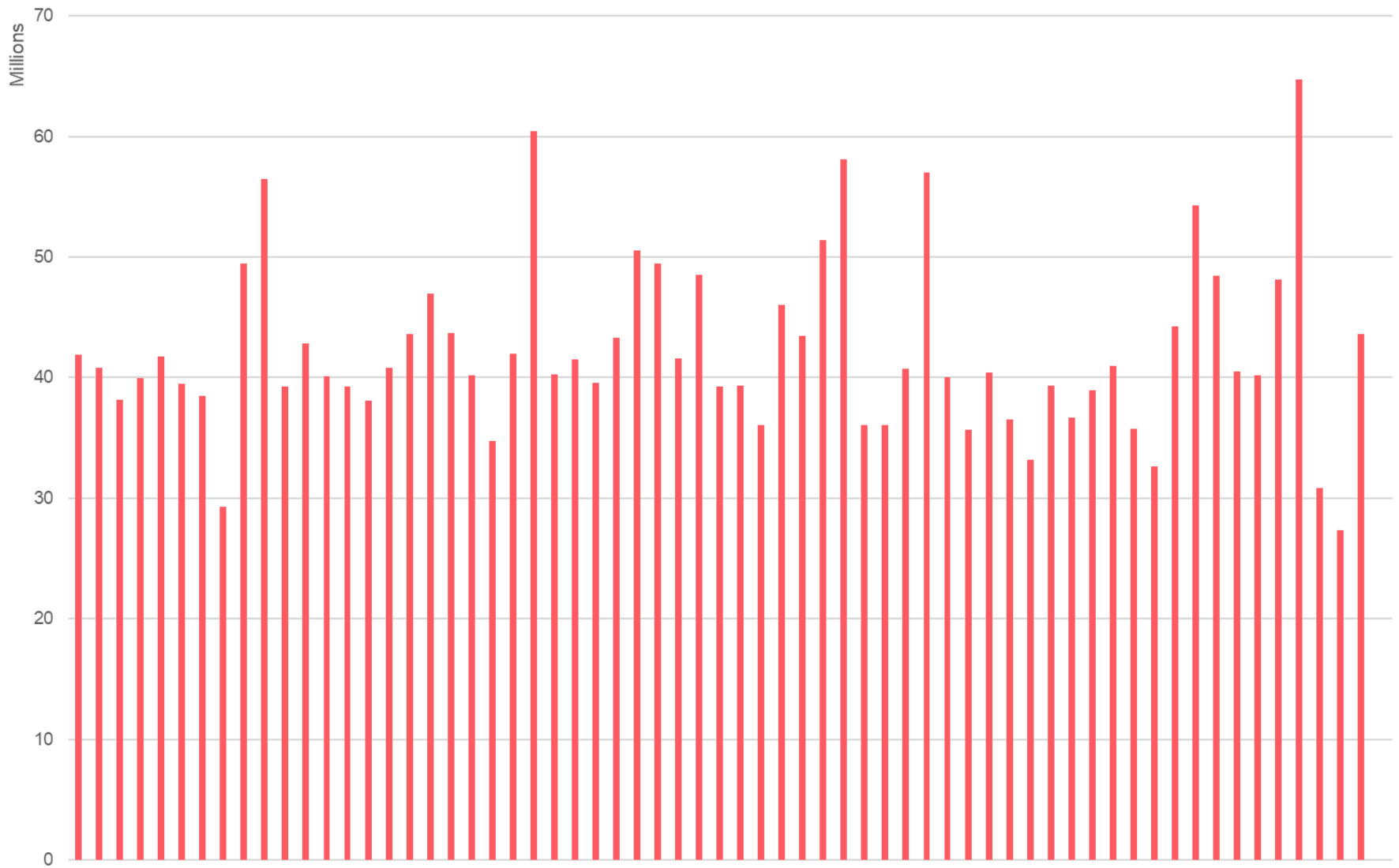


Figure 6.19 – Number of sequencing reads after trimming for each sample in the J20 dataset.

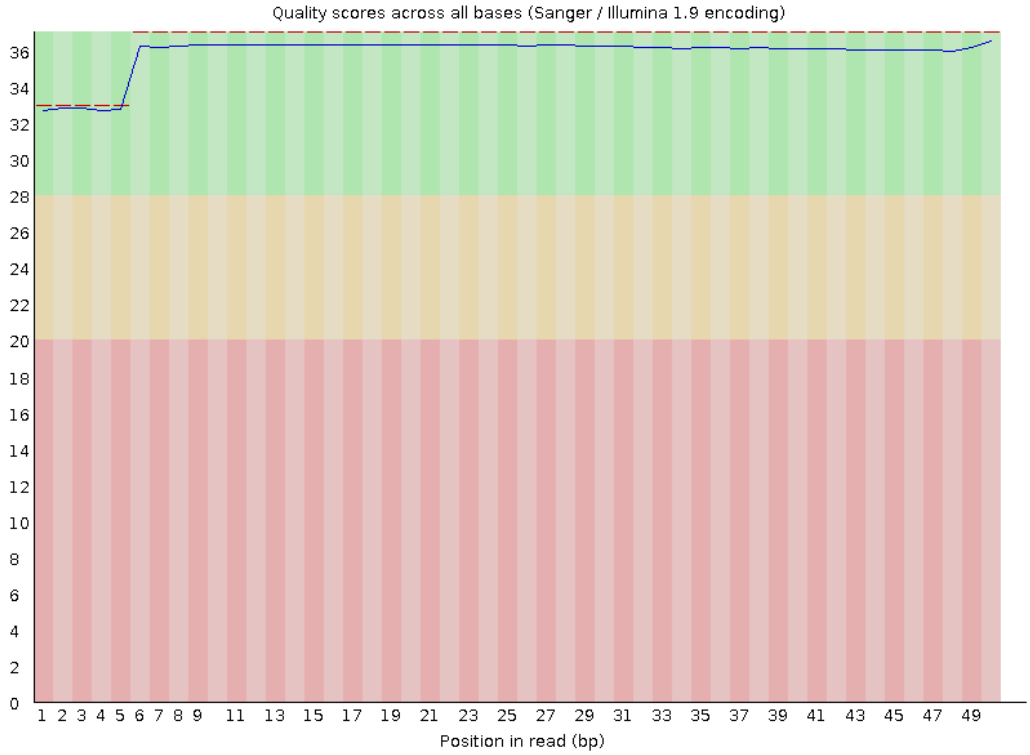


Figure 6.20 – Per base sequence quality from the FastQC report for trimmed RRBS reads.

Results for one sample (same sample as shown in **Figure 6.12**) are shown as an example.

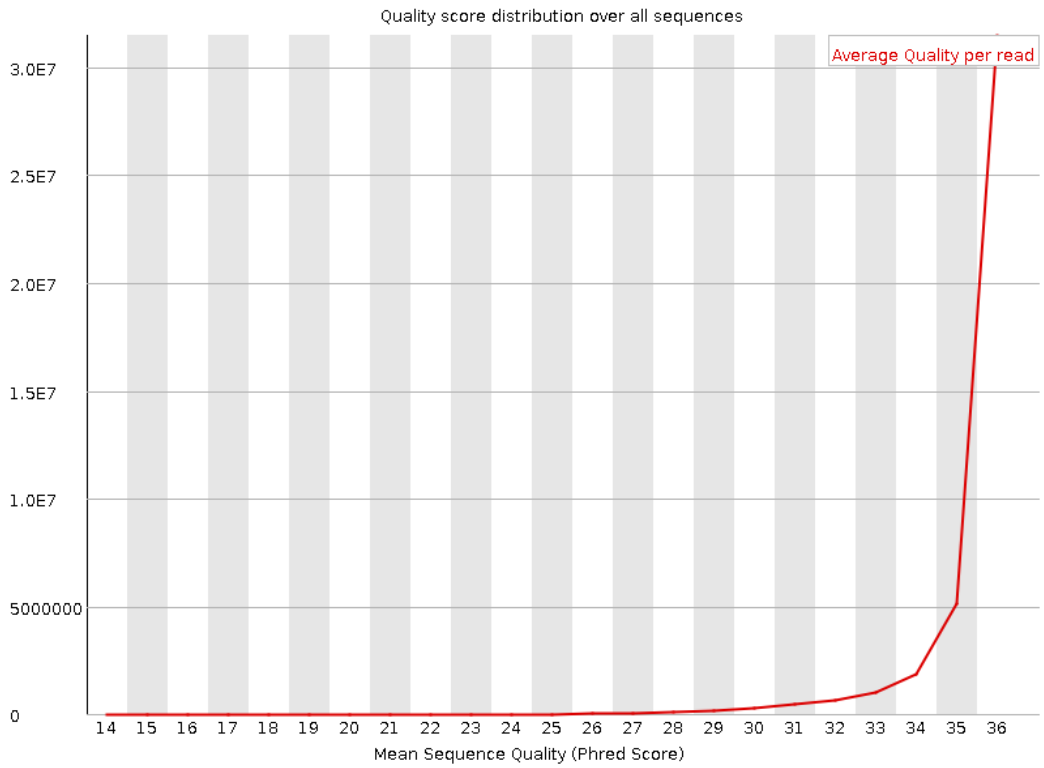


Figure 6.21 - Per sequencing quality scores from the FastQC report for trimmed RRBS reads.

Results for one sample (same sample as shown in **Figure 6.13**) are shown as an example.

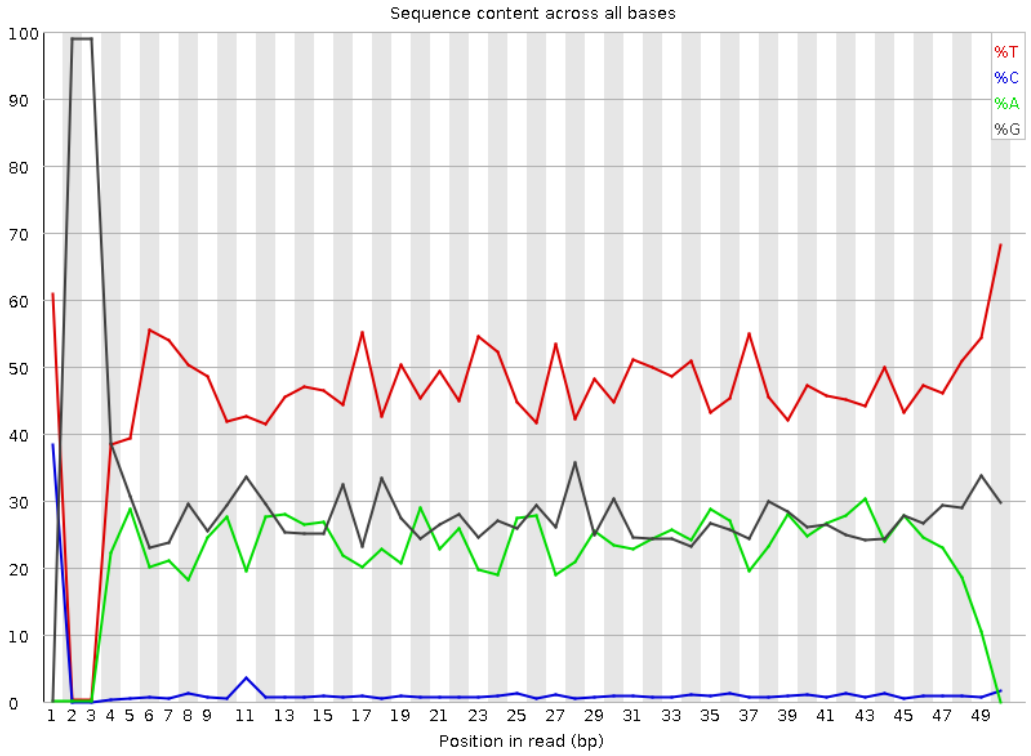


Figure 6.22 – Per base sequence content from the FastQC report for trimmed RRBS reads.

Results for one sample (same sample as shown in **Figure 6.14**) are shown as an example.

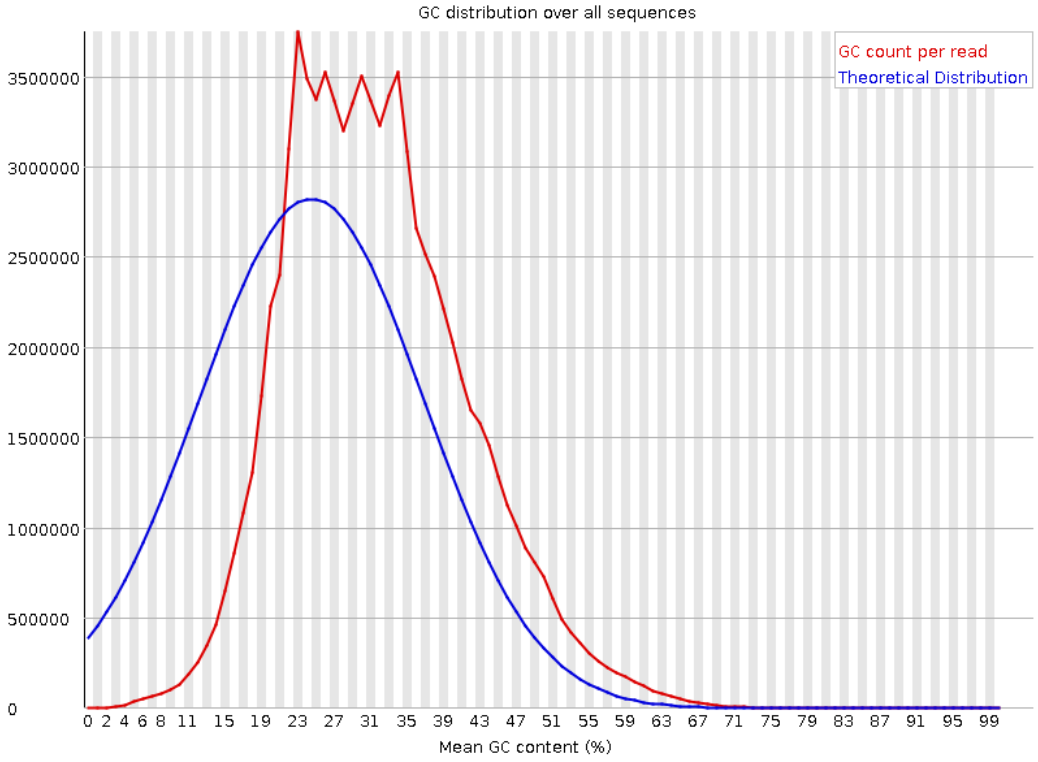


Figure 6.23 – Per sequence GC content from the FastQC report for trimmed RRBS reads.

Results for one sample (same sample as in **Figure 6.15**) are shown as an example.

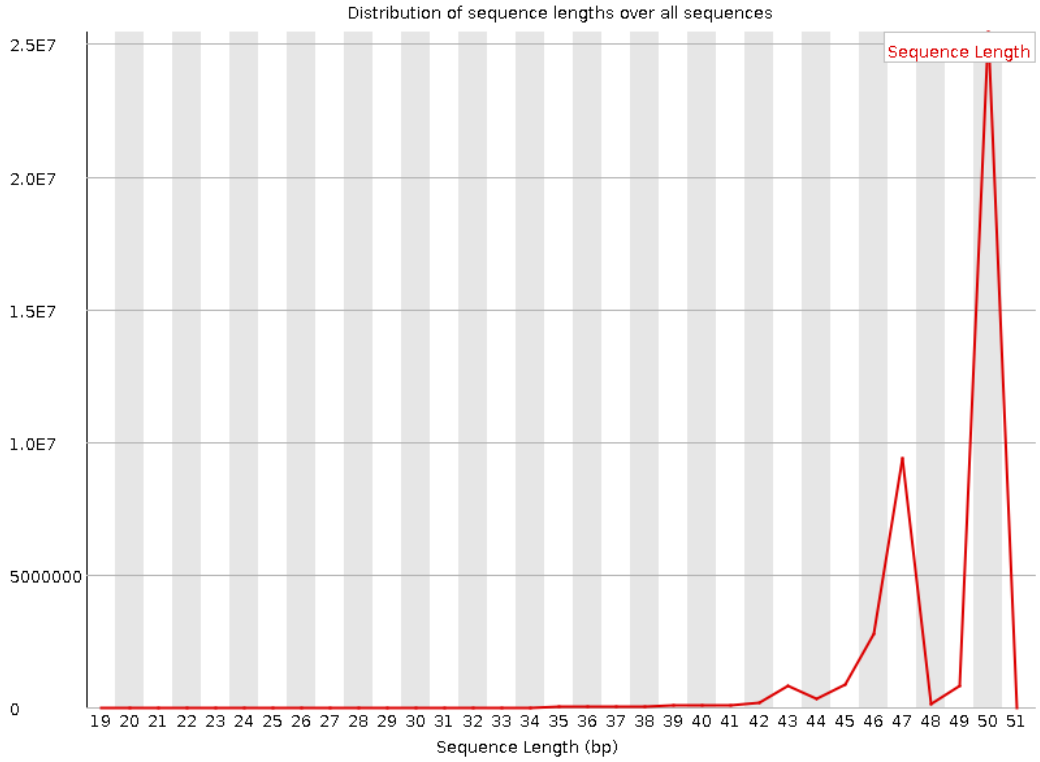


Figure 6.24 – Sequence length distribution from the FastQC report for trimmed RRBS reads.
Results for one sample (same sample as **Figure 6.16**) are shown as an example.

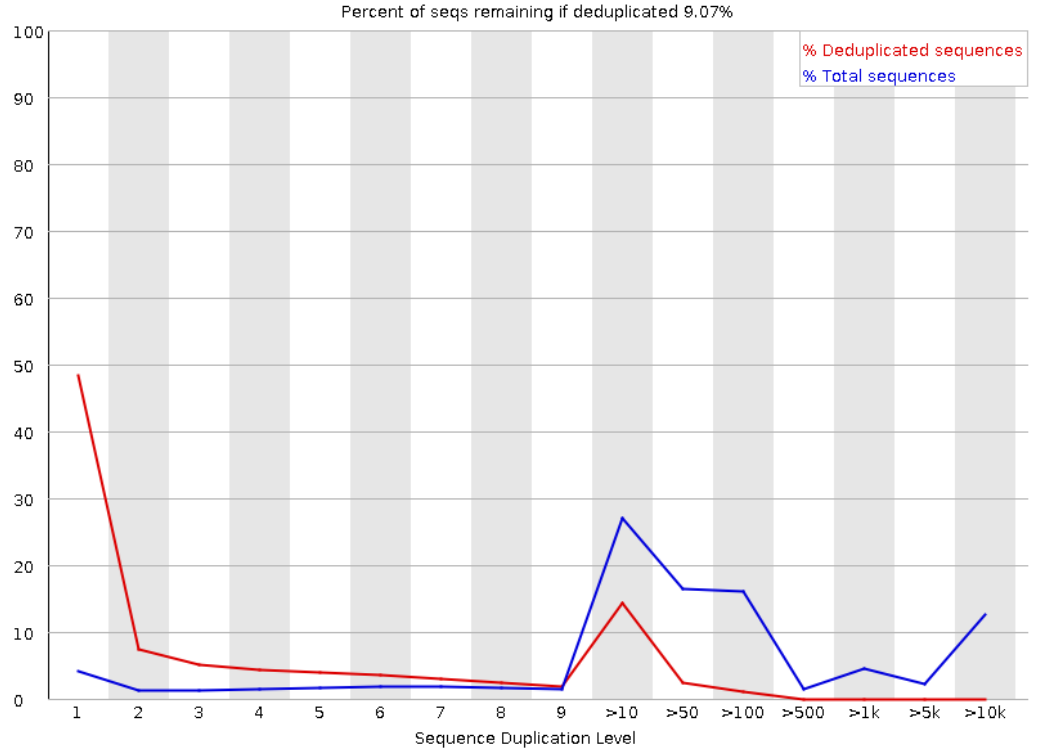


Figure 6.25 – Sequence duplication levels from the FastQC report for trimmed RRBS reads.
Results for one sample (same sample as in **Figure 6.17**) are shown as an example.

6.2.4.4. Alignment to reference genome and methylation calling

Trimmed samples were aligned to the mm10 (GRCm38.p4) reference mouse (*Mus musculus*) genome with *Bismark* (Krueger and Andrews, 2011, Krueger et al., 2012) (version 0.19.0), which uses *Bowtie2* (Langmead et al., 2009) (version 2.3.4.1) and *SAMtools* (Li et al., 2009) (version 1.8). Running *Bismark* involves five individual steps: (1) genome preparation, (2) alignment, (3) methylation extraction, (4) generating a *Bismark* processing report, and (5) generating a *Bismark* summary report. Step 1 includes bisulfite conversion of the genome of interest and indexing, to allow the alignments. In step 2, *Bismark* produces a combined alignment/methylation call output and a run statistics report. In step 3, methylation information is extracted from the *Bismark* alignment output. This methylation information can be split into different genomic contexts (CpG, CHG, and CHH), and can be strand-specific. Optional additional filtering is also possible during this step. In step 4, a graphical *HTML* report is produced for each sample, containing alignment statistics, deduplication, methylation extraction (CpG, CHG and CHH), and a methylation bias plot showing the methylation proportion across each possible position in the read. The *Bismark* summary report from step 5 corresponds to a graphical summary *HTML* report, containing alignment statistics and methylation extraction (CpG, CHG and CHH) for all samples. Both reports were used to further assess the quality of the samples and their respective alignment. Quality measures reported by *Bismark* for one representative sample are shown from **Figure 6.26** to **Figure 6.28**. For all samples, we obtained unique alignments ≥ 60% – a graphical summary for alignment statistics for the rTg4510 and J20 datasets is shown in **Figure 6.29** and **Figure 6.30**, respectively. All ECX samples (rTg4510 and J20, described in **Chapter 2**, sections **2.4** and **2.5**) passed QC at this stage.

Sequences analysed in total	41,880,099
Single-end alignments with a unique best hit	30,573,692
Sequences without alignments under any condition	3,795,509
Sequences that did not map uniquely	7,510,898
Genomic sequence context not extractable (edges of chromosomes)	0

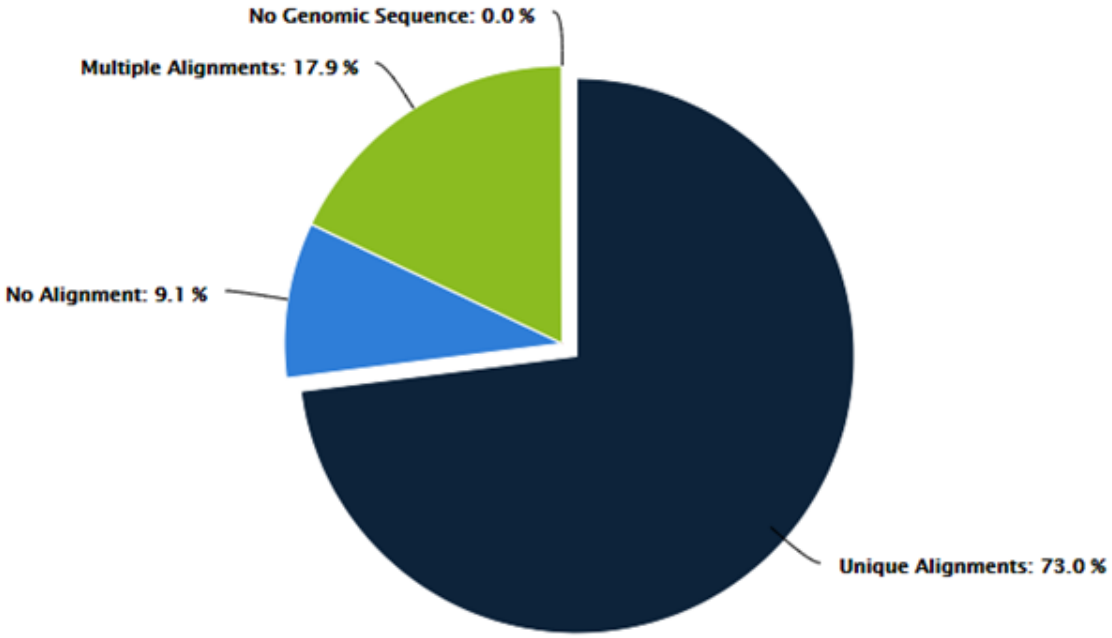


Figure 6.26 – Alignment statistics from Bismark report.
Results for one sample (same sample as shown in the previous figures) are shown as an example.

Total C's analysed	389,838,307
Methylated C's in CpG context	14,882,627
Methylated C's in CHG context	562,670
Methylated C's in CHH context	1,212,961
Unmethylated C's in CpG context	77,568,966
Unmethylated C's in CHG context	87,481,786
Unmethylated C's in CHH context	208,129,297
Percentage methylation (CpG context)	16.1%
Percentage methylation (CHG context)	0.6%
Percentage methylation (CHH context)	0.6%

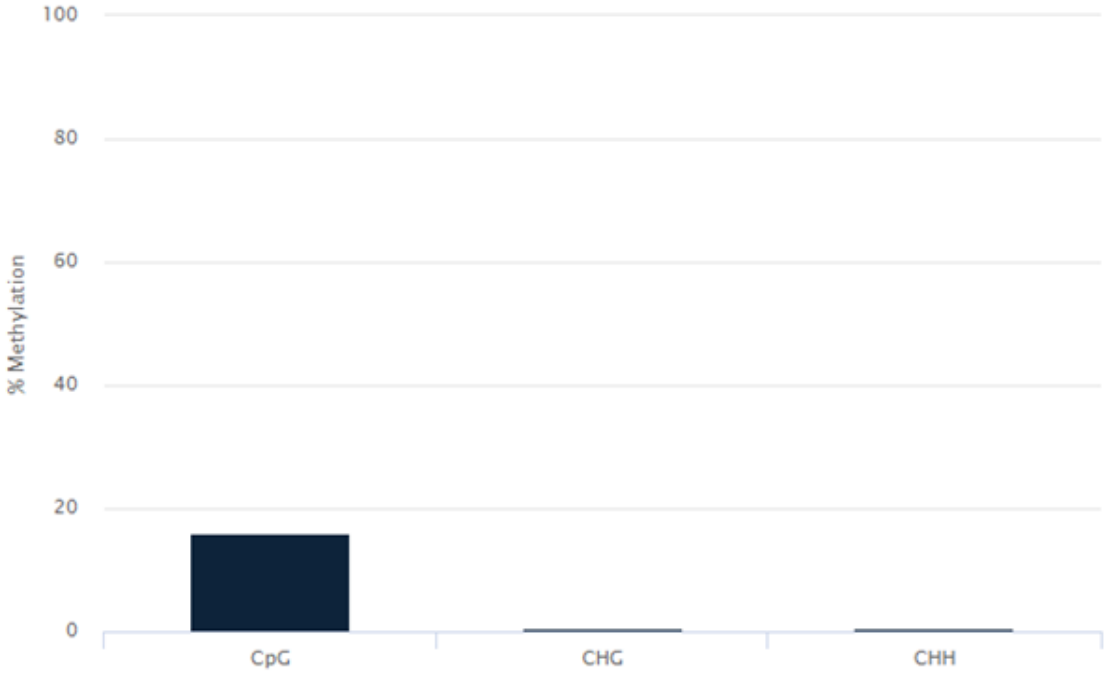


Figure 6.27 – Bismark report statistics for cytosine methylation after extraction. Results for one sample (same sample as shown in the previous figures) are shown as an example.

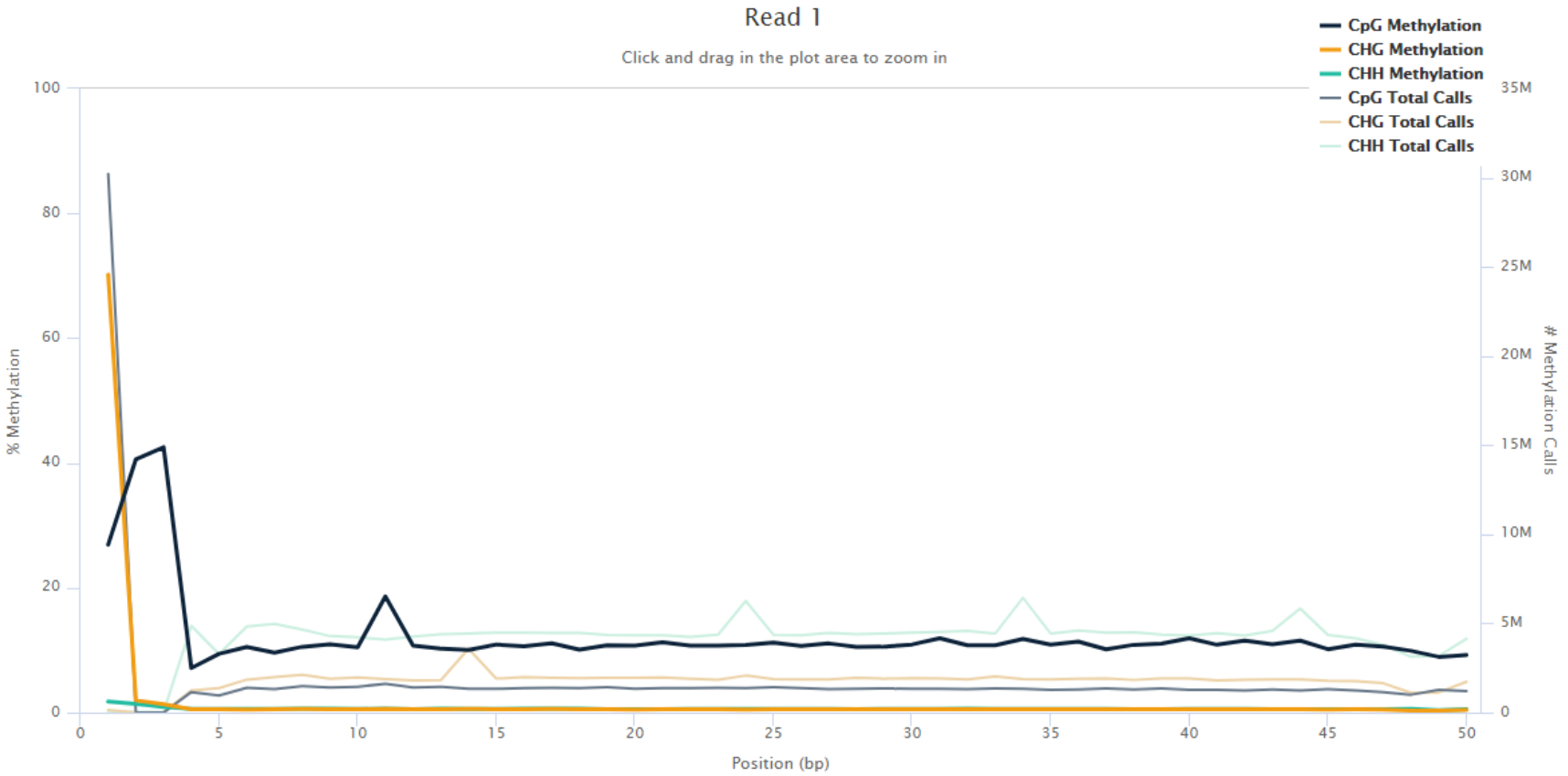


Figure 6.28 – M-bias plot from Bismark report, showing methylation levels averaged per position in the reads.
Results for one sample (same sample as shown in the previous figures) are shown as an example. Left axis: methylation (%) for CpG (dark blue), CHG (dark yellow), and CHH (green). Right axis: Total number of methylation calls for CpG (grey), CHG (light yellow), and CHH (light aqua green).

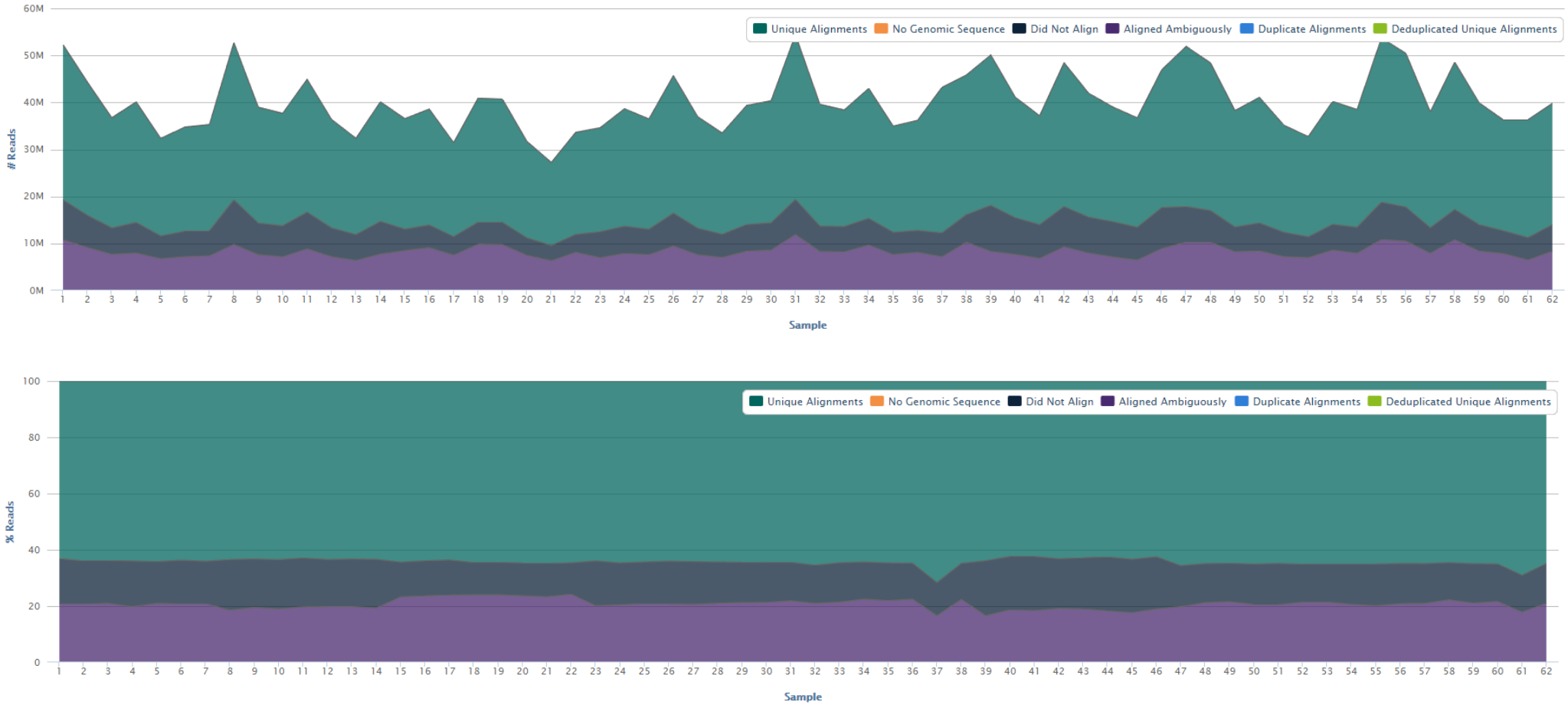


Figure 6.29 – Alignment statistics from Bismark summary report for rTg4510 samples.
The top panel shows the total number of reads per sample, and the panel on the bottom shows the percentage of reads per sample. Green corresponds to unique alignments; dark blue represents reads that did not align; and purple corresponds to reads that aligned ambiguously.

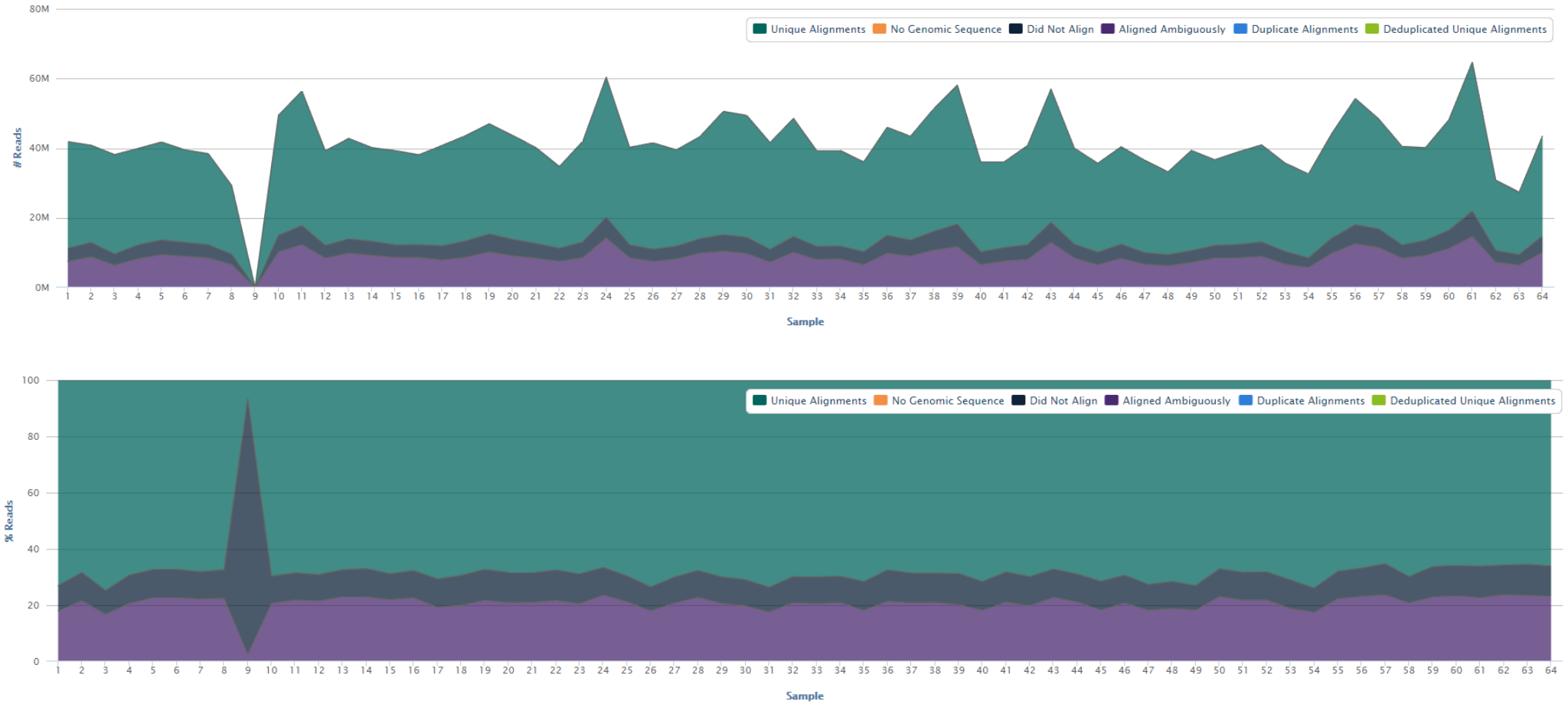


Figure 6.30 – Alignment statistics from Bismark summary report for J20 samples.
The top panel shows the total number of reads per sample, and the panel on the bottom shows the percentage of reads per sample. Green corresponds to unique alignments; dark blue represents reads that did not align; and purple corresponds to reads that aligned ambiguously. The sample in the 9th position refers to a negative control (containing elution buffer instead of DNA template) included in the experiment.

6.2.5. Differential DNA methylation analysis

All analyses were performed in R (version 3.4.3) unless otherwise stated. Dorothea Seiler Vellame (University of Exeter) provided assistance with data filtering (**Section 6.2.5.1**) and initial steps of the statistical analysis (**Section 6.2.5.2**).

6.2.5.1. Data filtering

RRBS data was filtered to exclude sites with 1) low read depth (i.e. those with a minimum of 10 reads per DNA methylation site) and 2) low standard deviation in DNA methylation values (because DNA methylation at most cytosine sites is invariable, sites that did not have a standard deviation greater than 5% of methylation were removed). Sites with DNA methylation data for fewer than five samples within each genotype (WT/TG) group were removed to ensure high quality data and maximise statistical power. The distribution of DNA methylation values across all sites for each sample was inspected by plotting the distribution of the data before and after filtering (**Figure 6.31**), revealing the expected binomial distribution of DNA methylation, with the majority of sites being highly or lowly methylated.

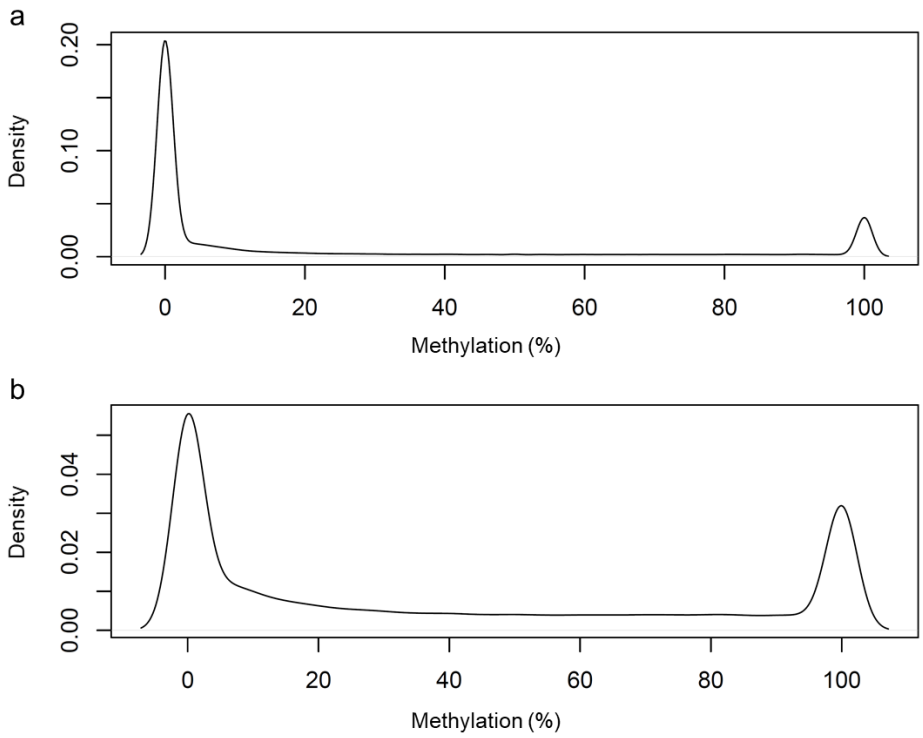


Figure 6.31 – Methylation distribution.
Distribution of methylation (%) for one sample (same representative sample shown from **Figure 6.19** to **Figure 6.28**) before **(a)** and after **(b)** filtering is shown as an example.

6.2.5.1. Statistical analysis

Given the complexity of DNA methylation data, the relatively small individual group sizes, and the large number of DNA methylation sites across the genome profiled by RRBS, I decided to focus my analyses on identifying genotype-associated differentially methylated sites. To achieve this, a t-test was used (using the R function *anova*), comparing all transgenic (TG) mice to all wild type (WT) controls for each DNA methylation site in both mouse models (rTg4510 (n = 62) and J20 (n = 63)) datasets. *P* values were adjusted for multiple testing, using the false discovery rate (FDR) method (also known as Benjamini and Hochberg correction (Benjamini and Hochberg, 1995)) implemented with the R function *p.adjust*; FDR-values < 0.05 were defined as 'significant'. Sites were annotated to the nearest gene with the R package *Granges* using the *biomaRt* (Durinck et al., 2005, Durinck et al., 2009) gene annotation database from *Ensembl* for mouse mm10 (GRCm38.p4) gene sets. *Enrichr* (Chen et al., 2013, Kuleshov et al., 2016) was used for gene ontology (GO) analysis of genes annotated to significant (FDR < 0.05) differentially methylated positions (DMPs).

6.3. Results

6.3.1. Sequencing metrics

There was no significant (P > 0.05) difference in RRBS read-depth between TG and WT controls for either the rTg4510 (**Table 6.1** and **Figure 6.32**) or J20 (**Table 6.2** and **Figure 6.33**) datasets.

Table 6.1 – Statistical results for RRBS read-depth in rTg4510 mice.

Mean ± SD (million reads)		t-value	P-value	N
WT	TG			
40.20 ± 7.13	40.38 ± 5.09	t(61) = 0.12	0.91	62

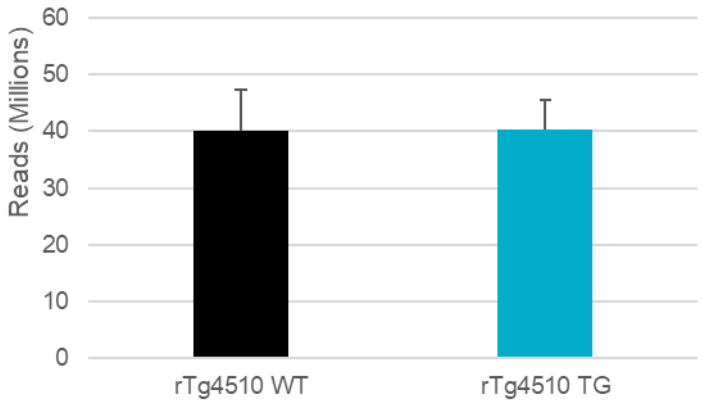


Figure 6.32 –Average number of raw RRBS reads for rTg4510 samples.
The number of raw sequencing reads did not differ between WT and TG mice (n = 62 animals, two-tailed unpaired t-test, t(61)= 0.12, P = 0.91). Bar plots represent mean and SD.

Table 6.2 – Statistical results for RRBS read-depth in J20 mice.

Mean ± SD (million reads)		t-value	P-value	N
WT	TG			
43.05 ± 7.92	41.66 ± 6.51	t(62) = 0.76	0.45	63

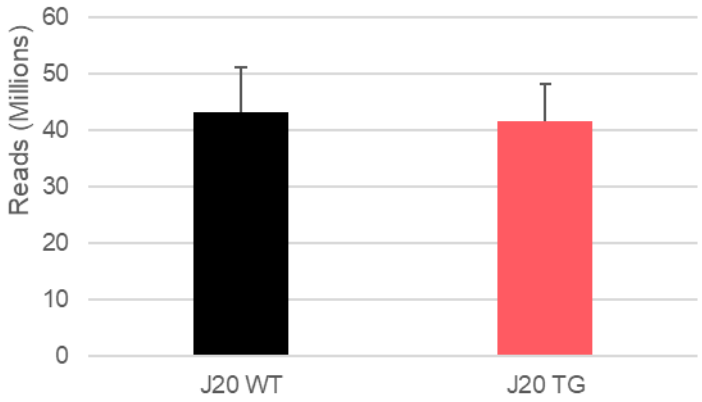


Figure 6.33 – Average number of raw RRBS reads for J20 samples.
The number of raw sequencing reads did not differ between WT and TG mice (n = 63 animals, two-tailed unpaired t-test, t(62)= -0.76, P = 0.45). Bar plots represent mean and SD.

6.3.2. Genotype-associated epigenetic wide methylation changes

In total, our filtered RRBS datasets included DNA methylation values for 1,066,467 sites in the rTg4510 model and 1,145,020 sites in the J20 model.

In the rTg4510 dataset, 20,070 (1.88%) sites were characterised by genotype-associated differential DNA methylation (FDR < 0.05) (**Figure 6.34**). Clusters of DMPs, seen as peaks (highlighted by pink arrows) in the Manhattan plot depicted in **Figure 6.34**, were annotated to major prion protein (*Prnp*) and protein tyrosine phosphatase, receptor type, N polypeptide 2 (*Ptprn2*). These differences likely reflect the effect of the transgene insertion and parallel the results of our RNA-seq analysis (see **Chapter 4**, **Section 4.3.2**). We therefore decided to remove all DMPs that could potentially reflect direct effects of the transgene insertion before running further analysis, with all sites annotated to regions spanning *Wdr60*, *Esyt2*, *Ncapg2*, *Ptprn2* or *Fgf14* being removed from further analyses. After

filtering for potentially transgene-associated sites, 20,023 (1.87%) sites were found to be differentially methylated between TG and WT mice (FDR <0.05). A filtered version of the Manhattan plot for rTg4510 genotype-associated methylation changes is shown in **Figure 6.35**; of note, we still observe differentially methylation in the proximity of the transgene-disrupted genes (black arrow in **Figure 6.35**), potentially reflecting cis effects on DNA methylation of the inserted transgene. Quantile-quantile (QQ) plots were used to assess the normality of distribution of test statistics from rTg4510 samples (**Figure 6.37** and **Figure 6.38**). Quantification of the extent of P-value inflation and the excess false positive rate was achieved by calculating the genomic inflation factor (lambda gc or λgc) (Yang et al., 2011) from FDR values. Λgc, which corresponds to the ratio of the median of the empirically observed distribution of the test statistic to the expected median (i.e., the median of the resulting chi-squared test statistics divided by the expected median of the chi-squared distribution), indicated very limited inflation in both the unfiltered (λgc = 1.05) and filtered data (λgc = 1.04) suggesting no major stratification in my results.

In the J20 dataset, 16,473 (1.44%) of profiled sites were found to be differentially methylated (FDR < 0.05) between TG and WT mice (**Figure 6.36**). A quantile-quantile plot to evaluate the normality of distribution of methylation data from J20 samples is shown in **Figure 6.39**; calculation of the genomic inflation factor from FDR values again indicated very limited P-value inflation (λgc = 0.92).

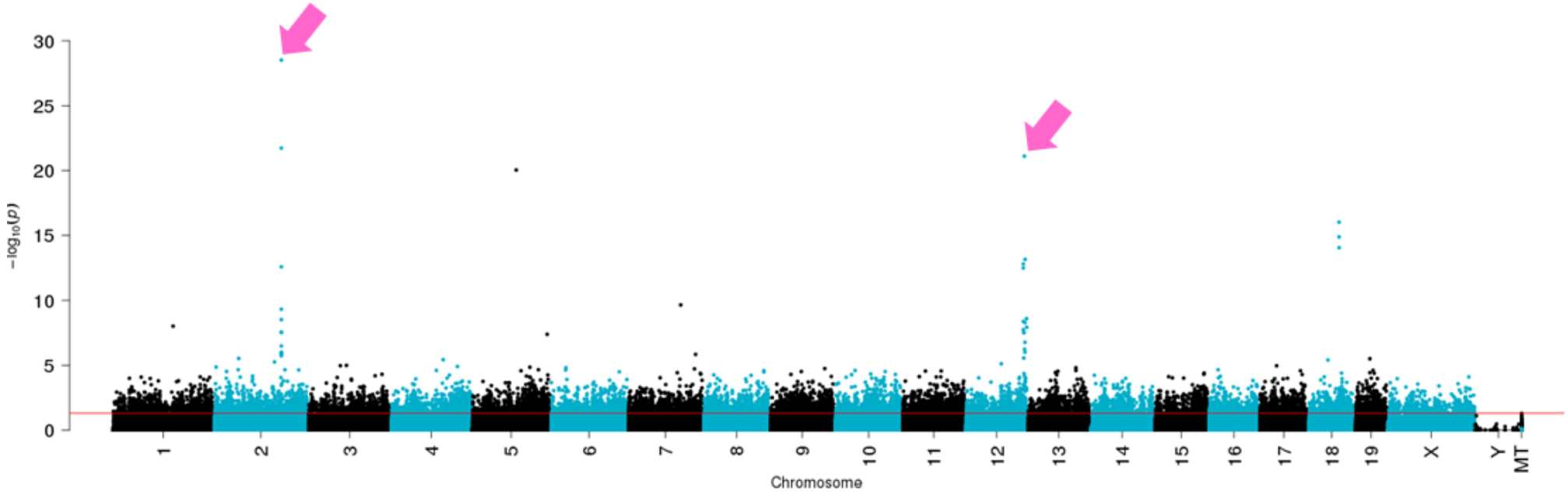


Figure 6.34 – Manhattan plot showing rTg4510 genotype-associated epigenetic wide methylation changes.
Pink arrows highlight changes likely to reflect transgene integration sites. N = 1,066,467 sites, 62 animals (WT = 31, TG = 31). The red line corresponds to FDR = 0.05.

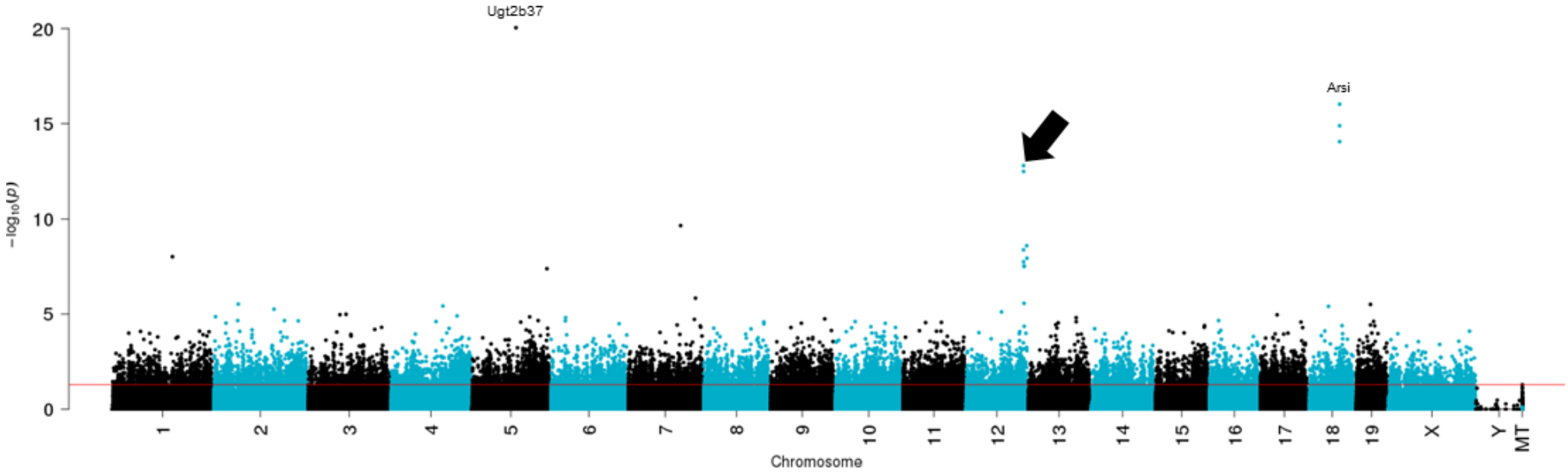


Figure 6.35 – Manhattan plot showing rTg4510 genotype-associated epigenetic wide methylation changes after removal of sites potentially reflecting transgene integration.

The two top differentially methylated sites are highlighted by the name of the gene that they were annotated to. The black arrow indicates a peak of methylation changes in several sites in the proximity of the genes disrupted by the insertion of the transgene. N = 1,066,420 sites, 62 animals (WT = 31, TG = 31). The red line corresponds to FDR = 0.05.

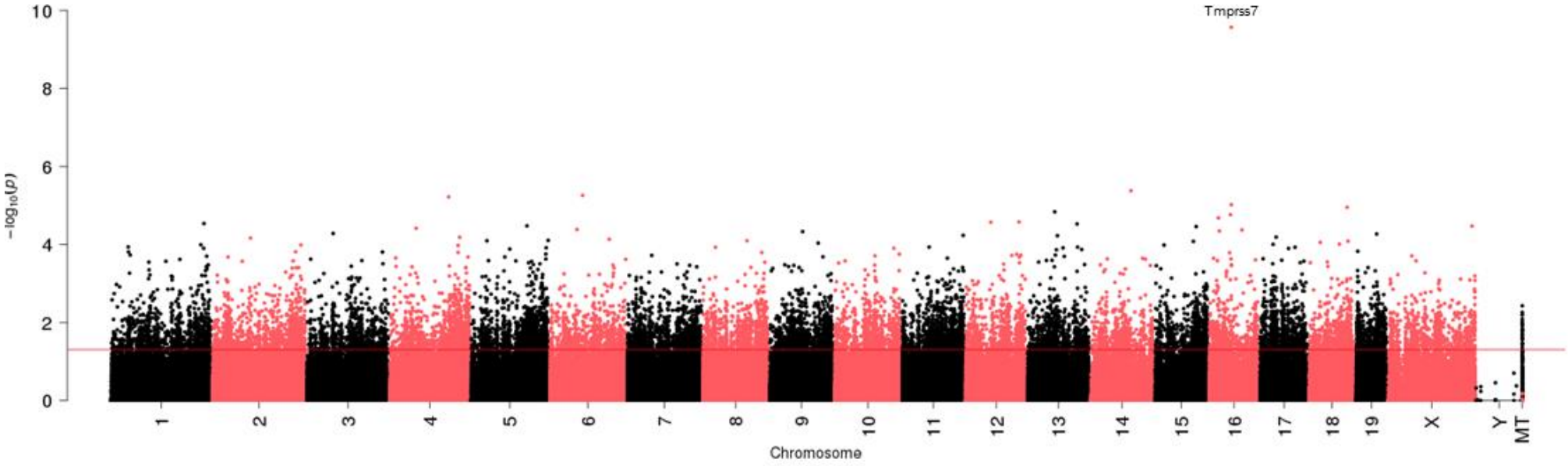


Figure 6.36 – Manhattan plot showing J20 genotype-associated epigenetic wide methylation changes.
The top differentially methylated site is highlighted by the name of the gene that it was annotated to N = 1,145,020 sites, 63 animals (WT = 32, TG = 31). The red line corresponds to FDR = 0.05.

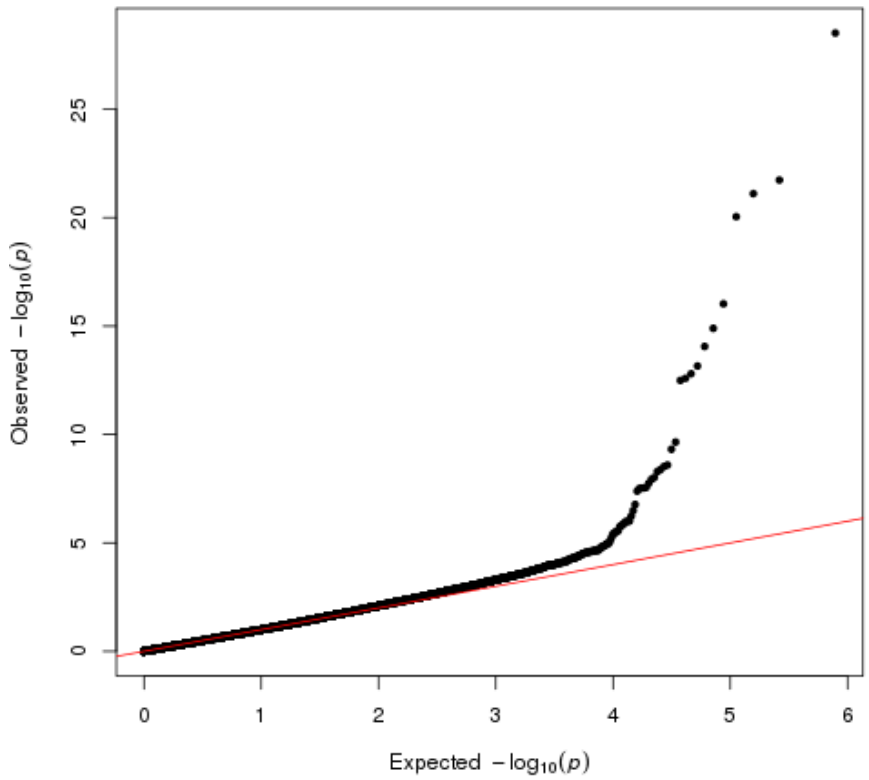


Figure 6.37 – Quantile-quantile (Q-Q) plot for rTg4510 RRBS samples before removal of transgene-associated sites.
N = 1,066,467 sites, 62 animals (WT = 31, TG = 31).

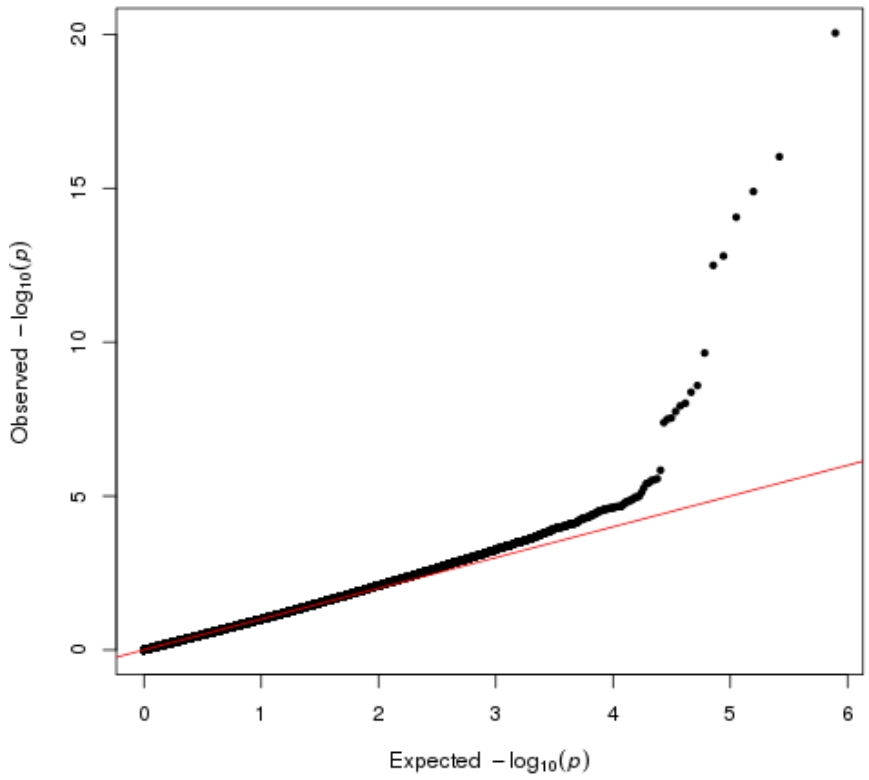


Figure 6.38 – Quantile-quantile (Q-Q) plot for rTg4510 RRBS samples after removal of transgene-associated sites.
N = 1,066,420 sites, 62 animals (WT = 31, TG = 31).

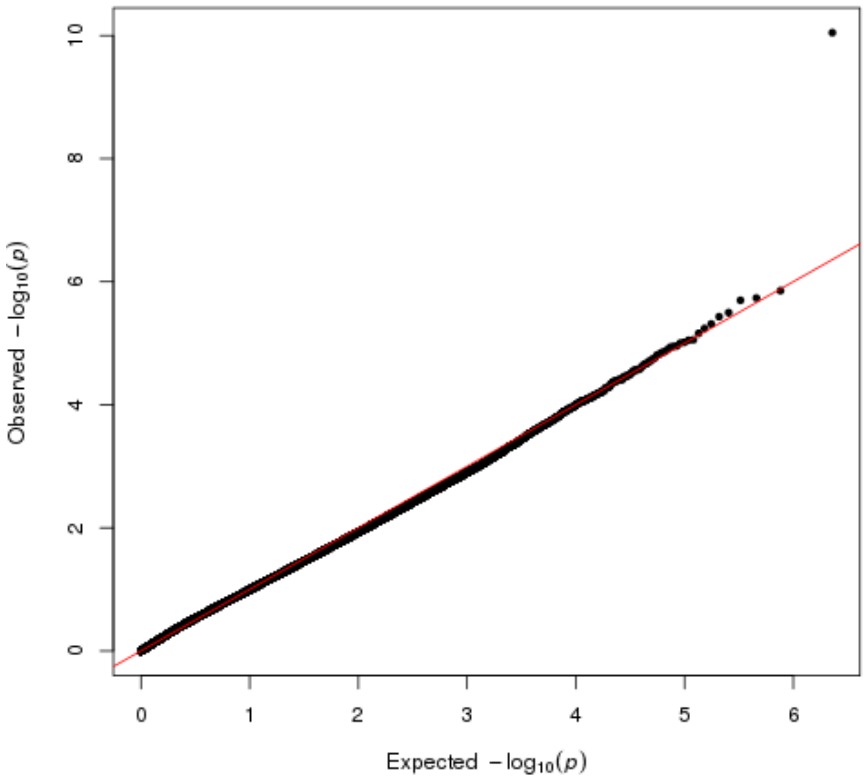


Figure 6.39 – Quantile-quantile (Q-Q) plot for J20 RRBS samples.
N = 1,066,467 sites, 63 animals (WT = 32, TG = 31).

6.3.3. DNA methylation at genotype-associated sites robustly distinguishes between TG and WT mice for both rTg4510 and J20 models

Across sites with DNA methylation data for all samples, hierarchical clustering of individual mouse samples based on methylation levels for genotype-associated sites was found to robustly discriminate between TG and WT groups for both Tg4510 (**Figure 6.40**) and J20 (**Figure 6.41**) models. In contrast to clustering between genotype groups using RNA-seq data (**Chapter 4**) – where samples also clustered by age within the rTg4510 TG group (**Figure 4.39**) – no such temporal clustering was seen for DMPs suggesting that temporal changes in DNA methylation are less pronounced. In rTg4510, there was a significant (exact binomial test, n = 20,023 DMPs, $P \leq$ 4.94E-324) enrichment of hypermethylated DMPs (n = 15,558 (77%) DMPs showing increased methylation in TG versus WT mice, compared to n = 4,465 (33%) of hypomethylated DMPs). In J20 mice, there was also a significant (exact binomial test, n = 16,473 DMPs, $P \leq$ 4.94E-324) enrichment of hypermethylated DMPs (n = 11,672 (71%) of hypermethylated DMPs, compared to n= 4,801 (29%) of hypomethylated DMPs, in TG compared to WT mice).

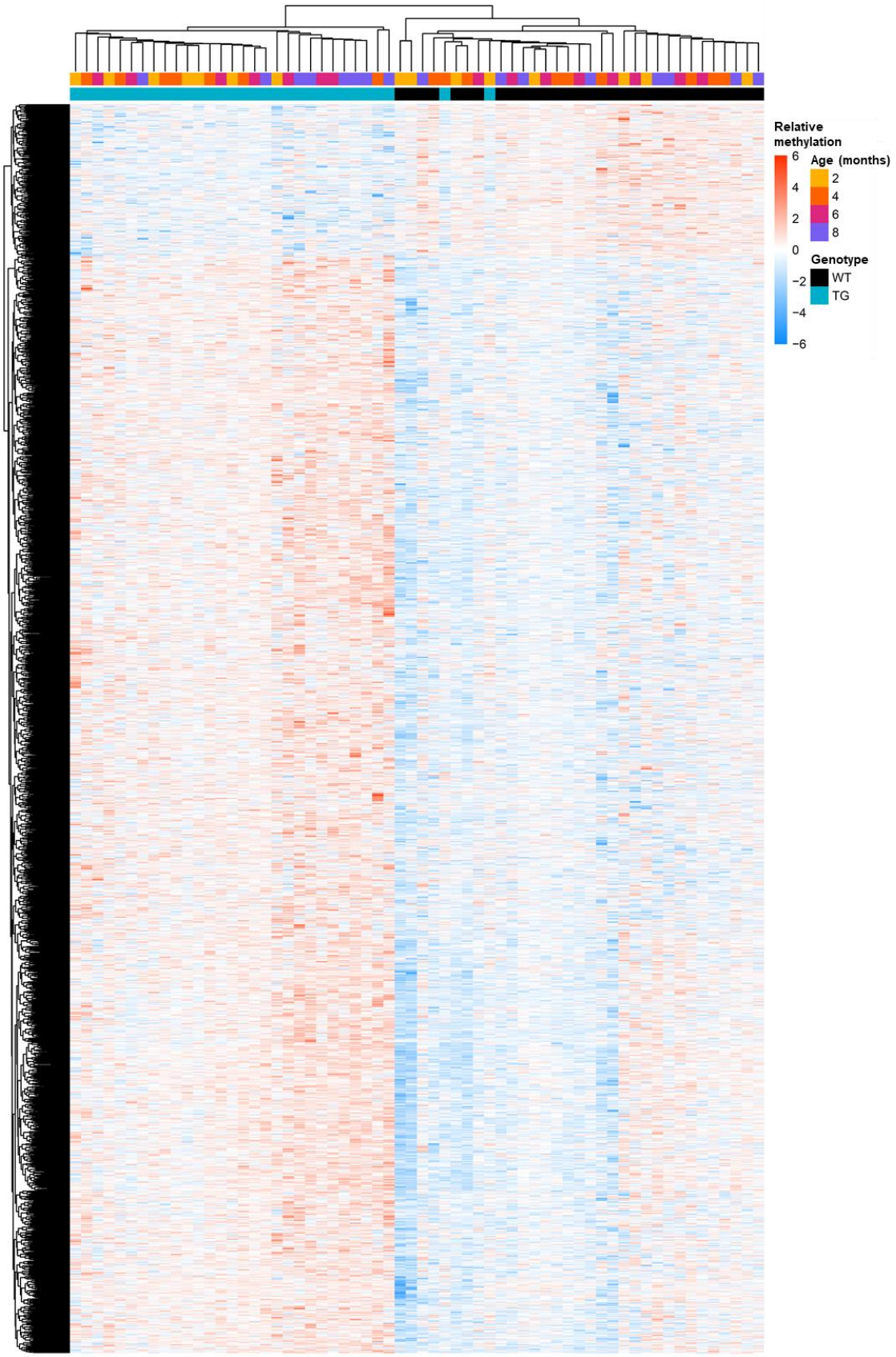


Figure 6.40 – Genotype-associated DNA methylation variation robustly discriminates between rTg4510 transgenic (TG) and wild type (WT) mice.
N= 3,625 sites, 62 animals (WT = 31, TG = 31). Direction of methylation (%), relative to mean levels of methylation across all individual mice ("relative methylation"), is represented in the heatmap (scaled) from high (red) to low (blue).

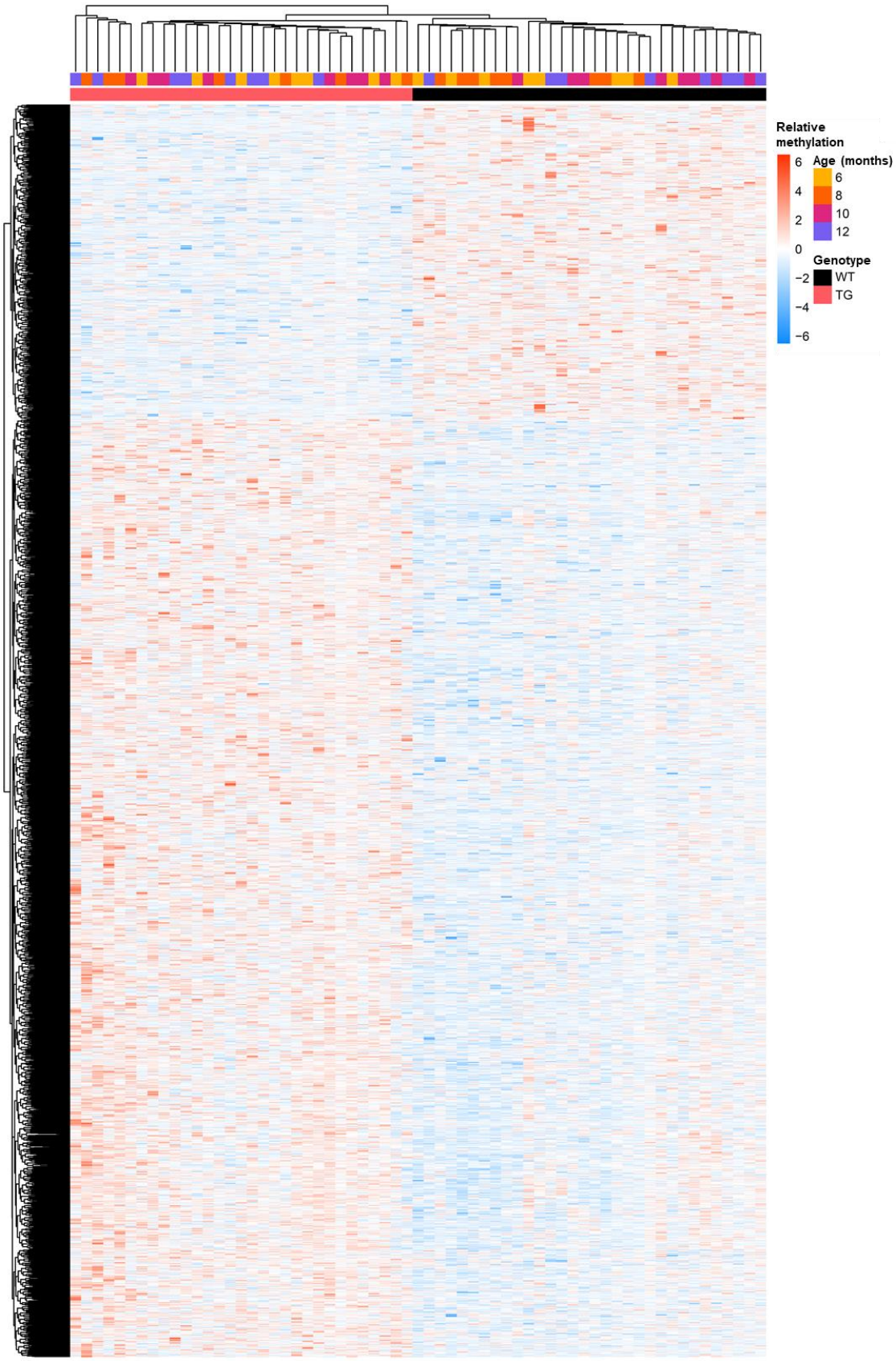


Figure 6.41 – Genotype-associated DNA methylation variation robustly discriminates between J20 transgenic (TG) and wild type (WT) mice.
N= 3,226 sites, 63 animals (WT = 32, TG = 31). Direction of methylation (%), relative to mean levels of methylation across all individual mice ("relative methylation"), is represented in the heatmap (scaled) from high (red) to low (blue).

6.3.4. Top genotype-associated differentially methylated sites

The top rTg450 genotype-associated DMP was annotated to *Ugt2b37 (***Figure 6.35***,* **Table 6.3** and **Figure 6.42a**, hypermethylated in TG mice compared to WT), which encodes for the member 37 of UDP glucuronosyltransferases (UGTs) subfamily 2B (UGT2B37). UGTs are cytosolic glycosyltransferases that catalyse glucuronidation (i.e., the transfer of the glucuronic acid component of UDP-glucuronic acid to a small hydrophobic molecule) and have a role in drug metabolism and detoxification (Kurita et al., 2017, Rowland et al., 2013). UGT2B37 in particular is involved in the metabolism of key endogenous compounds including bilirubin, bile acids, fatty acids, steroid hormones, thyroid hormones and fat soluble vitamins, and in glucuronidation of testosterone and dihydrotestosterone (Friedman et al., 2019, Zhou et al., 2015). *Ugt2b37* is highly expressed in the adult mammalian kidney and liver, with only moderate expression in the brain when compared to kidney and liver (Bohmdorfer et al., 2017, Sakamoto et al., 2015, Zhang et al., 2017). *Ugt2b37* has been shown to be expressed in neurons (Friedman et al., 2019), however, information about its expression in other brain cell types is still unavailable. Three DMPs were annotated to *Arsi* (**Figure 6.35**, **Table 6.3** and **Figure 6.42b-d**, all hypomethylated in TG mice compared to WT). This gene encodes arylsulfatase family member I (ARSI), which belongs to a large family of sulfatases involved in hormone biosynthesis, the modulation of cell signalling, and the degradation of macromolecules (Sardiello et al., 2005). Several sites located in chromosome 12 were found to be differentially methylated (black arrow in **Figure 6.35**), including three sites annotated to *Tmem212* (**Table 6.3** and **Figure 6.42e-g**, hypermethylated in TG compared to WT mice) that encodes transmembrane protein 212, and one site annotated to *Ighe* (**Table 6.3** and **Figure 6.42h**, hypermethylated in TG compared to WT) which encodes immunoglobulin heavy constant epsilon (constant region of heavy chain of immunoglobulin epsilon or IgE). As discussed in **Section 6.3.2**, these DMPs on chromosome 12 potentially reflect the insertion of the transgene and should be treated with caution.

Table 6.3 – Statistical results for top rTg4510 genotype-associated sites.
Sites are ordered by FDR. Δβ refers to difference in methylation (%) between TG and WT. Underlined sites are plotted in **Figure 6.42**.

Location	Nearest gene	t-value	Δβ	FDR	N
chr5:87231917	*Ugt2b37*	t(59) = 224.50	51.82	9.05E-21	58
chr18:60917719	*Arsi*	t(55) = 150.25	-47.78	9.44E-17	56
chr18:60917731	*Arsi*	t(55) = 131.59	-34.98	1.28E-15	56
chr18:60917761	*Arsi*	t(55) = 118.92	-40.60	8.74E-15	56
chr12:113190326	*Tmem121*	t(58) = 98.15	43.63	1.59E-13	59
chr12:113266015	*Ighe*	t(28) = 188.52	78.98	3.22E-13	29
chr7:103488891	*Olfr609*	t(37) = 82.59	-58.02	2.25E-10	38
chr12:119325190	*Gm6768*	t(48) = 59.74	36.32	2.57E-09	47
chr12:113199742	*Tmem121*	t(52) = 54.33	12.40	1.41E-09	53
chr1:118279473	*Tsn*	t(51) = 51.44	44.90	9.68E-09	52
chr12:119752693	*Gm25675*	t(57) = 48.56	33.06	1.17E-08	58
chr12:113190240	*Tmem121*	t(39) = 55.60	42.18	1.80E-08	40

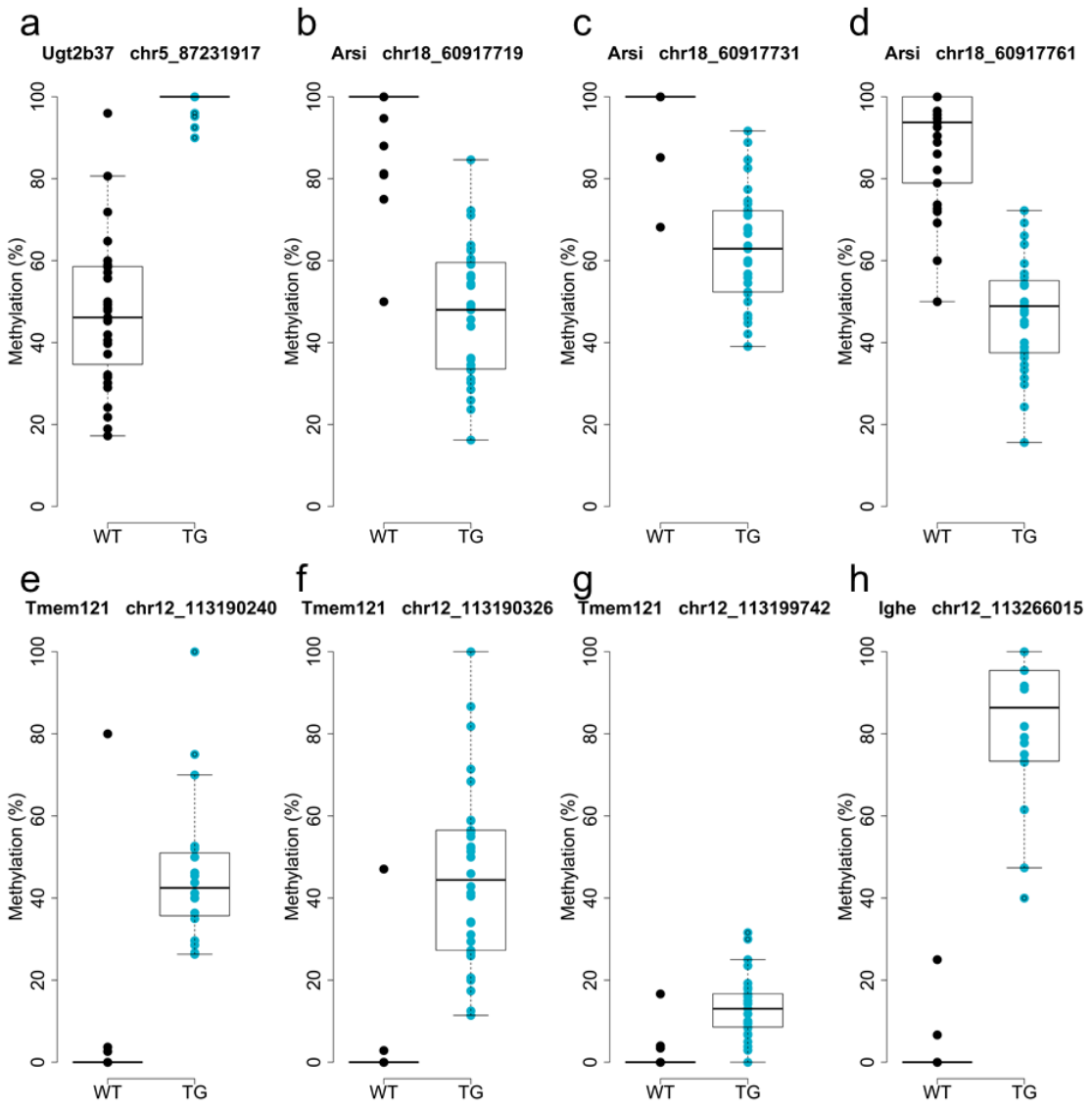


Figure 6.42 – Top rTg4510 genotype-associated sites.
Annotated genes and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

The top J20 genotype-associated DMP was annotated to *Tmprss7* (**Figure 6.36**, **Table 6.4** and **Figure 6.43a**, hypomethylated in TG compared to WT) that encodes for transmembrane serine protease 7 (also known as matriptase 3) – a type-II serine proteinase (TMPRSS7, located in the plasma membrane) of the peptidase S1 family, which preferentially hydrolyses peptides with Arg at the P1 position. In humans and in the mouse, expression of *Tmprss7* has been detected in several tissues, including brain (Szabo et al., 2005). Its biochemical functions include hydrolase activity, peptidase activity, and protein binding. *Tmprss7* was followed by *Gm9578* (**Table 6.4** and **Figure 6.43b**, hypermethylated in TG compared to WT) which encodes the glyceraldehyde-3-phosphate dehydrogenase pseudogene, i.e., non-coding, for which each function is yet to be uncovered. Another differentially methylated site was annotated to *Prdm5* (**Table 6.4** and **Figure 6.43c**, hypermethylated in TG compared to WT), encoding PR domain zinc finger protein 5, a transcription factor of the PR-domain protein family highly expressed in mouse embryonic stem cells, and hypothesised to regulate gene expression of a subset of developmental regulators during cell differentiation by affecting chromatin organisation (Galli et al., 2013). Another J20-associated DMP was annotated to *Mknk1* (**Table 6.4** and **Figure 6.43d**, hypermethylated in TG compared to WT), a protein coding gene encoding mitogen-activated protein kinase (MAPK) interacting serine/threonine kinase 1 (MKNK1), which belongs to a family of serine/threonine kinases, considered downstream effectors of MAPK signalling (Fukunaga and Hunter, 1997). MKNK1 is a serine/threonine protein kinase that interacts with p38 and extracellular-signal-regulated kinase (ERK) 1 and ERK2, suggesting a role in cellular stress responses (Waskiewicz et al., 1997). *Mknk1* is highly expressed in the spleen and modestly expressed in the liver and muscle, with relatively low levels expressed in the brain (Ueda et al., 2004, Waskiewicz et al., 1997). Despite the growing research on the involvement of MKNK1 in carcinogenic processes, only a limited number of studies have explored its role in the brain (Genheden et al., 2015, Moy et al., 2017), with Genheden et al. suggesting MNK1 as a key component of BDNF-mediated translational regulation in neurons (Genheden et al., 2015). A few brain-unrelated published studies have used mice in which the expression of MNK1 has been knocked out (Joshi et al., 2009, Moore et al., 2016, Moy et al., 2017) – research exploring the role of MNK1 in the mouse brain is certainly only one step away.

Table 6.4 – Statistical results for top J20 genotype-associated sites.
Sites are ordered by FDR. Δβ refers to difference in methylation (%) between TG and WT. Underlined sites are plotted in **Figure 6.43**.

Location	Nearest gene	t-value	Δβ	FDR	N
chr16:45667595	*Tmprss7*	t(33) = 89.22	-55.79	2.70E-10	34
chr14:80916197	*Gm9578*	t(31) = 35.98	42.05	4.20E-06	32
chr6:65870804	*Prdm5*	t(59) = 28.14	26.38	5.53E-06	60
chr4:115869229	*Mknk1*	t(25) = 118.92	44.99	6.02E-06	26
chr16:45740128	*Abhd10*	t(45) = 98.15	-46.16	9.53E-06	46
chr18:75831039	*Zbtb7c*	t(12) = 72.05	-69.67	1.11E-05	13
chr13:54346100	*Cplx2*	t(62) = 25.16	20.14	1.46E-05	63
chr16:43973665	*Zdhhc23*	t(53) = 25.52	-32.79	1.73E-05	54
chr16:20701672	*Fam131a*	t(61) = 24.26	24.39	2.08E-05	62
chr12:105535344	*Bdkrb2*	t(41) = 25.93	42.59	2.65E-05	42
chr12:51187188	*Gm22088*	t(50) = 24.57	28.46	2.69E-05	51
chr1:181583199	*Dnah14*	t(41) = 25.63	25.36	2.90E-05	42

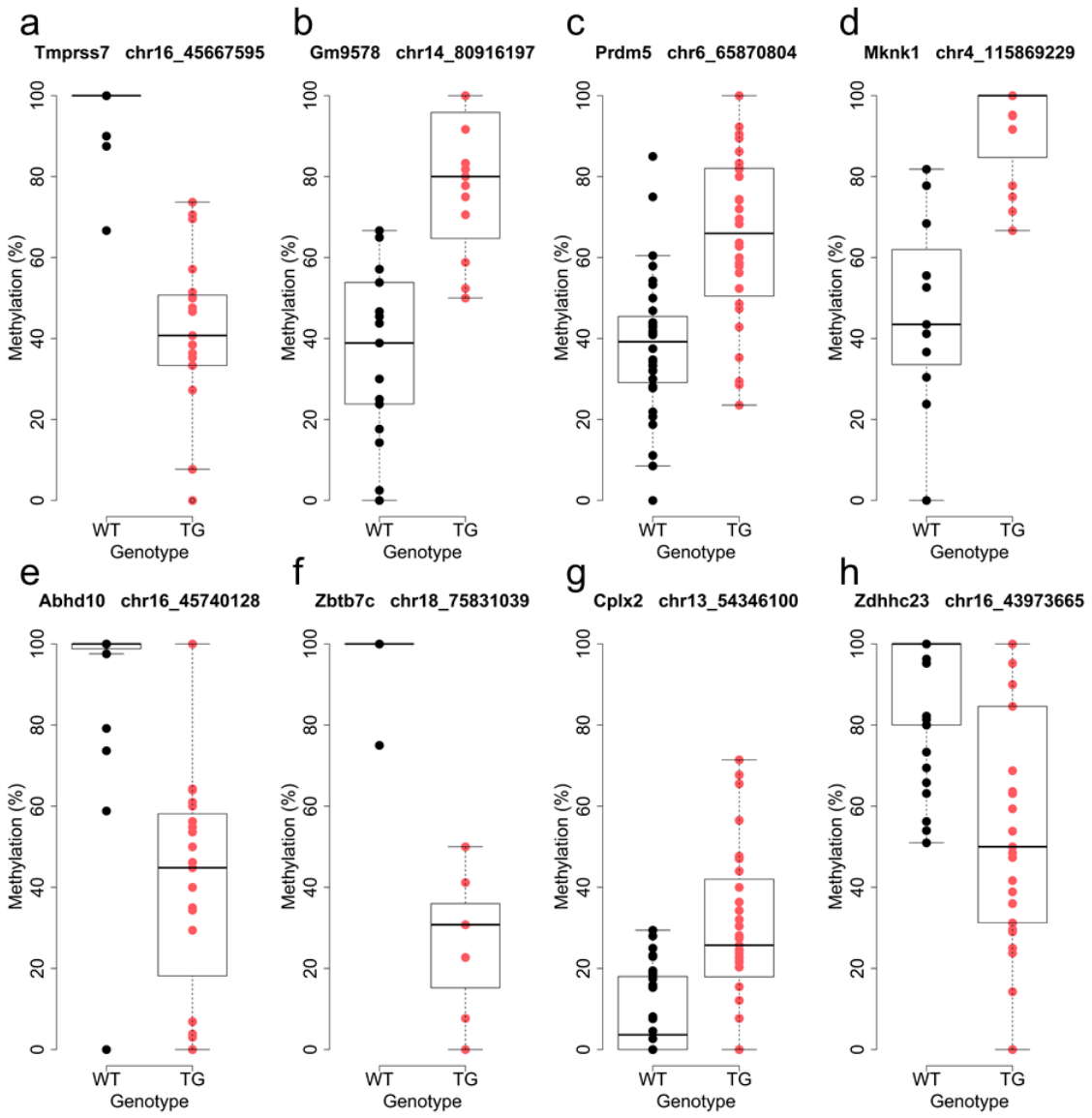


Figure 6.43 – Top J20 genotype-associated sites.
Annotated genes and chromosome positions are indicated in each individual plot. Shown are J20 transgenic (TG, red) female mice compared to wild type littermate controls (WT, black).

6.3.5. Several genotype-associated differentially methylated positions are located in the vicinity of *Mapt* and *App*

Interestingly, two rTg4510 genotype-associated differentially methylated sites were annotated to *Mapt* (**Table 6.5** and **Figure 6.44**), with both sites located in the gene body; one DMP was characterised by decreased methylation in TG mice compared to WT mice, and the other DMP was characterised by increased methylation in TG mice compared to WT mice. Furthermore, two genotype-associated differentially methylated sites in J20 mice were also located in the gene body of *Mapt*; one DMP was characterised by decreased methylation and the other by increased methylation in TG compared to WT mice (**Table 6.6** and **Figure 6.45**). *Mapt* DMPs identified in both models were located close to each other, suggesting potentially overlapping effects in both Tg4510 and J20 mice.

Table 6.5 – Significant DNA methylation differences at sites annotated to Mapt in rTg4510 samples.
Δβ refers to difference in methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr11:104247451	*Mapt*	t(24) = 8.12	-33.77	0.027	25
chr11:104299117	*Mapt*	t(50) = 6.82	17.70	0.036	51

Table 6.6 – Significant DNA methylation differences at sites annotated to Mapt in J20 samples.
Δβ refers to difference in methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr11:104310400	*Mapt*	t(56) = 6.61	-16.41	0.039	57
chr11:104316397	*Mapt*	t(24) = 8.24	34.69	0.026	25

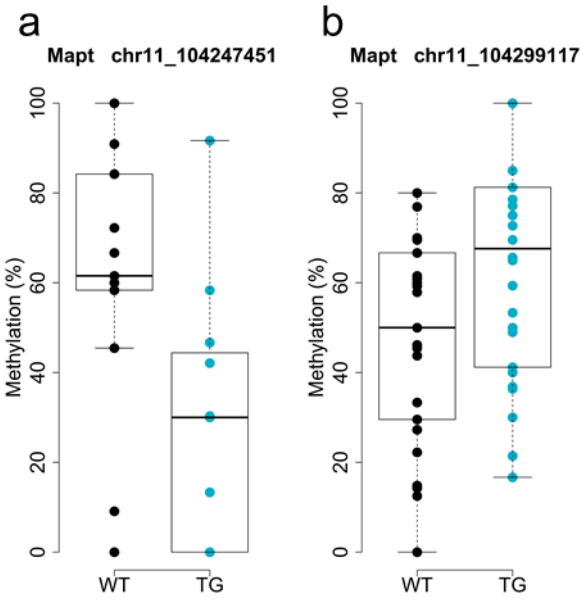


Figure 6.44 – Two DMPs associated with rTg4510 genotype were annotated to Mapt.
Annotated gene and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

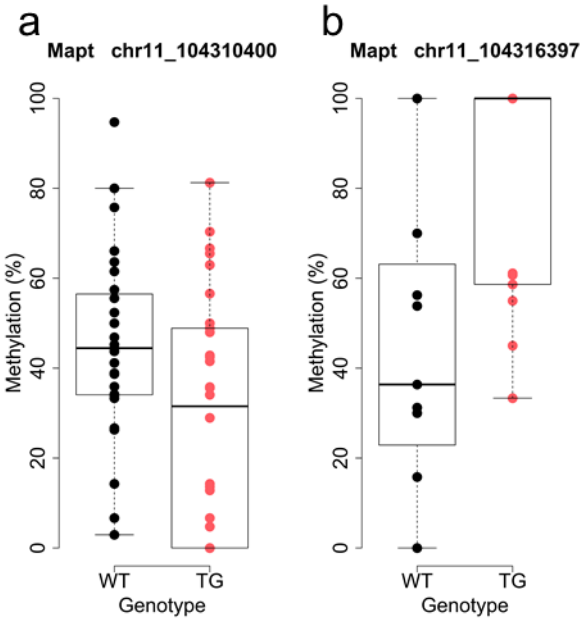


Figure 6.45 – Two DMPs associated with J20 genotype were annotated to Mapt.
Annotated gene and chromosome positions are indicated in each individual plot. Shown are J20 transgenic (TG, red) female mice compared to wild type littermate controls (WT, black).

In J20 mice, six genotype-associated DMPs were annotated to *App*; all were characterised by hypermethylation in J20 TG mice compared to WT (**Table 6.7** and **Figure 6.46**), with one DMP (chr16:84922199) being located ~35kb upstream of *App*, and the other five DMPs being located in the gene body. Interestingly, six DMPs associated with rTg4510 genotype were also annotated to *App*, again all hypermethylated in rTg4510 TG mice compared to WT (**Table 6.8** and **Figure 6.47**); five DMPs were located in the gene body of *App*, with an additional DMP (chr16:85269100) located ~95 kb downstream of the gene.

Table 6.7 – Significant DNA methylation differences at sites annotated to App in J20 samples.
Δβ refers to difference in methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr16:84922199	*App*	t(41) = 7.13	18.78	0.033	42
chr16:84995182	*App*	t(48) = 9.38	21.86	0.011	49
chr16:85084104	*App*	t(14) = 8.18	42.48	0.040	15
chr16:85086498	*App*	t(58) = 6.95	16.03	0.032	59
chr16:85087931	*App*	t(61) = 7.28	9.72	0.027	62
chr16:85120676	*App*	t(52) = 6.97	9.20	0.033	53

Table 6.8 – Significant DNA methylation differences at sites annotated to App in rTg4510 samples.
Δβ refers to difference in methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr16:85088148	*App*	t(59) = 6.91	10.99	0.033	60
chr16:85091216	*App*	t(58) = 12.76	12.37	0.0022	59
chr16:85091220	*App*	t(58) = 16.67	12.72	0.00042	59
chr16:85091240	*App*	t(58) = 9.19	9.32	0.011	59
chr16:85091540	*App*	t(58) = 6.81	9.55	0.035	59
chr16:85269100	*App*	t(53) = 14.20	11.09	0.0013	54

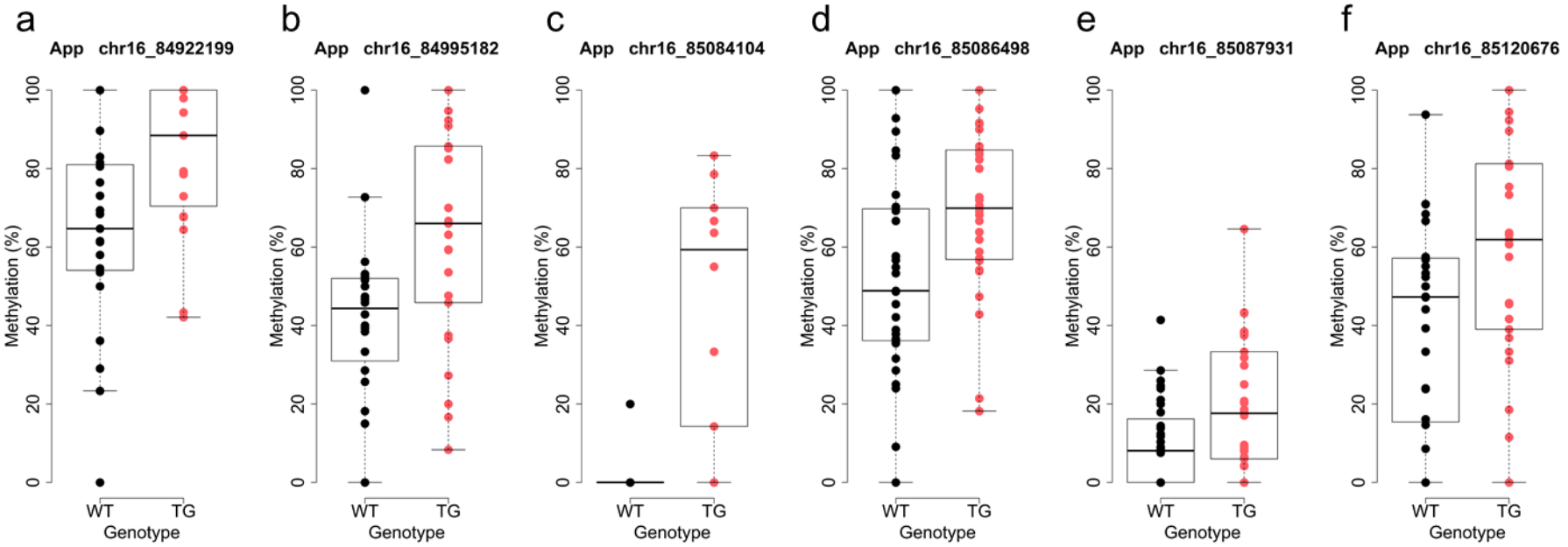


Figure 6.46 – Six DMPs associated with J20 genotype were annotated to App.
Annotated gene and chromosome positions are indicated in each individual plot. Shown are J20 transgenic (TG, red) female mice compared to wild type littermate controls (WT, black).

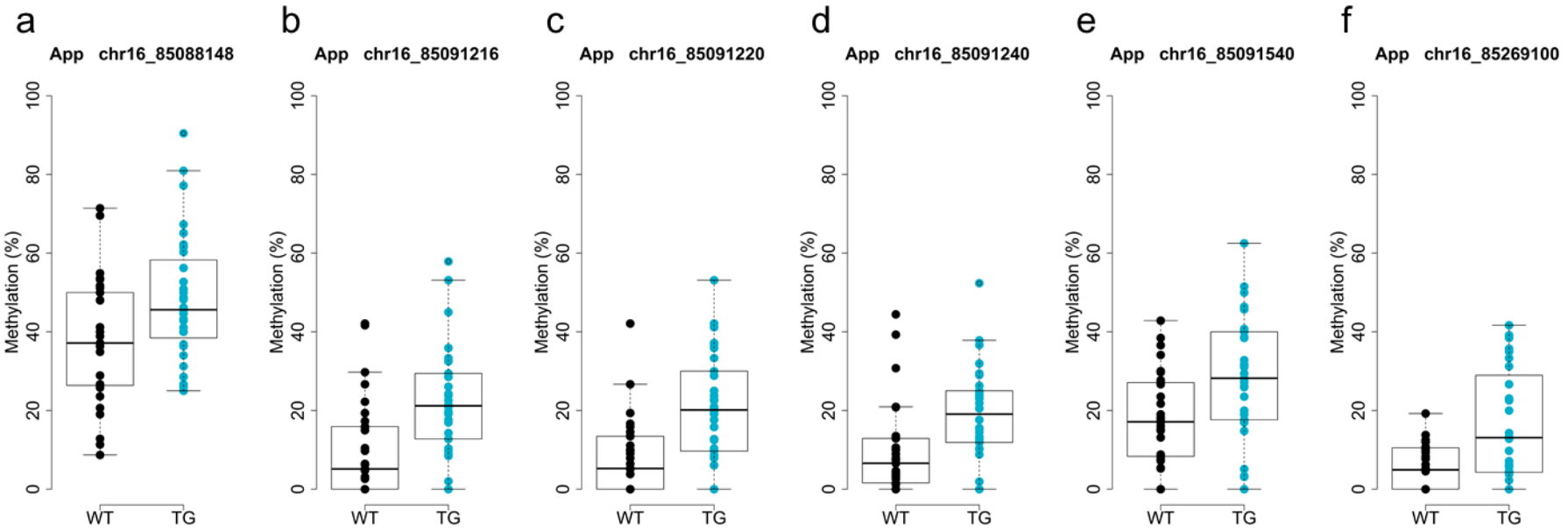


Figure 6.47 – Six DMPs associated with rTg4510 genotype were annotated to App.
Annotated gene and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

6.3.6. Annotated genes contain many genes associated with AD from EWAS

Recent work by our group and others have found that DNA methylation differences in human cortex annotated to four loci - *ANK1*, *RPL13*, *CDH23*, and *RHBDF2* – are robustly associated with AD neuropathology (**Figure 6.48**; see also **Chapter 1**, **Section 1.4**) (De Jager et al., 2014, Lunnon et al., 2014). Other loci annotated to AD-associated DMPs in these studies included *TIMM22-ABR*, *ACTR3BP2*, CLYBL-*TM9SF2* (Lunnon et al., 2014), *SLC2A1-FLJ32224*, *COQ7-ITPRIPL2*, *HOXA* region, *FOXK1-AP5Z1-RADIL*, *DIP2A*, *SERPINF1*, *SERPINF2* (De Jager et al., 2014) (**Figure 6.48**). I therefore explored whether DNA methylation differences in the mouse homologues of these genes were associated with either rTg4510 or J20 genotype. We identified genotype-associated DMPs annotated to *Ank1* (**Table 6.9**, **Figure 6.49**, **Table 6.10**, **Figure 6.50**) and *Cdh23* (**Table 6.11**, **Table 6.12**, **Figure 6.51**, **Figure 6.52**) in both rTg4510 and J20 ECX; of note, the vast majority of DMPs were found to be hypermethylated in TG mice compared to WT, with all DMPs located in the gene body, except one DMP (chr8:22974676) in J20 that was located 167 bp away from *Ank1*. In the rTg4510 mice, one genotype-associated DMP was annotated to *Rpl13* (**Figure 6.53**; $t(61) = 6.61$, $\Delta\beta = -3.16$, FDR = 0.038), and one DMP was annotated to *Rhbdf2* (**Figure 6.54**; $t(61) = 7.66$, $\Delta\beta = -19.79$, FDR = 0.038); no DMPs were annotated to *Rpl13* or *Rhbdf2* in J20 mice. Furthermore, several genotype-associated DMPs were annotated to *Abr* (**Table 6.13**, **Table 6.14**, **Figure 6.55**), *Clybl* (**Table 6.15**, **Table 6.16**, **Figure 6.56**, **Figure 6.57**), the Hoxa region (**Table 6.17**, **Table 6.18**, **Figure 6.58**, **Figure 6.59**), and a region spanning *Foxk1-Ap5z1-Radil* (**Table 6.19**, **Table 6.20**, **Figure 6.60**, **Figure 6.61**) in both the rTg4510 and J20 model. In the rTg4510 dataset, one DMP was annotated to *Coq7* (**Figure 6.62**, $t(56) = 6.50$, $\Delta\beta = -4.13$, FDR = 0.041), located 7665 bp from the gene. Furthermore, one DMP in each mouse model was annotated to *Dip2a* (**Figure 6.63**; rTg4510: $t(24) = 8.40$, $\Delta\beta = 6.04$, FDR = 0.024; J20: $t(12) = 13.27$, $\Delta\beta = 24.97$, FDR = 0.0039), and two DMPs in the rTg4510 mice and one DMP in J20 mice were annotated to *Serpinf2* (**Figure 6.64**). These results are interesting because they highlight potential overlaps in differentially methylated loci between studies of human AD and our mouse models of both amyloid and tau pathology.

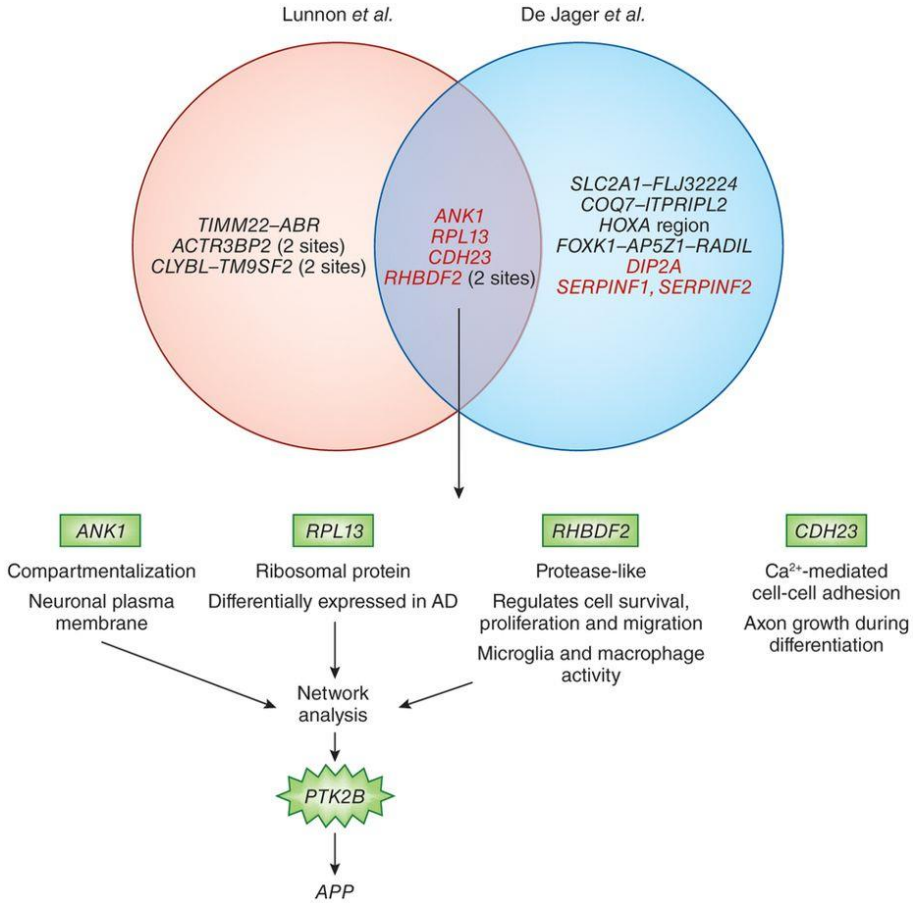


Figure 6.48 – Annotated genes for CpG sites shown to be associated with Alzheimer's disease pathology reported by Lunnon et al. (2014) and De Jager et al. (2014).

Genes in red were validated by De Jager et al. (2014) as showing differential mRNA expression. Figure and legend adapted from Lord and Cruchaga (2014).

Table 6.9 – Significant DNA methylation differences at sites annotated to Ank1 in rTg4510 samples.
Δβ corresponds to difference in methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr8:22978636	*Ank1*	t(44) = 7.78	23.86	0.024	45
chr8:22978666	*Ank1*	t(44) = 11.04	27.56	0.0055	45
chr8:23035040	*Ank1*	t(13) = 9.48	45.06	0.029	14
chr8:23035042	*Ank1*	t(13) = 18.40	54.11	0.0032	14
chr8:23045092	*Ank1*	t(28) = 9.75	13.96	0.013	29
chr8:23058852	*Ank1*	t(56) = 6.18	3.37	0.048	57
chr8:23143029	*Ank1*	t(45) = 8.08	19.15	0.020	46
chr8:23143041	*Ank1*	t(45) = 7.49	12.99	0.027	46

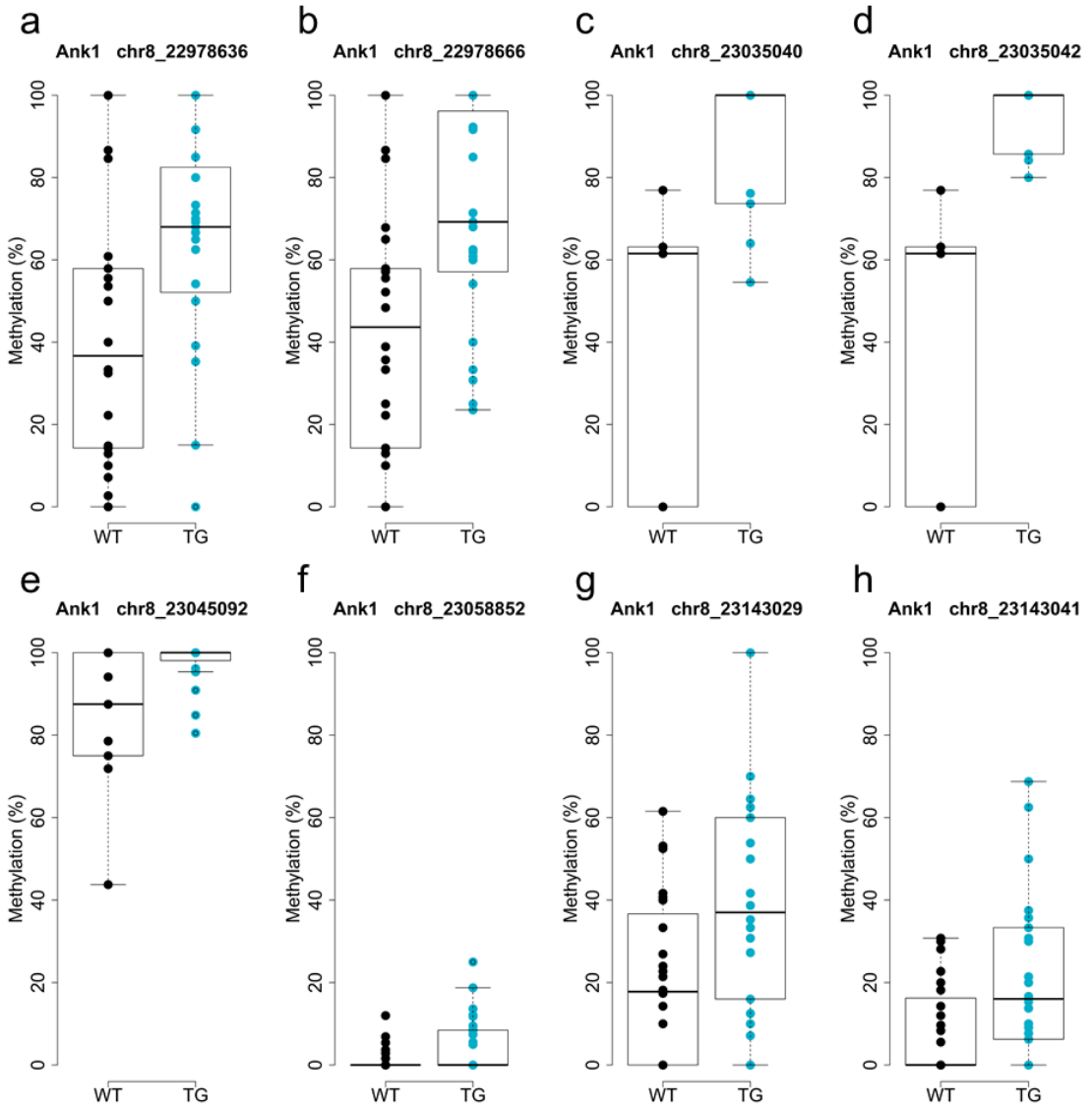


***Figure 6.49 – Eight genotype-associated DMPs were annotated to Ank1* in rTg4510 mice.**
Annotated gene and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

Table 6.10 – Significant DNA methylation differences at sites annotated to Ank1 in J20 samples.
Δβ corresponds to difference in methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr8:22974676	*Ank1*	t(13) = 12.20	-7.02	0.013	14
chr8:23034919	*Ank1*	t(61) = 9.67	8.82	0.0086	62
chr8:23116360	*Ank1*	t(41) = 10.60	-18.17	0.0069	42
chr8:23123856	*Ank1*	t(43) = 6.52	15.72	0.043	44
chr8:23123864	*Ank1*	t(43) = 9.12	25.54	0.013	44
chr8:23123900	*Ank1*	t(50) = 6.80	18.52	0.036	51

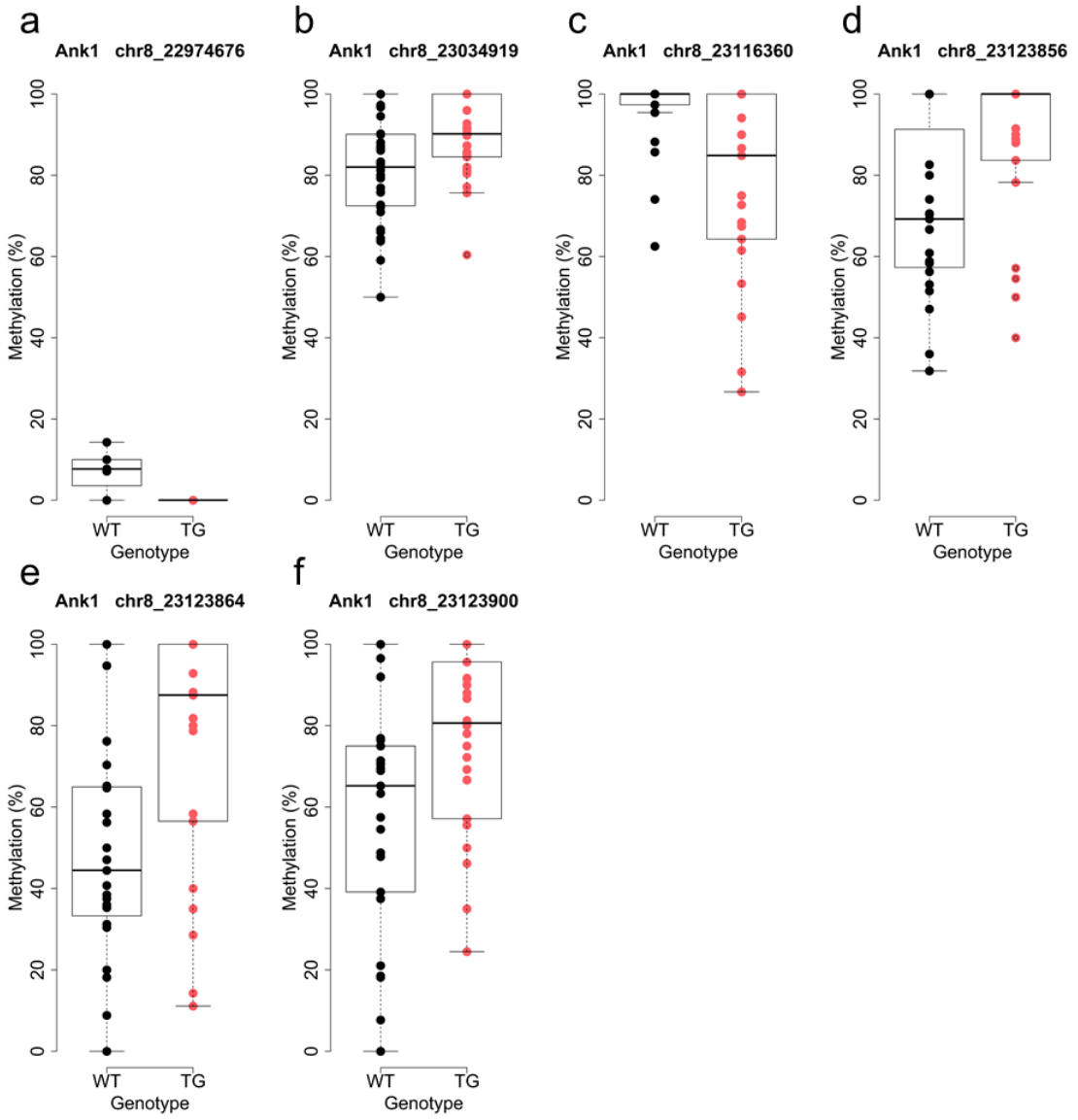


***Figure 6.50 – Six genotype-associated DMPs were annotated to Ank1* in J20 mice.**
Annotated gene and chromosome positions are indicated in each individual plot. Shown are J20 transgenic (TG, red) female mice compared to wild type littermate controls (WT, black).

Table 6.11 – Significant DNA methylation differences at sites annotated to Cdh23 in rTg4510 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr10:60321298	*Cdh23*	t(46) = 7.58	15.54	0.025	47
chr10:60346482	*Cdh23*	t(45) = 7.07	14.86	0.033	46
chr10:60455760	*Cdh23*	t(54) = 10.84	12.92	0.0053	55
chr10:60523066	*Cdh23*	t(16) = 7.83	10.04	0.040	17
chr10:60634392	*Cdh23*	t(61) = 11.76	12.99	0.003	62
chr10:60634564	*Cdh23*	t(42) = 6.29	21.50	0.049	43
chr10:60686029	*Cdh23*	t(61) = 6.35	4.97	0.043	62

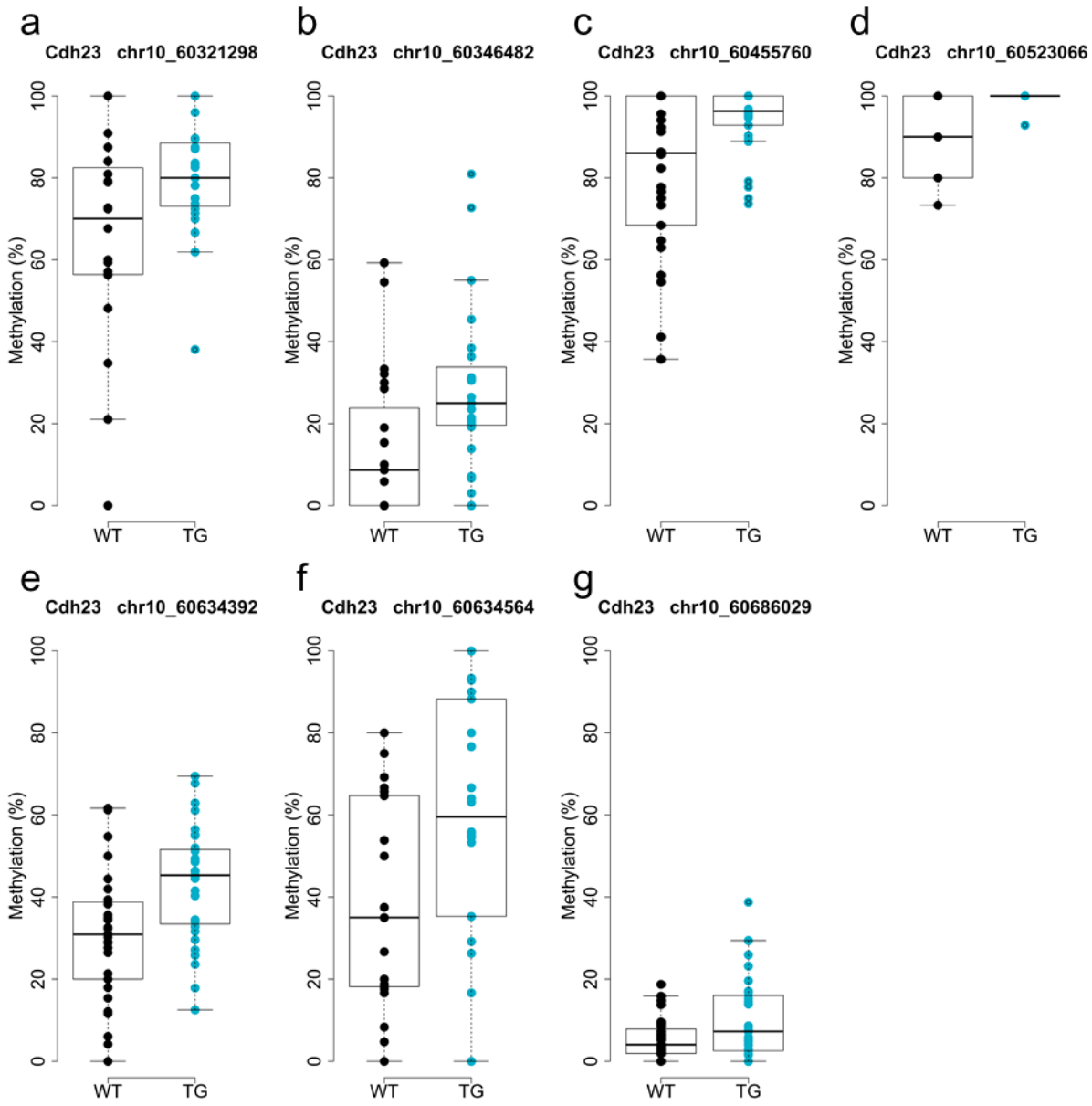


Figure 6.51 – Seven genotype-associated DMPs were annotated to Cdh23 in rTg4510 mice.
Annotated gene and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

Table 6.12 – Significant DNA methylation differences at sites annotated to Cdh23 in J20 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr10:60303537	*Cdh23*	t(50) = 7.69	20.90	0.024	51
chr10:60334144	*Cdh23*	t(40) = 7.04	-7.14	0.034	41
chr10:60346528	*Cdh23*	t(58) = 9.40	-19.40	0.010	59
chr10:60411179	*Cdh23*	t(59) = 8.06	12.14	0.019	60
chr10:60436753	*Cdh23*	t(61) = 8.44	12.81	0.015	62
chr10:60495199	*Cdh23*	t(30) = 16.18	14.02	0.0011	31
chr10:60495544	*Cdh23*	t(16) = 7.42	40.83	0.047	17
chr10:60540255	*Cdh23*	t(62) = 6.45	9.74	0.041	63
chr10:60595936	*Cdh23*	t(21) = 6.96	27.56	0.047	22
chr10:60596311	*Cdh23*	t(47) = 6.73	-19.17	0.038	48
chr10:60638890	*Cdh23*	t(60) = 7.49	9.24	0.025	61
chr10:60638910	*Cdh23*	t(60) = 7.96	9.14	0.019	61
chr10:60649040	*Cdh23*	t(60) = 7.58	14.91	0.024	61

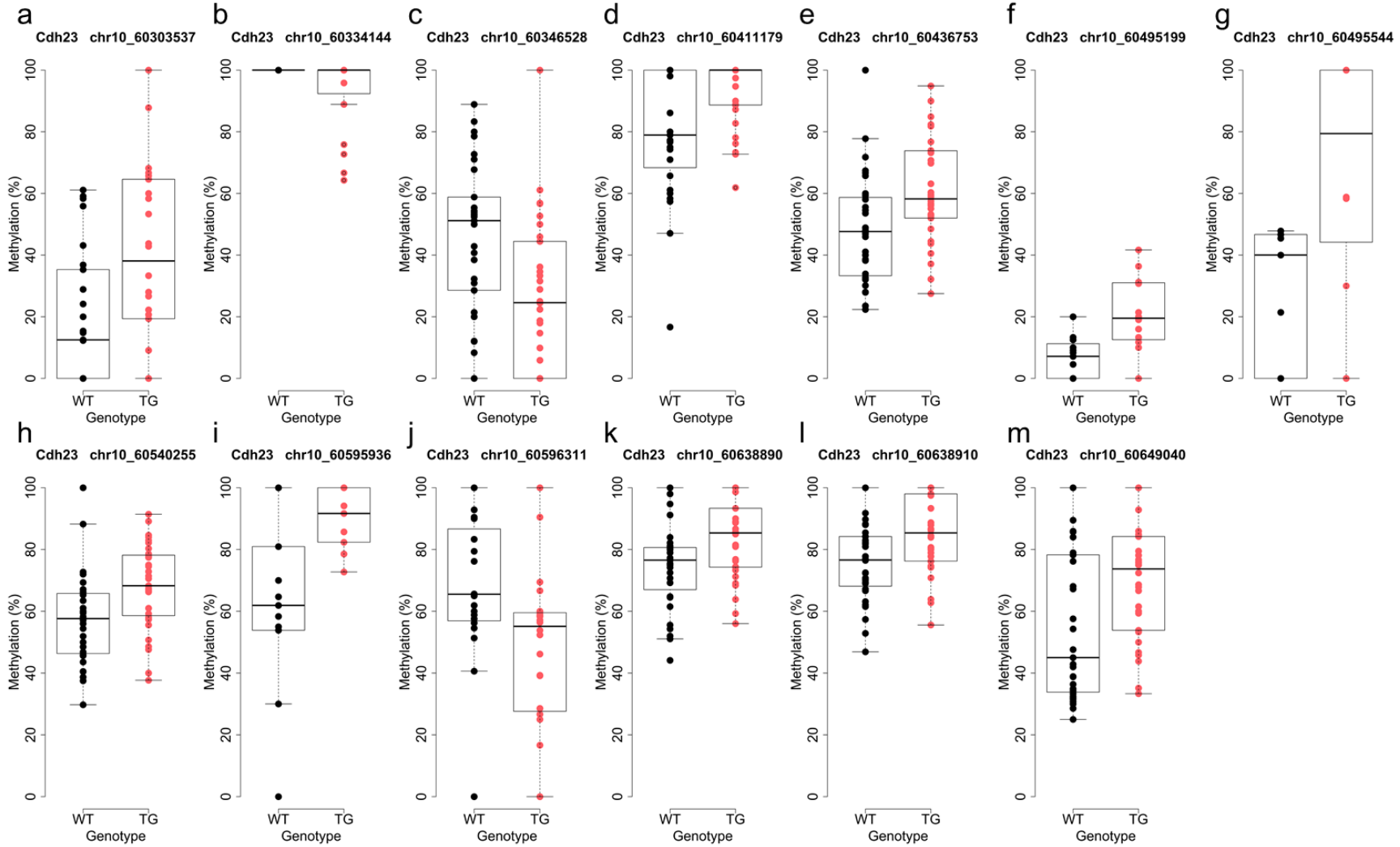


Figure 6.52 – Thirteen genotype-associated DMPs were annotated to Cdh23 in J20 mice.
Annotated gene and chromosome positions are indicated in each individual plot. Shown are J20 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

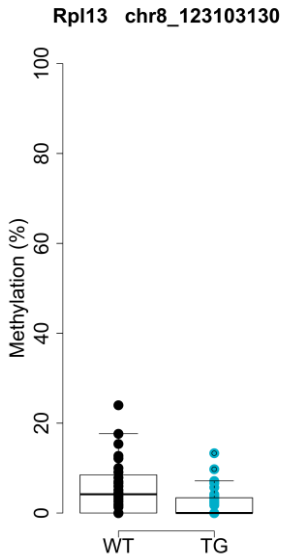


Figure 6.53 – One genotype-associated DMP was annotated to Rpl13 in rTg4510 mice.
Annotated gene and chromosome position are indicated. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

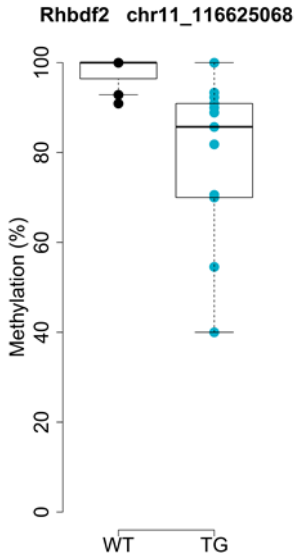


Figure 6.54 – One genotype-associated DMP was annotated to Rhbdf2 in rTg4510 mice.
Annotated gene and chromosome position are indicated. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

Table 6.13 – Significant DNA methylation differences at sites annotated to Abr in rTg4510 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr11:76508111	*Abr*	t(61) = 7.68	7.47	0.022	62
chr11:76508178	*Abr*	t(57) = 6.51	11.89	0.040	58
chr11:76542804	*Abr*	t(61) = 9.50	4.11	0.0092	62
chr11:76570619	*Abr*	t(61) = 6.43	6.99	0.041	62
chr11:76577662	*Abr*	t(46) = 6.84	13.56	0.036	47
chr11:76624326	*Abr*	t(52) = 7.13	16.64	0.030	53
chr11:76624371	*Abr*	t(52) = 9.14	19.43	0.012	53
chr11:76639885	*Abr*	t(55) = 7.71	11.66	0.023	56

Table 6.14 – Significant DNA methylation differences at sites annotated to Abr in J20 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr11:76473889	*Abr*	t(19) = 7.97	-14.47	0.034	20
chr11:76623004	*Abr*	t(62) = 7.37	-3.84	0.026	63

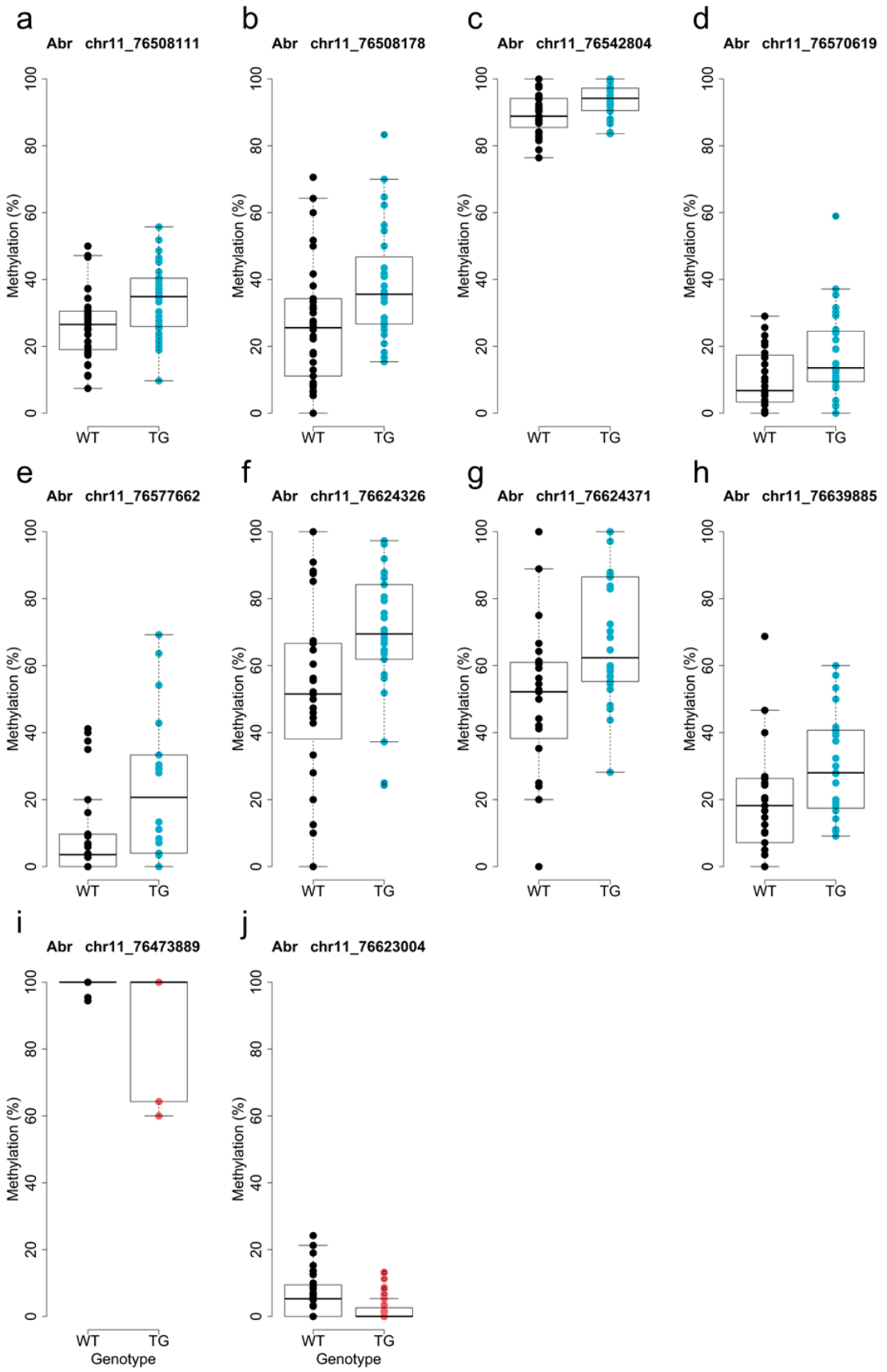


Figure 6.55 – Eight rTg4510 genotype-associated and two J20 genotype-associated DMPs were annotated to Abr.
Annotated gene and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) or J20 TG (red) female mice compared to their respective wild type littermate controls (WT, black).

Table 6.15 – Significant DNA methylation differences at sites annotated to Clybl in rTg4510 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr14:122172522	*Clybl*	t(30) = 15.82	35.38	0.0013	31
chr14:122172566	*Clybl*	t(30) = 19.04	34.30	0.00044	31
chr14:122257470	*Clybl*	t(52) = 6.22	14.19	0.048	53
chr14:122307062	*Clybl*	t(50) = 7.47	14.26	0.026	51
chr14:122406730	*Clybl*	t(57) = 6.77	7.57	0.035	58

Table 6.16 – Significant DNA methylation differences at sites annotated to Clybl in J20 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr14:122179634	*Clybl*	t(18) = 8.91	-20.13	0.025	19
chr14:122181645	*Clybl*	t(59) = 6.38	5.15	0.043	60
chr14:122200537	*Clybl*	t(62) = 6.82	13.82	0.034	63
chr14:122222613	*Clybl*	t(62) = 6.66	12.81	0.037	63
chr14:122222644	*Clybl*	t(61) = 7.23	10.72	0.028	62
chr14:122222740	*Clybl*	t(62) = 9.91	11.22	0.0076	63
chr14:122406878	*Clybl*	t(62) = 8.55	5.38	0.015	63
chr14:122407087	*Clybl*	t(57) = 8.23	-4.64	0.017	58

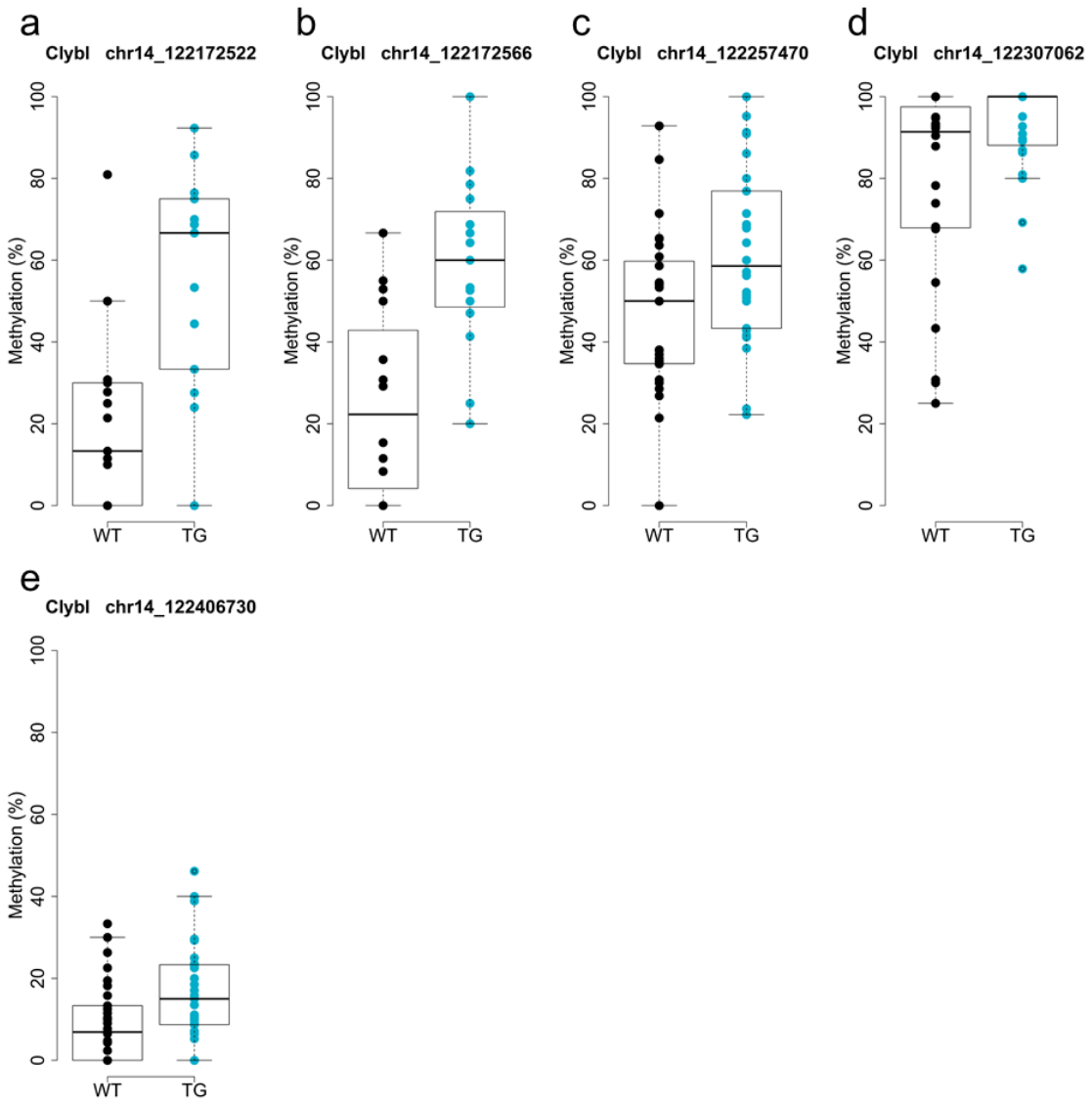


Figure 6.56 – Five genotype-associated DMPs were annotated to Clybl in rTg4510 mice.

Annotated gene and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

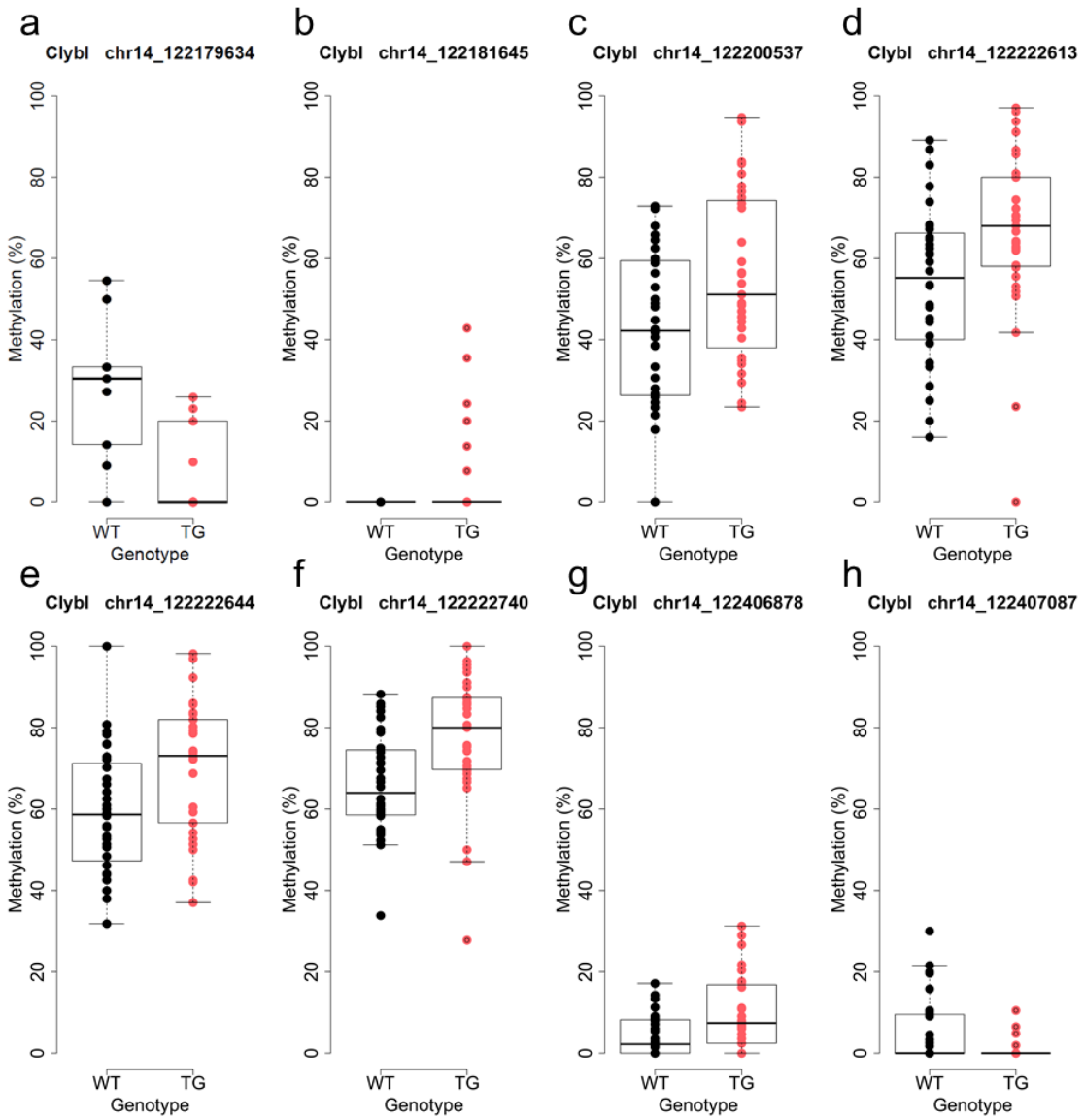


Figure 6.57 – Eight genotype-associated DMPs were annotated to Clybl in J20 mice.

Annotated gene and chromosome positions are indicated in each individual plot. Shown are J20 transgenic (TG, red) female mice compared to wild type littermate controls (WT, black).

Table 6.17 – Significant DNA methylation differences at sites annotated to the Hoxa region in rTg4510 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr6:52199921	*Hoxaas3*	t(61) = 7.55	4.01	0.024	62
chr6:52209539	*Hoxaas3*	t(61) = 9.76	9.34	0.0082	62
chr6:52217163	*Hoxa7*	t(55) = 8.39	-4.08	0.016	56
chr6:52224435	*Hoxa9*	t(59) = 6.49	-8.93	0.041	60
chr6:52225800	*Hoxa9*	t(12) = 8.87	11.35	0.038	13
chr6:52226299	*Hoxa9*	t(38) = 8.24	11.69	0.020	39
chr6:52244601	*Hoxa11os*	t(50) = 8.26	8.60	0.018	51
chr6:52252901	*Hoxa11os*	t(51) = 6.47	11.46	0.042	52
chr6:52259995	*Hoxa13*	t(55) = 7.14	-4.52	0.030	56
chr6:52261001	*Hoxa13*	t(13) = 7.77	-14.62	0.049	14

Table 6.18 – Significant DNA methylation differences at sites annotated to the Hoxa region in J20 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr6:52175647	*Hoxaas3*	t(61) = 7.04	3.50	0.030	62
chr6:52176077	*Hoxaas3*	t(46) = 7.71	6.02	0.024	47
chr6:52190581	*Hoxaas3*	t(62) = 7.77	3.65	0.021	63
chr6:52205277	*Hoxaas3*	t(52) = 7.25	12.93	0.029	53
chr6:52208860	*Hoxaas3*	t(62) = 6.22	-3.02	0.046	63
chr6:52226257	*Hoxa9*	t(50) = 7.46	3.94	0.026	51
chr6:52226267	*Hoxa9*	t(56) = 8.05	-9.87	0.019	57
chr6:52233069	*Hoxa10*	t(62) = 6.61	4.96	0.038	63
chr6:52244694	*Hoxa11os*	t(62) = 9.61	5.78	0.0089	63
chr6:52248587	*Hoxa11os*	t(62) = 6.24	-4.28	0.046	63
chr6:52252821	*Hoxa11os*	t(59) = 6.29	6.83	0.045	60
chr6:52260482	Hoxa13	t(58) = 10.35	-5.51	0.0064	59

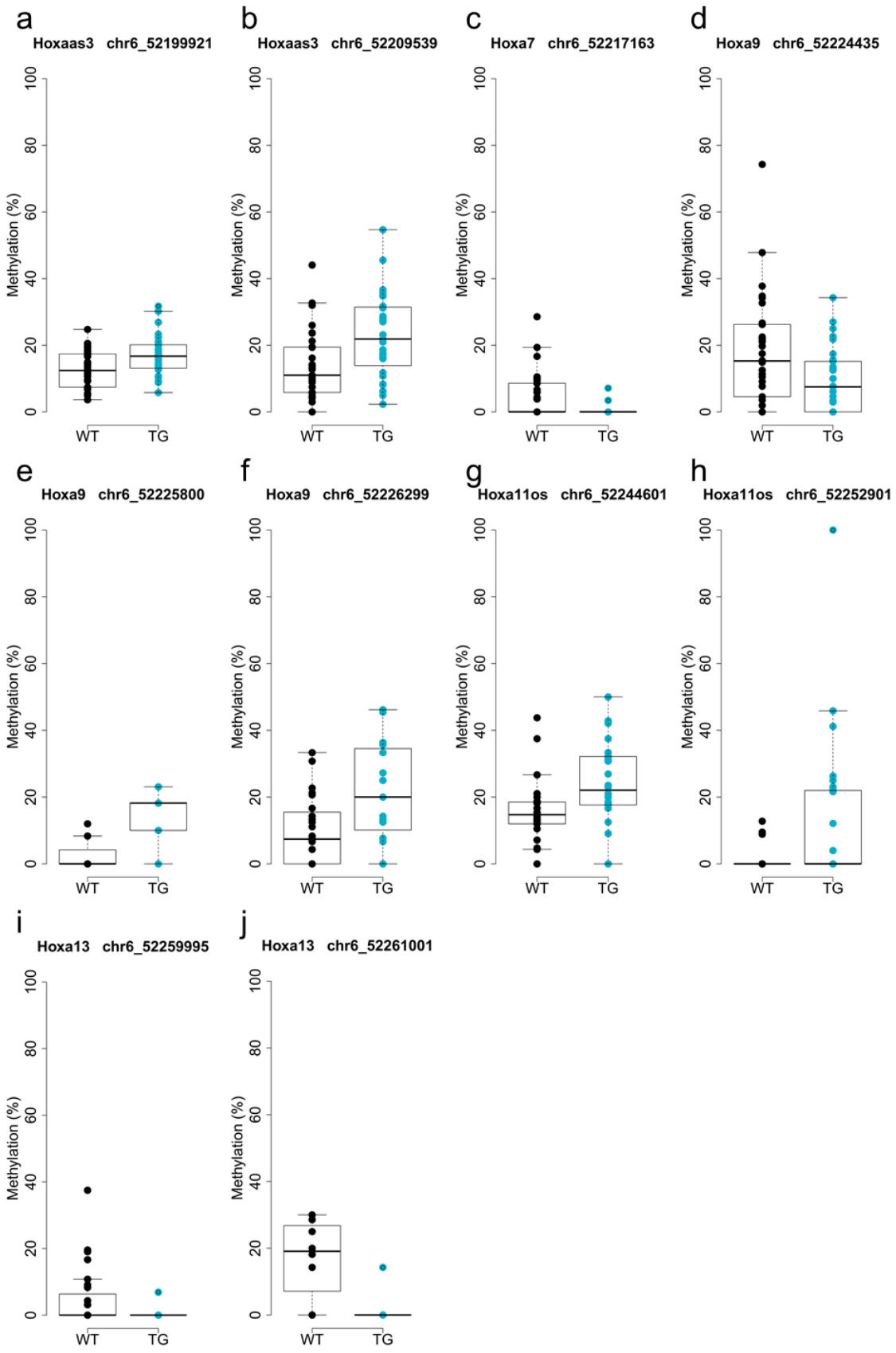


Figure 6.58 – Ten genotype-associated DMPs were annotated to the Hoxa region in rTg4510 mice.
Annotated genes and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

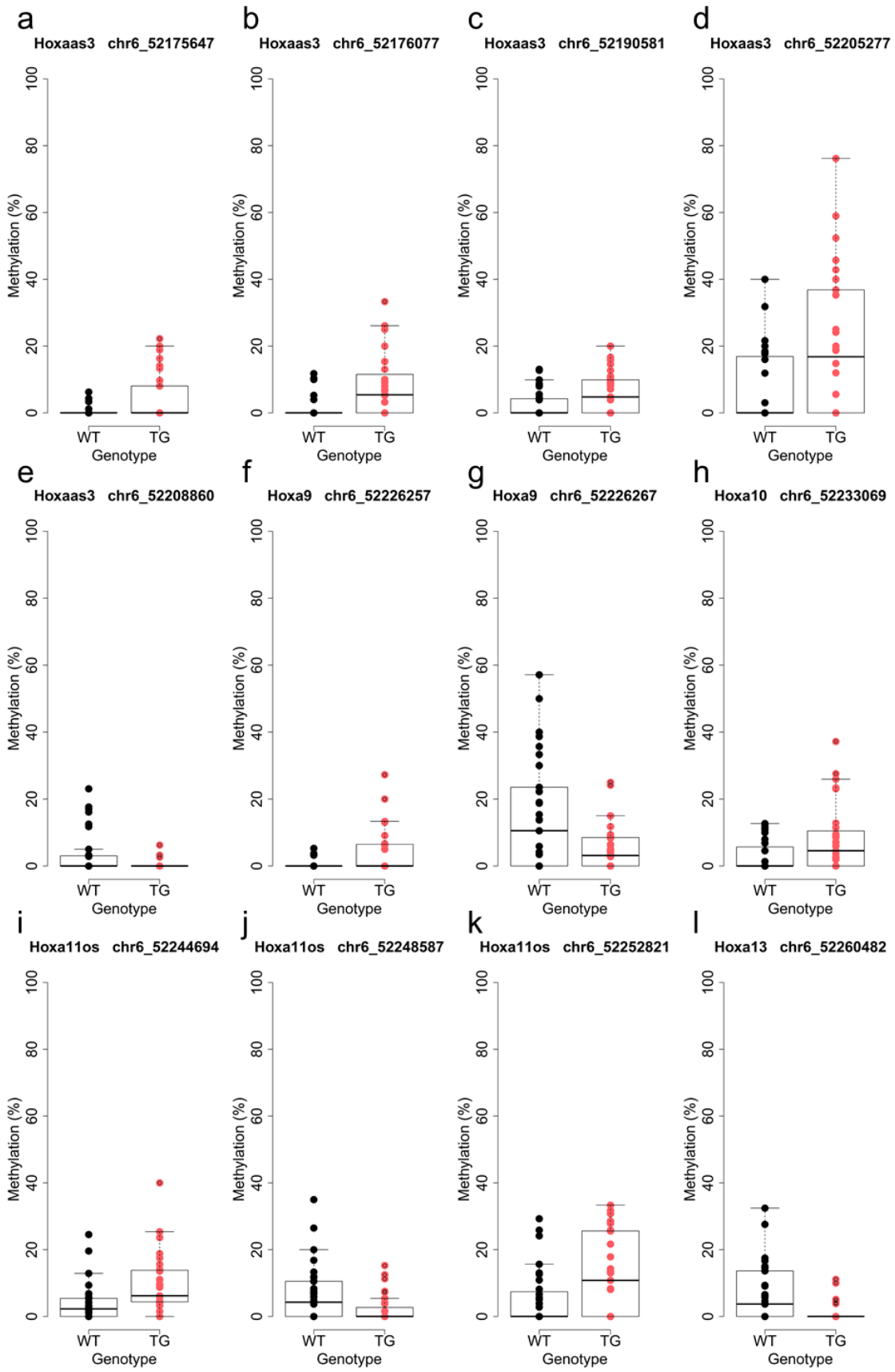


Figure 6.59 – Twelve genotype-associated DMPs were annotated to the Hoxa region in J20 mice.
Annotated genes and chromosome positions are indicated in each individual plot. Shown are J20 transgenic (TG, red) female mice compared to wild type littermate controls (WT, black).

Table 6.19 – Significant DNA methylation differences at sites annotated to the Foxk1-Ap5z1-Radil region in rTg4510 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr5:142400345	*Foxk1*	t() = 6.60	6.67	0.038083	62
chr5:142423143	*Foxk1*	t() = 7.37	26.19	0.034208	29
chr5:142474921	*Ap5z1*	t(49) = 7.06	15.33	0.031942	50
chr5:142477404	*Ap5z1*	t(57) = 8.15	-14.17	0.018051	58
chr5:142485494	*Radil*	t(34) = 9.29	8.92	0.013552	35
chr5:142546740	*Radil*	t(58) = 6.53	13.67	0.039903	59

Table 6.20 – Significant DNA methylation differences at sites annotated to the Foxk1-Ap5z1-Radil region in J20 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr5:142435122	*Foxk1*	t(50) = 7.25	9.83	0.029	51
chr5:142435276	*Foxk1*	t(55) = 7.05	13.49	0.031	56
chr5:142466639	*Ap5z1*	t(46) = 11.53	20.52	0.0043	47
chr5:142474598	*Ap5z1*	t(41) = 9.23	29.98	0.013	42
chr5:142506674	*Radil*	t(12) = 8.64	50.53	0.040	13
chr5:142435122	*Radil*	t(50) = 7.25	9.83	0.029	51

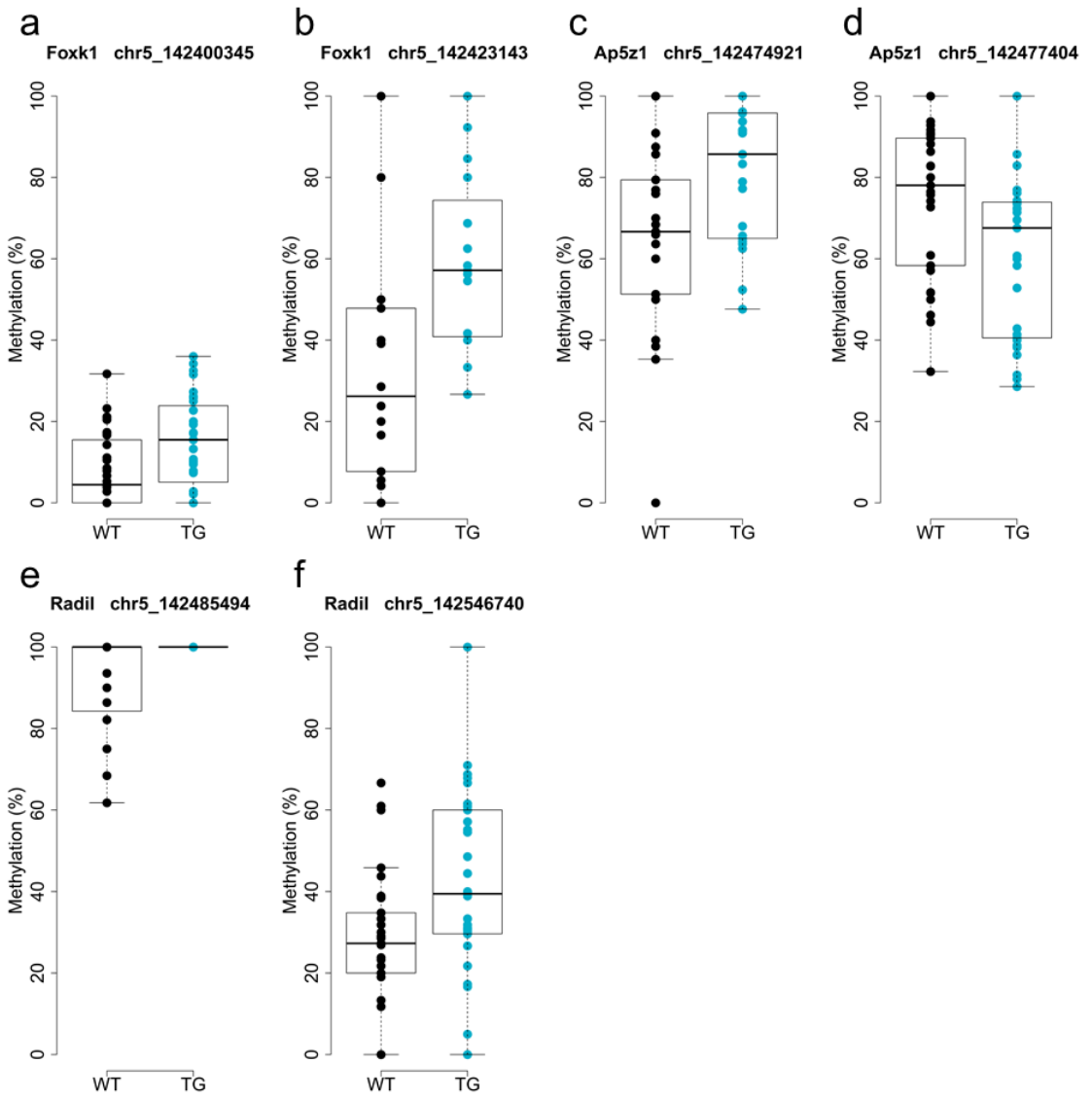


Figure 6.60 – Six genotype-associated DMPs were annotated to the Foxk1-Ap5z1-Radil region in rTg4510 mice.
Annotated genes and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

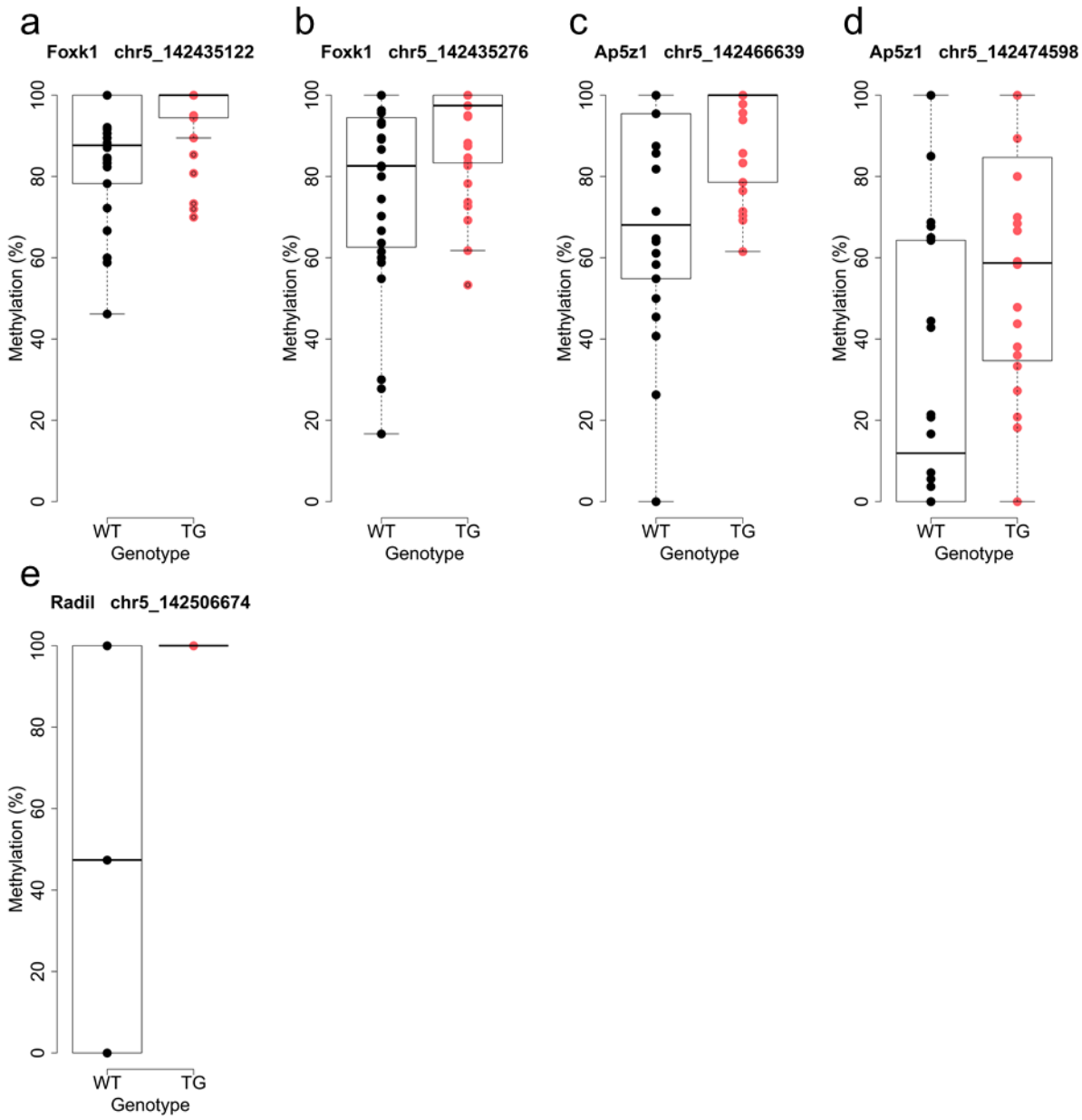


Figure 6.61 – Five genotype-associated DMPs were annotated to the Foxk1-Ap5z1-Radil region in J20 mice.
Annotated genes and chromosome positions are indicated in each individual plot. Shown are J20 transgenic (TG, red) female mice compared to wild type littermate controls (WT, black).

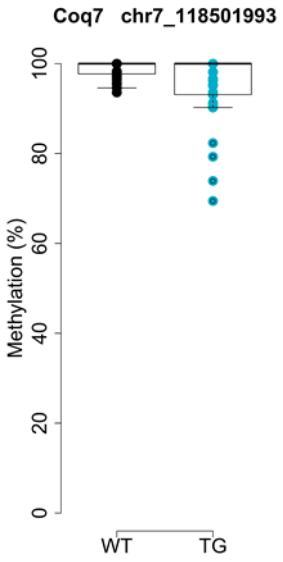


Figure 6.62 – One genotype-associated DMP was annotated to Coq7 in rTg4510 mice.
The annotated gene and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) female mice compared to wild type littermate controls (WT, black).

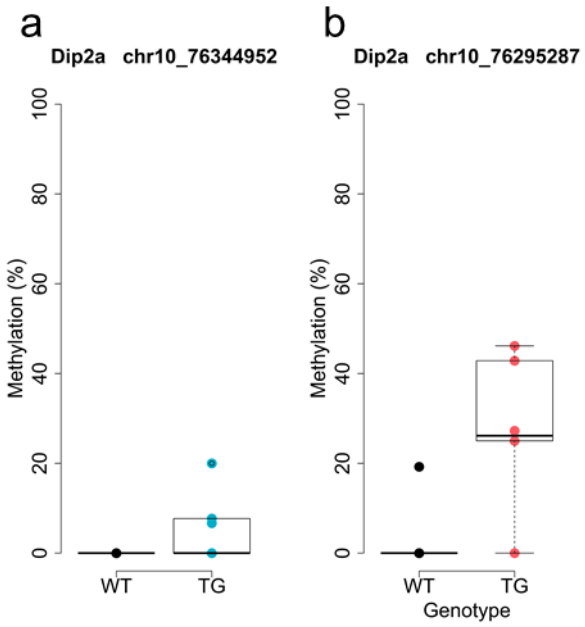


Figure 6.63 – One rTg4510 genotype-associated site and one J20 genotype-associated differentially methylated site were annotated to Dip2a.
The annotated gene and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) and J20 TG (red) female mice compared to their respective wild type littermate controls (WT, black).

Table 6.21 – Significant DNA methylation differences at sites annotated to Serpinf2 in rTg4510 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr11:75431357	*Serpinf2*	t(57) = 10.74	11.47	0.0054	58
chr11:75431375	*Serpinf2*	t(57) = 9.79	9.38	0.0083	58

Table 6.22 – Significant DNA methylation differences at sites annotated to Serpinf2 in J20 samples.
Δβ corresponds to difference in DNA methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr11:75436401	*Serpinf2*	t(53) = 10.16	20.10	0.0073	54

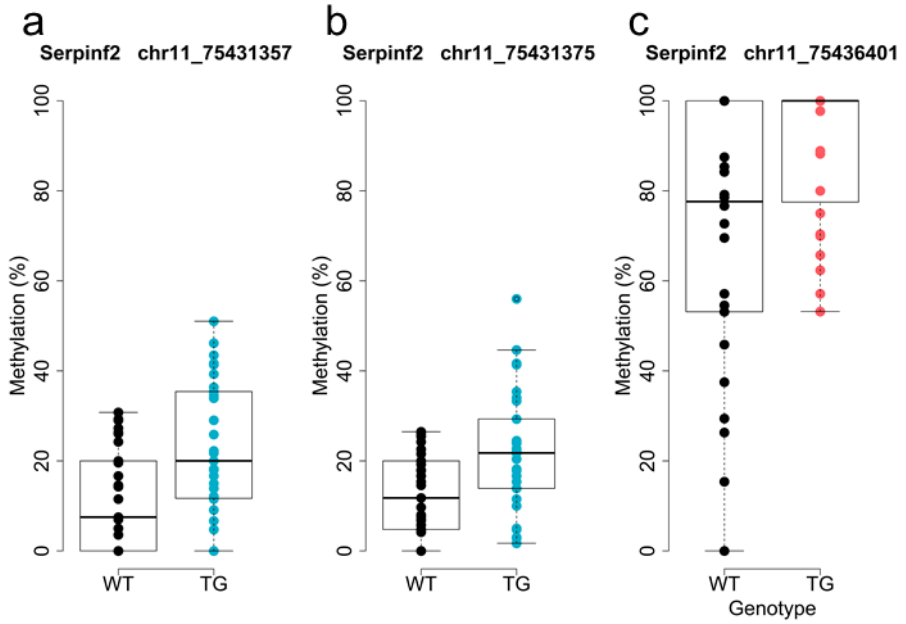


Figure 6.64 – Two rTg4510 genotype-associated DMPs and one J20 genotype-associated DMP were annotated to Serpinf2.
Annotated genes and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) and J20 TG (red) female mice compared to their respective wild type littermate controls (WT, black).

6.3.7. Genotype-associated differentially methylated positions included sites annotated to genes nominated from genetic studies of AD

A number of loci implicated in AD from GWAS and subsequent meta-analyses (discussed in detail in **Section 1.2** in **Chapter 1**) were found to also harbour DMPs in rTg4510 and J20 mice (**Table 6.23**, **Table 6.24**, and **Figure 6.65**). Three DMPs in rTg4510 and one DMP in J20 were annotated to *Abca7*; all hypermethylated in TG mice compared to WT, and all located in the gene body. DMPs annotated to *Clu* and *Bin1* were detected in the J20 model: from the three DMPs annotated to *Clu*, the two hypomethylated DMPs in TG compare to WT were located in the gene body, and the hypermethylated DMP in TG compared to WT was located 81 bp downstream of *Clu*; DMPs annotated to *Bin1* showed increased methylation in J20 TG compared to WT and were located in the gene body.

Table 6.23 – Significant DNA methylation differences at sites annotated to AD-associated genes from GWAS in rTg4510.
Δβ refers to difference in methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr10:80003740	*Abca7*	t(48) = 7.86	19.90	0.022	49
chr10:80013732	*Abca7*	t(16) =15.39	35.44	0.0041	17
chr10:80013994	*Abca7*	t(47) = 9.50	19.87	0.010	48

Table 6.24 – Significant DNA methylation differences at sites annotated to AD-associated genes from GWAS in J20.
Δβ refers to difference in methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr10:80015045	*Abca7*	t(59) = 8.73	12.20	0.014	60
chr14:65975708	*Clu*	t(20) = 7.11	-20.38	0.046	21
chr14:65981137	*Clu*	t(59) = 6.35	-16.66	0.043	60
chr14:65981630	*Clu*	t(37) = 7.73	22.03	0.026	38
chr18:32392554	*Bin1*	t(40) = 6.80	21.12	0.039	41
chr18:32392717	*Bin1*	t(47) = 7.46	18.79	0.027	48
chr18:32392818	*Bin1*	t(55) = 8.10	18.33	0.019	56

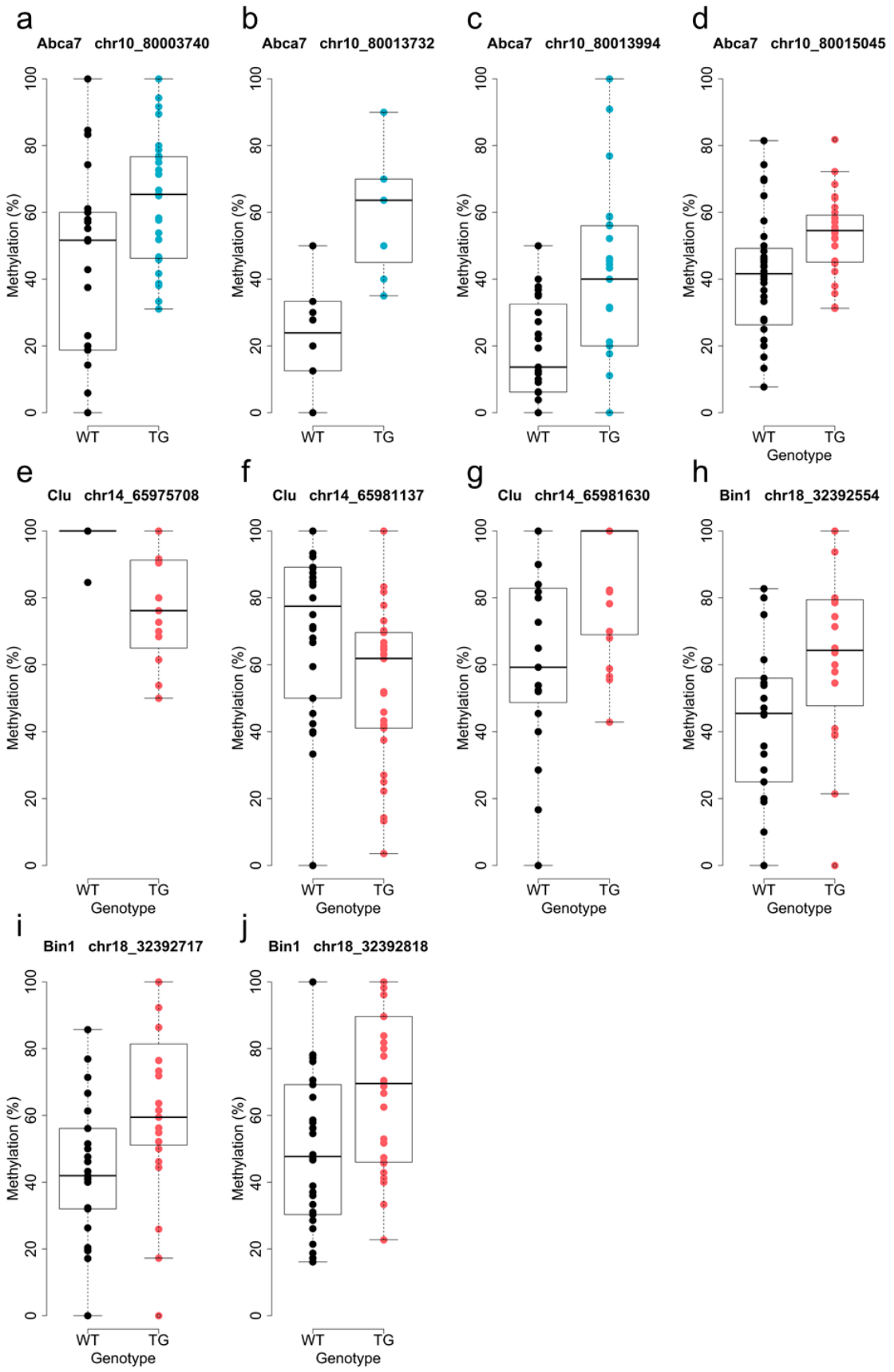


Figure 6.65 –Genotype-associated DMPs in rTg4510 and J20 ECX annotated to genes implicated in AD from GWAS.

Annotated genes and chromosome positions are indicated in each individual plot. Shown are rTg4510 transgenic (TG, blue) and J20 TG (red) female mice compared to their respective wild type littermate controls (WT, black).

6.3.8. Genotype-associated differentially methylated positions included sites annotated to genes differentially expressed in TG mice

I next explored whether DMPs were annotated to top-ranked genotype-associated DEGs identified in my RNA-seq analysis (**Chapter 4**, **Section 4.3.2** and **Section 4.3.3**). In rTg4510 mice, one DMP was annotated to *Car4* (located about 18 kb upstream of *Car4*), the most significant rTg4510 genotype-associated differentially-expressed transcript (**Chapter 4**, **Figure 4.41**). This DMP was hypermethylated in TG mice compared to WT (**Figure 6.66a**; $t(59) = 8.23$, $\Delta\beta = 16.53$, FDR = 0.017), and DNA methylation at this site was significantly positively correlated with *Car4* gene expression (**Figure 6.66b**; Pearson correlation, $r = 0.36$, $P = 0.0059$). One DMP was annotated to *Blnk* (located in the gene body), and showed decreased methylation in rTg4510 TG compared to WT (**Figure 6.67a**; $t(38) = 7.73$, $\Delta\beta = -16.11$, FDR = 0.025). At this site, there was no significant correlation between DNA methylation and gene expression (**Figure 6.67b**).

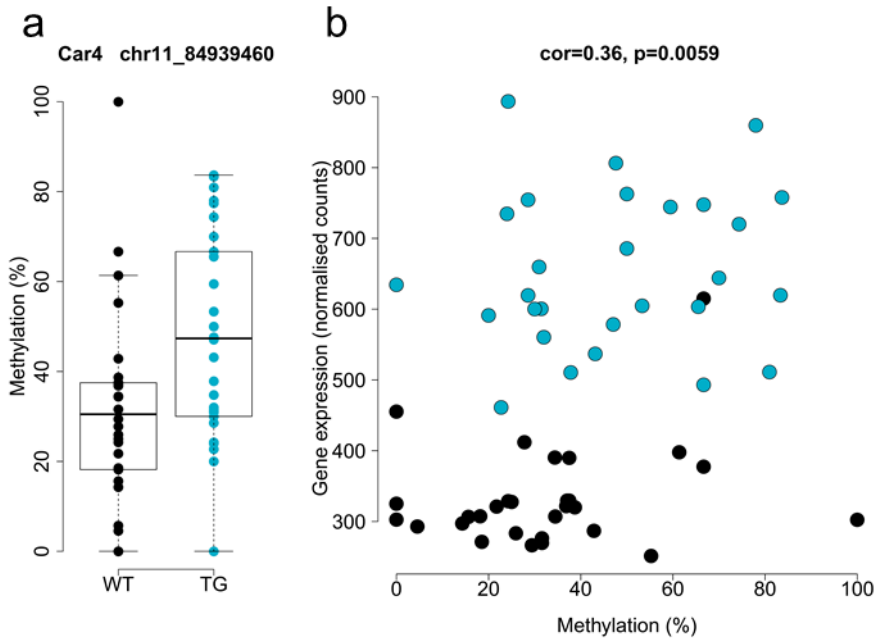


Figure 6.66 – One rTg4510 genotype-associated DMP was annotated to the top rTg4510 genotype-associated DEG Car4.
(a) Differential DNA methylation between rTg4510 WT (black) and TG (blue). **(b)** Significant ($\alpha < 0.05$) positive correlation between DNA methylation at this site and expression of the *Car4* gene.

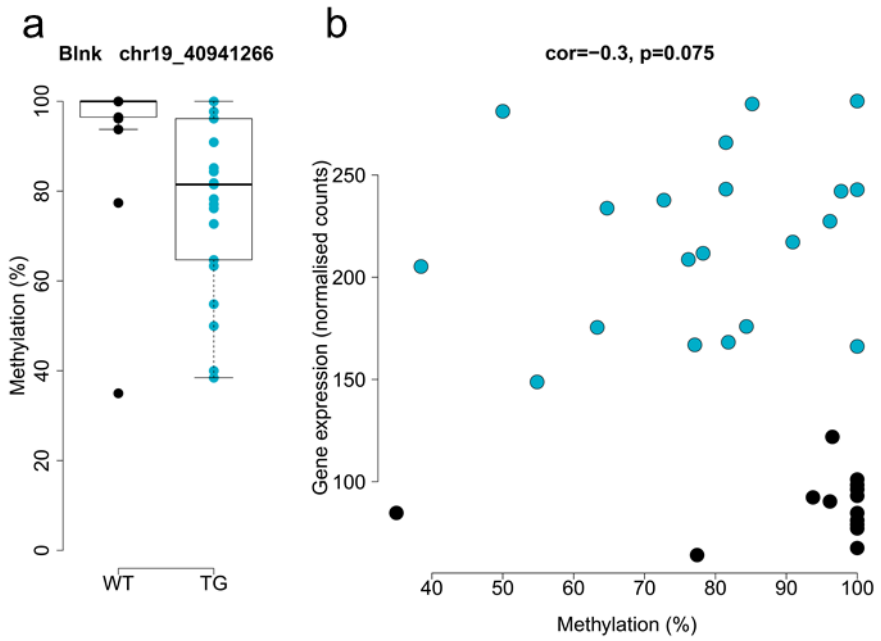


Figure 6.67 – One rTg4510 genotype-associated DMP was annotated to the rTg4510 genotype-associated DEG Blnk.
(a) Differential DNA methylation between rTg4510 WT (black) and TG (blue). **(b)** Correlation between DNA methylation at this site and expression of the *Blnk* gene.

In the J20 model, DMPs annotated to three of the four genes found to be differentially expressed in TG mice (**Chapter 4**, **Section 4.3.3**) were identified. One DMP was annotated to *Ccdc80* (located in the gene body), the most significant J20-associated differentially-expressed transcript (**Chapter 4**, **Figure 4.44**). This DMP showed decreased methylation in J20 TG mice compared to WT (**Figure 6.68a**; t(53) = 18.39, Δβ = -31.19, FDR = 0.00024), and was significantly negatively correlated with *Ccdc80* gene expression (**Figure 6.68b**; Pearson correlation, r = -0.31, *P* = 0.024). Two DMPs were annotated to *Abca8a* (**Table 6.25**, **Figure 6.69a,** and **Figure 6.69c**), both located in the gene body, with no significant correlation with gene expression (**Figure 6.69b** and **Figure 6.69d**). Two DMPs were annotated to *Htr1a* (**Table 6.26**, **Figure 6.70**) – one located about 73 kb upstream (chr13:105371119) and one located 199 kb downstream (chr13:105647323) of *Htr1a* – both hypermethylated in TG compared to WT (**Figure 6.70a** and **Figure 6.70c**), with one site being significantly negatively correlated with *Htr1a* gene expression (**Figure 6.70d**; Pearson correlation, r = -0.55, *P* = 0.0024).

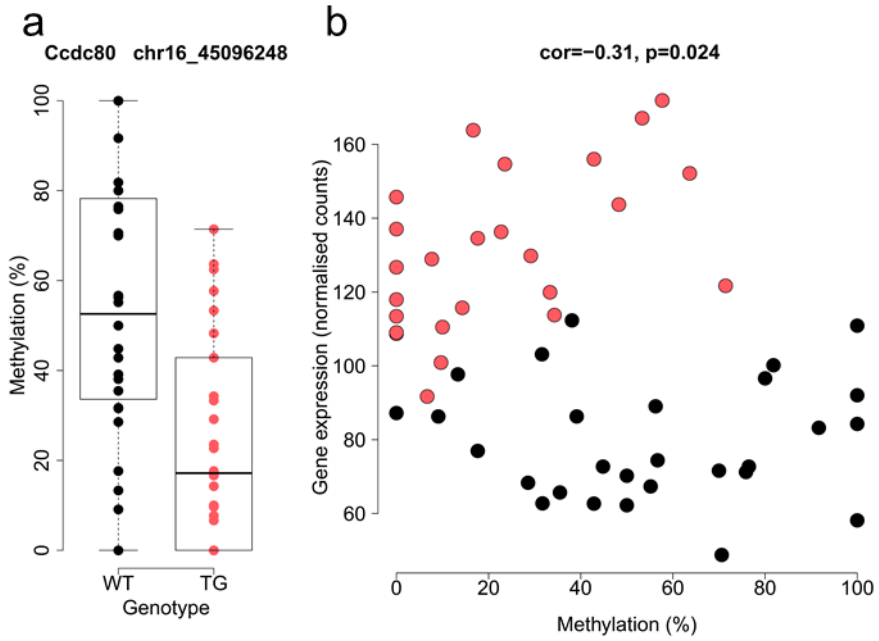


Figure 6.68 – One J20 genotype-associated DMP was annotated to the top J20 genotype-associated DEG Ccdc80.
(a) Differential DNA methylation between J20 WT (black) and TG (red). **(b)** Significant ($\alpha < 0.05$) negative correlation between DNA methylation at this site and expression of the *Ccdc80* gene.

Table 6.25 – Significant DNA methylation differences at sites annotated to Abca8a in J20.
Δβ refers to difference in methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr11:110056727	*Abca8a*	t(10) = 12.88	-29.40	0.018	11
chr11:110078353	*Abca8a*	t(57) = 8.79	-11.90	0.013	58

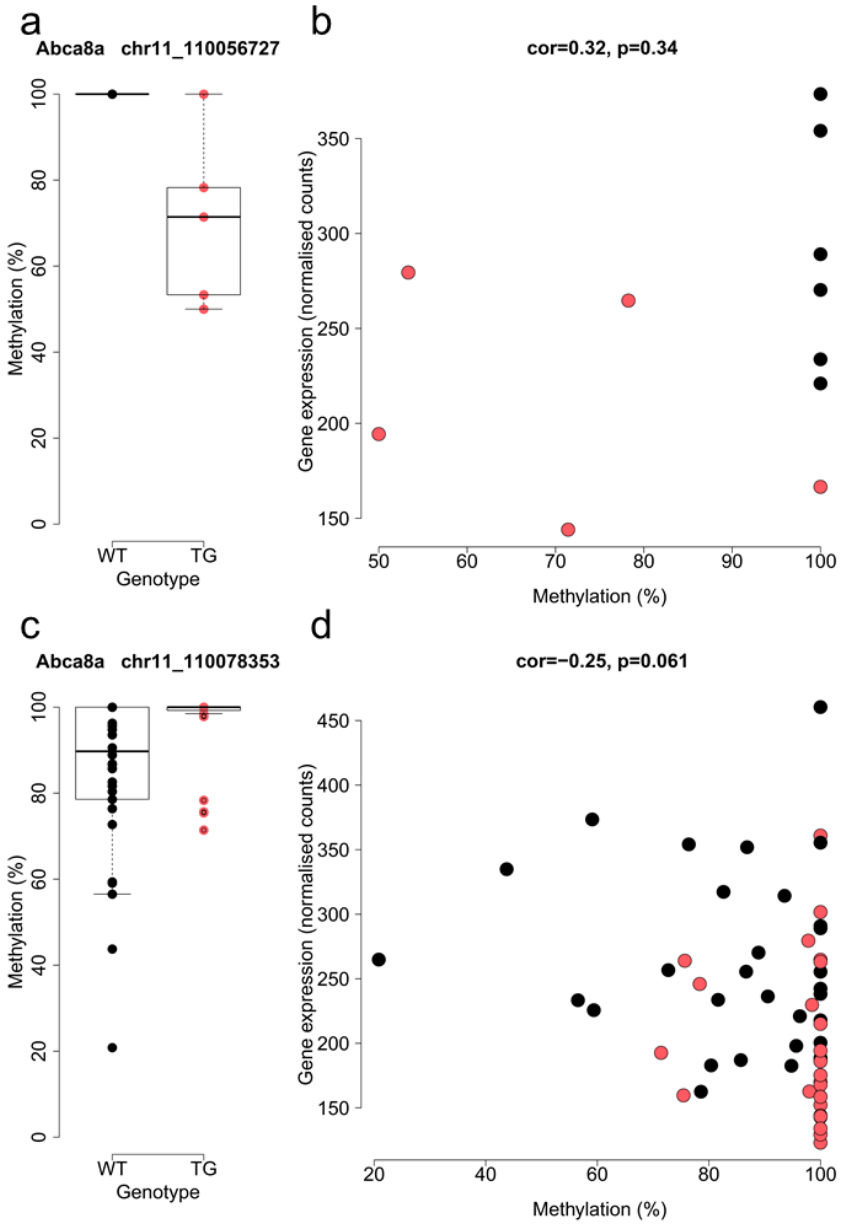


Figure 6.69 – Two J20 genotype-associated DMPs were annotated to the J20 genotype-associated DEG Abca8a.
(a, **c)** Differential DNA methylation between J20 WT (black) and TG (red). **(b**, **d)** Correlation between DNA methylation at these sites and expression of the *Abca8a* gene.

Table 6.26 – Significant DNA methylation differences at sites annotated to Htr1a in J20.
Δβ refers to difference in methylation (%) between TG and WT.

Location	Nearest gene	t-value	Δβ	FDR	N
chr13:105371119	*Htr1a*	t(62) = 8.08	5.92	0.018	63
chr13:105647323	*Htr1a*	t(27) = 6.73	7.44	0.046	28

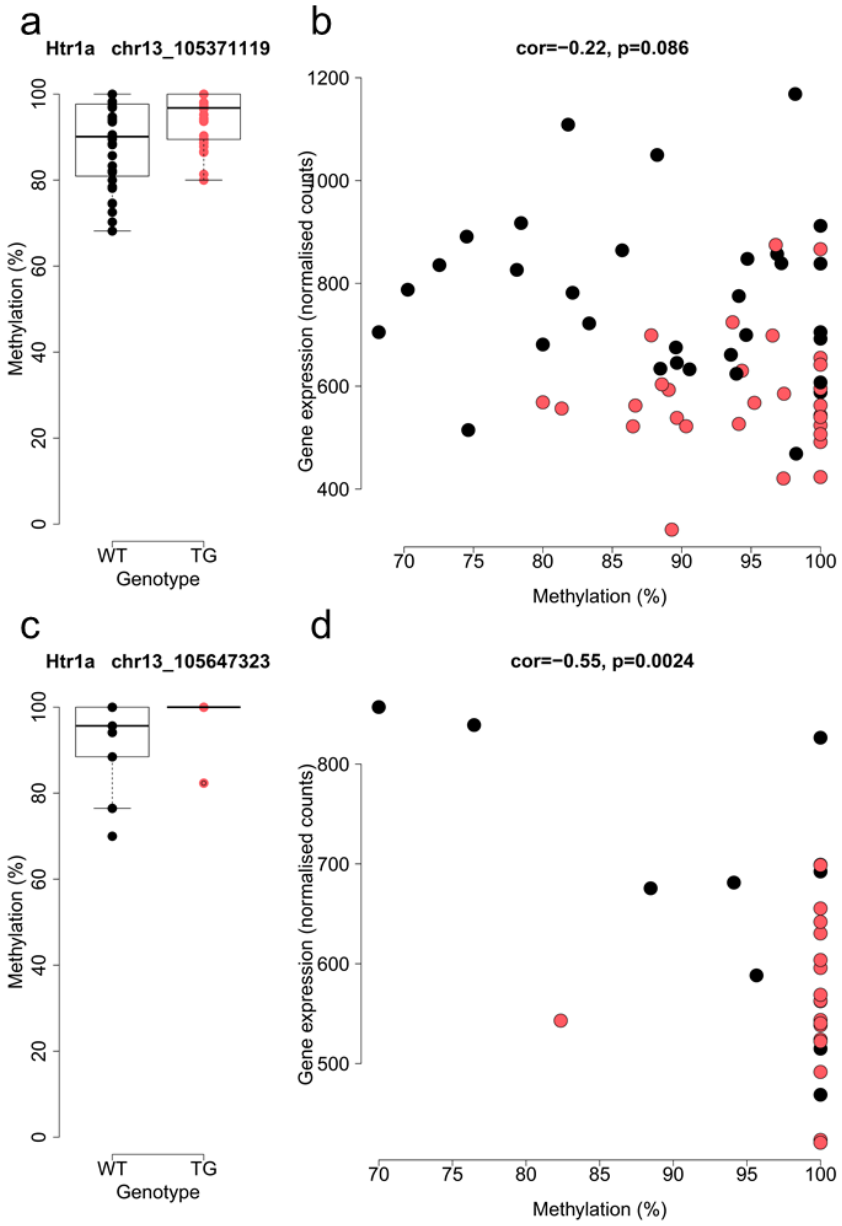


Figure 6.70 – Two J20 genotype-associated DMPs were annotated to the J20 genotype-associated DEG Htr1a.
(a, **c)** Differential DNA methylation between J20 WT (black) and TG (red). **(b)** No significant correlation between DNA methylation at this site and expression of the *Htr1a* gene. **(d)** Significant (α < 0.05) negative correlation between DNA methylation at this site and expression of the *Htr1a* gene.

6.3.9. Common differentially methylated positions and genes associated with both rTg4510 and J20 genotype

Considering sites present in both the final rTg4510 and J20 filtered RRBS datasets (see **Section 6.2.5.1**) (n = 867,012 sites), I explored the presence of significant (FDR < 0.05) DMPs associated with genotype in both models, identifying 364 overlapping DMPs (**Figure 6.71**), annotated to 342 genes. GO analysis using *Enrichr* identified disruption of biological processes (**Figure 6.72** and **Supplementary Table 24**) and molecular processes (**Figure 6.73** and **Supplementary Table 25**) including several relevant pathways related with ornithine transport, nervous system development, and lipid signalling (e.g. phosphatidylinositol). In parallel, from the DMPs identified associated with either the rTg4510 or J20 genotype, I explored common annotations and identified 5007 genes annotated to sites differentially methylated in both mouse models (**Figure 6.74**). Functional biological terms (**Figure 6.75** and **Supplementary Table 26**) and molecular function terms (**Figure 6.76** and **Supplementary Table 27**) from GO analysis using *Enrich* suggested disruption of axonal processes (e.g. axonal guidance, axonogenesis, axon extension), calcium transport, and nervous system development.

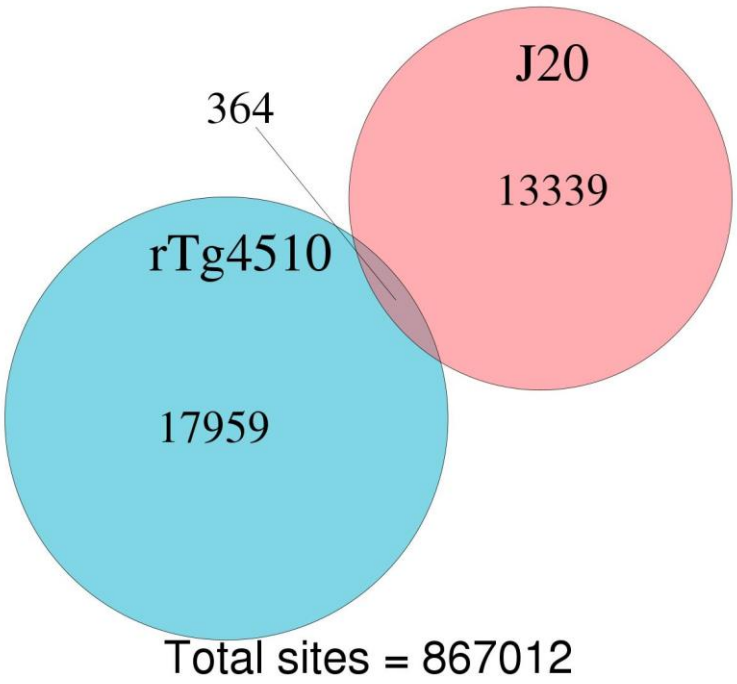


Figure 6.71 – Common sites differentially methylated associated with both rT4510 and J20 genotype.
All sites present in both rTg4510 and J20 filtered datasets were included.

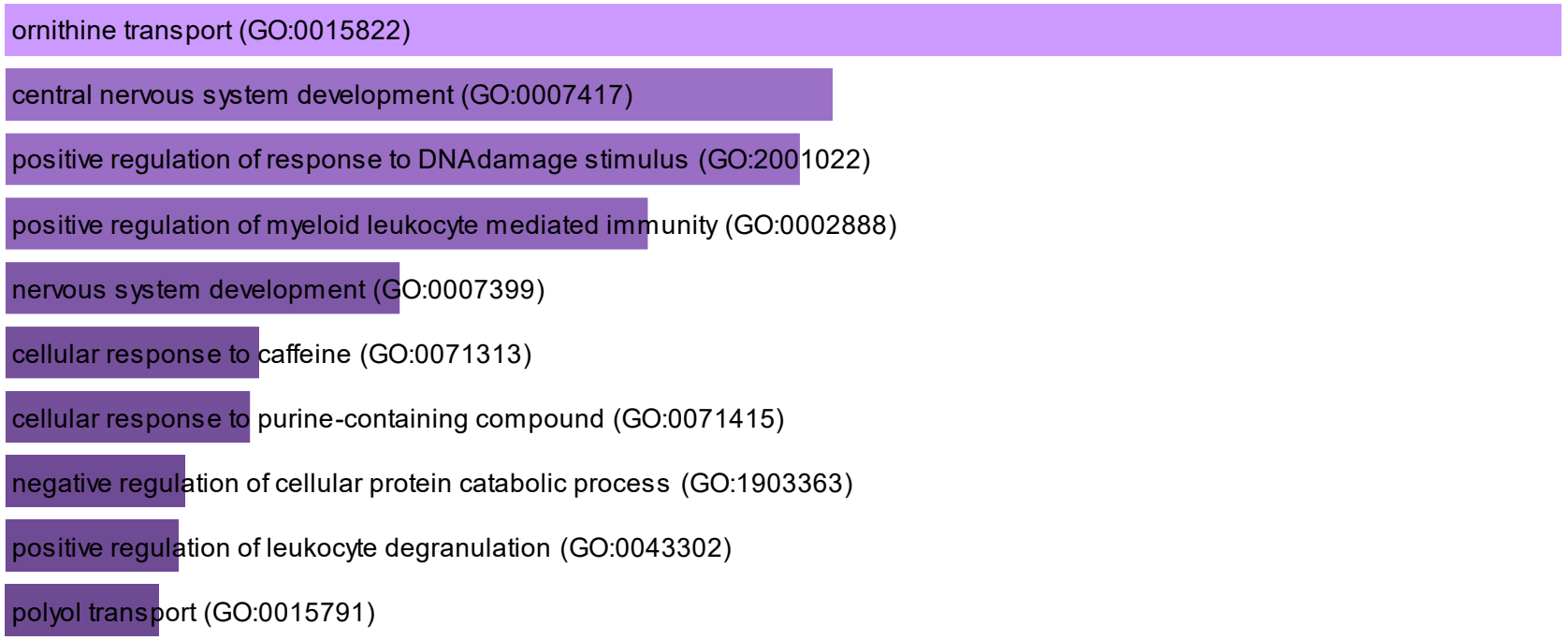


Figure 6.72 – Functional biological terms for annotated genes from common DMPs associated with genotype in both rTg4510 and J20 mouse models.
Gene ontology analysis performed using Enrichr.

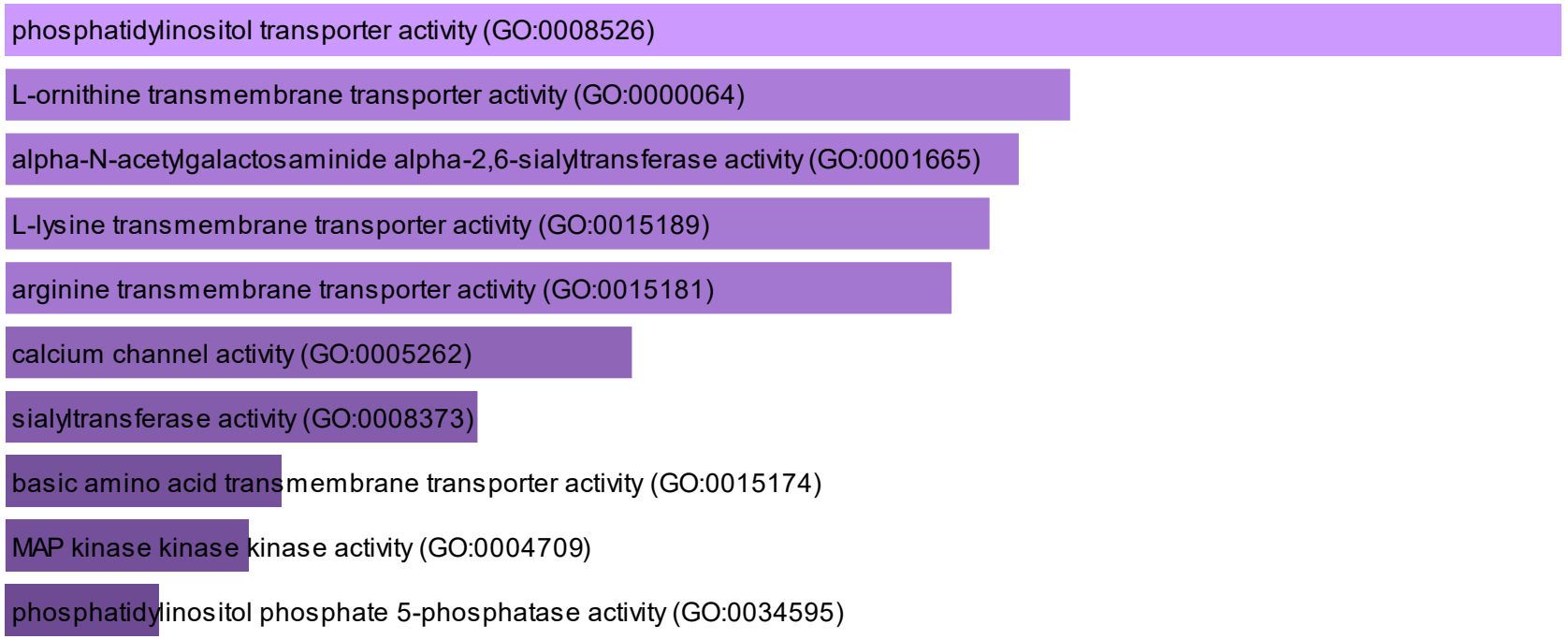


Figure 6.73 – Molecular function terms for annotated genes from common DMPs associated with genotype in both rTg4510 and J20 mouse models.
Gene ontology analysis performed using Enrichr.

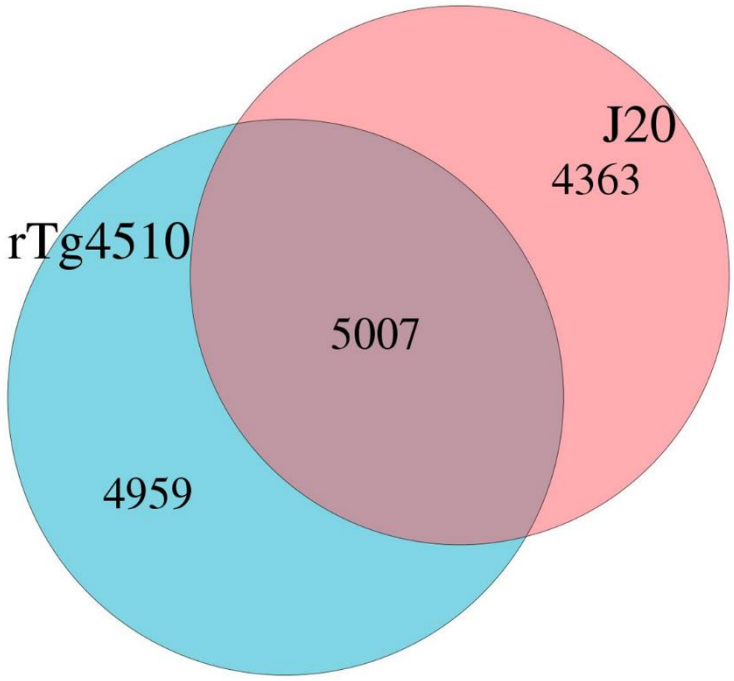


Figure 6.74 – Common annotated genes associated with both rTg4510 and J20 genotype.
All annotated genes from significant (FDR <0.05) DMPs for each dataset were included.

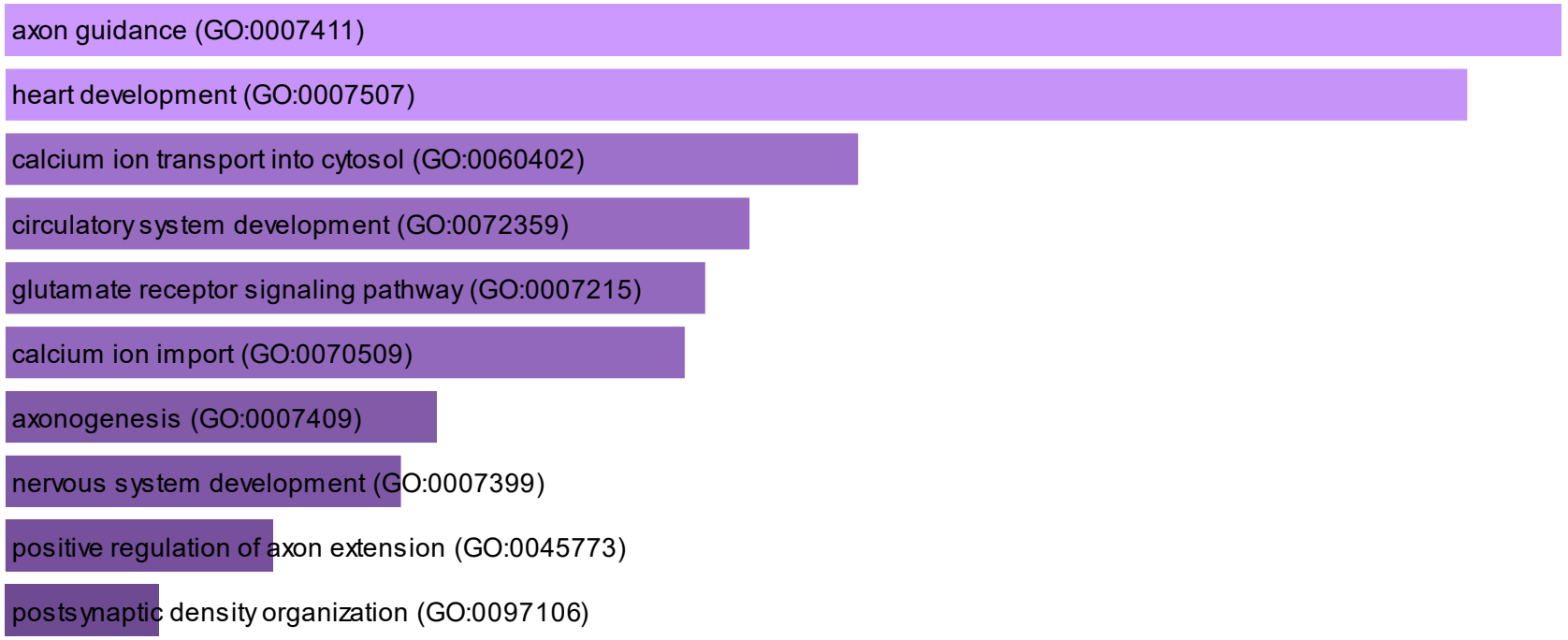


Figure 6.75 – Functional biological terms for common annotated genes associated with genotype in both rTg4510 and J20 mouse models.
Gene ontology analysis performed using Enrichr.

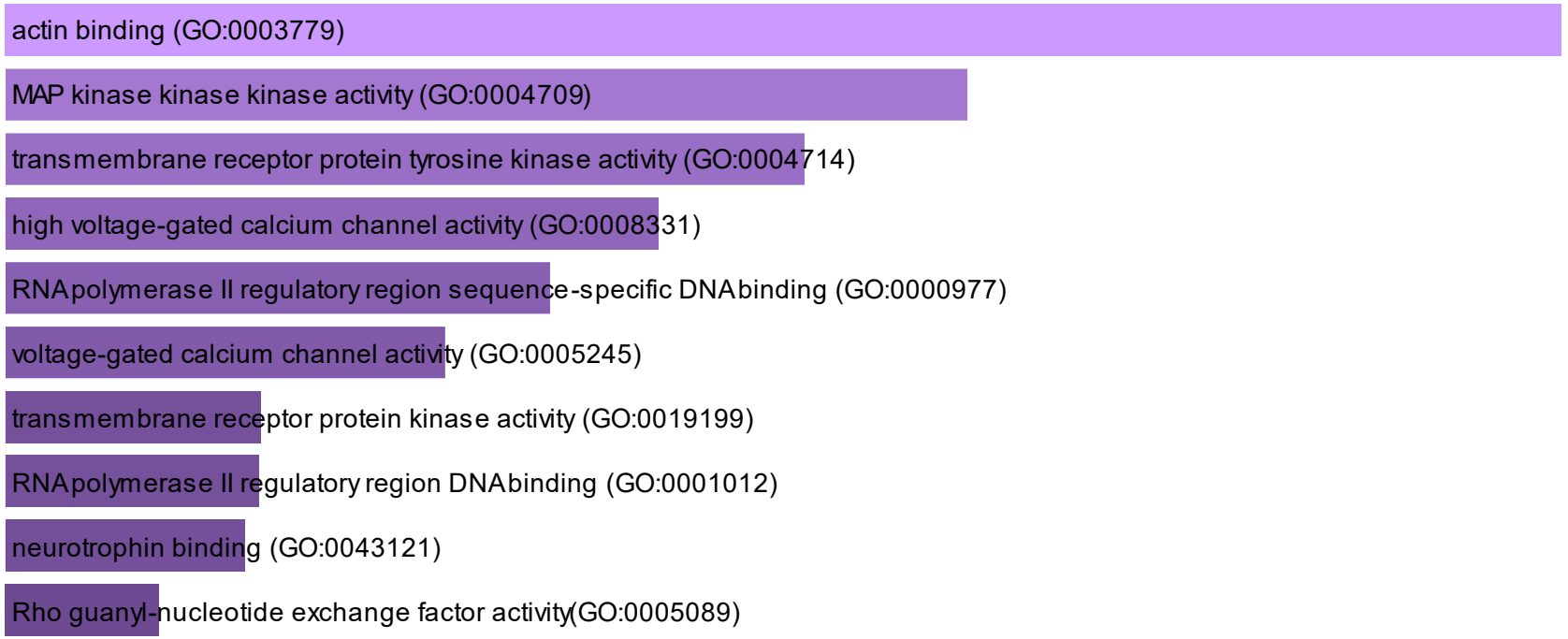


Figure 6.76 – Molecular function terms for common annotated genes associated with genotype in both rTg4510 and J20 mouse models.
Gene ontology analysis performed using Enrichr.

6.4. Discussion

6.4.1 Overview of results

In this chapter I have described the first sequencing-based analysis of genome-wide changes in DNA methylation in mouse models of AD pathology. I identified widespread DNA methylation differences in the entorhinal cortex from TG mice compared to WT, in both mouse models. These were more dramatic in the rTg4510 model, following a common theme in all my results – from neuropathology (**Chapter 3**), to gene expression (**Chapter 4**), and global DNA methylation (**Chapter 5**). In both rTg4510 and J20 models, the majority of DMPs were hypermethylated in TG compared to WT mice. Although this might appear to contrast with my finding of progressive global hypomethylation identified in the hippocampus of rTg4510 and J20 TG mice using immunohistochemistry (**Chapter 5**), this potentially reflects distinct changes in different brain regions. Also, the use of immunohistochemistry does not enable the analysis of DNA methylation at specific sites, and it captures variation by regions not assessed by RRBS (e.g. in intergenic regions covering most of the genome); while RRBS assesses mostly promoters, my immunohistochemistry results capture global shifts across all genomic regions. It also reflects results from studies of other complex disorders known to involve epigenetic dysfunction. Cancer, for example, is characterised by global DNA hypomethylation in conjunction with hypermethylation of specific sites (such as promoters of tumour suppressor genes and homeobox genes) (Ehrlich, 2009). As discussed in detail in the previous sections, the list of TG-associated DMPs overlapped with regions nominated from human AD EWAS and genetic studies, and also with genes identified in my gene expression RNA-seq study (**Chapter 4**).

The observation of hypomethylation in J20 TG mice compared to WT in the gene body of *Tmprss7* – that encodes for the protease matriptase 3 (see **Section 6.3.4** for details) – might be a response to the accumulating Aβ in this mouse model. Matriptase 3 has 31% amino acid identity with matriptase, a protein that has been shown to interact with and cleave APP, reducing its proteolytic cleavage to Aβ (Lanchec et al., 2017). Complex proteolytic pathways responsible for the generation and clearance of the abnormally folded protein fragments characteristic of AD (i.e., Aβ and tau) have been widely described (De Strooper,

2010); however, despite the recognition that other classes of proteases besides secretases can cleave APP, their roles in AD are yet to be fully understood. Furthermore, proteases produced by neurons, microglia, and astrocytes, and involved in the proteolysis of Aβ, have been shown to be active players in the neuroinflammatory processes observed in AD (Newcombe et al., 2018). For those reasons, the investigation of a possible role for matriptase 3 in Aβ processing would be an important follow-up for this study, particularly in parallel with the amyloid-associated increase in expression of the microglia-associated genes (i.e. neuroinflammatory) observed in J20 TG mice (**Chapter 4**).

The top DMP in the rTg4510 mouse model was located about 8kb upstream of the gene encoding the metabolic enzyme UGT2B37 (see **Section 6.3.4** for details); this site was hypermethylated in TG mice compared to WT. Growing evidence from the last decades support a link between central and peripheral metabolic dysfunction and AD, and drugs that modulate cell metabolism are currently being investigated for the treatment of AD (Clarke et al., 2018). Furthermore, a connection between AD and type 2 diabetes and obesity has been extensively described in recent years (Shalev and Arbuckle, 2017), with chronic inflammation, oxidative stress, reduced neuronal sensitivity to insulin, as well as impaired glucose uptake observed in this context (Clarke et al., 2018). The potential change in the regulation of *Ugt2b37* at the gene expression and protein levels as a result of the observed hypermethylated site may indicate a role for UGT2B37 in response to accumulation of tau and should be further investigated in the context of metabolic dysregulation. Interestingly, the transport of ornithine, a naturally occurring non-proteinogenic amino acid involved in the urea cycle, was the top functional biological term enriched amongst DMPs commonly identified in both mouse models (**Section 6.3.9**). Amino acids such as ornithine are involved in many crucial metabolic and signalling pathways in the cell; additionally, ornithine levels have been shown to be decreased in human AD brains, cerebrospinal fluid (CSF), and plasma (Griffin and Bradshaw, 2017, Fonteh et al., 2007, Shetty et al., 1996). Therefore, and given the limited literature available at this time, the study of the effects of altered amino acid metabolism in the context of AD is an area that requires additional research, particularly as it has been hypothesised as a potential compensatory response to the decreased neuronal glucose metabolism observed in AD brains (Griffin and Bradshaw, 2017).

6.4.2. Limitations

At this stage, I focused solely on genotype effects as I did not consider my study to be adequately powered to look at progressive changes in DNA methylation at individual CpG sites using the same statistical model ($y = Genotype + Age + Genotype * Age$) employed in the other chapters. In contrast to gene expression, the evaluation of DNA methylation is much more complex: 1) more than one methylation site can be associated to the same gene, i.e., DNA methylation may operate over broader regions, which means that looking at individual sites only might not be appropriate; 2) the presence of DNA methylation in different genomic contexts (transcriptional start sites, in gene bodies, at regulatory elements and at repeated sequences) can have very different relationships with gene regulation (Jones, 2012) (e.g. the presence of DNA methylation on gene regulatory sequences, such as promoters or enhancers, typically represses expression (Dor and Cedar, 2018), whilst the presence of DNA methylation in the gene body often promotes gene expression (Jones, 2012)); 3) sites are usually annotated to genes considering the distance between them, which can be quite arbitrary as their relationship is still not fully understood. For all these reasons, analysis packages (preferentially specific for RRBS data) that take this into account are more suitable than a 'simple' straightforward statistical analysis, which will be investigated in detail for the next steps of the statistical analysis (**Section 6.4.3**). Other analysis-related limitations to consider are associated with how RRBS works (where only CpG-rich regions of the genome are sequenced), as well as the filtering that we have performed – not every site of every sample has methylation data, with some sites only having measures in a few samples. Given the high number of significant DMPs and respective annotated genes identified in both mouse models, it is possible that the FDR threshold used might not be stringent enough to accurately exclude false positives; although this issue has been explored in detail during the last decade for genome-wide studies in human (Fadista et al., 2016), it has not been done for genome-wide studies in mice yet. All these aspects need to be (re)considered for future analyses. As in our gene expression RNA-seq experiments, we have used bulk tissue for RRBS DNA methylation assessments, and DNA methylation changes can be cell-specific.

Future work should address these issues by profiling additional samples at enhanced depth in purified cell populations.

6.4.3. Future directions

Despite the acknowledged limitations of these data, our preliminary results seem promising. In the next year, I will focus on developing novel computational methods to assess progressive DNA methylation changes in rTg4510 and J20 mouse models, as well as integrate these changes with gene expression changes (**Chapter 4**) and neuropathology data (**Chapter 3**) from the same individual mice (**Figure 6.77**). A few approaches have been described by others (Wreczycka et al., 2017, Yong et al., 2016), with a couple of methods showing encouraging strategies for systematic multi-omic data integration (Chen et al., 2018, Hansen et al., 2012, Park et al., 2014).

Figure 6.77 – Computational pipeline for genome-wide bisulfite sequencing data analysis.
Reads from bisulfite sequencing are first aligned to the reference genome. The alignment data may be visualised in different tracks for comparison. After methylation calling, the bulk methylation level and genome-wide methylation level can be calculated and plotted, and differentially methylated regions (DMRs) can be determined. DNA methylation data can be integrated with gene expression data (differentially expressed genes, DEGs) to delineate the regulatory role of DNA methylation. Figure and legend adapted from Yong et al. (2016).

Chapter 7. General discussion

7.1. Summary of results

In this PhD I have directly assessed transcriptomic and epigenomic variation associated with the progression of AD neuropathology in the mouse brain. I used cutting-edge genome-scale approaches to assess tissue dissected from two well validated transgenic mouse models – the rTg4510 line (to investigate genomic signatures of tau pathology) (Ramsden et al., 2005, Santacruz et al., 2005) and the J20 line (to investigate genomic signatures of amyloid pathology) (Mucke et al., 2000, Palop et al., 2007) – and related the identified genomic changes to histopathology information from the same mice.

In **Chapter 2** I outline the general methods common across the individual data chapters (**Chapters 4-6**) including all animal procedures, sample processing, and optimisation procedures for nucleic acid isolation and associated quality control. The experimental procedures specific to each individual chapter are described in detail in the methods sections of the corresponding chapter.

In **Chapter 3** I describe the progression of tau pathology in rTg4510 mice and of amyloid pathology in J20 mice, quantified using immunohistochemistry across multiple regions of the mouse brain.

In **Chapter 4** I describe an analysis of transcriptional variation in the entorhinal cortex associated with the development of both tau and amyloid pathology in rTg4510 and J20 mice, respectively. Using systems-level network analyses, I identified gene co-expression clusters of genes (or modules), related these to neuropathological burden from the same mice, and explored transcriptomic changes identified in the mouse models in datasets from human AD brains.

In **Chapter 5** I describe global changes in levels of two DNA modifications (5mC and 5mC) across multiple brain regions from tissue sections from rTg4510 and J20 mice using immunohistochemistry.

Finally, in **Chapter 6** I describe the quantification of genome-wide patterns of DNA methylation at single-base resolution in the entorhinal cortex from rTg4510 and J20 mice using reduced-representation bisulfite sequencing (RRBS).

Taking a global view of my results, a common theme is revealed across all the data modalities assessed (neuropathology (**Chapter 3**), gene expression (**Chapter 4**), global DNA modifications (**Chapter 5**), and genome-wide DNA methylation (**Chapter 6**)); I identified consistently larger changes in the rTg4510 samples compared to samples from the J20 model. This potentially reflects the slower and later accumulation of pathology in J20 compared to rTg4510 mice (described in detail in **Section 1.5.1.1** and **Section 1.5.1.2** of **Chapter 1** and discussed in **Section 4.3.5** of **Chapter 4**). Moreover, these findings were expected to a certain degree given the potential absence of neuronal cell loss (particularly in the entorhinal cortex) in J20 mice (Wright et al., 2013), in contrast widely present in rTg4510 mice (Blackmore et al., 2017, Booth et al., 2016, Ramsden et al., 2005, Santacruz et al., 2005).

My findings of altered gene expression in rTg4510 and J20 mice suggest upregulation of immune system genes with progressive AD pathology. In particular, I observed a common increase in expression of microglial genes – i.e. *Cst7* and *Itgax* (Cd11c) (see **Chapter 4**, **Section 4.3.5**) – associated with both tau and amyloid pathology. The functional role of microglia in AD is not yet fully understood, with both detrimental and beneficial effects resulting from microglia activation reported (Hansen et al., 2018). Currently the most accepted hypothesis is that beneficial consequences of microglial activity are important in the early stages of disease, whilst detrimental effects may take over in later stages. Microglia can engulf and remove abnormal neurons and synapses, likely via a complement-dependent mechanism; they can also exacerbate tau pathology and secrete inflammatory factors that can injure neurons directly or via activation of neurotoxic astrocytes (Hansen et al., 2018). Interestingly, the extracellular amyloid deposits typical of AD, mainly composed of aggregated Aβ, have been shown to be surrounded by activated microglia that express activation markers and pro-inflammatory proteins (Kamphuis et al., 2016). The inflammatory environment created by activated microglia has been suggested to increase Aβ production; however, there is also evidence showing that microglia-mediated inflammation reduces amyloid plaque deposition via upregulation of phagocytic and degradation pathways (Zuroff et al., 2017). The increased amounts of soluble and insoluble Aβ observed in human AD are hypothesised to result from an

imbalance between Aβ production and clearance (Mawuenyega et al., 2010). Whilst the literature indicates that the rare forms of EOAD caused by genetic alterations involve an increase in the production of Aβ, increased Aβ in sporadic LOAD has been suggested to be due to impairments in its clearance (Baranello et al., 2015, Mawuenyega et al., 2010, Ries and Sastre, 2016). Although the mechanisms involved in impaired clearance of Aβ in AD are believed to occur through interactions with ApoE4, decreased catabolism of Aβ via reduced proteolytic enzymes, impaired transport across the blood-brain barrier and impaired CSF transport (Mawuenyega et al., 2010), growing evidence supports a key role of microglia and monocyte-derived macrophages in this process (Mawuenyega et al., 2010, Zuroff et al., 2017). The accumulation of misfolded Aβ as a result of an imbalance between its production and removal may in turn stimulate synaptic and neuronal loss, as well as neuroinflammation (Zuroff et al., 2017). In fact, in addition to neuroinflammation, a large proportion of the risk genes for LOAD are associated with Aβ processing or trafficking, and with immunological processes related to myeloid cell-mediated Aβ clearance (Zuroff et al., 2017). For example, polymorphisms in the genes *CD33* and *TREM2*, shown to increase susceptibility to AD (**Chapter 1**, **Section 1.2**), have been associated with impairments in microglial- and macrophage-mediated phagocytosis of Aβ. Nevertheless, the question of whether the inflammation and myeloid cell disruption observed in AD are drivers for the disease or are a consequence of the neuropathology remains unanswered. Interestingly, the expression of many of these genes (e.g. *Cd33* and *Trem2*) was disrupted in my rTg4510 dataset, as a result of tau accumulation, which I did not observe in the J20 mice potentially due to the reasons stated elsewhere (**Chapter 4**, **Section 4.3.5**).

Together, my results suggest common pathways and mechanisms of response to the presence of tau and amyloid pathology. With the dramatic increase in cerebral Aβ beginning as early as twenty years prior to symptom manifestation (Zuroff et al., 2017), and the currently most accepted hypothesis stating tau pathology as downstream of amyloid (Selkoe and Hardy, 2016), targeting Aβ (aiming to reduce its levels) by promoting its clearance, for example, is a potential strategy for the treatment of AD. Currently, the study of the role(s) of microglia in AD is a hot topic within the scientific community, with a plethora of recent studies looking at sorted microglia or even single cells / nuclei; studies in the upcoming

years should clarify the role of these neural cells in Aβ and tau processing and clearance, as well as in overall AD toxicity. Building on the results of my thesis, it will be important to integrate the various layers of genomic regulation (e.g. DNA methylation and gene expression, as well as histone modifications), and protein levels, particularly in sorted microglia and this will be the focus of my ongoing research.

Another (potentially related) reason for insufficient clearance, and consequential accumulation and spread of oligomeric forms of neurotoxic proteins in AD, is related to the deficient extracellular clearance by proteases (Boland et al., 2018). Curiously, I observed hypomethylation in a region annotated to the gene that encodes the protease TMPRSS7 (or matriptase 3) in J20 TG compared to WT (see **Chapter 6**, **Section 6.3.4** and **Section 6.4.1**). It is important to perform functional follow-up studies (e.g. using cell culture approaches and/or knockout mouse models) aiming at exploring the role of matriptase 3 in Aβ processing. Additionally, the incorporation of CRISPR/Cas9 gene editing technology in these studies could be an interesting approach to take this further, with novel applications of this technology enabling the induction of epigenetic changes to facilitate the assessment of DNA methylation differences in this context (Liu et al., 2016, Liu et al., 2018).

7.2. Limitations

This study has a number of important limitations that should be considered when interpreting the results described in each of the data chapters.

First, to minimise the heterogeneity in my analysis, I only profiled female mice. However, a number of sex differences have been previously reported for these models, with females demonstrating elevated and more progressive pathology than males (Blackmore et al., 2017, Yue et al., 2011). Future work should focus on examining the extent to which the genomic changes identified in my work are consistent between male and female mice.

Second, my analysis was performed on bulk brain tissue, comprising a mix of different neural cell-types. Future work should focus on identifying changes specific to the different brain cell-types. Our laboratory, for example, has been

developing methods to purify nuclei from neurons and oligodendrocytes using the nuclear markers NeuN and Sox10, respectively, in conjunction with fluorescence-activated nuclei sorting (FANS) (Jeffries and Mill, 2017).

Third, compared to the rTg4510 model, relatively few changes were observed in J20 mice, potentially reflecting the slower and later accumulation of pathology (Harris et al., 2010), as well as the potential absence of neurodegeneration, in the entorhinal cortex in this model; future work looking at changes at the genomic level should focus on the additional analysis of other brain regions more directly affected in the early stages of amyloid pathology in J20 mice.

Fourth, I cannot exclude the possibility that my analysis did not cover an appropriate aging interval in order to being able to identify such changes; future work covering a more expanded aging interval, with samples collected from earlier and later time points, should be considered. In particular, an earlier baseline sample would enable a more accurate temporal assessment of genomic changes across the life course to be determined.

Finally, my PhD project only used two mouse models, and as discussed in **Chapter 1** (**Section 1.5.1**) other models of AD pathology are currently available. In order to validate my findings, future work should include other mouse models of tau or amyloid pathology, as well as models with characterised by both types of AD-associated neuropathology. In addition, it will be important to profile samples from knock-in mouse models, which are characterised by the accumulation of pathology without some of the additional effects observed in overexpression models.

7.3. Current challenges and future directions

Although studies of gene regulation in AD have demonstrated promising findings, more studies, particularly at the genome-wide level, are still required. There are however a number of challenges and caveats associated with performing such studies, and these need to be adequately addressed to allow the identification of novel genes, neurobiological pathways and pharmacological targets in AD.

7.3.1. Limitations on the use of AD animal models

A key issue with studying genomic variation in diseases of inaccessible tissues, such as the brain, is that human studies are restricted to post-mortem samples, which usually represent end-stage disease and so it is difficult to establish causality. Although murine models have utility for understanding disease progression, as discussed earlier (see **Section 1.5**), these also have their own caveats for studying a disease such as AD. For example, mouse models do not 'naturally' develop Aβ plaques, nor do they normally display overt neurodegeneration. In addition, the transgenic models described to date (see **Section 1.5.1**) are based on mutations identified in FAD and are not therefore really accurate models of SAD, which has both polygenic and environmental contributions to disease aetiology (Alzheimer's-Association, 2018, Selkoe, 2011) (see **Section 1.2** and **Section 1.4** in **Chapter 1**).

7.3.2. Cellular heterogeneity in the brain

Different tissues and cell types exhibit unique transcriptomic and epigenetic profiles, and therefore discriminating between disease-specific changes in cellular composition in heterogeneous tissues can be quite challenging. As discussed previously (**Chapter 1**, **Section 1.1**), some regions of the brain are more affected than others by AD neuropathology, with specific genomic changes reported in distinct brain regions and for specific cell types (see **Section 1.3**, **Section 1.4**, and **Section 1.5.2**). In addition, AD is characterised by progressive changes in brain cell composition, including neuronal loss due to neurodegeneration, an increased number of activated microglia (microgliosis), and the abnormal hypertrophy of astrocytes (reactive astrogliosis), which is

particularly true in brain regions highly susceptible to neuropathology such as the hippocampus and entorhinal cortex (see **Chapter 1**, **Section 1.1**). The vast majority of AD genomic studies, including the studies performed for this thesis, have used 'bulk' brain tissue; therefore, it becomes difficult to draw conclusions about cell-specific AD-associated genomic changes. Furthermore, it is plausible that some important disease associated genomic variation is missed if genomic changes in one cell type might be masked by opposing changes in a different cell type. The optimal analysis approach would use single cell profiling to investigate epigenomic and transcriptomic changes in individual cells; however, the costs associated with performing such experiments with adequate power to detect disease-associated differences are still prohibitive. Nonetheless, studies utilising methods aimed at isolating specific cell populations, utilising LCM, FACS, magnetic affinity cell sorting (MACS), and density gradients, are now feasible.

7.3.3. Beyond DNA methylation: incorporating other epigenetic modifications

An issue with current DNA methylation studies, including the published EWAS analyses of human AD and my RRBS analysis presented in this thesis (**Chapter 6**), is that they have utilised sodium bisulfite conversion to enable quantification of DNA methylation across the genome. This method converts unmodified cytosines to uracil (explained in detail in **Chapter 6**, **Section 6.1**), whilst both 5-methylcytosines (DNA methylation) and 5-hydroxymethylcytosines (DNA hydroxymethylation; see **Chapter 1**, **Section 1.4.1** for details) are protected from the conversion; the presence of a DNA modification can be then determined in downstream analyses as a difference in the sequence between modified [C] and unmodified [T] DNA. In this way, these approaches cannot differentiate between DNA methylation and hydroxymethylation; given the relatively high abundance of 5hmC in the brain this may represent a major confounder in current analyses. A recent adaptation to the bisulfite conversion protocol, where the DNA is oxidised prior to bisulfite treatment, converts 5hmC to 5-formylcytosine (5fC) (which is converted to uracil), meaning that data generated from oxidative bisulfite treated DNA gives a measure of 'true' DNA methylation (5mC). Furthermore, by running bisulfite and oxidative bisulfite treatment in parallel one can calculate 5hmC levels through the subtraction of 5mC from total modifications. In conjunction with array- or sequencing- based assays this allows for the evaluation of 5mC and 5hmC in a genome-wide fashion (Lunnon et al., 2016, Zhao et al., 2017). Although this method has not yet been applied to studies in AD, future analyses of DNA modifications in human and mouse brain tissue should employ these novel approaches. Other epigenetic processes other than DNA modifications, such as histone modifications, should also be profiled in future work; for example, H3K27ac – a robust mark of active enhancers/promoters (Creyghton et al., 2010, Wang et al., 2008) that is strongly correlated with gene expression and transcription factor binding (Kumar et al., 2013) has been shown to be dysregulated in human AD entorhinal cortex (Marzi et al., 2018).

7.3.4. Sequencing-based approaches as the way forward

Many of the existing genomic studies of AD pathology have utilised microarray-based technologies, particularly in transcriptomic studies (discussed in **Section 1.3** and **Section 1.5.2** of **Chapter 1**). Although microarrays are a cost-effective method for analysing the transcriptome and epigenome, especially when profiling large numbers of samples, their content is relatively constrained, and they only cover a small proportion of known transcripts (or DNA modification sites). As such, pathology-associated genomic variation in other regions in the genome will not be detected. Sequencing-based technologies are likely to offer a more optimal approach for uncovering AD-associated genomic variation. For example, RNA sequencing (RNA-seq), the approach that I used to evaluate transcriptomic changes in this thesis (**Chapter 4**), enables the quantification of all expressed transcripts, without the limitations/assumptions associated with pre-designed arrays (Sahraeian et al., 2017). Likewise, whole genome bisulfite sequencing (WGBS-Seq), allows the identification of DNA methylation changes at single nucleotide resolution across the entire genome, including at non-CpG sites (Yong et al., 2016); however, the depth of sequencing required to accurately profile DNA methylation using this approach is still associated with prohibitively high costs to enable a large number of samples to be profiled (Ziller et al., 2015). In this study I used RRBS as a compromise approach to enable the cost-effective profiling of DNA methylation at single base resolution (Gu et al., 2011, Meissner et al., 2005) (**Chapter 6**). Looking to the future, third generation (or long-read) sequencing technologies, currently under active development, will enable the simultaneous identification of both genetic and epigenetic variation in the same sample (Hirst and Marra, 2010).

7.3.5. The ‘big data’ era

Our understanding of the role of altered gene regulation in AD pathology has expanded considerably in the last few years; novel technological advances and the reduction in the cost of genomic sequencing mean that even bigger datasets are likely to be on the horizon. The increase in “multi-omics” analyses, including the integration of different data types using network analyses will require extensive collaboration between multidisciplinary teams and the development of

novel analytical strategies. We are currently in the ‘big data’ era; where large ‘omics’ data-sets with extensive clinical and demographic information have been amassed, which offers extraordinary possibilities to better understand AD aetiology. Although these studies are still in their infancy, the integration of genetic, epigenetic, transcriptomic, proteomic and clinical data, from the same individuals will enable a fully comprehensive analysis of dysfunctional mechanisms and pathways involved in AD. Further, by comparing similar data-sets generated in mouse models to the human AD data-sets a more comprehensive analyses of disease mechanisms will be achievable. Mouse models will continue to be important as they overcome a number of caveats associated with studying human samples; the integration of human and mouse genomic data will allow the temporal/spatial analysis of disease loci and ultimately nominate novel targets for therapeutic intervention. A handful of recent papers have pioneered this approach, exploring findings from AD mouse models in human data (Gjoneska et al., 2015, Wang et al., 2017).

7.3.6. Beyond AD

It should be noted that many other diseases besides AD are characterised by the symptoms and pathology associated with dementia. Other forms of dementia include VaD, DLB, HD, PD dementia (PDD) and FTD. Although the field has mainly focused on AD, translational genomic studies in other dementias are also needed especially given the likely aetiological overlap between them. Moreover, most studies have focused on regions in the temporal lobe (particularly the hippocampus) and the frontal lobe (mostly the prefrontal cortex), and studies assessing other brain regions will also be important. Ultimately, if these cross-region analyses were performed in the same individuals, it would allow the mapping of spatial genomic changes in the development of neuropathology.

7.4. Future research directions

The results presented in this PhD represent the foundation for a broader and larger follow-up study – recently funded by Alzheimer's Research UK (ARUK) – in which we will expand the diversity of mouse models to be profiled. In addition to profiling gene expression and DNA methylation, we will include long read sequencing (to assess gene isoforms and splicing), other DNA modifications (i.e. DNA hydroxymethylation), as well as histone modifications (e.g. H3K27ac). As in this study, genomic profiling data (RNA-seq, DNA modifications and histone modifications) will be integrated with histopathological, biochemical and neurophysiological phenotypes across multiple brain regions through a systems-level network analysis. Additionally, we will tackle some of the challenges described in the previous sections; for example, we will purify specific neural cell types to test the cellular specificity of pathology-associated genomic changes.

Importantly, epigenetic and transcriptional modifications represent potential novel therapeutic targets, particularly considering that many are dynamic, modifiable and potentially reversible. Compounds that can target these modifications might potentially help reverse abnormal gene expression typical of disease. Given that there are currently no therapeutics in AD that can reverse, or slow, the development of pathology, the possibility of specifically targeting epigenetic modifications to treat AD is very promising. Together with our collaborators at Eli Lilly, our ultimate goal is the identification of pathways involved in AD-associated neuropathology that can be potential targets for treatment and we will undertake targeted studies of specific test compounds.

7.5. Conclusions

In conclusion, the work presented in my thesis represents a comprehensive study of neuropathology, transcriptomic changes and alterations to DNA modifications in two mouse models of AD pathology. My data represent an important contribution to the field of Alzheimer's disease research. First, my neuropathology findings corroborate the extensive characterisation of tau pathology in the rTg4510 mouse model, and amyloid pathology in the J20 mouse model. Second, my transcriptomic data highlight widespread changes in gene expression paralleling the development of AD-associated neuropathology, particularly tau pathology in rTg4510 mice, and confirm alterations observed in other models of pathology. Additionally, these data provide further support for an immune-response component in the progression of AD-associated accumulation, particularly of tau pathology, in addition to nominating novel molecular pathways associated with the progression of AD neuropathology. Finally, my DNA methylation data, representing the first analysis of genome-wide changes in DNA methylation in any mouse model of AD pathology, highlights changes in regions of the genome annotated to genes nominated from human AD EWAS and genetic studies, as well as genes identified in my gene expression study. This data provides an excellent foundation for follow-up investigations of genomic regulation involved in the progression of AD, which I am highly motivated to continue contributing for its understanding.

Supplementary information

Data availability

Raw RNA-seq data has been deposited in GEO under accession number GSE125957.

Supplementary tables are available for download from our website (www.epigenomicslab.com/admice-rna-seq/).

Bibliography

AHMED, Z., COOPER, J., MURRAY, T. K., GARN, K., MCNAUGHTON, E., CLARKE, H., PARHIZKAR, S., WARD, M. A., CAVALLINI, A., JACKSON, S., BOSE, S., CLAVAGUERA, F., TOLNAY, M., LAVENIR, I., GOEDERT, M., HUTTON, M. L. & O'NEILL, M. J. 2014. A novel in vivo model of tau propagation with rapid and progressive neurofibrillary tangle pathology: the pattern of spread is determined by connectivity, not proximity. *Acta Neuropathol,* 127**,** 667-83.

AKIYAMA, H. & MCGEER, P. L. 1990. Brain microglia constitutively express beta-2 integrins. *J Neuroimmunol,* 30**,** 81-93.

ALLEN, B., INGRAM, E., TAKAO, M., SMITH, M. J., JAKES, R., VIRDEE, K., YOSHIDA, H., HOLZER, M., CRAXTON, M., EMSON, P. C., ATZORI, C., MIGHELI, A., CROWTHER, R. A., GHETTI, B., SPILLANTINI, M. G. & GOEDERT, M. 2002. Abundant tau filaments and nonapoptotic neurodegeneration in transgenic mice expressing human P301S tau protein. *J Neurosci,* 22**,** 9340-51.

ALZHEIMER'S-ASSOCIATION 2018. 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia,* 14**,** 367-429.

ALZHEIMER, A. 1898. Neuere Arbeiten über die Dementia senilis und die auf atheromatöser Gefässerkrankung basierenden Hirnkrankheiten. *Monatsschr Psychiat Neurol,* 3**,** 101–115.

ALZHEIMER, A. 1907. Über eine eigenartige Erkrankung der Hirnrinde. *Allgemeine Zeitschrift fur Psychiatrie und Psychisch-gerichtliche Medizin,* 64**,** 146–148.

ANDREWS, S. 2010. FastQC: A quality control tool for high throughput sequence data. 0.11.4 ed.

ARONESTY, E. 2011. ea-utils: Command-line tools for processing biological sequencing data.

BAKULSKI, K. M., DOLINOY, D. C., SARTOR, M. A., PAULSON, H. L., KONEN, J. R., LIEBERMAN, A. P., ALBIN, R. L., HU, H. & ROZEK, L. S. 2012. Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex. *J Alzheimers Dis,* 29**,** 571-88.

BARANELLO, R. J., BHARANI, K. L., PADMARAJU, V., CHOPRA, N., LAHIRI, D. K., GREIG, N. H., PAPPOLLA, M. A. & SAMBAMURTI, K. 2015. Amyloid-beta protein clearance and degradation (ABCD) pathways and their role in Alzheimer's disease. *Curr Alzheimer Res,* 12**,** 32-46.

BARRETT, T., WILHITE, S. E., LEDOUX, P., EVANGELISTA, C., KIM, I. F., TOMASHEVSKY, M., MARSHALL, K. A., PHILLIPPY, K. H., SHERMAN, P. M., HOLKO, M., YEFANOV, A., LEE, H., ZHANG, N., ROBERTSON, C. L., SEROVA, N., DAVIS, S. & SOBOLEVA, A. 2013. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res,* 41**,** D991-5.

BEN HAIM, L., CARRILLO-DE SAUVAGE, M. A., CEYZERIAT, K. & ESCARTIN, C. 2015. Elusive roles for reactive astrocytes in neurodegenerative diseases. *Front Cell Neurosci,* 9**,** 278.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological),* 57**,** 289-300.

BILSLAND, C. A., DIAMOND, M. S. & SPRINGER, T. A. 1994. The leukocyte integrin p150,95 (CD11c/CD18) as a receptor for iC3b. Activation by a heterologous beta subunit and localization of a ligand recognition site to the I domain. *J Immunol,* 152**,** 4582-9.

BLACKMORE, T., MEFTAH, S., MURRAY, T. K., CRAIG, P. J., BLOCKEEL, A., PHILLIPS, K., EASTWOOD, B., O'NEILL, M. J., MARSTON, H., AHMED, Z., GILMOUR, G. & GASTAMBIDE, F. 2017. Tracking progressive pathological and functional decline in the rTg4510 mouse model of tauopathy. *Alzheimers Res Ther,* 9**,** 77.

BLALOCK, E. M., GEDDES, J. W., CHEN, K. C., PORTER, N. M., MARKESBERY, W. R. & LANDFIELD, P. W. 2004. Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc Natl Acad Sci U S A,* 101**,** 2173-8.

BLENNOW, K., DE LEON, M. J. & ZETTERBERG, H. 2006. Alzheimer's disease. *The Lancet,* 368**,** 387-403.

BOHMDORFER, M., SZAKMARY, A., SCHIESTL, R. H., VAQUERO, J., RIHA, J., BRENNER, S., THALHAMMER, T., SZEKERES, T. & JAGER, W. 2017.

Involvement of UDP-Glucuronosyltransferases and Sulfotransferases in the Excretion and Tissue Distribution of Resveratrol in Mice. *Nutrients,* 9.

BOLAND, B., YU, W. H., CORTI, O., MOLLEREAU, B., HENRIQUES, A., BEZARD, E., PASTORES, G. M., RUBINSZTEIN, D. C., NIXON, R. A., DUCHEN, M. R., MALLUCCI, G. R., KROEMER, G., LEVINE, B., ESKELINEN, E. L., MOCHEL, F., SPEDDING, M., LOUIS, C., MARTIN, O. R. & MILLAN, M. J. 2018. Promoting the clearance of neurotoxic proteins in neurodegenerative disorders of ageing. *Nat Rev Drug Discov,* 17**,** 660-688.

BOLLATI, V., GALIMBERTI, D., PERGOLI, L., DALLA VALLE, E., BARRETTA, F., CORTINI, F., SCARPINI, E., BERTAZZI, P. A. & BACCARELLI, A. 2011. DNA methylation in repetitive elements and Alzheimer disease. *Brain Behav Immun,* 25**,** 1078-83.

BOOTH, C. A., WITTON, J., NOWACKI, J., TSANEVA-ATANASOVA, K., JONES, M. W., RANDALL, A. D. & BROWN, J. T. 2016. Altered intrinsic pyramidal neuron properties and pathway-specific synaptic dysfunction underlie aberrant hippocampal network function in a mouse model of tauopathy. *J Neurosci,* 36**,** 350-63.

BORCHELT, D. R., RATOVITSKI, T., VAN LARE, J., LEE, M. K., GONZALES, V., JENKINS, N. A., COPELAND, N. G., PRICE, D. L. & SISODIA, S. S. 1997. Accelerated amyloid deposition in the brains of transgenic mice coexpressing mutant presenilin 1 and amyloid precursor proteins. *Neuron,* 19**,** 939-45.

BORCHELT, D. R., THINAKARAN, G., ECKMAN, C. B., LEE, M. K., DAVENPORT, F., RATOVITSKY, T., PRADA, C. M., KIM, G., SEEKINS, S., YAGER, D., SLUNT, H. H., WANG, R., SEEGER, M., LEVEY, A. I., GANDY, S. E., COPELAND, N. G., JENKINS, N. A., PRICE, D. L., YOUNKIN, S. G. & SISODIA, S. S. 1996. Familial Alzheimer's disease-linked presenilin 1 variants elevate Abeta1-42/1-40 ratio in vitro and in vivo. *Neuron,* 17**,** 1005-13.

BOSSERS, K., WIRZ, K. T., MEERHOFF, G. F., ESSING, A. H., VAN DONGEN, J. W., HOUBA, P., KRUSE, C. G., VERHAAGEN, J. & SWAAB, D. F. 2010. Concerted changes in transcripts in the prefrontal cortex precede neuropathology in Alzheimer's disease. *Brain,* 133**,** 3699-723.

BRAAK, H., ALAFUZOFF, I., ARZBERGER, T., KRETZSCHMAR, H. & DEL TREDICI, K. 2006. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol,* 112**,** 389-404.

BRAAK, H. & BRAAK, E. 1991. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol,* 82**,** 239-59.

BRADLEY-WHITMAN, M. A. & LOVELL, M. A. 2013. Epigenetic changes in the progression of Alzheimer's disease. *Mech Ageing Dev,* 134**,** 486-95.

CADENA-DEL-CASTILLO, C., VALDES-QUEZADA, C., CARMONA-ALDANA, F., ARIAS, C., BERMUDEZ-RATTONI, F. & RECILLAS-TARGA, F. 2014. Age-dependent increment of hydroxymethylation in the brain cortex in the triple-transgenic mouse model of Alzheimer's disease. *J Alzheimers Dis,* 41**,** 845-54.

CARRASQUILLO, M. M., BELBIN, O., HUNTER, T. A., MA, L., BISCEGLIO, G. D., ZOU, F., CROOK, J. E., PANKRATZ, V. S., DICKSON, D. W., GRAFF-RADFORD, N. R., PETERSEN, R. C., MORGAN, K. & YOUNKIN, S. G. 2010. Replication of CLU, CR1, and PICALM associations with alzheimer disease. *Arch Neurol,* 67**,** 961-4.

CASAS, C. 2017. GRP78 at the centre of the stage in cancer and neuroprotection. *Front Neurosci,* 11**,** 177.

CASTANHO, I. & LUNNON, K. 2019. Epigenetic processes in Alzheimer's disease. *In:* BINDA, O. (ed.) *Chromatin Signaling and Neurological Disorders.* Elsevier.

CASTANHO, I., MURRAY, T. K., HANNON, E., JEFFRIES, A., WALKER, E., LAING, E., BAULF, H., HARVEY, J., RANDALL, A., MOORE, K., O'NEILL, P., LUNNON, K., COLLIER, D. A., AHMED, Z., O'NEILL, M. J. & MILL, J. 2019. Transcriptional signatures of progressive neuropathology in transgenic tau and amyloid mouse models. *bioRxiv***,** 548578.

CASTILLO, E., LEON, J., MAZZEI, G., ABOLHASSANI, N., HARUYAMA, N., SAITO, T., SAIDO, T., HOKAMA, M., IWAKI, T., OHARA, T., NINOMIYA, T., KIYOHARA, Y., SAKUMI, K., LAFERLA, F. M. & NAKABEPPU, Y. 2017. Comparative profiling of cortical gene expression in Alzheimer's disease patients

and mouse models demonstrates a link between amyloidosis and neuroinflammation. *Sci Rep,* 7**,** 17762.

CHEN, E. Y., TAN, C. M., KOU, Y., DUAN, Q., WANG, Z., MEIRELLES, G. V., CLARK, N. R. & MA'AYAN, A. 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics,* 14**,** 128.

CHEN, Y., PAL, B., VISVADER, J. E. & SMYTH, G. K. 2018. Differential methylation analysis of reduced representation bisulfite sequencing experiments using edgeR. *F1000Res,* 6**,** 2055.

CHEN, Y., WU, H., WANG, S., KOITO, H., LI, J., YE, F., HOANG, J., ESCOBAR, S. S., GOW, A., ARNETT, H. A., TRAPP, B. D., KARANDIKAR, N. J., HSIEH, J. & LU, Q. R. 2009. The oligodendrocyte-specific G protein-coupled receptor GPR17 is a cell-intrinsic timer of myelination. *Nat Neurosci,* 12**,** 1398-406.

CHOULIARAS, L., MASTROENI, D., DELVAUX, E., GROVER, A., KENIS, G., HOF, P. R., STEINBUSCH, H. W., COLEMAN, P. D., RUTTEN, B. P. & VAN DEN HOVE, D. L. 2013. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol Aging,* 34**,** 2091-9.

CHOULIARAS, L., RUTTEN, B. P., KENIS, G., PEERBOOMS, O., VISSER, P. J., VERHEY, F., VAN OS, J., STEINBUSCH, H. W. & VAN DEN HOVE, D. L. 2010. Epigenetic regulation in the pathophysiology of Alzheimer's disease. *Prog Neurobiol,* 90**,** 498-510.

CLARKE, J. R., RIBEIRO, F. C., FROZZA, R. L., DE FELICE, F. G. & LOURENCO, M. V. 2018. Metabolic Dysfunction in Alzheimer's Disease: From Basic Neurobiology to Clinical Approaches. *J Alzheimers Dis,* 64**,** S405-S426.

CONDLIFFE, D., WONG, A., TROAKES, C., PROITSI, P., PATEL, Y., CHOULIARAS, L., FERNANDES, C., COOPER, J., LOVESTONE, S., SCHALKWYK, L., MILL, J. & LUNNON, K. 2014. Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain. *Neurobiol Aging,* 35**,** 1850-4.

CONESA, A., MADRIGAL, P., TARAZONA, S., GOMEZ-CABRERO, D., CERVERA, A., MCPHERSON, A., SZCZESNIAK, M. W., GAFFNEY, D. J., ELO,

L. L., ZHANG, X. & MORTAZAVI, A. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol,* 17**,** 13.

CONG, L., JIA, J., QIN, W., REN, Y. & SUN, Y. 2014. Genome-wide analysis of DNA methylation in an APP/PS1 mouse model of Alzheimer's disease. *Acta Neurol Belg,* 114**,** 195-206.

COPPIETERS, N., DIERIKS, B. V., LILL, C., FAULL, R. L., CURTIS, M. A. & DRAGUNOW, M. 2014. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol Aging,* 35**,** 1334-44.

CORDER, E. H., SAUNDERS, A. M., STRITTMATTER, W. J., SCHMECHEL, D. E., GASKELL, P. C., SMALL, G. W., ROSES, A. D., HAINES, J. L. & PERICAK-VANCE, M. A. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science,* 261**,** 921-3.

CORNEVEAUX, J. J., MYERS, A. J., ALLEN, A. N., PRUZIN, J. J., RAMIREZ, M., ENGEL, A., NALLS, M. A., CHEN, K., LEE, W., CHEWNING, K., VILLA, S. E., MEECHOOVET, H. B., GERBER, J. D., FROST, D., BENSON, H. L., O'REILLY, S., CHIBNIK, L. B., SHULMAN, J. M., SINGLETON, A. B., CRAIG, D. W., VAN KEUREN-JENSEN, K. R., DUNCKLEY, T., BENNETT, D. A., DE JAGER, P. L., HEWARD, C., HARDY, J., REIMAN, E. M. & HUENTELMAN, M. J. 2010. Association of CR1, CLU and PICALM with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. *Hum Mol Genet,* 19**,** 3295-301.

COUPLAND, K., KIM, W. S., HALLIDAY, G., DOBSON-STONE, C. & KWOK, J. B. J. 2014. Mapt methylation in Alzheimer's disease. *Alzheimer's & Dementia,* 10**,** P317-P318.

COUPLAND, K. G., KIM, W. S., HALLIDAY, G. M., HALLUPP, M., DOBSON-STONE, C. & KWOK, J. B. 2015. Effect of PSEN1 mutations on MAPT methylation in early-onset Alzheimer's disease. *Curr Alzheimer Res,* 12**,** 745-51.

CREWS, L. & MASLIAH, E. 2010. Molecular mechanisms of neurodegeneration in Alzheimer's disease. *Hum Mol Genet,* 19**,** R12-20.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M.,

SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A,* 107**,** 21931-6.

CRUCHAGA, C., KARCH, C. M., JIN, S. C., BENITEZ, B. A., CAI, Y., GUERREIRO, R., HARARI, O., NORTON, J., BUDDE, J., BERTELSEN, S., JENG, A. T., COOPER, B., SKORUPA, T., CARRELL, D., LEVITCH, D., HSU, S., CHOI, J., RYTEN, M., SASSI, C., BRAS, J., GIBBS, R. J., HERNANDEZ, D. G., LUPTON, M. K., POWELL, J., FORABOSCO, P., RIDGE, P. G., CORCORAN, C. D., TSCHANZ, J. T., NORTON, M. C., MUNGER, R. G., SCHMUTZ, C., LEARY, M., DEMIRCI, F. Y., BAMNE, M. N., WANG, X., LOPEZ, O. L., GANGULI, M., MEDWAY, C., TURTON, J., LORD, J., BRAAE, A., BARBER, I., BROWN, K., ALZHEIMER'S RESEARCH, U. K. C., PASTOR, P., LORENZO-BETANCOR, O., BRKANAC, Z., SCOTT, E., TOPOL, E., MORGAN, K., ROGAEVA, E., SINGLETON, A., HARDY, J., KAMBOH, M. I., GEORGE-HYSLOP, P. S., CAIRNS, N., MORRIS, J. C., KAUWE, J. S. K. & GOATE, A. M. 2014. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature,* 505**,** 550-554.

DATTA, S. & NETTLETON, D. (eds.) 2014. *Statistical analysis of next generation sequencing data*: Springer International Publishing.

DE JAGER, P. L., SRIVASTAVA, G., LUNNON, K., BURGESS, J., SCHALKWYK, L. C., YU, L., EATON, M. L., KEENAN, B. T., ERNST, J., MCCABE, C., TANG, A., RAJ, T., REPLOGLE, J., BRODEUR, W., GABRIEL, S., CHAI, H. S., YOUNKIN, C., YOUNKIN, S. G., ZOU, F., SZYF, M., EPSTEIN, C. B., SCHNEIDER, J. A., BERNSTEIN, B. E., MEISSNER, A., ERTEKIN-TANER, N., CHIBNIK, L. B., KELLIS, M., MILL, J. & BENNETT, D. A. 2014. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci,* 17**,** 1156-63.

DE STROOPER, B. 2010. Proteases and proteolysis in Alzheimer disease: a multifactorial view on the disease process. *Physiol Rev,* 90**,** 465-94.

DE STROOPER, B. & KARRAN, E. 2016. The cellular phase of Alzheimer's disease. *Cell,* 164**,** 603-15.

DEMATTOS, R. B., LU, J., TANG, Y., RACKE, M. M., DELONG, C. A., TZAFERIS, J. A., HOLE, J. T., FORSTER, B. M., MCDONNELL, P. C., LIU, F.,

KINLEY, R. D., JORDAN, W. H. & HUTTON, M. L. 2012. A plaque-specific antibody clears existing beta-amyloid plaques in Alzheimer's disease mice. *Neuron,* 76**,** 908-20.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

DOHERTY, R. & COULDREY, C. 2014. Exploring genome wide bisulfite sequencing for DNA methylation analysis in livestock: a technical assessment. *Front Genet,* 5**,** 126.

DOR, Y. & CEDAR, H. 2018. Principles of DNA methylation and their implications for biology and medicine. *Lancet,* 392**,** 777-786.

DUNCKLEY, T., BEACH, T. G., RAMSEY, K. E., GROVER, A., MASTROENI, D., WALKER, D. G., LAFLEUR, B. J., COON, K. D., BROWN, K. M., CASELLI, R., KUKULL, W., HIGDON, R., MCKEEL, D., MORRIS, J. C., HULETTE, C., SCHMECHEL, D., REIMAN, E. M., ROGERS, J. & STEPHAN, D. A. 2006. Gene expression correlates of neurofibrillary tangles in Alzheimer's disease. *Neurobiol Aging,* 27**,** 1359-71.

DURINCK, S., MOREAU, Y., KASPRZYK, A., DAVIS, S., DE MOOR, B., BRAZMA, A. & HUBER, W. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics,* 21**,** 3439-40.

DURINCK, S., SPELLMAN, P. T., BIRNEY, E. & HUBER, W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc,* 4**,** 1184-91.

EDGAR, R., DOMRACHEV, M. & LASH, A. E. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res,* 30**,** 207-10.

EHRLICH, M. 2009. DNA hypomethylation in cancer cells. *Epigenomics,* 1**,** 239-59.

EMILSSON, L., SAETRE, P. & JAZIN, E. 2006. Alzheimer's disease: mRNA expression profiles of multiple patients show alterations of genes involved with calcium signaling. *Neurobiol Dis,* 21**,** 618-25.

FADISTA, J., MANNING, A. K., FLOREZ, J. C. & GROOP, L. 2016. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet,* 24**,** 1202-5.

FELSKY, D., ROOSTAEI, T., NHO, K., RISACHER, S. L., BRADSHAW, E. M., PETYUK, V., SCHNEIDER, J. A., SAYKIN, A., BENNETT, D. A. & DE JAGER, P. L. 2019. Neuropathological correlates and genetic architecture of microglial activation in elderly human brain. *Nat Commun,* 10**,** 409.

FONSECA, N. A., RUNG, J., BRAZMA, A. & MARIONI, J. C. 2012. Tools for mapping high-throughput sequencing data. *Bioinformatics,* 28**,** 3169-77.

FONTEH, A. N., HARRINGTON, R. J., TSAI, A., LIAO, P. & HARRINGTON, M. G. 2007. Free amino acid and dipeptide changes in the body fluids from Alzheimer's disease subjects. *Amino Acids,* 32**,** 213-24.

FOUSE, S. D., NAGARAJAN, R. O. & COSTELLO, J. F. 2010. Genome-scale DNA methylation analysis. *Epigenomics,* 2**,** 105-17.

FRANCIS, Y. I., FA, M., ASHRAF, H., ZHANG, H., STANISZEWSKI, A., LATCHMAN, D. S. & ARANCIO, O. 2009. Dysregulation of histone acetylation in the APP/PS1 mouse model of Alzheimer's disease. *J Alzheimers Dis,* 18**,** 131-9.

FRIEDMAN, L. K., OSEI-TUTU, N. & ZHANG, B. 2019. Exposing immature hippocampal neurons to excitotoxins reveals distinct transcriptome and protein regulation with induction of common survival signaling pathways. *Mol Cell Neurosci*.

FUKUNAGA, R. & HUNTER, T. 1997. MNK1, a new MAP kinase-activated protein kinase, isolated by a novel expression screening method for identifying protein kinase substrates. *EMBO J,* 16**,** 1921-33.

FUMAGALLI, M., LECCA, D. & ABBRACCHIO, M. P. 2016. CNS remyelination as a novel reparative approach to neurodegenerative diseases: The roles of purinergic signaling and the P2Y-like receptor GPR17. *Neuropharmacology,* 104**,** 82-93.

GALLI, G. G., CARRARA, M., FRANCAVILLA, C., DE LICHTENBERG, K. H., OLSEN, J. V., CALOGERO, R. A. & LUND, A. H. 2013. Genomic and proteomic analyses of Prdm5 reveal interactions with insulator binding proteins in embryonic stem cells. *Mol Cell Biol,* 33**,** 4504-16.

GASPARONI, G., BULTMANN, S., LUTSIK, P., KRAUS, T. F. J., SORDON, S., VLCEK, J., DIETINGER, V., STEINMAURER, M., HAIDER, M., MULHOLLAND, C. B., ARZBERGER, T., ROEBER, S., RIEMENSCHNEIDER, M., KRETZSCHMAR, H. A., GIESE, A., LEONHARDT, H. & WALTER, J. 2018. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. *Epigenetics Chromatin,* 11**,** 41.

GATZ, M., REYNOLDS, C. A., FRATIGLIONI, L., JOHANSSON, B., MORTIMER, J. A., BERG, S., FISKE, A. & PEDERSEN, N. L. 2006. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry,* 63**,** 168-74.

GENHEDEN, M., KENNEY, J. W., JOHNSTON, H. E., MANOUSOPOULOU, A., GARBIS, S. D. & PROUD, C. G. 2015. BDNF stimulation of protein synthesis in cortical neurons requires the MAP kinase-interacting kinase MNK1. *J Neurosci,* 35**,** 972-84.

GJONESKA, E., PFENNING, A. R., MATHYS, H., QUON, G., KUNDAJE, A., TSAI, L. H. & KELLIS, M. 2015. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature,* 518**,** 365-9.

GOODWIN, L. O., SPLINTER, E., DAVIS, T. L., URBAN, R., HE, H., BRAUN, R. E., CHESLER, E. J., KUMAR, V., VAN MIN, M., NDUKUM, J., PHILIP, V. M., REINHOLDT, L. G., SVENSON, K., WHITE, J. K., SASNER, M., LUTZ, C. & MURRAY, S. A. 2017. Large-scale discovery of mouse transgenic integration sites reveals frequent structural variation and insertional mutagenesis. *bioRxiv***,** 236307.

GOTZ, J., BODEA, L. G. & GOEDERT, M. 2018. Rodent models for Alzheimer disease. *Nat Rev Neurosci,* 19**,** 583-598.

GOTZ, J. & ITTNER, L. M. 2008. Animal models of Alzheimer's disease and frontotemporal dementia. *Nat Rev Neurosci,* 9**,** 532-44.

GRIFFIN, J. W. & BRADSHAW, P. C. 2017. Amino Acid Catabolism in Alzheimer's Disease Brain: Friend or Foe? *Oxid Med Cell Longev,* 2017**,** 5472792.

GSTIR, R., SCHAFFERER, S., SCHEIDELER, M., MISSLINGER, M., GRIEHL, M., DASCHIL, N., HUMPEL, C., OBERMAIR, G. J., SCHMUCKERMAIR, C.,

STRIESSNIG, J., FLUCHER, B. E. & HUTTENHOFER, A. 2014. Generation of a neuro-specific microarray reveals novel differentially expressed noncoding RNAs in mouse models for neurodegenerative diseases. *RNA,* 20**,** 1929-43.

GU, H., SMITH, Z. D., BOCK, C., BOYLE, P., GNIRKE, A. & MEISSNER, A. 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc,* 6**,** 468-81.

GUERREIRO, R., WOJTAS, A., BRAS, J., CARRASQUILLO, M., ROGAEVA, E., MAJOUNIE, E., CRUCHAGA, C., SASSI, C., KAUWE, J. S., YOUNKIN, S., HAZRATI, L., COLLINGE, J., POCOCK, J., LASHLEY, T., WILLIAMS, J., LAMBERT, J. C., AMOUYEL, P., GOATE, A., RADEMAKERS, R., MORGAN, K., POWELL, J., ST GEORGE-HYSLOP, P., SINGLETON, A., HARDY, J. & ALZHEIMER GENETIC ANALYSIS, G. 2013. TREM2 variants in Alzheimer's disease. *N Engl J Med,* 368**,** 117-27.

GUERREIRO, R. J., GUSTAFSON, D. R. & HARDY, J. 2012. The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. *Neurobiol Aging,* 33**,** 437-56.

GUO, H., ZHU, P., GUO, F., LI, X., WU, X., FAN, X., WEN, L. & TANG, F. 2015. Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nat Protoc,* 10**,** 645-59.

HAMILTON, G., COLBERT, J. D., SCHUETTELKOPF, A. W. & WATTS, C. 2008. Cystatin F is a cathepsin C-directed protease inhibitor regulated by proteolysis. *EMBO J,* 27**,** 499-508.

HANSEN, D. V., HANSON, J. E. & SHENG, M. 2018. Microglia in Alzheimer's disease. *J Cell Biol,* 217**,** 459-472.

HANSEN, K. D., BRENNER, S. E. & DUDOIT, S. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res,* 38**,** e131.

HANSEN, K. D., LANGMEAD, B. & IRIZARRY, R. A. 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol,* 13**,** R83.

HAROLD, D., ABRAHAM, R., HOLLINGWORTH, P., SIMS, R., GERRISH, A., HAMSHERE, M. L., PAHWA, J. S., MOSKVINA, V., DOWZELL, K., WILLIAMS,

A., JONES, N., THOMAS, C., STRETTON, A., MORGAN, A. R., LOVESTONE, S., POWELL, J., PROITSI, P., LUPTON, M. K., BRAYNE, C., RUBINSZTEIN, D. C., GILL, M., LAWLOR, B., LYNCH, A., MORGAN, K., BROWN, K. S., PASSMORE, P. A., CRAIG, D., MCGUINNESS, B., TODD, S., HOLMES, C., MANN, D., SMITH, A. D., LOVE, S., KEHOE, P. G., HARDY, J., MEAD, S., FOX, N., ROSSOR, M., COLLINGE, J., MAIER, W., JESSEN, F., SCHURMANN, B., HEUN, R., VAN DEN BUSSCHE, H., HEUSER, I., KORNHUBER, J., WILTFANG, J., DICHGANS, M., FROLICH, L., HAMPEL, H., HULL, M., RUJESCU, D., GOATE, A. M., KAUWE, J. S., CRUCHAGA, C., NOWOTNY, P., MORRIS, J. C., MAYO, K., SLEEGERS, K., BETTENS, K., ENGELBORGHS, S., DE DEYN, P. P., VAN BROECKHOVEN, C., LIVINGSTON, G., BASS, N. J., GURLING, H., MCQUILLIN, A., GWILLIAM, R., DELOUKAS, P., AL-CHALABI, A., SHAW, C. E., TSOLAKI, M., SINGLETON, A. B., GUERREIRO, R., MUHLEISEN, T. W., NOTHEN, M. M., MOEBUS, S., JOCKEL, K. H., KLOPP, N., WICHMANN, H. E., CARRASQUILLO, M. M., PANKRATZ, V. S., YOUNKIN, S. G., HOLMANS, P. A., O'DONOVAN, M., OWEN, M. J. & WILLIAMS, J. 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet,* 41**,** 1088-93.

HARRIS, J. A., DEVIDZE, N., HALABISKY, B., LO, I., THWIN, M. T., YU, G. Q., BREDESEN, D. E., MASLIAH, E. & MUCKE, L. 2010. Many neuronal and behavioral impairments in transgenic mouse models of Alzheimer's disease are independent of caspase cleavage of the amyloid precursor protein. *J Neurosci,* 30**,** 372-81.

HASANBASIC, S., JAHIC, A., KARAHMET, E., SEJRANIC, A. & PRNJAVORAC, B. 2016. The Role of Cysteine Protease in Alzheimer Disease. *Mater Sociomed,* 28**,** 235-8.

HEFFNER, T. G., HARTMAN, J. A. & SEIDEN, L. S. 1980. A rapid method for the regional dissection of the rat brain. *Pharmacol Biochem Behav,* 13**,** 453-6.

HIRST, M. & MARRA, M. A. 2010. Next generation sequencing based approaches to epigenomics. *Brief Funct Genomics,* 9**,** 455-65.

HOKAMA, M., OKA, S., LEON, J., NINOMIYA, T., HONDA, H., SASAKI, K., IWAKI, T., OHARA, T., SASAKI, T., LAFERLA, F. M., KIYOHARA, Y. &

NAKABEPPU, Y. 2014. Altered expression of diabetes-related genes in Alzheimer's disease brains: the Hisayama study. *Cereb Cortex,* 24**,** 2476-88.

HOLLINGWORTH, P., HAROLD, D., SIMS, R., GERRISH, A., LAMBERT, J. C., CARRASQUILLO, M. M., ABRAHAM, R., HAMSHERE, M. L., PAHWA, J. S., MOSKVINA, V., DOWZELL, K., JONES, N., STRETTON, A., THOMAS, C., RICHARDS, A., IVANOV, D., WIDDOWSON, C., CHAPMAN, J., LOVESTONE, S., POWELL, J., PROITSI, P., LUPTON, M. K., BRAYNE, C., RUBINSZTEIN, D. C., GILL, M., LAWLOR, B., LYNCH, A., BROWN, K. S., PASSMORE, P. A., CRAIG, D., MCGUINNESS, B., TODD, S., HOLMES, C., MANN, D., SMITH, A. D., BEAUMONT, H., WARDEN, D., WILCOCK, G., LOVE, S., KEHOE, P. G., HOOPER, N. M., VARDY, E. R., HARDY, J., MEAD, S., FOX, N. C., ROSSOR, M., COLLINGE, J., MAIER, W., JESSEN, F., RUTHER, E., SCHURMANN, B., HEUN, R., KOLSCH, H., VAN DEN BUSSCHE, H., HEUSER, I., KORNHUBER, J., WILTFANG, J., DICHGANS, M., FROLICH, L., HAMPEL, H., GALLACHER, J., HULL, M., RUJESCU, D., GIEGLING, I., GOATE, A. M., KAUWE, J. S., CRUCHAGA, C., NOWOTNY, P., MORRIS, J. C., MAYO, K., SLEEGERS, K., BETTENS, K., ENGELBORGHS, S., DE DEYN, P. P., VAN BROECKHOVEN, C., LIVINGSTON, G., BASS, N. J., GURLING, H., MCQUILLIN, A., GWILLIAM, R., DELOUKAS, P., AL-CHALABI, A., SHAW, C. E., TSOLAKI, M., SINGLETON, A. B., GUERREIRO, R., MUHLEISEN, T. W., NOTHEN, M. M., MOEBUS, S., JOCKEL, K. H., KLOPP, N., WICHMANN, H. E., PANKRATZ, V. S., SANDO, S. B., AASLY, J. O., BARCIKOWSKA, M., WSZOLEK, Z. K., DICKSON, D. W., GRAFF-RADFORD, N. R., PETERSEN, R. C., et al. 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet,* 43**,** 429-35.

HOLMES, H. E., COLGAN, N., ISMAIL, O., MA, D., POWELL, N. M., O'CALLAGHAN, J. M., HARRISON, I. F., JOHNSON, R. A., MURRAY, T. K., AHMED, Z., HEGGENES, M., FISHER, A., CARDOSO, M. J., MODAT, M., WALKER-SAMUEL, S., FISHER, E. M., OURSELIN, S., O'NEILL, M. J., WELLS, J. A., COLLINS, E. C. & LYTHGOE, M. F. 2016. Imaging the accumulation and suppression of tau pathology using multiparametric MRI. *Neurobiol Aging,* 39**,** 184-94.

HOOZEMANS, J. J., VEERHUIS, R., VAN HAASTERT, E. S., ROZEMULLER, J. M., BAAS, F., EIKELENBOOM, P. & SCHEPER, W. 2005. The unfolded protein response is activated in Alzheimer's disease. *Acta Neuropathol,* 110**,** 165-72.

HOPPERTON, K. E., MOHAMMAD, D., TREPANIER, M. O., GIULIANO, V. & BAZINET, R. P. 2018. Markers of microglia in post-mortem brain samples from patients with Alzheimer's disease: a systematic review. *Mol Psychiatry,* 23**,** 177-198.

HORESH, Y., KATSEL, P., HAROUTUNIAN, V. & DOMANY, E. 2011. Gene expression signature is shared by patients with Alzheimer's disease and schizophrenia at the superior temporal gyrus. *Eur J Neurol,* 18**,** 410-24.

HSIAO, K., CHAPMAN, P., NILSEN, S., ECKMAN, C., HARIGAYA, Y., YOUNKIN, S., YANG, F. & COLE, G. 1996. Correlative memory deficits, Abeta elevation, and amyloid plaques in transgenic mice. *Science,* 274**,** 99-102.

ILLUMINA 2017. TruSeq stranded mRNA - Reference guide.

IWATA, A., NAGATA, K., HATSUTA, H., TAKUMA, H., BUNDO, M., IWAMOTO, K., TAMAOKA, A., MURAYAMA, S., SAIDO, T. & TSUJI, S. 2014. Altered CpG methylation in sporadic Alzheimer's disease is associated with APP and MAPT dysregulation. *Hum Mol Genet,* 23**,** 648-56.

JANSEN, I. E., SAVAGE, J. E., WATANABE, K., BRYOIS, J., WILLIAMS, D. M., STEINBERG, S., SEALOCK, J., KARLSSON, I. K., HAGG, S., ATHANASIU, L., VOYLE, N., PROITSI, P., WITOELAR, A., STRINGER, S., AARSLAND, D., ALMDAHL, I. S., ANDERSEN, F., BERGH, S., BETTELLA, F., BJORNSSON, S., BRAEKHUS, A., BRATHEN, G., DE LEEUW, C., DESIKAN, R. S., DJUROVIC, S., DUMITRESCU, L., FLADBY, T., HOHMAN, T. J., JONSSON, P. V., KIDDLE, S. J., RONGVE, A., SALTVEDT, I., SANDO, S. B., SELBAEK, G., SHOAI, M., SKENE, N. G., SNAEDAL, J., STORDAL, E., ULSTEIN, I. D., WANG, Y., WHITE, L. R., HARDY, J., HJERLING-LEFFLER, J., SULLIVAN, P. F., VAN DER FLIER, W. M., DOBSON, R., DAVIS, L. K., STEFANSSON, H., STEFANSSON, K., PEDERSEN, N. L., RIPKE, S., ANDREASSEN, O. A. & POSTHUMA, D. 2019. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet,* 51**,** 404-413.

JEFFRIES, A. R. & MILL, J. 2017. Profiling regulatory variation in the brain: Methods for exploring the neuronal epigenome. *Biol Psychiatry,* 81**,** 90-91.

JICHA, G. A., WEAVER, C., LANE, E., VIANNA, C., KRESS, Y., ROCKWOOD, J. & DAVIES, P. 1999. cAMP-dependent protein kinase phosphorylations on tau in Alzheimer's disease. *J Neurosci,* 19**,** 7486-94.

JONES, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet,* 13**,** 484-92.

JONSSON, T., STEFANSSON, H., STEINBERG, S., JONSDOTTIR, I., JONSSON, P. V., SNAEDAL, J., BJORNSSON, S., HUTTENLOCHER, J., LEVEY, A. I., LAH, J. J., RUJESCU, D., HAMPEL, H., GIEGLING, I., ANDREASSEN, O. A., ENGEDAL, K., ULSTEIN, I., DJUROVIC, S., IBRAHIM-VERBAAS, C., HOFMAN, A., IKRAM, M. A., VAN DUIJN, C. M., THORSTEINSDOTTIR, U., KONG, A. & STEFANSSON, K. 2013. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med,* 368**,** 107-16.

JOSHI, S., KAUR, S., REDIG, A. J., GOLDSBOROUGH, K., DAVID, K., UEDA, T., WATANABE-FUKUNAGA, R., BAKER, D. P., FISH, E. N., FUKUNAGA, R. & PLATANIAS, L. C. 2009. Type I interferon (IFN)-dependent activation of Mnk1 and its role in the generation of growth inhibitory responses. *Proc Natl Acad Sci U S A,* 106**,** 12097-102.

JOUANNE, M., RAULT, S. & VOISIN-CHIRET, A. S. 2017. Tau protein aggregation in Alzheimer's disease: An attractive target for the development of novel therapeutic agents. *Eur J Med Chem,* 139**,** 153-167.

JUN, G., NAJ, A. C., BEECHAM, G. W., WANG, L. S., BUROS, J., GALLINS, P. J., BUXBAUM, J. D., ERTEKIN-TANER, N., FALLIN, M. D., FRIEDLAND, R., INZELBERG, R., KRAMER, P., ROGAEVA, E., ST GEORGE-HYSLOP, P., ALZHEIMER'S DISEASE GENETICS, C., CANTWELL, L. B., DOMBROSKI, B. A., SAYKIN, A. J., REIMAN, E. M., BENNETT, D. A., MORRIS, J. C., LUNETTA, K. L., MARTIN, E. R., MONTINE, T. J., GOATE, A. M., BLACKER, D., TSUANG, D. W., BEEKLY, D., CUPPLES, L. A., HAKONARSON, H., KUKULL, W., FOROUD, T. M., HAINES, J., MAYEUX, R., FARRER, L. A., PERICAK-VANCE, M. A. & SCHELLENBERG, G. D. 2010. Meta-analysis confirms CR1, CLU, and

PICALM as alzheimer disease risk loci and reveals interactions with APOE genotypes. *Arch Neurol,* 67**,** 1473-84.

KAMPHUIS, W., KOOIJMAN, L., SCHETTERS, S., ORRE, M. & HOL, E. M. 2016. Transcriptional profiling of CD11c-positive microglia accumulating around amyloid plaques in a mouse model for Alzheimer's disease. *Biochim Biophys Acta,* 1862**,** 1847-60.

KAN, M. J., LEE, J. E., WILSON, J. G., EVERHART, A. L., BROWN, C. M., HOOFNAGLE, A. N., JANSEN, M., VITEK, M. P., GUNN, M. D. & COLTON, C. A. 2015. Arginine deprivation and immune suppression in a mouse model of Alzheimer's disease. *J Neurosci,* 35**,** 5969-82.

KARIMI, M., JOHANSSON, S., STACH, D., CORCORAN, M., GRANDER, D., SCHALLING, M., BAKALKIN, G., LYKO, F., LARSSON, C. & EKSTROM, T. J. 2006. LUMA (LUminometric Methylation Assay)--a high throughput method to the analysis of genomic DNA methylation. *Exp Cell Res,* 312**,** 1989-95.

KARIMI, M., LUTTROPP, K. & EKSTROM, T. J. 2011. Global DNA methylation analysis using the Luminometric Methylation Assay. *Methods Mol Biol,* 791**,** 135-44.

KATSEL, P., LI, C. & HAROUTUNIAN, V. 2007. Gene expression alterations in the sphingolipid metabolism pathways during progression of dementia and Alzheimer's disease: a shift toward ceramide accumulation at the earliest recognizable stages of Alzheimer's disease? *Neurochem Res,* 32**,** 845-56.

KENT, W. J. 2002. BLAT - the BLAST-like alignment tool. *Genome Res,* 12**,** 656-64.

KEREN-SHAUL, H., SPINRAD, A., WEINER, A., MATCOVITCH-NATAN, O., DVIR-SZTERNFELD, R., ULLAND, T. K., DAVID, E., BARUCH, K., LARA-ASTAISO, D., TOTH, B., ITZKOVITZ, S., COLONNA, M., SCHWARTZ, M. & AMIT, I. 2017. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell,* 169**,** 1276-1290 e17.

KHARE, T., PAI, S., KONCEVICIUS, K., PAL, M., KRIUKIENE, E., LIUTKEVICIUTE, Z., IRIMIA, M., JIA, P., PTAK, C., XIA, M., TICE, R., TOCHIGI, M., MORERA, S., NAZARIANS, A., BELSHAM, D., WONG, A. H., BLENCOWE, B. J., WANG, S. C., KAPRANOV, P., KUSTRA, R., LABRIE, V.,

KLIMASAUSKAS, S. & PETRONIS, A. 2012. 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat Struct Mol Biol,* 19**,** 1037-43.

KORPELAINEN, E., TUIMALA, J., SOMERVUO, P., HUSS, M. & WONG, G. 2014. *RNA-seq data analysis: A practical approach*, Chapman and Hall/CRC

KOSS, D. J., JONES, G., CRANSTON, A., GARDNER, H., KANAAN, N. M. & PLATT, B. 2016. Soluble pre-fibrillar tau and beta-amyloid species emerge in early human Alzheimer's disease and track disease progression and cognitive decline. *Acta Neuropathol,* 132**,** 875-895.

KRUEGER, F. 2017. Trim Galore. 0.4.5 ed.

KRUEGER, F. & ANDREWS, S. R. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics,* 27**,** 1571-2.

KRUEGER, F., KRECK, B., FRANKE, A. & ANDREWS, S. R. 2012. DNA methylome analysis using short bisulfite sequencing data. *Nature Methods,* 9**,** 145-151.

KULESHOV, M. V., JONES, M. R., ROUILLARD, A. D., FERNANDEZ, N. F., DUAN, Q., WANG, Z., KOPLEV, S., JENKINS, S. L., JAGODNIK, K. M., LACHMANN, A., MCDERMOTT, M. G., MONTEIRO, C. D., GUNDERSEN, G. W. & MA'AYAN, A. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res,* 44**,** W90-7.

KUMAR, V., MURATANI, M., RAYAN, N. A., KRAUS, P., LUFKIN, T., NG, H. H. & PRABHAKAR, S. 2013. Uniform, optimal signal processing of mapped deep-sequencing data. *Nat Biotechnol,* 31**,** 615-22.

KUNINAKA, N., KAWAGUCHI, M., OGAWA, M., SATO, A., ARIMA, K., MURAYAMA, S. & SAITO, Y. 2015. Simplification of the modified Gallyas method. *Neuropathology,* 35**,** 10-5.

KUNKLE, B. W., GRENIER-BOLEY, B., SIMS, R., BIS, J. C., DAMOTTE, V., NAJ, A. C., BOLAND, A., VRONSKAYA, M., VAN DER LEE, S. J., AMLIE-WOLF, A., BELLENGUEZ, C., FRIZATTI, A., CHOURAKI, V., MARTIN, E. R., SLEEGERS, K., BADARINARAYAN, N., JAKOBSDOTTIR, J., HAMILTON-NELSON, K. L., MORENO-GRAU, S., OLASO, R., RAYBOULD, R., CHEN, Y., KUZMA, A. B., HILTUNEN, M., MORGAN, T., AHMAD, S., VARDARAJAN, B. N.,

EPELBAUM, J., HOFFMANN, P., BOADA, M., BEECHAM, G. W., GARNIER, J. G., HAROLD, D., FITZPATRICK, A. L., VALLADARES, O., MOUTET, M. L., GERRISH, A., SMITH, A. V., QU, L., BACQ, D., DENNING, N., JIAN, X., ZHAO, Y., DEL ZOMPO, M., FOX, N. C., CHOI, S. H., MATEO, I., HUGHES, J. T., ADAMS, H. H., MALAMON, J., SANCHEZ-GARCIA, F., PATEL, Y., BRODY, J. A., DOMBROSKI, B. A., NARANJO, M. C. D., DANIILIDOU, M., EIRIKSDOTTIR, G., MUKHERJEE, S., WALLON, D., UPHILL, J., ASPELUND, T., CANTWELL, L. B., GARZIA, F., GALIMBERTI, D., HOFER, E., BUTKIEWICZ, M., FIN, B., SCARPINI, E., SARNOWSKI, C., BUSH, W. S., MESLAGE, S., KORNHUBER, J., WHITE, C. C., SONG, Y., BARBER, R. C., ENGELBORGHS, S., SORDON, S., VOIJNOVIC, D., ADAMS, P. M., VANDENBERGHE, R., MAYHAUS, M., CUPPLES, L. A., ALBERT, M. S., DE DEYN, P. P., GU, W., HIMALI, J. J., BEEKLY, D., SQUASSINA, A., HARTMANN, A. M., ORELLANA, A., BLACKER, D., RODRIGUEZ-RODRIGUEZ, E., LOVESTONE, S., GARCIA, M. E., DOODY, R. S., MUNOZ-FERNADEZ, C., SUSSAMS, R., LIN, H., FAIRCHILD, T. J., BENITO, Y. A., et al. 2019. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet,* 51**,** 414-430.

KURITA, A., MIYAUCHI, Y., IKUSHIRO, S., MACKENZIE, P. I., YAMADA, H. & ISHII, Y. 2017. Comprehensive characterization of mouse UDP-glucuronosyltransferase (Ugt) belonging to the Ugt2b subfamily: Identification of Ugt2b36 as the predominant isoform involved in morphine glucuronidation. *J Pharmacol Exp Ther,* 361**,** 199-208.

LAMBERT, J. C., HEATH, S., EVEN, G., CAMPION, D., SLEEGERS, K., HILTUNEN, M., COMBARROS, O., ZELENIKA, D., BULLIDO, M. J., TAVERNIER, B., LETENNEUR, L., BETTENS, K., BERR, C., PASQUIER, F., FIEVET, N., BARBERGER-GATEAU, P., ENGELBORGHS, S., DE DEYN, P., MATEO, I., FRANCK, A., HELISALMI, S., PORCELLINI, E., HANON, O., EUROPEAN ALZHEIMER'S DISEASE INITIATIVE, I., DE PANCORBO, M. M., LENDON, C., DUFOUIL, C., JAILLARD, C., LEVEILLARD, T., ALVAREZ, V., BOSCO, P., MANCUSO, M., PANZA, F., NACMIAS, B., BOSSU, P., PICCARDI, P., ANNONI, G., SERIPA, D., GALIMBERTI, D., HANNEQUIN, D., LICASTRO, F., SOININEN, H., RITCHIE, K., BLANCHE, H., DARTIGUES, J. F., TZOURIO, C., GUT, I., VAN BROECKHOVEN, C., ALPEROVITCH, A., LATHROP, M. &

AMOUYEL, P. 2009. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet,* 41**,** 1094-9.

LAMBERT, J. C., IBRAHIM-VERBAAS, C. A., HAROLD, D., NAJ, A. C., SIMS, R., BELLENGUEZ, C., DESTAFANO, A. L., BIS, J. C., BEECHAM, G. W., GRENIER-BOLEY, B., RUSSO, G., THORTON-WELLS, T. A., JONES, N., SMITH, A. V., CHOURAKI, V., THOMAS, C., IKRAM, M. A., ZELENIKA, D., VARDARAJAN, B. N., KAMATANI, Y., LIN, C. F., GERRISH, A., SCHMIDT, H., KUNKLE, B., DUNSTAN, M. L., RUIZ, A., BIHOREAU, M. T., CHOI, S. H., REITZ, C., PASQUIER, F., CRUCHAGA, C., CRAIG, D., AMIN, N., BERR, C., LOPEZ, O. L., DE JAGER, P. L., DERAMECOURT, V., JOHNSTON, J. A., EVANS, D., LOVESTONE, S., LETENNEUR, L., MORON, F. J., RUBINSZTEIN, D. C., EIRIKSDOTTIR, G., SLEEGERS, K., GOATE, A. M., FIEVET, N., HUENTELMAN, M. W., GILL, M., BROWN, K., KAMBOH, M. I., KELLER, L., BARBERGER-GATEAU, P., MCGUINESS, B., LARSON, E. B., GREEN, R., MYERS, A. J., DUFOUIL, C., TODD, S., WALLON, D., LOVE, S., ROGAEVA, E., GALLACHER, J., ST GEORGE-HYSLOP, P., CLARIMON, J., LLEO, A., BAYER, A., TSUANG, D. W., YU, L., TSOLAKI, M., BOSSU, P., SPALLETTA, G., PROITSI, P., COLLINGE, J., SORBI, S., SANCHEZ-GARCIA, F., FOX, N. C., HARDY, J., DENIZ NARANJO, M. C., BOSCO, P., CLARKE, R., BRAYNE, C., GALIMBERTI, D., MANCUSO, M., MATTHEWS, F., EUROPEAN ALZHEIMER'S DISEASE, I., GENETIC, ENVIRONMENTAL RISK IN ALZHEIMER'S, D., ALZHEIMER'S DISEASE GENETIC, C., COHORTS FOR, H., AGING RESEARCH IN GENOMIC, E., MOEBUS, S., MECOCCI, P., DEL ZOMPO, M., MAIER, W., HAMPEL, H., PILOTTO, A., BULLIDO, M., PANZA, F., CAFFARRA, P., et al. 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet,* 45**,** 1452-8.

LANCHEC, E., DESILETS, A., BELIVEAU, F., FLAMIER, A., MAHMOUD, S., BERNIER, G., GRIS, D., LEDUC, R. & LAVOIE, C. 2017. The type II transmembrane serine protease matriptase cleaves the amyloid precursor protein and reduces its processing to beta-amyloid peptide. *J Biol Chem,* 292**,** 20669-20682.

LANDEL, V., BARANGER, K., VIRARD, I., LORIOD, B., KHRESTCHATISKY, M., RIVERA, S., BENECH, P. & FERON, F. 2014. Temporal gene profiling of the

5XFAD transgenic mouse model highlights the importance of microglial activation in Alzheimer's disease. *Mol Neurodegener,* 9**,** 33.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* 9**,** 559.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol,* 10**,** R25.

LARDENOIJE, R., VAN DEN HOVE, D. L. A., HAVERMANS, M., VAN CASTEREN, A., LE, K. X., PALMOUR, R., LEMERE, C. A. & RUTTEN, B. P. F. 2018. Age-related epigenetic changes in hippocampal subregions of four animal models of Alzheimer's disease. *Mol Cell Neurosci,* 86**,** 1-15.

LASHLEY, T., GAMI, P., VALIZADEH, N., LI, A., REVESZ, T. & BALAZS, R. 2015. Alterations in global DNA methylation and hydroxymethylation are not detected in Alzheimer's disease. *Neuropathol Appl Neurobiol,* 41**,** 497-506.

LEE, C. Y. D., DAGGETT, A., GU, X., JIANG, L. L., LANGFELDER, P., LI, X., WANG, N., ZHAO, Y., PARK, C. S., COOPER, Y., FERANDO, I., MODY, I., COPPOLA, G., XU, H. & YANG, X. W. 2018. Elevated TREM2 gene dosage reprograms microglia responsivity and ameliorates pathological phenotypes in Alzheimer's disease models. *Neuron,* 97**,** 1032-1048 e5.

LEE, Y. K., JIN, S., DUAN, S., LIM, Y. C., NG, D. P., LIN, X. M., YEO, G. S. & DING, C. 2014. Improved reduced representation bisulfite sequencing for epigenomic profiling of clinical samples. *Biol Proced Online,* 16**,** 1.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LIANG, W. S., DUNCKLEY, T., BEACH, T. G., GROVER, A., MASTROENI, D., RAMSEY, K., CASELLI, R. J., KUKULL, W. A., MCKEEL, D., MORRIS, J. C., HULETTE, C. M., SCHMECHEL, D., REIMAN, E. M., ROGERS, J. & STEPHAN, D. A. 2008. Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. *Physiol Genomics,* 33**,** 240-56.

LIAO, Y., SMYTH, G. K. & SHI, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-30.

LIDDELOW, S. A., GUTTENPLAN, K. A., CLARKE, L. E., BENNETT, F. C., BOHLEN, C. J., SCHIRMER, L., BENNETT, M. L., MUNCH, A. E., CHUNG, W. S., PETERSON, T. C., WILTON, D. K., FROUIN, A., NAPIER, B. A., PANICKER, N., KUMAR, M., BUCKWALTER, M. S., ROWITCH, D. H., DAWSON, V. L., DAWSON, T. M., STEVENS, B. & BARRES, B. A. 2017. Neurotoxic reactive astrocytes are induced by activated microglia. *Nature,* 541**,** 481-487.

LISANTI, S., OMAR, W. A., TOMASZEWSKI, B., DE PRINS, S., JACOBS, G., KOPPEN, G., MATHERS, J. C. & LANGIE, S. A. 2013. Comparison of methods for quantification of global DNA methylation in human cells and tissues. *PLoS One,* 8**,** e79044.

LISTER, R., O'MALLEY, R. C., TONTI-FILIPPINI, J., GREGORY, B. D., BERRY, C. C., MILLAR, A. H. & ECKER, J. R. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell,* 133**,** 523-36.

LIU, X. S., WU, H., JI, X., STELZER, Y., WU, X., CZAUDERNA, S., SHU, J., DADON, D., YOUNG, R. A. & JAENISCH, R. 2016. Editing DNA methylation in the mammalian genome. *Cell,* 167**,** 233-247 e17.

LIU, X. S., WU, H., KRZISCH, M., WU, X., GRAEF, J., MUFFAT, J., HNISZ, D., LI, C. H., YUAN, B., XU, C., LI, Y., VERSHKOV, D., CACACE, A., YOUNG, R. A. & JAENISCH, R. 2018. Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. *Cell,* 172**,** 979-992 e6.

LOGSDON, B., PERUMAL, T. M., SWARUP, V., WANG, M., FUNK, C., GAITERI, C., ALLEN, M., WANG, X., DAMMER, E., SRIVASTAVA, G., MUKHERJEE, S., SIEBERTS, S. K., OMBERG, L., DANG, K. D., EDDY, J. A., SNYDER, P., CHAE, Y., AMBERKAR, S., WEI, W., HIDE, W., PREUSS, C., ERGUN, A., EBERT, P. J., AIREY, D. C., CARTER, G. W., MOSTAFAVI, S., YU, L., KLEIN, H.-U., COLLIER, D. A., GOLDE, T., LEVEY, A., BENNETT, D. A., ESTRADA, K., DECKER, M., LIU, Z., SHULMAN, J. M., ZHANG, B., SCHADT, E., DE JAGER, P. L., PRICE, N. D., ERTEKIN-TANER, N. & MANGRAVITE, L. M. 2019. Meta-analysis of the human brain transcriptome identifies heterogeneity across human

AD coexpression modules robust to sample collection and methodological approach. *bioRxiv*, 510420.

LORD, J. & CRUCHAGA, C. 2014. The epigenetic landscape of Alzheimer's disease. *Nat Neurosci,* 17, 1138-40.

LOVE, M. I., ANDERS, S. & HUBER, H. 2016a. Differential analysis of count data – the DESeq2 package.

LOVE, M. I., ANDERS, S., KIM, V. & HUBER, W. 2016b. RNA-seq workflow: gene-level exploratory analysis and differential expression. *F1000Res,* 4, 1070.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15, 550.

LU, Y., SHEN, Y., WARREN, W. & WALTER, R. 2015. Next generation sequencing in aquatic models. *In:* KULSKI, J. (ed.) *Next generation sequencing - advances, applications and challenges.* IntechOpen.

LUNNON, K., HANNON, E., SMITH, R. G., DEMPSTER, E., WONG, C., BURRAGE, J., TROAKES, C., AL-SARRAJ, S., KEPA, A., SCHALKWYK, L. & MILL, J. 2016. Variation in 5-hydroxymethylcytosine across human cortex and cerebellum. *Genome Biol,* 17, 27.

LUNNON, K. & MILL, J. 2013. Epigenetic studies in Alzheimer's disease: current findings, caveats, and considerations for future studies. *Am J Med Genet B Neuropsychiatr Genet,* 162B, 789-99.

LUNNON, K., SMITH, R., HANNON, E., DE JAGER, P. L., SRIVASTAVA, G., VOLTA, M., TROAKES, C., AL-SARRAJ, S., BURRAGE, J., MACDONALD, R., CONDLIFFE, D., HARRIES, L. W., KATSEL, P., HAROUTUNIAN, V., KAMINSKY, Z., JOACHIM, C., POWELL, J., LOVESTONE, S., BENNETT, D. A., SCHALKWYK, L. C. & MILL, J. 2014. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci,* 17, 1164-70.

LUTTROPP, K., SJOHOLM, L. K. & EKSTROM, T. J. 2015. Global Analysis of DNA 5-Methylcytosine Using the Luminometric Methylation Assay, LUMA. *Methods Mol Biol,* 1315, 209-19.

MAGISTRI, M., VELMESHEV, D., MAKHMUTOVA, M. & FAGHIHI, M. A. 2015. Transcriptomics profiling of Alzheimer's disease reveal neurovascular defects,

altered amyloid-beta homeostasis, and deregulated expression of long noncoding RNAs. *J Alzheimers Dis,* 48**,** 647-65.

MALIK, M., SIMPSON, J. F., PARIKH, I., WILFRED, B. R., FARDO, D. W., NELSON, P. T. & ESTUS, S. 2013. CD33 Alzheimer's risk-altering polymorphism, CD33 expression, and exon 2 splicing. *J Neurosci,* 33**,** 13320-5.

MARIONI, R. E., HARRIS, S. E., ZHANG, Q., MCRAE, A. F., HAGENAARS, S. P., HILL, W. D., DAVIES, G., RITCHIE, C. W., GALE, C. R., STARR, J. M., GOATE, A. M., PORTEOUS, D. J., YANG, J., EVANS, K. L., DEARY, I. J., WRAY, N. R. & VISSCHER, P. M. 2018. GWAS on family history of Alzheimer's disease. *Transl Psychiatry,* 8**,** 99.

MARQUES, S. C., LEMOS, R., FERREIRO, E., MARTINS, M., DE MENDONCA, A., SANTANA, I., OUTEIRO, T. F. & PEREIRA, C. M. 2012. Epigenetic regulation of BACE1 in Alzheimer's disease patients and in transgenic mice. *Neuroscience,* 220**,** 256-66.

MARZI, S. J., LEUNG, S. K., RIBARSKA, T., HANNON, E., SMITH, A. R., PISHVA, E., POSCHMANN, J., MOORE, K., TROAKES, C., AL-SARRAJ, S., BECK, S., NEWMAN, S., LUNNON, K., SCHALKWYK, L. C. & MILL, J. 2018. A histone acetylome-wide association study of Alzheimer's disease identifies disease-associated H3K27ac differences in the entorhinal cortex. *Nat Neurosci,* 21**,** 1618-1627.

MASTROENI, D., GROVER, A., DELVAUX, E., WHITESIDE, C., COLEMAN, P. D. & ROGERS, J. 2010. Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. *Neurobiol Aging,* 31**,** 2025-37.

MASTROENI, D., MCKEE, A., GROVER, A., ROGERS, J. & COLEMAN, P. D. 2009. Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer's disease. *PLoS One,* 4**,** e6617.

MASTROENI, D., SEKAR, S., NOLZ, J., DELVAUX, E., LUNNON, K., MILL, J., LIANG, W. S. & COLEMAN, P. D. 2017. ANK1 is up-regulated in laser captured microglia in Alzheimer's brain; the importance of addressing cellular heterogeneity. *PLoS One,* 12**,** e0177814.

MATARIN, M., SALIH, D. A., YASVOINA, M., CUMMINGS, D. M., GUELFI, S., LIU, W., NAHABOO SOLIM, M. A., MOENS, T. G., PAUBLETE, R. M., ALI, S. S.,

PERONA, M., DESAI, R., SMITH, K. J., LATCHAM, J., FULLEYLOVE, M., RICHARDSON, J. C., HARDY, J. & EDWARDS, F. A. 2015. A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology. *Cell Rep,* 10**,** 633-44.

MATHYS, H., ADAIKKAN, C., GAO, F., YOUNG, J. Z., MANET, E., HEMBERG, M., DE JAGER, P. L., RANSOHOFF, R. M., REGEV, A. & TSAI, L. H. 2017. Temporal tracking of microglia activation in neurodegeneration at single-cell resolution. *Cell Rep,* 21**,** 366-380.

MATLOCK, B. 2015. Technical note: Assessment of nucleic acid purity. *Thermo Fisher Scientific*.

MAWUENYEGA, K. G., SIGURDSON, W., OVOD, V., MUNSELL, L., KASTEN, T., MORRIS, J. C., YARASHESKI, K. E. & BATEMAN, R. J. 2010. Decreased clearance of CNS beta-amyloid in Alzheimer's disease. *Science,* 330**,** 1774.

MEISSNER, A., GNIRKE, A., BELL, G. W., RAMSAHOYE, B., LANDER, E. S. & JAENISCH, R. 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res,* 33**,** 5868-77.

MILL, J. 2011. Toward an integrated genetic and epigenetic approach to Alzheimer's disease. *Neurobiol Aging,* 32**,** 1188-91.

MILL, J. & HEIJMANS, B. T. 2013. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet,* 14**,** 585-94.

MILLER, J. A., OLDHAM, M. C. & GESCHWIND, D. H. 2008. A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J Neurosci,* 28**,** 1410-20.

MILLER, J. A., WOLTJER, R. L., GOODENBOUR, J. M., HORVATH, S. & GESCHWIND, D. H. 2013. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med,* 5**,** 48.

MOORE, C. E., PICKFORD, J., CAGAMPANG, F. R., STEAD, R. L., TIAN, S., ZHAO, X., TANG, X., BYRNE, C. D. & PROUD, C. G. 2016. MNK1 and MNK2 mediate adverse effects of high-fat feeding in distinct ways. *Sci Rep,* 6**,** 23476.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods,* 5**,** 621-8.

MOSTAFAVI, S., GAITERI, C., SULLIVAN, S. E., WHITE, C. C., TASAKI, S., XU, J., TAGA, M., KLEIN, H. U., PATRICK, E., KOMASHKO, V., MCCABE, C., SMITH, R., BRADSHAW, E. M., ROOT, D. E., REGEV, A., YU, L., CHIBNIK, L. B., SCHNEIDER, J. A., YOUNG-PEARSE, T. L., BENNETT, D. A. & DE JAGER, P. L. 2018. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci,* 21**,** 811-819.

MOY, J. K., KHOUTORSKY, A., ASIEDU, M. N., BLACK, B. J., KUHN, J. L., BARRAGAN-IGLESIAS, P., MEGAT, S., BURTON, M. D., BURGOS-VEGA, C. C., MELEMEDJIAN, O. K., BOITANO, S., VAGNER, J., GKOGKAS, C. G., PANCRAZIO, J. J., MOGIL, J. S., DUSSOR, G., SONENBERG, N. & PRICE, T. J. 2017. The MNK-eIF4E Signaling Axis Contributes to Injury-Induced Nociceptive Plasticity and the Development of Chronic Pain. *J Neurosci,* 37**,** 7481-7499.

MUCKE, L., MASLIAH, E., YU, G. Q., MALLORY, M., ROCKENSTEIN, E. M., TATSUNO, G., HU, K., KHOLODENKO, D., JOHNSON-WOOD, K. & MCCONLOGUE, L. 2000. High-level neuronal expression of abeta 1-42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *J Neurosci,* 20**,** 4050-8.

MUELLER, O., LIGHTFOOT, S. & SCHROEDER, A. 2016. RNA integrity number (RIN) – Standardization of RNA quality control. *Agilent Technologies*.

MURPHY, C. 2019. Olfactory and other sensory impairments in Alzheimer disease. *Nat Rev Neurol,* 15**,** 11-24.

NEE, L. E. & LIPPA, C. F. 1999. Alzheimer's disease in 22 twin pairs--13-year follow-up: hormonal, infectious and traumatic factors. *Dement Geriatr Cogn Disord,* 10**,** 148-51.

NEWCOMBE, E. A., CAMATS-PERNA, J., SILVA, M. L., VALMAS, N., HUAT, T. J. & MEDEIROS, R. 2018. Inflammation: the link between comorbidities, genetics, and Alzheimer's disease. *J Neuroinflammation,* 15**,** 276.

NI, J., FERNANDEZ, M. A., DANIELSSON, L., CHILLAKURU, R. A., ZHANG, J., GRUBB, A., SU, J., GENTZ, R. & ABRAHAMSON, M. 1998. Cystatin F is a glycosylated human low molecular weight cysteine proteinase inhibitor. *J Biol Chem,* 273**,** 24797-804.

O'BRIEN, R. J. & WONG, P. C. 2011. Amyloid precursor protein processing and Alzheimer's disease. *Annu Rev Neurosci,* 34**,** 185-204.

O'LEARY, E. E., MAZURKIEWICZ-MUNOZ, A. M., ARGETSINGER, L. S., MAURES, T. J., HUYNH, H. T. & CARTER-SU, C. 2013. Identification of steroid-sensitive gene-1/Ccdc80 as a JAK2-binding protein. *Mol Endocrinol,* 27**,** 619-34.

OAKLEY, H., COLE, S. L., LOGAN, S., MAUS, E., SHAO, P., CRAFT, J., GUILLOZET-BONGAARTS, A., OHNO, M., DISTERHOFT, J., VAN ELDIK, L., BERRY, R. & VASSAR, R. 2006. Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J Neurosci,* 26**,** 10129-40.

ODDO, S., CACCAMO, A., SHEPHERD, J. D., MURPHY, M. P., GOLDE, T. E., KAYED, R., METHERATE, R., MATTSON, M. P., AKBARI, Y. & LAFERLA, F. M. 2003. Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Abeta and synaptic dysfunction. *Neuron,* 39**,** 409-21.

OFENGEIM, D., MAZZITELLI, S., ITO, Y., DEWITT, J. P., MIFFLIN, L., ZOU, C., DAS, S., ADICONIS, X., CHEN, H., ZHU, H., KELLIHER, M. A., LEVIN, J. Z. & YUAN, J. 2017. RIPK1 mediates a disease-associated microglial response in Alzheimer's disease. *Proc Natl Acad Sci U S A,* 114**,** E8788-E8797.

OHSAWA, K., IMAI, Y., SASAKI, Y. & KOHSAKA, S. 2004. Microglia/macrophage-specific protein Iba1 binds to fimbrin and enhances its actin-bundling activity. *J Neurochem,* 88**,** 844-56.

OSHLACK, A. & WAKEFIELD, M. J. 2009. Transcript length bias in RNA-seq data confounds systems biology. *Biol Direct,* 4**,** 14.

PAL, S. & TYLER, J. K. 2016. Epigenetics and aging. *Sci Adv,* 2**,** e1600584.

PALOP, J. J., CHIN, J., ROBERSON, E. D., WANG, J., THWIN, M. T., BIEN-LY, N., YOO, J., HO, K. O., YU, G. Q., KREITZER, A., FINKBEINER, S., NOEBELS, J. L. & MUCKE, L. 2007. Aberrant excitatory neuronal activity and compensatory

remodeling of inhibitory hippocampal circuits in mouse models of Alzheimer's disease. *Neuron,* 55**,** 697-711.

PANTER, S. S., MCSWIGAN, J. D., SHEPPARD, J. R., EMORY, C. R. & FREY, W. H., 2ND 1985. Glial fibrillary acidic protein and Alzheimer's disease. *Neurochem Res,* 10**,** 1567-76.

PARK, Y., FIGUEROA, M. E., ROZEK, L. S. & SARTOR, M. A. 2014. MethylSig: a whole genome DNA methylation analysis pipeline. *Bioinformatics,* 30**,** 2414-22.

PAXINOS, G. & FRANKLIN, K. B. J. 2001. *The mouse brain in stereotaxic coordinates*, Academic Press.

PIEHLER, A. P., OZCURUMEZ, M. & KAMINSKI, W. E. 2012. A-subclass ATP-binding cassette proteins in brain lipid homeostasis and neurodegeneration. *Front Psychiatry,* 3**,** 17.

PLOMIN, R., OWEN, M. J. & MCGUFFIN, P. 1994. The genetic basis of complex human behaviors. *Science,* 264**,** 1733-9.

PRINCE, M., COMAS-HERRERA, A., KNAPP, M., GUERCHET, M. & KARAGIANNIDOU, M. 2016. World Alzheimer report 2016. London.

PRINCE, M., KNAPP, M., GUERCHET, M., MCCRONE, P., PRINA, M., COMAS-HERRERA, A., WITTENBERG, R., ADELAJA, B., HU, B., KING, D., REHILL, A. & SALIMKUMAR, D. 2014. Dementia UK update. 2nd ed.

RAFF, M. C., FIELDS, K. L., HAKOMORI, S. I., MIRSKY, R., PRUSS, R. M. & WINTER, J. 1979. Cell-type-specific markers for distinguishing and studying neurons and the major classes of glial cells in culture. *Brain Res,* 174**,** 283-308.

RAJ, T., LI, Y. I., WONG, G., HUMPHREY, J., WANG, M., RAMDHANI, S., WANG, Y. C., NG, B., GUPTA, I., HAROUTUNIAN, V., SCHADT, E. E., YOUNG-PEARSE, T., MOSTAFAVI, S., ZHANG, B., SKLAR, P., BENNETT, D. A. & DE JAGER, P. L. 2018. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat Genet,* 50**,** 1584-1592.

RAJ, T., RYAN, K. J., REPLOGLE, J. M., CHIBNIK, L. B., ROSENKRANTZ, L., TANG, A., ROTHAMEL, K., STRANGER, B. E., BENNETT, D. A., EVANS, D. A., DE JAGER, P. L. & BRADSHAW, E. M. 2014. CD33: increased inclusion of exon

2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. *Hum Mol Genet,* 23**,** 2729-36.

RAJAN, K. B., WILSON, R. S., WEUVE, J., BARNES, L. L. & EVANS, D. A. 2015. Cognitive impairment 18 years before clinical diagnosis of Alzheimer disease dementia. *Neurology,* 85**,** 898-904.

RAMOS-VARA, J. A. & MILLER, M. A. 2014. When tissue antigens and antibodies get along: revisiting the technical aspects of immunohistochemistry--the red, brown, and blue technique. *Vet Pathol,* 51**,** 42-87.

RAMSDEN, M., KOTILINEK, L., FORSTER, C., PAULSON, J., MCGOWAN, E., SANTACRUZ, K., GUIMARAES, A., YUE, M., LEWIS, J., CARLSON, G., HUTTON, M. & ASHE, K. H. 2005. Age-dependent neurofibrillary tangle formation, neuron loss, and memory impairment in a mouse model of human tauopathy (P301L). *J Neurosci,* 25**,** 10637-47.

RAO, J. S., KELESHIAN, V. L., KLEIN, S. & RAPOPORT, S. I. 2012. Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients. *Transl Psychiatry,* 2**,** e132.

RIDGE, P. G., MUKHERJEE, S., CRANE, P. K., KAUWE, J. S. & ALZHEIMER'S DISEASE GENETICS, C. 2013. Alzheimer's disease: analyzing the missing heritability. *PLoS One,* 8**,** e79771.

RIES, M. & SASTRE, M. 2016. Mechanisms of Abeta Clearance and Degradation by Glial Cells. *Front Aging Neurosci,* 8**,** 160.

RISSO, D., NGAI, J., SPEED, T. P. & DUDOIT, S. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol,* 32**,** 896-902.

ROTHMAN, S. M., TANIS, K. Q., GANDHI, P., MALKOV, V., MARCUS, J., PEARSON, M., STEVENS, R., GILLILAND, J., WARE, C., MAHADOMRONGKUL, V., O'LOUGHLIN, E., ZEBALLOS, G., SMITH, R., HOWELL, B. J., KLAPPENBACH, J., KENNEDY, M. & MIRESCU, C. 2018. Human Alzheimer's disease gene expression signatures and immune profile in APP mouse models: a discrete transcriptomic view of Abeta plaque pathology. *J Neuroinflammation,* 15**,** 256.

ROWLAND, A., MINERS, J. O. & MACKENZIE, P. I. 2013. The UDP-glucuronosyltransferases: their role in drug metabolism and detoxification. *Int J Biochem Cell Biol,* 45**,** 1121-32.

SAHARA, N., PEREZ, P. D., LIN, W. L., DICKSON, D. W., REN, Y., ZENG, H., LEWIS, J. & FEBO, M. 2014. Age-related decline in white matter integrity in a mouse model of tauopathy: an in vivo diffusion tensor magnetic resonance imaging study. *Neurobiol Aging,* 35**,** 1364-74.

SAHRAEIAN, S. M. E., MOHIYUDDIN, M., SEBRA, R., TILGNER, H., AFSHAR, P. T., AU, K. F., BANI ASADI, N., GERSTEIN, M. B., WONG, W. H., SNYDER, M. P., SCHADT, E. & LAM, H. Y. K. 2017. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat Commun,* 8**,** 59.

SAITO, T., MATSUBA, Y., MIHIRA, N., TAKANO, J., NILSSON, P., ITOHARA, S., IWATA, N. & SAIDO, T. C. 2014. Single App knock-in mouse models of Alzheimer's disease. *Nat Neurosci,* 17**,** 661-3.

SAKAMOTO, M., ITOH, T., TUKEY, R. H. & FUJIWARA, R. 2015. Nicotine regulates the expression of UDP-glucuronosyltransferase (UGT) in humanized UGT1 mouse brain. *Drug Metab Pharmacokinet,* 30**,** 269-75.

SALIH, D. A., BAYRAM, S., GUELFI, M. S., REYNOLDS, R. H., SHOAI, M., RYTEN, M., BRENTON, J., ZHANG, D., MATARIN, M., BOTIA, J., SHAH, R., BROOKES, K., GUETTA-BARANES, T., MORGAN, K., BELLOU, E., CUMMINGS, D. M., HARDY, J., EDWARDS, F. A. & ESCOTT-PRICE, V. 2018. Genetic variability in response to Aβ deposition influences Alzheimer's risk. *bioRxiv***,** 437657.

SANCHEZ-MUT, J. V., ASO, E., PANAYOTIS, N., LOTT, I., DIERSSEN, M., RABANO, A., URDINGUIO, R. G., FERNANDEZ, A. F., ASTUDILLO, A., MARTIN-SUBERO, J. I., BALINT, B., FRAGA, M. F., GOMEZ, A., GURNOT, C., ROUX, J. C., AVILA, J., HENSCH, T. K., FERRER, I. & ESTELLER, M. 2013. DNA methylation map of mouse and human brain identifies target genes in Alzheimer's disease. *Brain,* 136**,** 3018-27.

SANTACRUZ, K., LEWIS, J., SPIRES, T., PAULSON, J., KOTILINEK, L., INGELSSON, M., GUIMARAES, A., DETURE, M., RAMSDEN, M., MCGOWAN,

E., FORSTER, C., YUE, M., ORNE, J., JANUS, C., MARIASH, A., KUSKOWSKI, M., HYMAN, B., HUTTON, M. & ASHE, K. H. 2005. Tau suppression in a neurodegenerative mouse model improves memory function. *Science,* 309**,** 476-81.

SARDIELLO, M., ANNUNZIATA, I., ROMA, G. & BALLABIO, A. 2005. Sulfatases and sulfatase modifying factors: an exclusive and promiscuous relationship. *Hum Mol Genet,* 14**,** 3203-17.

SASAGURI, H., NILSSON, P., HASHIMOTO, S., NAGATA, K., SAITO, T., DE STROOPER, B., HARDY, J., VASSAR, R., WINBLAD, B. & SAIDO, T. C. 2017. APP mouse models for Alzheimer's disease preclinical studies. *EMBO J,* 36**,** 2473-2487.

SEKAR, S., MCDONALD, J., CUYUGAN, L., ALDRICH, J., KURDOGLU, A., ADKINS, J., SERRANO, G., BEACH, T. G., CRAIG, D. W., VALLA, J., REIMAN, E. M. & LIANG, W. S. 2015. Alzheimer's disease is associated with altered expression of genes involved in immune response and mitochondrial processes in astrocytes. *Neurobiol Aging,* 36**,** 583-91.

SELKOE, D. J. 2011. Alzheimer's disease. *Cold Spring Harb Perspect Biol,* 3.

SELKOE, D. J. & HARDY, J. 2016. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med,* 8**,** 595-608.

SEN, P., SHAH, P. P., NATIVIO, R. & BERGER, S. L. 2016. Epigenetic mechanisms of longevity and aging. *Cell,* 166**,** 822-839.

SERRANO-POZO, A., FROSCH, M. P., MASLIAH, E. & HYMAN, B. T. 2011. Neuropathological alterations in Alzheimer disease. *Cold Spring Harb Perspect Med,* 1**,** a006189.

SHALEV, D. & ARBUCKLE, M. R. 2017. Metabolism and Memory: Obesity, Diabetes, and Dementia. *Biol Psychiatry,* 82**,** e81-e83.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res,* 13**,** 2498-504.

SHETTY, H. U., HOLLOWAY, H. W. & SCHAPIRO, M. B. 1996. Cerebrospinal fluid and plasma distribution of myo-inositol and other polyols in Alzheimer disease. *Clin Chem,* 42**,** 298-302.

SIMPSON, J. E., INCE, P. G., SHAW, P. J., HEATH, P. R., RAMAN, R., GARWOOD, C. J., GELSTHORPE, C., BAXTER, L., FORSTER, G., MATTHEWS, F. E., BRAYNE, C., WHARTON, S. B., FUNCTION, M. R. C. C. & AGEING NEUROPATHOLOGY STUDY, G. 2011. Microarray analysis of the astrocyte transcriptome in the aging brain: relationship to Alzheimer's pathology and APOE genotype. *Neurobiol Aging,* 32**,** 1795-807.

SIMS, R., VAN DER LEE, S. J., NAJ, A. C., BELLENGUEZ, C., BADARINARAYAN, N., JAKOBSDOTTIR, J., KUNKLE, B. W., BOLAND, A., RAYBOULD, R., BIS, J. C., MARTIN, E. R., GRENIER-BOLEY, B., HEILMANN-HEIMBACH, S., CHOURAKI, V., KUZMA, A. B., SLEEGERS, K., VRONSKAYA, M., RUIZ, A., GRAHAM, R. R., OLASO, R., HOFFMANN, P., GROVE, M. L., VARDARAJAN, B. N., HILTUNEN, M., NOTHEN, M. M., WHITE, C. C., HAMILTON-NELSON, K. L., EPELBAUM, J., MAIER, W., CHOI, S. H., BEECHAM, G. W., DULARY, C., HERMS, S., SMITH, A. V., FUNK, C. C., DERBOIS, C., FORSTNER, A. J., AHMAD, S., LI, H., BACQ, D., HAROLD, D., SATIZABAL, C. L., VALLADARES, O., SQUASSINA, A., THOMAS, R., BRODY, J. A., QU, L., SANCHEZ-JUAN, P., MORGAN, T., WOLTERS, F. J., ZHAO, Y., GARCIA, F. S., DENNING, N., FORNAGE, M., MALAMON, J., NARANJO, M. C. D., MAJOUNIE, E., MOSLEY, T. H., DOMBROSKI, B., WALLON, D., LUPTON, M. K., DUPUIS, J., WHITEHEAD, P., FRATIGLIONI, L., MEDWAY, C., JIAN, X., MUKHERJEE, S., KELLER, L., BROWN, K., LIN, H., CANTWELL, L. B., PANZA, F., MCGUINNESS, B., MORENO-GRAU, S., BURGESS, J. D., SOLFRIZZI, V., PROITSI, P., ADAMS, H. H., ALLEN, M., SERIPA, D., PASTOR, P., CUPPLES, L. A., PRICE, N. D., HANNEQUIN, D., FRANK-GARCIA, A., LEVY, D., CHAKRABARTY, P., CAFFARRA, P., GIEGLING, I., BEISER, A. S., GIEDRAITIS, V., HAMPEL, H., GARCIA, M. E., WANG, X., LANNFELT, L., MECOCCI, P., EIRIKSDOTTIR, G., CRANE, P. K., PASQUIER, F., BOCCARDI, V., et al. 2017. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet,* 49**,** 1373-1384.

SMITH, A. R., MILL, J., SMITH, R. G. & LUNNON, K. 2016a. Elucidating novel dysfunctional pathways in Alzheimer's disease by integrating loci identified in genetic and epigenetic studies. *Neuroepigenetics,* 6**,** 32-50.

SMITH, A. R., SMITH, R. G., BURRAGE, J., TROAKES, C., AL-SARRAJ, S., KALARIA, R. N., SLOAN, C., ROBINSON, A. C., MILL, J. & LUNNON, K. 2019. A cross-brain regions study of ANK1 DNA methylation in different neurodegenerative diseases. *Neurobiol Aging,* 74**,** 70-76.

SMITH, A. R., SMITH, R. G., CONDLIFFE, D., HANNON, E., SCHALKWYK, L., MILL, J. & LUNNON, K. 2016b. Increased DNA methylation near TREM2 is consistently seen in the superior temporal gyrus in Alzheimer's disease brain. *Neurobiol Aging,* 47**,** 35-40.

SMITH, R. G., HANNON, E., DE JAGER, P. L., CHIBNIK, L., LOTT, S. J., CONDLIFFE, D., SMITH, A. R., HAROUTUNIAN, V., TROAKES, C., AL-SARRAJ, S., BENNETT, D. A., POWELL, J., LOVESTONE, S., SCHALKWYK, L., MILL, J. & LUNNON, K. 2018. Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. *Alzheimers Dement*.

SMITH, R. G. & LUNNON, K. 2017. DNA modifications and Alzheimer's disease. *In:* DELGADO-MORALES, R. (ed.) *Neuroepigenomics in Aging and Disease.* Springer.

SPIERS, H., HANNON, E., SCHALKWYK, L. C., BRAY, N. J. & MILL, J. 2017. 5-hydroxymethylcytosine is highly dynamic across human fetal brain development. *BMC Genomics,* 18**,** 738.

SPIRES, T. L., ORNE, J. D., SANTACRUZ, K., PITSTICK, R., CARLSON, G. A., ASHE, K. H. & HYMAN, B. T. 2006. Region-specific dissociation of neuronal loss and neurofibrillary pathology in a mouse model of tauopathy. *Am J Pathol,* 168**,** 1598-607.

STRITTMATTER, W. J., SAUNDERS, A. M., SCHMECHEL, D., PERICAK-VANCE, M., ENGHILD, J., SALVESEN, G. S. & ROSES, A. D. 1993. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci U S A,* 90**,** 1977-81.

SWARUP, V., HINZ, F. I., REXACH, J. E., NOGUCHI, K. I., TOYOSHIBA, H., ODA, A., HIRAI, K., SARKAR, A., SEYFRIED, N. T., CHENG, C., HAGGARTY, S. J., INTERNATIONAL FRONTOTEMPORAL DEMENTIA GENOMICS, C., GROSSMAN, M., VAN DEERLIN, V. M., TROJANOWSKI, J. Q., LAH, J. J., LEVEY, A. I., KONDOU, S. & GESCHWIND, D. H. 2019. Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia. *Nat Med,* 25**,** 152-164.

SZABO, R., NETZEL-ARNETT, S., HOBSON, J. P., ANTALIS, T. M. & BUGGE, T. H. 2005. Matriptase-3 is a novel phylogenetically preserved membrane-anchored serine protease with broad serpin reactivity. *Biochem J,* 390**,** 231-42.

TAN, M. G., CHUA, W. T., ESIRI, M. M., SMITH, A. D., VINTERS, H. V. & LAI, M. K. 2010. Genome wide profiling of altered gene expression in the neocortex of Alzheimer's disease. *J Neurosci Res,* 88**,** 1157-69.

THAL, D. R., RUB, U., ORANTES, M. & BRAAK, H. 2002. Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology,* 58**,** 1791-800.

TOHGI, H., UTSUGISAWA, K., NAGANE, Y., YOSHIMURA, M., GENDA, Y. & UKITSU, M. 1999a. Reduction with age in methylcytosine in the promoter region -224 approximately -101 of the amyloid precursor protein gene in autopsy human cortex. *Brain Res Mol Brain Res,* 70**,** 288-92.

TOHGI, H., UTSUGISAWA, K., NAGANE, Y., YOSHIMURA, M., UKITSU, M. & GENDA, Y. 1999b. The methylation status of cytosines in a tau gene promoter region alters with age to downregulate transcriptional activity in human cerebral cortex. *Neurosci Lett,* 275**,** 89-92.

TOLLERVEY, J. R., WANG, Z., HORTOBAGYI, T., WITTEN, J. T., ZARNACK, K., KAYIKCI, M., CLARK, T. A., SCHWEITZER, A. C., ROT, G., CURK, T., ZUPAN, B., ROGELJ, B., SHAW, C. E. & ULE, J. 2011. Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. *Genome Res,* 21**,** 1572-82.

TOSH, J. L., RICKMAN, M., RHYMES, E., NORONA, F. E., CLAYTON, E., MUCKE, L., ISAACS, A. M., FISHER, E. M. C. & WISEMAN, F. K. 2017. The

integration site of the APP transgene in the J20 mouse model of Alzheimer's disease. *Wellcome Open Res,* 2**,** 84.

TRAPNELL, C., PACHTER, L. & SALZBERG, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* 25**,** 1105-11.

TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc,* 7**,** 562-78.

TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol,* 28**,** 511-5.

UEDA, T., WATANABE-FUKUNAGA, R., FUKUYAMA, H., NAGATA, S. & FUKUNAGA, R. 2004. Mnk2 and Mnk1 are essential for constitutive and inducible phosphorylation of eukaryotic initiation factor 4E but not for cell growth or development. *Mol Cell Biol,* 24**,** 6539-49.

UGAROVA, T. P. & YAKUBENKO, V. P. 2001. Recognition of fibrinogen by leukocyte integrins. *Ann N Y Acad Sci,* 936**,** 368-85.

VAN GURP, T. P., MCINTYRE, L. M. & VERHOEVEN, K. J. 2013. Consistent errors in first strand cDNA due to random hexamer mispriming. *PLoS One,* 8**,** e85583.

WALKER, M. P., LAFERLA, F. M., ODDO, S. S. & BREWER, G. J. 2013. Reversible epigenetic histone modifications and Bdnf expression in neurons with aging and from a mouse model of Alzheimer's disease. *Age (Dordr),* 35**,** 519-31.

WANG, H., LI, Y., RYDER, J. W., HOLE, J. T., EBERT, P. J., AIREY, D. C., QIAN, H. R., LOGSDON, B., FISHER, A., AHMED, Z., MURRAY, T. K., CAVALLINI, A., BOSE, S., EASTWOOD, B. J., COLLIER, D. A., DAGE, J. L., MILLER, B. B., MERCHANT, K. M., O'NEILL, M. J. & DEMATTOS, R. B. 2018. Genome-wide RNAseq study of the molecular mechanisms underlying microglia activation in response to pathological tau perturbation in the rTg4510 tau transgenic animal model. *Mol Neurodegener,* 13**,** 65.

WANG, L. L., MIN, L., GUO, Q. D., ZHANG, J. X., JIANG, H. L., SHAO, S., XING, J. G., YIN, L. L., LIU, J. H., LIU, R. & GUO, S. L. 2017. Profiling microRNA from Brain by Microarray in a Transgenic Mouse Model of Alzheimer's Disease. *Biomed Res Int,* 2017**,** 8030369.

WANG, Y. & MANDELKOW, E. 2016. Tau in physiology and pathology. *Nat Rev Neurosci,* 17**,** 5-21.

WANG, Z., ZANG, C., ROSENFELD, J. A., SCHONES, D. E., BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., PENG, W., ZHANG, M. Q. & ZHAO, K. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet,* 40**,** 897-903.

WASKIEWICZ, A. J., FLYNN, A., PROUD, C. G. & COOPER, J. A. 1997. Mitogen-activated protein kinases activate the serine/threonine kinases Mnk1 and Mnk2. *EMBO J,* 16**,** 1909-20.

WATSON, C. T., ROUSSOS, P., GARG, P., HO, D. J., AZAM, N., KATSEL, P. L., HAROUTUNIAN, V. & SHARP, A. J. 2016. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Med,* 8**,** 5.

WES, P. D., EASTON, A., CORRADI, J., BARTEN, D. M., DEVIDZE, N., DECARR, L. B., TRUONG, A., HE, A., BARREZUETA, N. X., POLSON, C., BOURIN, C., FLYNN, M. E., KEENAN, S., LIDGE, R., MEREDITH, J., NATALE, J., SANKARANARAYANAN, S., CADELINA, G. W., ALBRIGHT, C. F. & CACACE, A. M. 2014. Tau overexpression impacts a neuroinflammation gene expression network perturbed in Alzheimer's disease. *PLoS One,* 9**,** e106050.

WEST, R. L., LEE, J. M. & MAROUN, L. E. 1995. Hypomethylation of the amyloid precursor protein gene in the brain of an Alzheimer's disease patient. *J Mol Neurosci,* 6**,** 141-6.

WRECZYCKA, K., GOSDSCHAN, A., YUSUF, D., GRUNING, B., ASSENOV, Y. & AKALIN, A. 2017. Strategies for analyzing bisulfite sequencing data. *J Biotechnol,* 261**,** 105-115.

WRIGHT, A. L., ZINN, R., HOHENSINN, B., KONEN, L. M., BEYNON, S. B., TAN, R. P., CLARK, I. A., ABDIPRANOTO, A. & VISSEL, B. 2013.

Neuroinflammation and neuronal loss precede Abeta plaque deposition in the hAPP-J20 mouse model of Alzheimer's disease. *PLoS One,* 8**,** e59586.

XU, P. T., LI, Y. J., QIN, X. J., SCHERZER, C. R., XU, H., SCHMECHEL, D. E., HULETTE, C. M., ERVIN, J., GULLANS, S. R., HAINES, J., PERICAK-VANCE, M. A. & GILBERT, J. R. 2006. Differences in apolipoprotein E3/3 and E4/4 allele-specific gene expression in hippocampus in Alzheimer disease. *Neurobiol Dis,* 21**,** 256-75.

YANG, J., WEEDON, M. N., PURCELL, S., LETTRE, G., ESTRADA, K., WILLER, C. J., SMITH, A. V., INGELSSON, E., O'CONNELL, J. R., MANGINO, M., MAGI, R., MADDEN, P. A., HEATH, A. C., NYHOLT, D. R., MARTIN, N. G., MONTGOMERY, G. W., FRAYLING, T. M., HIRSCHHORN, J. N., MCCARTHY, M. I., GODDARD, M. E., VISSCHER, P. M. & CONSORTIUM, G. 2011. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet,* 19**,** 807-12.

YONG, W. S., HSU, F. M. & CHEN, P. Y. 2016. Profiling genome-wide DNA methylation. *Epigenetics Chromatin,* 9**,** 26.

YOUN, H., JEOUNG, M., KOO, Y., JI, H., MARKESBERY, W. R., JI, I. & JI, T. H. 2007. Kalirin is under-expressed in Alzheimer's disease hippocampus. *J Alzheimers Dis,* 11**,** 385-97.

YOUNG, M. D., WAKEFIELD, M. J., SMYTH, G. K. & OSHLACK, A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol,* 11**,** R14.

YU, L., CHIBNIK, L. B., SRIVASTAVA, G. P., POCHET, N., YANG, J., XU, J., KOZUBEK, J., OBHOLZER, N., LEURGANS, S. E., SCHNEIDER, J. A., MEISSNER, A., DE JAGER, P. L. & BENNETT, D. A. 2015. Association of brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. *JAMA Neurol,* 72**,** 15-24.

YUE, M., HANNA, A., WILSON, J., RODER, H. & JANUS, C. 2011. Sex difference in pathology and memory decline in rTg4510 mouse model of tauopathy. *Neurobiol Aging,* 32**,** 590-603.

ZAJKOWICZ, A., GDOWICZ-KLOSOK, A., KRZESNIAK, M., JANUS, P., LASUT, B. & RUSIN, M. 2018. The Alzheimer's disease-associated TREM2 gene is regulated by p53 tumor suppressor protein. *Neurosci Lett,* 681**,** 62-67.

ZERBINO, D. R., ACHUTHAN, P., AKANNI, W., AMODE, M. R., BARRELL, D., BHAI, J., BILLIS, K., CUMMINS, C., GALL, A., GIRON, C. G., GIL, L., GORDON, L., HAGGERTY, L., HASKELL, E., HOURLIER, T., IZUOGU, O. G., JANACEK, S. H., JUETTEMANN, T., TO, J. K., LAIRD, M. R., LAVIDAS, I., LIU, Z., LOVELAND, J. E., MAUREL, T., MCLAREN, W., MOORE, B., MUDGE, J., MURPHY, D. N., NEWMAN, V., NUHN, M., OGEH, D., ONG, C. K., PARKER, A., PATRICIO, M., RIAT, H. S., SCHUILENBURG, H., SHEPPARD, D., SPARROW, H., TAYLOR, K., THORMANN, A., VULLO, A., WALTS, B., ZADISSA, A., FRANKISH, A., HUNT, S. E., KOSTADIMA, M., LANGRIDGE, N., MARTIN, F. J., MUFFATO, M., PERRY, E., RUFFIER, M., STAINES, D. M., TREVANION, S. J., AKEN, B. L., CUNNINGHAM, F., YATES, A. & FLICEK, P. 2018. Ensembl 2018. *Nucleic Acids Res,* 46**,** D754-D761.

ZHANG, B., GAITERI, C., BODEA, L. G., WANG, Z., MCELWEE, J., PODTELEZHNIKOV, A. A., ZHANG, C., XIE, T., TRAN, L., DOBRIN, R., FLUDER, E., CLURMAN, B., MELQUIST, S., NARAYANAN, M., SUVER, C., SHAH, H., MAHAJAN, M., GILLIS, T., MYSORE, J., MACDONALD, M. E., LAMB, J. R., BENNETT, D. A., MOLONY, C., STONE, D. J., GUDNASON, V., MYERS, A. J., SCHADT, E. E., NEUMANN, H., ZHU, J. & EMILSSON, V. 2013. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell,* 153**,** 707-20.

ZHANG, B. & HORVATH, S. 2005. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol,* 4**,** Article17.

ZHANG, H., ZHAO, T., LI, Z., YAN, M., ZHAO, H., ZHU, B. & LI, P. 2017. Transcriptional Profile of Kidney from Type 2 Diabetic db/db Mice. *J Diabetes Res,* 2017**,** 8391253.

ZHANG, J., HU, M., TENG, Z., TANG, Y. P. & CHEN, C. 2014. Synaptic and cognitive improvements by inhibition of 2-AG metabolism are through upregulation of microRNA-188-3p in a mouse model of Alzheimer's disease. *J Neurosci,* 34**,** 14919-33.

ZHANG, Q., YU, J. T., ZHU, Q. X., ZHANG, W., WU, Z. C., MIAO, D. & TAN, L. 2010. Complement receptor 1 polymorphisms and risk of late-onset Alzheimer's disease. *Brain Res,* 1348**,** 216-21.

ZHANG, Y., SLOAN, S. A., CLARKE, L. E., CANEDA, C., PLAZA, C. A., BLUMENTHAL, P. D., VOGEL, H., STEINBERG, G. K., EDWARDS, M. S., LI, G., DUNCAN, J. A., 3RD, CHESHIER, S. H., SHUER, L. M., CHANG, E. F., GRANT, G. A., GEPHART, M. G. & BARRES, B. A. 2016. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron,* 89**,** 37-53.

ZHAO, J., ZHU, Y., YANG, J., LI, L., WU, H., DE JAGER, P. L., JIN, P. & BENNETT, D. A. 2017. A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. *Alzheimers Dement,* 13**,** 674-688.

ZHOU, Q., XIONG, Y., HUANG, X. R., TANG, P., YU, X. & LAN, H. Y. 2015. Identification of Genes Associated with Smad3-dependent Renal Injury by RNA-seq-based Transcriptome Analysis. *Sci Rep,* 5**,** 17901.

ZILLER, M. J., HANSEN, K. D., MEISSNER, A. & ARYEE, M. J. 2015. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat Methods,* 12**,** 230-2, 1 p following 232.

ZUROFF, L., DALEY, D., BLACK, K. L. & KORONYO-HAMAOUI, M. 2017. Clearance of cerebral Abeta in Alzheimer's disease: reassessing the role of microglia and monocytes. *Cell Mol Life Sci,* 74**,** 2167-2201.