

The Molecular Determinants of IGF-1 (Mecasermin) Action in Rett Syndrome Patients

This thesis submitted to University of Dublin, Trinity College

For the degree of MSc (Masters) by Research



Stephen Shovlin

31/10/2017

Supervisor: Dr Daniela Tropea

Department of Psychiatry,

School of Medicine,

Trinity Translation Medicine Institute,

St. James's Hospital.

Table of Contents:	ii
Declaration and Statement of Plagiarism	iv
Acknowledgements	v
Abbreviations	vi
List of Figures and Tables	xv
List of Appendices	xvii
Abstract	1
Chapter One: Introduction	
1.1 Background and History of Rett Syndrome	2
1.2 Comorbidities of Rett Syndrome	4
1.3 Mutations in Rett Syndrome	5
1.4 <i>MECP2</i> Functions in Rett Syndrome	6
1.5 <i>IGF-1</i> Functioning and Molecular Mechanisms	8
1.6 Transcriptome-Level Analysis of Rett Patient's Tissues	10
1.7 Transcriptome Analysis in Post-Mortem Brain Samples	12
1.8 Transcriptome Analysis in Blood Tissue Samples	14
1.9 Transcriptome Analysis in Induced Pluripotent Stem Cells Derived from Rett Patient's Tissues	16
1.10 IGF-1 in Clinical Trials	18
1.11 Aim, Hypotheses and Interpretation of Literature	19
Chapter Two: Methods	
2.1 Samples from Boston Children's Hospital	22
2.2 Samples from Versilia Hospital Italy	27
2.3 Determining Response of Mecasermin Trial Participants	31
2.4 Laboratory and Analysis Methods	36
Chapter Three: Results	
3.1 Summary of Results	52
3.2 Time-Point Comparisons	54
3.3 Time-Point Comparison in Responders and Non-Responders of BCH	63

3.4	Response Based Comparisons	65
Chapter Four: Discussion		
4.1	Main Findings	69
4.2	Context of Literature	73
4.3	Limitations	76
4.4	Future Directions	78
4.5	Conclusion	80
References		82
Appendices		
Appendix A		105
Appendix B		108

Declaration and Statement of Plagiarism

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library condition of use and acknowledgement

Signed

Date

Acknowledgements

This author would like to acknowledge Dr Daniela Tropea for her tireless efforts, guidance and advice, Albert Sanfeliu, Dr Elaine Kenny and Kevin Harte for their technical assistance and all the staff of the Institute of Molecular Medicine (IMM), and Trinity Biomedical Sciences Institute (TBSI).

I would also thank my parents Ian Shovlin and Ursula Keegan for their great support and encouragement. They were always there to help.

Abbreviations:

A

AD	-	Alzheimer's disease
ADA	-	Adenosine Deaminase
ADAMS	-	Anxiety Depression and Mood Scale
ADP	-	Adenosine Di-Phosphate
ADTRP	-	Androgen Dependent TFPI Regulating Protein
Akt	-	Protein Kinase B
AKT	-	AKT Serine/Threonine Kinase 1
ASD	-	Autism Spectrum Disorder
ATP	-	Adenosine Tri-Phosphate

B

BA	-	Brodmann's Area
BAM	-	Binary Alignment/Map
BBS2	-	Bardet-Biedl Syndrome type 2
BCH	-	Boston Children's Hospital
BCH-t0	-	Sample from start of MAD phase in Boston Children's Hospital
BCH-t1	-	Sample from the end of MAD phase in Boston Children's Hospital
BCH-t2	-	Sample from start of OLE phase in Boston Children's Hospital
BCH-t3	-	Sample from end of OLE phase in Boston Children's Hospital
BCL2A1	-	BCL2 Related Protein A1
BID	-	Twice Daily (<i>bis in die</i>)
BP	-	Base Pair
BR1	-	Resuspension buffer
BR2	-	Binding buffer

BR3	-	Wash buffer
BR4	-	Wash buffer
BR5	-	Elution buffer
BRSK1	-	BR Serine/Threonine Kinase 1
C		
CAG	-	Cytosine-Adenine-Guanine
CCDC85c	-	Coiled-Coil Domain Containing 85c
CDKL-5	-	Cyclin Dependent Kinase Like-5
cDNA	-	Complementary Deoxyribonucleic Acid
CGI	-	Clinical Global Impression
ChIP-Seq	-	Chromatin Immunoprecipitation-Sequencing
CNS	-	Central Nervous System
COX	-	Cytochrome C Oxidase
CO2	-	Carbon Dioxide
CRYAB	-	Crystallin Alpha B
CSBS-DP Profile	-	Communication and Symbolic Behaviour Scales-Development- Profile
CSF	-	Cerebrospinal Fluid
CSV	-	Comma Separated Value
C1Q	-	Complement C1Q
C1QA	-	Complement C1Q subcomponent subunit A
C1QB	-	Complement C1Q subcomponent subunit B
C1QC	-	Complement C1Q subcomponent subunit C
D		
DG	-	Dentate Gyrus

DGE - Differential Gene Expression

DMTN - Dematin Actin Binding Protein

DNA - Deoxyribonucleic Acid

DNM1 - Dynamin 1

E

EAAT1 - Solute Carrier family 1 Member 3

EEF1A1P7 - Elongation Factor 1 Alpha 1 Pseudogene 7

EEG - Electroencephalogram

ERCC5 - Excision Repair 5 Endonuclease

ESAT - End Sequence Analysis ToolKit

EtOH - Ethanol

F

FDA - United States Food and Drug Administration

FDR - False Discovery Rate

FOXG1 - Forkhead box G1

FPKM - Fragments per Kilobase of million Mapped reads

FU - Fluorescence

FXYP1 - FXYP Domain Containing Ion Transport Regulator 1

G

GFAP - Glial Fibrillary Acidic Protein

GH - Growth Hormone

GSK3 β - glycogen synthase 3 β

GRCh38 - Genome Reference Consortium Human Build 38

GTF - Gene Transfer Format

GTF2IRD2B - GTF2I Repeat Domain Containing 2B

H

HDAC	-	Histone Deacetylase
HDH ^{Q111}	-	Huntingtin Knock-in mice
HLA-DRB5	-	Major Histocompatibility Complex Class II, DR Beta 5
Htt	-	Huntingtin

I

ID	-	Inter Domain
IGF-1	-	Insulin-like Growth Factor-1
IGF-1R	-	Insulin-like Growth Factor-1 Receptor
IL1	-	Interleukin-1
IL1R1	-	Interleukin-1 Receptor Type 1
IL7R	-	Interleukin 7 Receptor
IMI	-	Innovative Medicines Initiative
IPSC	-	Induced Pluripotent Stem Cells
IR	-	Insulin Receptor
IRS-1	-	Insulin Receptor Substrate-1
ISS	-	International Scoring System

J

JCHAIN	-	Joining Chain of Multimeric IgA and IgM
--------	---	---

K

KEGG	-	Kyoto Encyclopaedia of Genes and Genomes
KLF4	-	Krupple Like Factor 4
KO	-	Knockout

L

LEAP	-	Longitudinal European Autism Project
------	---	--------------------------------------

LIMMA	-	Linear Models for MicroArray
Log ₂ FC	-	Log ₂ Fold Change
M		
MAPK	-	Mitogen-Activated Protein Kinase
MAD	-	Multiple Ascending Dosage
MBD	-	Methyl Binding Domain
MCEMP1	-	Mast Cell Expressed Membrane Protein 1
mCAC	-	methyl-Cytosine Adenine Cytosine
mCG	-	methyl-Cytosine Guanosine
mCH	-	methyl-Cytosine Guanosine/Adenine/Thymine
MELAS	-	Mitochondrial Encephalomyopathy, lactic acidosis and stroke-like episodes
Mecasermin	-	Humanised Recombinant IGF-1
MECP2	-	Methyl CpG binding protein 2
MHC	-	Major Histocompatibility Complex
miRNA	-	microRNA
MIT	-	Massachusetts Institute of Technology
mRNA	-	Messenger RNA
Mut-	-	Mutant-
MT-ATP6	-	Mitochondrially encoded ATP Synthase 6
MT-CO1	-	Mitochondrially encoded Cytochrome C Oxidase Subunit 1
MT-CO2	-	Mitochondrially encoded Cytochrome C Oxidase Subunit 2
MYC	-	MYC Protein Oncogene, BHLH Transcription Factor
N		
NaOH	-	Sodium Hydroxide

NARP	-	Neurogenic Muscle Weakness Ataxia and Retinitis Pigmentosa
NARS2	-	Asparaginyl-tRNA Synthetase 2, Mitochondrial
NCoR/SMRT	-	Nuclear Co-Repressor/Silencing Mediator of Retinoic acid Thyroid hormone receptor
NDUFA1	-	NADH: Ubiquinone Oxidoreductase Subunit A1
NER	-	Nucleotide Excision Repair
NFAT	-	Nuclear Factor of Activated T-cells
NFκB	-	Nuclear Factor of kappa B subunit
NLS	-	Nuclear Localisation Signal
NMDA	-	N-Methyl- D-Aspartic acid

O

OCT4	-	Octamer binding transcription factor 4
OE	-	Over-Expressing
OLE	-	Open Label Extension
OS	-	Oxidative Stress

P

PBMC	-	Peripheral Blood lymphomonocyte cell
PBRT	-	PAXgene Blood RNA Tubes
PBZ	-	Pini Bonuccelli Zappella Scale
pCO ₂	-	Partial Pressure of Carbon Dioxide
PCR	-	Polymerase Chain Reaction
PI3K	-	Phosphatidylinositol-3 kinase
PKO	-	Peripheral Knock-Out
PLLP	-	Plasmolipin
PMB	-	Post-Mortem Brain

PMDS	-	Phelan-McDermid Syndrome
PNRC2	-	Proline Rich Nuclear Receptor Coactivator 2
PSV	-	Preserved Speech Variant
PTSVAS	-	Parent Targeted Symptom Visual Analogue Scale

Q

R

RAB37	-	RAS Oncogene Family, member 37
RFID	-	Radio-Frequency Identification
RNA	-	Ribonucleic acid
RNA-Seq	-	RNA-Sequencing
ROS	-	Reactive Oxygen Species
RPL27	-	Ribosomal protein L27
RSBQ	-	Rett Syndrome Behavioural Questionnaire
RSS	-	Rett Severity Score
RTT	-	Rett Syndrome

S

SAM	-	Significance Analysis of Microarrays
SCRB-Seq	-	Single-Cell RNA Barcoding and Sequencing
SEC11C	-	SEC11 Homolog C Signal Peptidase Complex Subunit
SLX1A	-	Structure-Specific Endonuclease subunit, Homologs A
SLX1B	-	Structure-Specific Endonuclease subunit, Homologs B
Sin3a	-	Sin Transcriptional Regulator Family A
SMCR5	-	Smith-Magenis Syndrome Chromosome Region Candidate 5
SOX2	-	Sex Determining Region Y-Box 2
SVZ	-	Subventricular Zone

S100A8	-	S100 Calcium Binding Protein A8
S100A9	-	S100 Calcium Binding Protein A9
S100A12	-	S100 Calcium Binding Protein A12
S100A13	-	S100 Calcium Binding Protein A13
T		
TBI	-	Traumatic Brain Injury
TE	-	Tris-Ethylenediaminetetraacetic Acid
TFPI	-	Tissue Factor Pathway Inhibitor
TGF β	-	Transforming growth factor β
TMEM176A	-	Transmembrane Protein 176A
TMEM176B	-	Transmembrane Protein 176B
TRD	-	Transcriptional Repressor Domain
TRD-NLS	-	Transcriptional Repressor Domain-Nuclear Localisation Signal
Tris-HCl	-	Tris-Hydrochloride
U		
UPS	-	Ubiquitin-Proteasome System
UTR	-	Untranslated Regions
V		
VH-t0	-	Untreated samples from start of Versilia Hospital trial
VH-t1	-	Treated samples at the end of Versilia Hospital trial
VH-t2	-	Treated samples over a year after the end of Versilia Hospital trial
W		
WT-	-	Wild Type
X		
XCI	-	X-Chromosome Inactivation

Y

Z

Numbers

5mC - 5-methyl Cytosine

5mCG - 5-methyl Cytosine Guanosine

List of Figures and Tables

Figures

Figure 2.1 – Time points of whole blood samples for BCH mecamermin trial

Figure 2.2 – Dosage and structure of BCH mecamermin trial

Figure 2.3 – Structure of the whole blood samples from Versilia Hospital mecamermin trial

Figure 2.4 - Graph showing the change in apnoeic index in BCH mecamermin trial

Figure 2.5 - Plot of PBZ scores for RTT samples from Versilia Hospital mecamermin trial

Figure 2.6 – Bioanalyser graph for contaminated library preparation

Figure 2.7 – Bioanalyser graph of clean pooled library preparations

Figure 2.8 – Example of FASTQC plot

Figure 3.1 – Schematic of the blood samples analysed for BCH mecamermin trial

Figure 3.2 – Schematic of nucleotide excision repair pathway (KEGG: 03420)

Figure 3.3 – Schematic of the Fanconi anemia pathway (KEGG: 03460)

Figure 3.4 – Schematic of the oxidative phosphorylation pathway (KEGG: 00190)

Figure 3.5 – Box plot of absolute perturbation values for over-represented genes in the oxidative phosphorylation pathway

Figure 3.6 – Schematic of the Huntington's disease pathway (KEGG: 05016)

Figure 3.7 – Expression of overlapping genes from response comparison based on FPKM values of Versilia Hospital mecamermin trial

Figure 3.8 – Expression of overlapping genes from response comparisons based on log₂ counts values of BCH mecamermin trial

Tables

Table 2.1 – Blood sample data for mecamermin trial in Boston Children's Hospital

Table 2.2 – Criteria for Rett severity score (RSS)

Table 2.3 – Blood sample data for mecamermin trial based in Versilia Hospital

Table 2.4 – Summary of participants of mecasermin trial from Versilia Hospital and BCH data

Table 2.5 - Apnoeic index data for RTT trial participants over time points of the BCH mecasermin trial

Table 2.6 – Criteria for PBZ scale

Table 2.7 – Dates for PBZ assessments and mecasermin treatment for Versilia hospital trial

Table 2.8 – PCR cycles for library enrichment

Table 2.9 – miSeq sample sheet

Table 3.1 – Significant results from time point comparisons from mecasermin trials from Versilia Hospital and BCH

Table 3.2 – Significant results from responders and non-responders from BCH mecasermin trial

Table 3.3 – Significant results from response based comparisons from Versilia Hospital and BCH mecasermin trial

Table 3.4 – Overlapping significant genes from edgeR analysis based on response comparisons in BCH and Versilia Hospital trials

Table 4.1 – Summary of the main findings from analyses

List of the Appendices:

- Appendix A - Full List of Significant DGE Results from edgeR Analysis
- Appendix B - Full List of Significant DGE Results from DESeq2 Analysis

Abstract

This is a preliminary investigation of effects of insulin-like growth factor-1 (IGF-1) treatment on differential gene expression (DGE) of whole blood samples from Rett syndrome (RTT) patients. Samples were obtained from two independent open label clinical trials testing the safety and efficacy of recombinant human IGF-1 (mecasermin).

We used patient matched samples to compare DGE for different time intervals across these trials. Due to heterogeneous design the only time interval that could be compared across the studies was the untreated samples.

Our initial findings showed a clear down regulation of two genes that encode core subunits in the mitochondrial complex enzymes after treatment with mecasermin. Mitochondrial dysfunctions have been shown in RTT patients in the past and recent lines of evidence suggests that mitochondrial dysfunction causes increased oxidative stress in RTT.

We also found a number of immune response functioning genes to be differentially expressed, including genes related to antigen presenting cells and dendritic cells. Immune responses have been identified in RTT, including autoimmune inflammation and glial cell activity.

Our analysis identified a set of significantly differentiated genes that were found in untreated RTT patient blood samples that showed clinical improvements when treated relative to those patients without improvements. These genes warrant further assessment to analyse as potential blood biomarkers for positively responding RTT patients treated with mecasermin.

As a clinical tool blood biomarkers may aid clinicians in treatment of RTT in the future. Considering the variability of mecasermin as a treatment, blood biomarkers for RTT may provide a better criteria for inclusion in clinical trials or guide treatment options. Despite the need for further validation of these genes *S100A8* and *S100A12* in particular have been identified as biomarkers in other neurological conditions including stroke and traumatic brain injury.

In summary this analysis revealed that ribonucleic acid (RNA) from extracted whole blood was of sufficient quality for further application in high-throughput sequencing. It also provided a number of targets for further investigation in their potential roles in the RTT molecular mechanism.

Chapter One: Introduction

1.1 Background and History of Rett Syndrome

Rett syndrome (RTT) is an X-linked, neurodevelopmental disorder found to nearly exclusively affect female patients. Although infants have an apparently normal early post-natal development, signs and symptoms of RTT begin to present around six and 18 months after birth. RTT classically is characterised as a loss of purposeful hand movements and acquired skills, loss of language, hand stereotypies such as wringing or clapping and gait abnormalities. These symptoms generally progress in a defined manner across four stages, including a period of stabilisation after initial regression. When a patient fulfils these main criteria (loss of acquired skills, loss of language, hand stereotypies and gait abnormalities) and experience a stabilisation period after regression the patient is clinically diagnosed with typical or classic RTT as described in Neul and colleagues in “Rett Syndrome: Revised Diagnostic Criteria and Nomenclature” (Neul et al. 2010).

RTT is named after Dr Andreas Rett whose first paper on RTT, “On a remarkable syndrome of cerebral atrophy associated with hyperammonemia in childhood” described his early observations about a cohort of 22 patients. These patients had a disorder with a clear neurological and psychological profile that was not yet described in the literature until its publication in 1966 (Rett 1966). However the disorder remained in relative obscurity until 1983 when Hagberg and colleagues further characterised another 35 cases of RTT (Zoghbi 2016). In 1999 Amir and colleagues elevated the level of understanding of RTT by describing *Methyl CpG binding Protein 2 (MECP2)* as a causative genetic mutation underpinning the disorder (Amir et al. 1999). Subsequently, *MECP2* loss of function mutations have been described as being the causative mutation in over 95% of cases of typical RTT cases (Krishnaraj et al. 2017). Clearly the discovery of *MECP2*’s relationship to RTT was seminal in RTT research. However not all people with *MECP2* loss of function mutations have RTT (Krishnaraj et al. 2017) and some patients with clinically diagnosed RTT have no observable *MECP2* mutations (Milani et al. 2005). The role of this gene should be thought of generally as highly associated with RTT rather than causative.

Typical RTT appears to occur in four distinct stages. These stages are:

Stage one. Early onset of disease after birth.

Stage two. Developmental regression.

Stage three. Pseudostationary stage where stabilisation after regression period occurs.

Stage four. Late motor deterioration stage.

As well as typical RTT there are a number of atypical forms of the disorder that have been grouped together and simplified by Neul and Colleagues in their revision of RTT criteria (Neul et al. 2010). Atypical RTT is defined as having at least two of four main criteria of typical RTT (i.e. partial or complete loss of language, partial or complete loss of purposeful hand skills, gait abnormality and hand stereotypies) with five out of 11 supportive criteria (supportive criteria includes: breathing disturbance while awake, bruxism/teeth grinding while awake, impaired sleep, abnormal muscle tone, peripheral vasomotor disturbance, scoliosis, growth retardation, small cold hands and feet, inappropriate laughing or screaming, diminished response to pain, and intense eye communication). These criteria were set to simplify and standardise the classification of atypical RTT patients who up until then were a dispirited group of variant disorders that had distinct clusters of symptoms (Hagberg 2002).

Common RTT variants include: early onset seizure variant RTT which is associated with mutations to the *cyclin dependent kinase like-5 (CDKL-5)* (Archer et al. 2006; Evans et al. 2005). However the *CDKL-5* variant has recently been distanced from RTT and is now considered its own clinical condition known as CDKL-5 disorder (Fehr et al. 2011; Mangatt et al. 2016). Preserved Speech Variant (PSV) or Zapella variant describes RTT patients with less severe speech impairments and acquired hand functions. PSV patients have even been able to draw and write words (Renieri et al. 2009). Congenital RTT which is associated with mutations to the *forkhead box G1 (FOXP1)* gene, is a variant where the patients display decreased muscle tone and show signs of mental retardation from birth (Renieri et al. 2009). The “forme fruste” RTT describes an incomplete RTT variant where although it roughly follows the course of RTT patients, patients can retain some motor function (Hagberg & Rasmussen 1986). It is important to consider typical and atypical RTT as distinct diagnoses especially for researchers who primarily look at typical RTT. Generally researchers observe cohorts of patients with confirmed *MECP2* mutations, though this is not always possible.

1.2 Comorbidities of Rett Syndrome

In typical RTT the main criteria involve the classic neurological and motor components. However there are a number comorbidities that can play a large role in the lives of RTT patients and caregivers. Respiratory, cardiac, and autonomic dysfunctions, scoliosis, epilepsy, gastrointestinal disruption, disturbed sleep and low bone density are all common among RTT patients (Weese-Mayer et al. 2006; Ager et al. 2006; Motil et al. 2012; Dolce et al. 2013; G. Pini et al. 2016; Downs et al. 2016).

Seizures represent one of the most common comorbidities of RTT though statistics vary between studies substantially ranging from 50-90% of patients (Dolce et al. 2013). Glaze and colleagues showed a cohort of 528 patients with classic RTT and 60% were reported as having seizures (Glaze et al. 2010). While Steffenburg and colleagues found that 94% of their smaller cohort of RTT patients, most of whom had been followed for many years, had a history of seizures. This higher proportion may have to do with the inclusion of variant forms of RTT in their data which included nine girls with early seizure onset. Of the 53 girls studied here nine had died at a date of prevalence where the remaining 44 girls partook in a follow-up study. Of these nine only two RTT patient's deaths were closely associated with epileptic events. This shows that although particularly stressful for patients and caregivers epilepsy is generally not fatal in RTT patients (Steffenburg et al. 2001).

Scoliosis is another common comorbidity that can be particularly difficult for patients and often requires corrective surgery. The Australian Rett Syndrome database conducted a 20 year follow-up study of 396 RTT patients, 86% of whom suffered from scoliosis with 40% having undergone corrective surgery (Anderson et al. 2014). Like epilepsy different studies have shown variation in prevalence. Downs and colleagues found the number of RTT patients with scoliosis to be closer to 66% (Downs et al. 2015) and an even larger study by Percy and colleagues found that only half of typical RTT had scoliosis (Percy et al. 2010). It has also been found that spinal fusion for treatment of scoliosis in RTT is less successful than that of adolescent idiopathic scoliosis with a higher risk of respiratory failure and an increased time on assisted ventilation systems post-surgery (Rumbak et al. 2016).

Autonomic dysfunction in RTT is an umbrella term encompassing problems with cardiac function, respiratory function, and breathing as well as sleep disturbance. Evidence of such problems come from studies linking breath holding patterns with heart rate and

respiratory dysregulation (Julu et al. 2001; Weese-Mayer et al. 2006). Efforts to understand if the autonomic dysfunctions were made by testing vagal tone of RTT patients and found that sympathovagal imbalance (Julu et al. 1997). Differences in the periaqueductal grey and nucleus of the solitary tract of the brainstem showed strong genotype-dependent changes in *Mecp2* mutant mice (Kron et al. 2012). These structures have functions in autonomic homeostasis. The nucleus of the solitary tract particularly has a tendency toward hyper-excitability which supports the hypothesis that medullary immaturity is responsible for autonomic dysfunction in RTT (Kline et al. 2010; Julu et al. 1997).

Interestingly De Felice and colleagues identified gas exchange abnormalities in RTT patients that were associated with oxidative stress markers, linking oxidative stress and autonomic disorder (Cortelazzo, De Felice, Guerranti, et al. 2014). The disturbed breathing patterns elicited by RTT presents in three distinct phenotypes. These include forceful breathing, feeble breathing and apneusitic breathers, each breathing phenotype leads to specific disruptions to homeostasis in the patient. Forceful breathers tend have low fixed concentrations of partial pressure of carbon dioxide (pCO₂) which leads to chronic respiratory alkalosis, while feeble breathers will have high concentrations of pCO₂ which leads to chronic respiratory acidosis. Apneusitic breathers tend to accumulate carbon dioxide (CO₂) due to expiration dysregulation (Julu 2008).

Despite the prevalence and severity of these comorbidities in patient's epidemiology studies has revealed that the most common cause of death in RTT is pneumonia (Laurvick 2006) although the rates of sudden death are much higher in RTT than in the normal population (Crosson et al. 2017). Taken together it is clear that these comorbidities play a substantial role in the everyday difficulties of RTT patients and carers. A better understanding of these mechanisms of comorbidities can help to guide drug designs to better facilitate and treat RTT patients.

1.3 Mutations in Rett Syndrome

Typical RTT is strongly correlated to loss of function mutations in the global transcriptional regulator *MECP2* (*Methyl-CpG-binding protein 2*) and represent over 95% of typical cases and 75% of variant RTT cases (Krishnaraj et al. 2017). Despite high coincidence there are roughly 5% of typical RTT cases that have a different mutation other than *MECP2* and more importantly there patients who do not have RTT but have *MECP2* mutations (Suter et al. 2014; Carney et al. 2003). These reported cases

of the patients with neurodevelopmental abnormalities other than RTT positive for *MECP2* mutations are important because they help illustrate that RTT is a clinical diagnosis. The relationship between RTT phenotype and *MECP2* mutations should be seen as highly correlated but two distinctly different things. The monogenic nature of RTT makes it an interesting case study for researchers as theoretically it limits the number of genetic risk factors influencing the pathology, however *MECP2*'s complex functional role adds to this.

Even though 99.5% of cases of RTT are de novo mutations (Trappe et al. 2001) there are seven common mutations of *MECP2* that are thought to represent approximately 70% of all RTT cases. These are R106W, R133C, T158M, R168X, R255X, R270X, R294X and R306C (Colvin et al. 2004). Attempts have been made to determine whether certain mutations confer more severe forms of RTT and have found some success. Cuddapah and colleagues found that mutations R133C, R294X, R306C and 3' truncations to the *MECP2* gene were relatively less severe and mutations R106W, R168X, R255X, R270X, splice sites large deletions, insertions and deletions (these include deletions to exon 3 and 4, complex insertions or rearrangements) to the *MECP2* gene were significantly more severe mutations (Cuddapah et al. 2014). Colvin and colleagues observed that nuclear localisation signal region mutations were significantly more severe and mutations in the region after transcriptional repression domain-nuclear localisation signal (TRD-NLS) region including the C-terminal region conferred a less severe phenotype (Cuddapah et al. 2014). In general it has been found that larger deletions have more severe phenotypes while C-terminal truncation mutations tend to produce a less severe phenotype (Bebbington et al. 2012; Neul et al. 2008; Shah & Bird 2017).

1.4 MECP2 Functions in Rett Syndrome

The *MECP2* gene itself codes for the protein Methyl CpG Binding Protein 2 which is expressed in a number of different tissues but most highly in the central nervous system, lung and spleen with lower amounts in the heart and kidney (Mona D Shahbazian et al. 2002). In fact neuronal concentrations of *MECP2* were found to be near to the histone-octamer levels in the nuclei of neurons prepared from mouse brains. This means there is one molecule of *MECP2* to every two nucleosomes in the brain (Skene et al. 2010).

The understanding of glial expression of *MECP2* protein in RTT has changed over time with initial reports of no detectable expression (Mona D Shahbazian et al. 2002; Chen et al. 2001). However recent findings have shown that *MECP2* is not only expressed in

glia but could even potentially play an integral role in determining the RTT-phenotype of neurons in non-cell autonomous fashion (Ballas et al. 2009).

The functional mechanism of MECP2 has also been disputed over time and has led to different models being proposed for the pathological mechanism in RTT. The first studies of the MECP2 protein's function considered it a transcriptional repressor. It was shown to bind to methylated promotor sites of a certain density in a non-linear manner across genomic deoxyribonucleic acid (DNA) (Nan et al. 1997). Structurally the main functional domains in MECP2 are the Methyl Binding Domain (MBD) and the transcriptional repression domain (TRD) which contains one of two nuclear localisation signals (NLS) the other is in the inter domain (ID) region. The MBD (Residues 78-163) is the primary domain involved in the establishing the binding of MECP2 to the methylated protein, however both the ID and TRD have also been shown to affect binding of MECP2 to methylated proteins (Asmita 2008). The TRD region (residues 205-310) of the MECP2 protein functions as long range repressor of transcription. This is likely due to its interaction with transcriptional factors as opposed to direct influencing of transcriptional elements (Nan et al. 1997). The TRD also contains a specific NLS thought to be located between residues 255 and 271 that allows for the protein to be transported to the nucleus (Nan et al. 1997).

Although this initial work described MECP2 as exclusively a transcriptional repressor through the recruitment of sin transcriptional regulator family A (Sin3a)/Histone deacetylase (HDAC) complex (Nan et al. 1998) there has been much research published that supports and that argues against this hypothesis. MECP2 has been found to up-regulate gene expression in a number of studies (Chahrour et al. 2008; Li et al. 2013; Zhao et al. 2013). Early work in mice by Lewis and colleagues showed that MECP2 bound to symmetrically to methyl Cytosine -Guanine (mCG) which is highly concentrated in the heterochromatin accounting for roughly 40% of all 5 methyl-Cytosine (5mC) (Lewis et al. 1992). RTT mutations have been shown to ablate the proper functioning nuclear co-repressor/silencing mediator of retinoic acid thyroid hormone receptor (NCoR/SMRT) co-repressor by disrupting its' protein-protein interaction with MECP2 (Lyst et al. 2013). Recently it has been shown that MECP2 does not specifically bind to 5 methyl Cytosine Guanosine (5mCG) (Guo et al. 2014). MECP2 was shown to bind to methyl-Cytosine Guanine (mCG), methyl-Cytosine Guanine/Thymine/Adenine (mCH) and non-methylated DNA although non-methylated DNA binding has a significantly lower affinity than mCG and mCH (Chen et al. 2015).

Lagger and colleagues found non-mCG binding occurred in methyl-Cytosine Adenine Cytosine (mCAC) site (Lagger et al. 2017). They showed that MECP2 binding was proportional to the concentration of mCAC + mCG density in the mouse brain through a chromatin immunoprecipitation-sequencing (ChIP-Seq) technique.

A recent paper by Ross and colleagues is of particular relevance to our study. They generated a peripheral knock-out (pKO) of *Mecp2* in a mouse model that shows nearly normal levels MECP2 in the nervous system but severe depletion outside of the nervous system (Ross et al. 2016). Although many of the RTT-like symptoms that occur in global knockout (KO) of *Mecp2* were rescued in the PKO there was a subset of symptoms that were detected with further physiological tests. The *Mecp2* PKO showed a propensity for the mouse to exercise fatigue and in general hypoactivity as well as bone abnormalities. This has important implications for our research and also gives further credence to exploration of non-central nervous system abnormalities in RTT. Metabolic differences especially of cholesterol metabolism have been shown to be altered in RTT (Segatto et al. 2014; Lopez et al. 2016). Various studies have also identified mitochondrial dysfunctions in RTT animal models and RTT patients (Kriaucionis et al. 2006; Pecorelli et al. 2013; De Felice et al. 2014).

1.5 IGF-1 Functioning and Molecular Mechanisms

IGF-1 binds to IGF-1 receptor (IGF-1R) which is a heterotetrameric glycoprotein that is widely expressed in the neural and epidermal tissues (Liu et al. 1993; Wrigley et al. 2017). In the rat nervous system, expression of IGF-1 decreases postnatally over time. In humans IGF-1 has been proposed to have functions in neurogenesis and also decreases postnatally (Baron-Van Evercooren et al. 1991; Wrigley et al. 2017). The binding of IGF-1 to IGF-1R causes auto-phosphorylation through tyrosine kinase activation which activates insulin receptor substrate 1 (IRS-1). This causes further activation of downstream signalling cascades including phosphatidylinositol-3 kinase (PI3K) and Ras-Raf-Mitogen-activated protein kinase (MAPK) (Ye et al. 2010). The PI3K pathway signalling promotes cell growth and maturation through protein kinase B (Akt) activation and inhibits the pro-apoptotic protein glycogen synthase 3 β (GSK3 β) leading to increased glucose uptake by neurons (Bondy & Cheng 2004). Ras-Raf-MAPK signalling is involved in cell cycle regulation (Molina & Adjei 2006).

IGF-1 is part of the insulin superfamily that has a wide number of biological functions associated with it. Murine models that KO IGF-1 were shown to cause growth problems

particularly muscle hypoplasia and reduce lifespan (Liu et al. 1993). In humans IGF-1 deficiency causes a type of dwarfism called Laron syndrome (Laron 2001). This deficiency is a result of growth hormone (GH) resistance. Counterintuitively IGF-1 KO mice have been associated with increased longevity which raises questions of the relationship of IGF-1 and aging (Laron 2008; Laron 2005). The IGF-1 lifespan relationship appears to be conserved in different species. When the insulin receptor (IR) signalling was decreased in drosophila fly there was an 85% (Tatar 2001) increase in lifespan and mouse model IGF-1 KOs had 26% increase of lifespan (Holzenberger et al. 2003). IGF-1 has been developed as a treatment option in people with growth failure (Chernausk et al. 2007). The drug mecasermin or human recombinant IGF-1 which was approved by United States food and drug administration (FDA) is an attractive option for treatment of RTT. Researchers have shown it has a good safety profile (Khwaja et al. 2014a; Giorgio Pini, Congiu, Benincasa, et al. 2016) and ameliorates a number of symptoms in RTT animal models (Castro et al. 2014).

As well as its roles in growth and longevity IGF-1 has also shown to have integral functions in the central nervous system (CNS) including promoting cell survival, myelination and brain growth (Barres et al. 1992; Carson et al. 1993). Specifically IGF-1 was shown to induce proliferation, neural differentiation and neurogenesis in the adult rat hippocampus (Anderson et al. 2002). IGF-1 has been shown to be expressed in all cell types of CNS especially in the cerebral cortex, hippocampus, and cerebellum (Bondy & Lee 1993). Nestin driven overexpression of IGF-1 in transgenic mice has resulted in increased brain weight due to greater cortical plate volume and cell numbers (Popken et al. 2004). Together these findings highlight the role of IGF-1 in brain development, maturation and maintenance.

Recent evidence has indicated that IGF-1 plays a strong role in neural plasticity. Neural plasticity is the capacity of neural tissues such as neurons, glia and blood vesicles to reorganise due to functional requirements during an individual's lifespan (Llorens-Martín et al. 2009). The “neurogenic reserve” hypothesis of neurogenesis as stated by Kempermann says that the ability of the brain to develop new neurons allows for long-lasting activity-dependent adaptations and cumulative network adaptations to be made over an individual life time. These adaptations provide the flexibility and plasticity in the hippocampus to cope with constant stream of novel external stimuli (Kempermann 2008). In rats IGF-1 has been shown to decrease postnatally after birth but expression remains steady in areas associated with neurogenesis such as subventricular zone (SVZ)

and dentate gyrus (DG) before further decreasing at a time scale matching the eventual decrease of neurogenesis (Anderson et al. 2002). Ciucci and colleagues showed that IGF-1 mediates the effects of an enriched environmental stimuli in visual cortex of rat brains (Ciucci et al. 2007). Enriched environments have been used as an experimental paradigm to show that cages with more complicated, stimulating structures like tunnels, platforms, running wheels and stairs cause activity dependent changes in the brains of rodent animal models such as acceleration of visual acuity (Cancedda 2004).

Further evidence of the IGF-1s functions in neurogenesis have been demonstrated through overexpression of IGF-1 in mice which displayed overgrowth and dysregulation of neuronal and oligodendrocyte numbers (D'Ercole et al. 2002) and a persistent increase in number of neurons and synapses in the DG (O'Kusky et al. 2000). IGF-1 infusions were shown to attenuate age related deficits in working and reference memory in water maze and object recognition in rats (Markowska et al. 1998). GH administration has shown increased micro-vascularisation which has been linked to an IGF-1 expression increase as well as altering the regulation of N-methyl- D-aspartic acid (NMDA) receptors in aged animal brains (Sonntag et al. 2000). These findings demonstrate that IGF-1 has an integral function in regulation of neurogenesis and this property could contribute to the beneficial effects that IGF-1 treatment has elicited in RTT mice models (Tropea et al. 2009) and RTT patients (Pini et al. 2012; Giorgio Pini, Congiu, Benincasa, et al. 2016).

1.6 Transcriptome Level Analysis of Rett Patient's Tissues

RTT syndrome research occupies a unique position in the study of neurological disorders. The prevalence of *MECP2* mutations in typical RTT is an advantage in that researchers can wholly focus on the functions and effects of a single gene. *Mecp2* KO mouse models can accurately mimic the neurological deficits found in RTT (Guy et al. 2001; Chen et al. 2001). However a number different RTT mouse models have been developed such as *Mecp2*^{308/y} truncation, CNS specific knock out (Nestin-Cre-mediated deletion) and PKO (hemizygous *Mecp2*^{stop/y}, *Nestin-Cre*) which all have different sets of RTT-like features (Mona D. Shahbazian et al. 2002; Ross et al. 2016; Chen et al. 2001). As well as the type of disruption to *Mecp2* sex differences also occur between the models, a topic which has been reviewed by Ricceri and colleagues (Ricceri et al. 2008).

MECP2 functions as a global transcriptional regulator and there is a wide range of potentially effected genes, proteins and tissues that result from *MECP2* mutations. This

understanding of *MECP2*'s functions has changed over time. Initially *MECP2* was thought to be a transcriptional repressor (Nan et al. 1997; Ballestar et al. 2000; Coy et al. 1999), however a study in hypothalamus of *Mecp2* null and overexpressing mice has shown that there was a large amount gene activation due to *Mecp2* (approximately 85%). This activation demonstrates a dual role for *MECP2* as both an activator and a repressor of transcription (Chahrour et al. 2008). A more nuanced view of the role of *MECP2* came when Skene and colleagues found that *MECP2* expression in brain was almost as high as the level of the histone octamer. They hypothesised that *MECP2*'s repression of transcription might function so as to dampen transcriptional noise across the genome (Skene et al. 2010). Further exploration of the transcriptional dampener function of *MECP2* has come through studying the ability of *MECP2* to bind to non-mCG sites (Lagger et al. 2017). It was found that non-mCG binding mainly occurred as mCAC and binding was proportional to mCAC + mCG density in genome regions of the mouse brain. Lagger and colleagues therefore propose *MECP2* is a negative modulator of neuronal transcription in the brain. These studies suggest that the exact role of *MECP2* in transcription is not exactly defined however on the current evidence it would appear that repression on a genome wide scale is more likely.

Considering the complex role of *MECP2* (Maxwell et al. 2013; Skene et al. 2010), transcriptomic studies provide data on gene expression at a level appropriate for understanding the global effects of loss of function from *MECP2* mutations. Although gene expression cannot in and of itself be used to infer the physiological actions, it provides an important framework for describing the transcriptional regulation in RTT tissues. Understanding gene expression in different tissue types in RTT will be vital to how we treat the disorder in the future. Whether researchers can identify reliable biomarkers that can predict prognosis of patients or tailor drug prescription to individual patients may be as important as finding new therapeutics.

As RTT is primarily a neurodevelopmental disorder the sources that can be used for experimentation with human patients is very limited. Post-mortem brain samples are the only source that includes tissues with the primary pathology of the disorder. Conducting research with post-mortem brain samples requires age matched, clinically and genetically comparable samples from the similar brain regions. The rarity of the disorder makes it very difficult to assemble large cohorts of post-mortem brains while controlling for confounding factors such as age, sex and ethnicity. Even if samples are gathered there is still the problem of heterogeneity of cell populations across brain samples which can bias

gene expression (Olson et al. 2014; Mona D Shahbazian et al. 2002). Sugino and colleagues addressed this problem by devising a manual cell sorting method to negate this effect in RTT mouse model (Sugino et al. 2014).

Another interesting aspect to these studies is the evolving nature of the transcriptomic technology. RNA-sequencing (RNA-Seq) techniques are beginning to overtake microarray as the technique of choice when comparing lowly expressed differential changes. Dynamic range of RNA-Seq is improved compared to microarray meaning that resolution of higher fold-changes is improved (Xu et al. 2014; Zhao et al. 2014). One of the benefits of RNA-Seq methodology arise from the direct reading of the transcript sequences as opposed to microarray which is based on the hybridisation samples to a number of predesigned probes. These probesets have a limited number of transcripts of each gene and depend on the manufacturers. Later versions of microarray chips have looked to increase detection of genes by increasing the number of transcripts tested for with strong replication and reliability of transcript detection (Piper et al. 2002).

In the following sections we will look at some of the major findings of transcriptome level analysis in a number of different tissues including post-mortem brain tissues from RTT patients, blood tissues from RTT patients and induced pluripotent stem cell (iPSC) lines derived from RTT patient tissues.

1.7 Transcriptome Analysis in Post-Mortem Brain Samples

Often considered the gold standard of transcriptomic analysis of neuropsychiatric and neurodegenerative disorders post-mortem brain (PMB) samples based studies are understandably rare to come across (Costa et al. 2012). This is especially obvious in a disorder such as RTT where development and symptoms are not linear and different effects such as X-chromosome inactivation (XCI) may affect severity (Archer et al. 2007). The logistical and financial burden of RTT has meant that there has only been four transcriptome level studies using post-mortem brain samples from human patients (Colantuoni et al. 2001; Lin et al. 2016; Deng et al. 2007; Gibson et al. 2010). Each of these studies has had to contend with different limiting factors in their analysis. This means that none of the studies are compatible enough to conduct a meaningful meta-analysis. Age groups, mutation type, control comparisons, analysis techniques and brain region used all varied across the studies.

Colantuoni and colleagues provided the first evidence of DGE in PMB samples. They compared six typical RTT patients with six age and sex matched controls across a

number of microarray platforms and using a number of statistical software programs to increase sensitivity, detection and reliability. Brain tissues utilised from the patients were taken from Brodmann's areas (BA) 1-5 at the junction of the parietal and frontal cortices of the cerebrum. Their results identified three major areas of DGE: neuronal maturation genes, presynaptic marker genes and glial specific genes. These results pointed to a disruption of synaptic connections (specifically presynaptically), synapse maturation, synaptic transmission and increased reactivity of glia in the RTT brain.

Deng and colleagues conducted a study using microarray to examine DGE in both human RTT patients and an *MECP2* null model. They subsequently used their results to generate a novel mouse model with overexpression of the *FXYD domain containing ion transport regulator 1 (FXYD1)* gene which was found consistently over expressed in the PMB samples. *FXYD1* is a FXYD domain ion transport regulator which modulates the Na⁺/K⁺ ATPase channel activity. The *FXYD1* over-expressing (OE) mice neurons were cultured and showed reductions in the dendrite spine density by Scholl analysis. To support the relationship between these findings confirmation of two methylation promotor sites were found on the *FXYD1* gene which were shown to be binding sites for Mecp2. Their PMB samples were taken from the superior frontal gyrus which was adjacent to where Colantuoni and colleagues harvested their samples in BA1-5. Although these are distinct regions of the cerebral cortex *MECP2* expression in the human RTT brain is based the maturity of neurons across the brain (Mona D Shahbazian et al. 2002).

Gibson and colleagues compared frontal and occipital cortices of six RTT patients and six healthy sex-matched controls to examine differential expression with microarray analysis. However the experiment lacked age matched controls which is clearly a confounding factor. To compensate for this Gibson and colleagues used *MECP2* knockdown cell line and a four-way comparison between RTT/control/frontal/occipital cortex samples to confirm expression levels. For a gene to be considered to be significantly differentially expressed five out of six patients had to agree with the microarray analysis. Gibson and colleagues found an up-regulation of *Dynamin 1*, *Amyloid like protein 1*, *Clusterin* and *Dihydropyrimidinase related protein 1* with a decrease in *cytochrome c oxidase subunit 1* in the frontal cortex. Interestingly Clusterin protein appears to be induced by HDAC inhibitors (Nuutinen et al. 2005). This result makes sense considering *MEPC2*'s molecular mechanism is thought to work through recruitment of HDAC (Jones et al. 1998).

Lin and colleagues analysed samples from four RTT patients and four age-, sex- and ethnically matched controls and conducted both RNA-Seq and microarray analysis of frontal and temporal cortices. They used a murine model of RTT to provide further independent verification of DGE. The findings from this paper were confirmed with an independent dataset of DGE which were conducted by Ben-Shachar and colleagues in *Mecp2* KO murine model of RTT (Ben-Shachar et al. 2009). Notably the complement C1Q (C1Q) complex genes *C1QA*, *C1QB* and *C1QC* were all found to be decreased. Further pathway enrichment analysis showed that from the significantly down-regulated genes a number of other genes functionally involved in immune response and complement cascade were over-represented.

1.8 Transcriptome Analysis in Blood Tissue Samples

Blood is a tissue source that is freely available, easily replenished and causes minimal adverse effects on patients health when taken. This accessibility allows blood samples to be taken at multiple time points from the same patient. A patient matched approach is much more robust model for examining the effects of a given drug treatment as it removes genetic background differences and gives a direct comparison before and after treatment. Here we review two DGE studies in RTT blood tissues that use case control design (Pecorelli et al. 2013; Colak et al. 2011).

Pecorelli et al. 2013 used a microarray technique to analyse DGE of peripheral blood lymphomonocyte cells (PBMC) of 12 RTT patients compared to seven age and sex matched controls. This study used two tools, linear models for microarray (LIMMA) and significance analysis of mircoarrays (SAM) as microarray analysis models to find significant differential expression between RTT and control samples. LIMMA and SAM were highly congruent with each other, 480 genes were significantly differentially expressed with only 11 genes being found by just one of the programs. Gene ontology term and clustering analysis was then performed to identify pathways either up or down regulated, it showed four major pathway components. The analysis revealed; genes relating to chromatin folding functions were down-regulated, genes relating to mitochondrial functioning were up-regulated, genes relating to antioxidant defence were up-regulated, and genes with ubiquitin-proteasome system (UPS) functions up-regulated. Pecorelli and colleagues hypothesised that the antioxidant defence and UPS related gene upregulation was a feedback mechanism due to the increase in oxidative stress produced by the increase in mitochondrial functioning. Increased mitochondrial functioning

causes an increase in the generation of reactive oxygen species (ROS). This would require increased anti-oxidative defence and UPS mechanisms to compensate for the ROS and the damaged proteins (Park et al. 2014). They speculate that this change in mitochondrial activity would have an effect on the adenosine tri-phosphate (ATP) levels, although ATP levels are not measured in this study. However *MECP2* null mice have shown a decrease in ATP levels (Saywell et al. 2006). A more recent study which used a murine model more reflective of the human RTT condition, a *MECP2*-308 truncation male mice (De Filippis et al. 2010). They showed an over production of ROS attributed to the mitochondrial complex II as well as a reduction in the overall production of ATP when compared to wildtype littermates (De Filippis et al. 2015).

Colak and colleagues used microarray analysis of whole blood samples from both typical RTT patients (n = 3) and RTT-like patients (n = 2) (Colak et al. 2011). The purpose of their work was to identify potential shared mechanisms between typical RTT and RTT-like phenotypes. Their data suggested that RTT-like patients have dysregulation in oxidative phosphorylation, mitochondrial functioning, tumour suppressor p53 signalling and docosaheptaenoic acid signalling. Of particular note here is that RTT-like patients showed a dysregulation to mitochondrial functioning but this was not observed in the classic RTT patients. Network analysis identified potential regulatory critical regulatory function of the following pathways, interleukin-1 (IL1), interleukin-1 receptor type 1 (IL1R1), transforming growth factor β (TGF β), interferon- α , interferon- β and nuclear factor kappa B subunit 1 (NF κ B) in both classic RTT and RTT-like patients. The NF κ B pathway is thought to have a critical role in synapse development (Meffert et al. 2003). As well as NF κ B calcium homeostasis, cholesterol metabolism and nuclear factor of activated T-cells (NFAT)/calcineurin signalling were implicated through the bioinformatics analysis in both phenotypes.

In our analysis we hope to find differential expression of genes in whole blood of RTT patients treated with mecasermin. This may not inform our understanding of the pathology in the CNS but could provide us with potential genes to investigate as biomarkers to better understand prognosis or response to treatment. Recently a pilot study by Hardy and colleagues attempted to determine the clinical status of patients with post-concussion syndrome using the molecular profile generated from blood RNA sequencing techniques with a reasonably high accuracy (Hardy et al. 2017). They used 32 military service members with history of non-acute concussion (>90 days) and used RNA-Seq of whole blood to determine an exon expression profile for those still suffering

chronic post-concussion syndrome. This allowed them to validate 14 patients as having either non-acute or chronic post-concussion syndrome with an accuracy of 86%. Though they admit that the low sample size meant that they could not assess confounding factors such as race, sex etc. this study provides preliminary data on the utility of RNA-Seq in whole blood for diagnostic purposes. Post-concussion syndrome was confirmed by certified neuropsychologist based on US Veteran Association/Department of Defence clinical practice guideline for management of concussion/mild traumatic brain injury (Cifu et al. 2009).

Another pilot study has shown that phosphorylated-tau 181 to be a promising biomarker for the detection using an ultra-sensitive immunoassay for Alzheimer's disease (AD) and Down's syndrome (Tatebe et al. 2017). In fact phosphorylated-tau 181 and amyloid Beta 42 (A β 42) are being analysed as biomarkers for Alzheimer's in seven clinical trials at various stages of analysis (NCT03300726, NCT01142336, NCT02646982, NCT01595646, NCT02843529, NCT02460783 and NCT03402659). Of these trials the only significant result was a possible disease modifying effect in AD patients with hypercholesterolemia when treated with simvastatin associated with reduced cerebrospinal fluid (CSF) phosphorylated Tau 181 levels (Li et al. 2017).

Together the transcriptional findings in RTT patients' blood along with the ongoing clinical investigations in other neurological disorders such Alzheimer's disease and concussion highlights the utility of DGE analysis in patient's blood samples.

1.9 Transcriptome Analysis in Induced Pluripotent Stem Cells Derived from Rett Patient's Tissues

Induced pluripotent stem cells (iPSC) are a relatively recent technique that represent an intriguing solution to study directly RTT neuronal cells in vivo. iPSCs are cultured from primary sources in patients and are treated in such a way as to reverse cell fate using a number of overexpressed reprogramming factors octamer binding transcription factor 4 (OCT4), sex determining region Y-box 2 (SOX2), Kruppel like factor 4 (KLF4), and MYC protein oncogene, BHLH transcription factor (MYC). These reprogrammed cells have pluripotency and the ability to be cultured and differentiated into specific cell types depending on the microenvironment (Park et al. 2008). Recently these techniques were employed using RTT patient fibroblasts to reprogram and then differentiate iPSCs into RTT phenotypic neurons. A number of different groups have attempted this (Marchetto et al. 2010; Kim et al. 2010; Cheung et al. 2011; Pomp et al. 2011). The findings from

these studies showed that differentiated neuronal cells from cultured RTT-IPSCs were consistent with the RTT phenotype. Interestingly some of these studies found that IPSCs produced cell lines with non-random highly skewed X chromosome inactivation (XCI) (Pomp et al. 2011; Ananiev et al. 2011; Cheung et al. 2012) while others have shown a reactivation of the fibroblasts inactive X chromosomes (Marchetto et al. 2010; Kim et al. 2010). IPSC lines with non-random skewing allow for isolation of populations of IPSCs that only display either mutant *MECP2* (mut-*MECP2*) or wild type *MECP2* (WT-*MECP2*). This is an advantage as researchers can compare exactly how the mutation to *MECP2* effects the gene expression in those cells. However this skewed pattern of XCI is not reflective of what occurs in RTT patients. Clinically RTT patients have more mosaic patterns of some neurons with mut-*MECP2* and some neurons with WT-*MECP2* expression which can influence surrounding neurons in a non-cell-autonomous manner (Braunschweig et al. 2004).

Although the mosaic expression are more clinically relevant, results from experiments using mosaic expression are more difficult to interpret. This is because wild type and mutant neurons cannot readily be discerned without further immunostaining and western blotting to detect WT- and Mut-*MECP2* status (Marchetto et al. 2010). Specific antibodies for this many not be available. As well as the phenotypic questions surrounding the highly skewed XCI IPSCs the fact that reactivation of the X chromosome does not occur begs the question did the cells actually reach full pluripotency or were they simply partially reprogrammed (Cheung et al. 2012). A further consideration in the use of IPSC is that fact that lower numbers of passages through cell culturing has a greater risk of residual epigenetic signatures remaining from the cell's original pre-reprogrammed state (Polo et al. 2010). Pomp and colleagues found that neurons derived from RTT IPSCs with isogenic populations of WT- or Mut-*MECP2* cells at lower passages, implying that at lower passages the cells were more likely to have skewed XCI. While RTT IPCs at higher passages were mosaic in *MECP2* expression (Pomp et al. 2011). Clearly there are advantages and disadvantages to the use of both isogenic and mosaic systems. Considering the difficulty in studying the primary RTT pathology both these models could prove powerful tools for research.

Tanaka and colleagues completed a recent study investigating DGE in undifferentiated RTT-IPSCs to examine the regulatory role of *MECP2* loss of function mutations on early cell development. They used both WT- and Mut-*MECP2* expressing cell lines from the fibroblasts of RTT patients and sequenced the samples using RNA-Seq (Tanaka et al.

2014). They found that mutant RTT-IPSCs showed de-repression of X-linked genes. Several pathways were found to be affected in mutant RTT-IPSCs but importantly each *MECP2* mutation appeared to give a different pathway profile. Stem cell development processes did not appear to be affected by *MECP2* mutant expression, although maturation marker *NOTCH1* was decreased in RTT cell lines and *AKT serine/threonine kinase 1 (AKT1)* was found to be significantly lower in two out of three mutant RTT-IPSC lines. Altogether these results show that even at undifferentiated stage of cell development that RTT cells are beginning to diverge from normal development. The unique profiles of each *MECP2* mutation highlight the importance that developing methods of accurate treatment screening for RTT may play in future medical care.

1.10 IGF-1 in Clinical Trials

Clinical trials based on humanised IGF-1 treatment (mecasermin) have been developed for a number of neurodevelopmental conditions including Autism Spectrum Disorder (ASD) (NCT01970345), Phelan-McDermid Syndrome (PMDS) (NCT01525901) and RTT (Phase one: NCT01253317, phase two: NCT1777542). In the cases of the ASD and PMDS the trials are still ongoing, while the RTT mecasermin trials have been completed.

A phase one open label trial for mecasermin was performed to analyse safety, pharmacokinetics and preliminary efficacy of the drug in RTT girls (Khwaja et al. 2014a). The design of the study used 12 patients (nine Classic RTT patients and three *MECP2* related disorder patients) and treated them with increasing doses (40/80/120 µg/kg) of mecasermin for four weeks to assess safety, tolerability and pharmacokinetics of treatment. This was termed the multiple ascending dose (MAD) phase. After a wash out period nine RTT patients were placed on an open label extension (OLE) at 120 µg/kg for a further 20 weeks to give preliminary efficacy data. The same increasing dosages were used to reach maximum dose in OLE as in the MAD. The trial found that the treatment had an acceptable safety and tolerability profile in all 12 patients. For the nine patients with *MECP2* mutations apnoeic index and anxiety depression and mood scale (ADAMS) were assessed in the OLE. Khwaja and colleagues found significant improvements to the apnoeic index of RTT patients, while ADAMS did not significantly improve. The DGE data of the whole blood samples from these patients across this study were made available to us and forms part of the analyses of this thesis. Further description of the patient samples will be discussed in the methods chapter.

Most recently the phase two trial has been completed for mecasermin treatment in RTT (O’Leary et al. 2018). O’Leary and colleagues found no significant improvement to primary outcomes with some worsening of the patient’s condition. However there was significant improvement to hand stereotypies and an improvement in the social subscale of the communication and symbolic behaviour scales-development profile (CSBS-DP) in RTT patients. Despite the lack of significant improvements to the primary outcomes these secondary improvements are not trivial. O’Leary and colleagues point to certain limitations as potential factors for the failure of the trial including inadequate dose/administration and low statistical power as well as the possibility the drug treatment was ineffective.

Another trial assessing the treatment of RTT patients with mecasermin was conducted at Versilia Hospital in Italy, where clinicians and independent observers (with International Scoring System and Rett Severity Score respectively) assessed patients after 20 weeks of treatment with a similar dose of mecasermin (100 µg/kg) that was used by Khwaja and colleague (120 µg/kg). The duration of the treatment is comparable to the OLE of the phase of Khwaja and colleagues trial. The findings for this trial showed significant improvement to both measures, the International Scoring System (p-value =0.0106) and the Rett Severity Score (p-value = 0.0274). Blood samples from the patients in this trial from before and after treatment were provided for sequencing analysis and will form a second cohort of analysis for this thesis.

1.11 Aim, Hypotheses and Interpretation of Literature

The aim of this project was to investigate the expression profiles (DGE) of whole blood samples from RTT patients in two independent open label phase I clinical trials using humanised recombinant IGF-1 (mecasermin) treatment. Through this investigation we hope to validate two hypotheses on the expression profiles of patients. The hypotheses are as follows:

1. The mecasermin treatment induces changes in the DGE in peripheral blood of patients. When comparing whole blood samples of RTT patients before and after treatment with mecasermin we expect to detect differences in expression profiles (DGE) before and after treatment.
2. Patients that show improved phenotype in response to treatment have a different expression profile from patients who do not respond to treatment. We considered the clinical assessment of the trial and we used apnoeic index (in BCH trial) and

PBZ behavioural scores (in Versilia hospital trial) before and after treatment, and we divided the patients into responders and non-responders. We then compared the different expression profiles (DGE) in responders versus non-responders before mecamsermin treatment.

We have been provided with DGE data from the phase one clinical trial by Khwaja and colleagues in Boston Children's Hospital (BCH) and a set of whole blood samples taken before and after mecamsermin treatment in a trial by Pini and colleagues based in Versilia Hospital Italy. Both studies used mecamsermin (humanised recombinant IGF-1) to treat RTT in classic RTT patients with stage two or three disease using a similar dosage and period of treatment. As this is a novel study exploring the safety and preliminary efficacy of mecamsermin there is no prior evidence of the effects of treatment on RTT patients. Instead we must focus on the transcriptome level case control experiments that we discussed previously in the literature review. These studies () collectively showed three common mechanisms in human RTT brain, blood and iPSC derived tissues. These mechanisms were abnormal expression in mitochondrial functioning genes, reactivation of glial cell genes and abnormal synaptic functioning genes.

Evidence for both increased and decreased mitochondrial genes in RTT patients has been observed in the literature so it is difficult to determine whether we would expect to see an increase or decrease in analysis. Pecorelli and colleagues' found mitochondrial functioning genes were significantly increased and a number of mitochondrial pathways were significantly over-represented. While transcriptional and cellular evidence from Gibson and colleagues found mitochondrial complex IV gene *mitochondrially encoded cytochrome c oxidase subunit 1 (MT-CO1)* was significantly downregulated in RTT frontal cortex compared to control frontal cortex and RTT occipital cortex in microarray analysis. SH-SY5Y cultures showed that *MECP2* knockdown suppressed the normal increase in *MT-CO1* expression observed when SH-SY5Y cells are differentiated. Direct activity of cytochrome c oxidase (COX) was tested by measuring the rate of oxidation of cytochrome c after SH-SY5Y cultures were knocked down for *MECP2* indicating a significant decrease.

A common thread found in a number of studies in post mortem brain samples (Colantuoni et al. 2001; Deng et al. 2007; Gibson et al. 2010) was the abnormality in the synapse functioning and dendritic arborisation. Colantuoni and colleagues found a number of presynaptic marker genes significantly decreased and postsynaptic markers genes significantly increased. Gibson and colleagues identified a significant increase in

Dynamin 1 (DNM1). De novo mutations in this gene cause epileptic encephalopathy and has been suggested disrupt synaptic vesicle endocytosis in a mouse model (Dhindsa et al. 2015). Deng and colleagues identified the *FXYD1* gene, a modulator of Na⁺ K⁺ ATPase activity and target of *MECP2*, as being overexpressed in RTT PMB samples. A mouse model overexpressing *FXYD1* showed dendritic reduction in the frontal cortex neurons. These data support the hypothesis that aberrant synaptic functioning and connectivity contribute to abnormal RTT brain phenotype (Pohodich & Zoghbi 2015; Katz et al. 2016).

The final common mechanism we identified to affect RTT primary tissues is glial cell reactivation. Lin and colleagues found *C1Q* complement genes were decreased in RTT PMB samples and could also contribute to the abnormal synaptic functioning through changes in microglial reactivity. Colantuoni and colleagues to found a number of specific glial markers increased in the RTT PMB samples implicating glial reactivity including *glial fibrillary acidic protein (GFAP)*, *solute carrier family 1 member 3 (EAAT1)*, *S100 calcium binding protein A13 (S100A13)*, and *crystallin alpha B (CRYAB)*.

We hope to investigate whether these three mechanisms were effected in whole blood samples treated with mecasermin. To do this firstly we will look difference in expression comparing from before treatment samples (BCH-t0) to time points after mecasermin treatment (VH-t1 and VH-t2 in Versilia Hospital trial and BCH-t1, BCH-t2 and BCH-t3 in BCH trial) in all patients.

Secondly we will compare the expression of profiles of the patients that showed positive response to treatment. Response will be determined either by decrease in apnoeic index or through improvements in severity score in the BCH and Versilia Hospital trials respectively. We will compare the more positive responding trial participants with less responsive trial participants before treatment and for the BCH trial we will repeat the time point comparisons for responders only and non-responders only.

Chapter Two: Methods

Clinical Methods

2.1 Samples from Boston Children's Hospital

In collaboration with Khwaja and colleague's clinical trial based in Boston Children's Hospital (BCH) we received data from a bulk form of single cell RNA barcoding and sequencing (bulk SCRB-Seq) experiment conducted on whole blood of patients (Khwaja et al. 2014a). The trial recruited 12 female patients (nine with classic RTT and three with *MECP2* related disorders) between the ages of two and 12 years of age with 10 subjects being Caucasian and two subjects being Asian. The patients were recruited from BCH under principle investigator Dr Mustafa Sahin getting informed parental consent for each patient. Patients had to fulfil the following inclusion/exclusion criteria:

Inclusion

1. Participants were female
2. Participants suffer with typical or variant RTT as defined by the 2010 RTT diagnostic criteria (Neul et al 2010)
3. Participants had defined genetic mutation or deletion to the *MECP2* gene
4. Participants were prepubertal as defined by Tanner stage 1 or 2 breast development, Tanner stage 1 or 2 pubic development and younger than 12 years by bone age
5. Participants were 2 years or older

Exclusion

1. Participants must not have had prior therapeutic use of IGF-1, growth hormone, Lupron or other sex steroids
2. Participants must not have any allergy to the treatment
3. Participants must not have the following co-morbidities of RTT: diabetes mellitus, fatty acid oxidation disorder, syndrome associated with higher risk of malignancy or history of malignancy, chromosomal aneuploidy or any exposure to spinal irradiation either current or previous.
4. Participants must not have severe scoliosis with a spinal curve of 70 degrees or more by clinical or radiological examination

Participants were treated with mecasermin in two phases. Firstly the multiple ascending dosage (MAD) phase where participants received increasing doses of 40, 80 and 120 µg/kg of mecasermin and were treated for a total of 4 weeks. This was followed by a wash out period of 12 to 30 weeks. Participants with classic RTT progressed to the second treatment phase the OLE and received 120 µg/kg of mecasermin for a further 20 weeks. The clinical endpoints for the BCH trial were primarily safety and pharmacokinetic assessment while secondary measures were also made to examine the preliminary efficacy of the treatment. Safety and tolerability were measured in both phases of the trial while levels of IGF-1 and IGF binding protein 3 were measured in the MAD phase only. Preliminary efficacy was measured by recording the apnoeic index of participants using BioRadio device and measuring the social avoidance subscale of the ADAMS from pre-MAD to post-OLE.

As part of the study during the in-patient sessions patient bloods were taken by IV throughout the trial. The study was approved by the institutional review board of Boston Children's Hospital (NCT01253317). The inclusion/exclusion criteria and the primary and secondary outcomes were stated online at clinicaltrials.gov (<https://clinicaltrials.gov/ct2/show/NCT01253317?term=mecasermin&cond=Rett+Syndrome&rank=1>). All other information in this section was taken from Khwaja and colleagues publication based on the results of the trial (Khwaja et al. 2014a; Khwaja et al. 2014b). Table 2.1 below displays the sample codes for each participant in the BCH trial along with the dates where whole blood samples were taken. Participants with *MECP2* related disorders participated only in the MAD phase of the trial but did not progress to the OLE.

Table 2.1 - Blood sample data for mecaseimerin trial in Boston Children's Hospital. Sample dates for all blood samples that were used for the bulk SCRB-Seq DGE analysis. These were categorised as BCH-t0 = screening of MAD, BCH-t1 = end of MAD, BCH-t2 = baseline of OLE and BCH-t3 = final visit 5 of the OLE.

Patient code	Sample date Screening (BCH-t0)	Sample date End of MAD (BCH-t1)	Sample date Baseline of OLE (BCH-t2)	Sample date Visit 5 of OLE (BCH-t3)
RTT1	31/05/2011	05/07/2011	31/01/2012	19/06/2012
RTT2	19/08/2011	20/09/2011		
RTT3	15/08/2011	19/09/2011		
RTT4	25/03/2011	24/05/2011	07/02/2012	26/06/2012
RTT5	26/07/2012	30/08/2011		06/07/2012
RTT6	26/04/2011	02/06/2011	13/01/2012	08/06/2012
RTT7	07/03/2011	25/04/2011	07/02/2012	02/07/2012
RTT8	09/05/2011	14/06/2011	07/02/2012	02/07/2012
RTT9	01/08/2011	29/08/2011	13/01/2012	08/06/2012
RTT10	23/08/2011	07/10/2011		
RTT11	22/07/2011	06/09/2011	10/02/2012	03/07/2012
RTT12	09/09/2011	11/10/2011		19/06/2012

Figure 2.1 below displays a schematic representation of the time points where blood samples were taken during the trial and are coded from BCH-t0 to BCH-t3. Figure 2.2 on the next page shows the dosage and structure of the BCH mecasermin trial displaying the escalating doses that occur at the beginning of the MAD and OLE phases.

Figure 2.1 - Time points of whole blood samples for BCH mecasermin trial. Where BCH-t0 corresponds to screening of MAD, BCH-t1 corresponds to end of MAD, BCH-t2 corresponds to baseline of OLE and BCH-t3 corresponds to visit 5 of OLE.

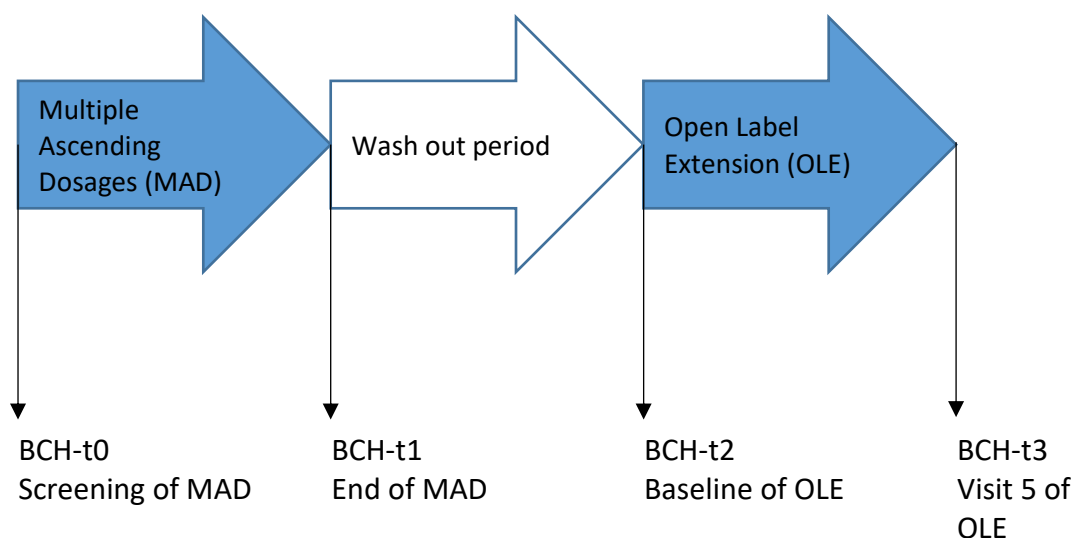
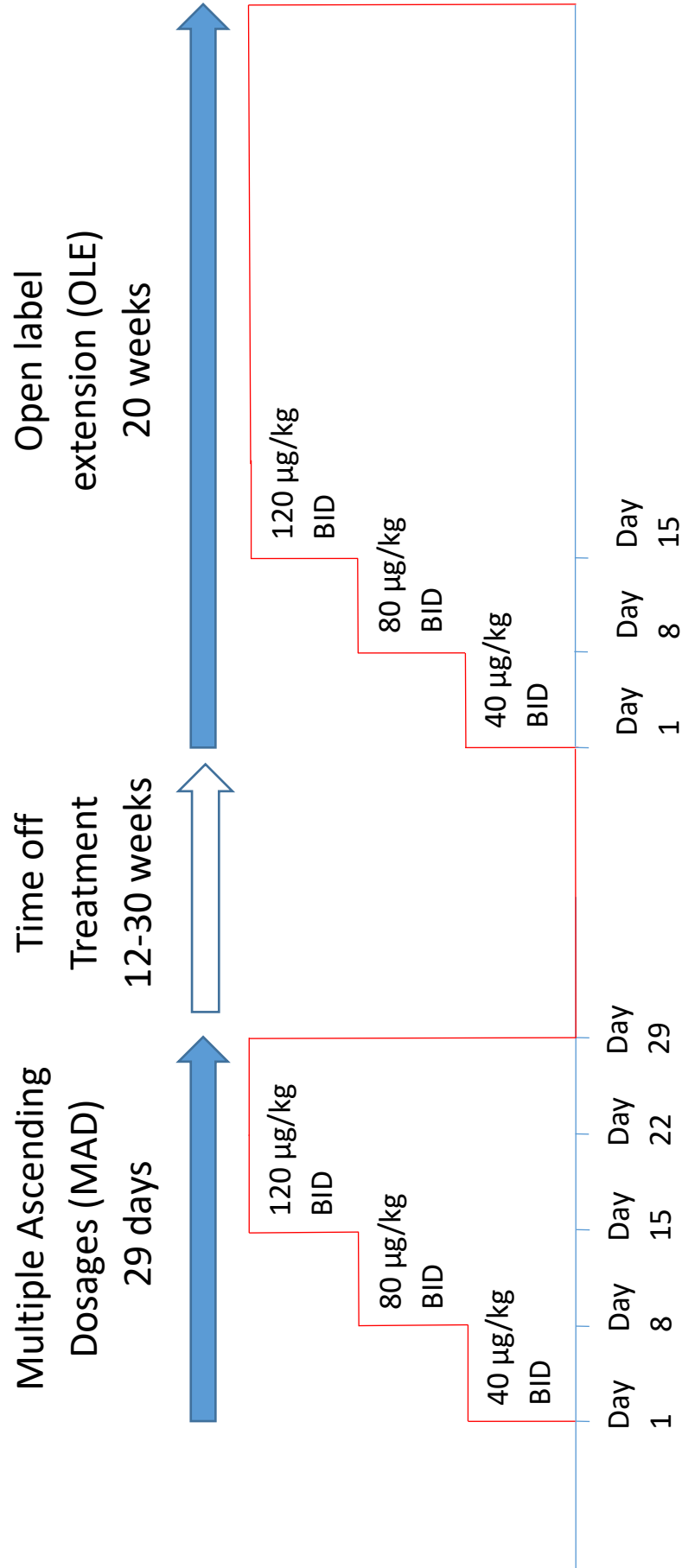


Figure 2.2 – Dosage and structure of BCH mecasermin trial. Schematic representation of the dosages and duration of the treatment periods in the Boston study. The study had two phases of IGF-1 treatment with a 12 to 30 week intermittent period off treatment. The first IGF-1 treatment phase in the study was the Multiple Ascending Dose (MAD) phase where patients were administered subcutaneously twice daily (BID) with ascending doses from 40 µg/Kg to 120 µg/Kg for three weeks in increments of 40 µg/Kg twice daily. For the final the fourth week of MAD (Day 21–29) the patients remained on 120 µg/Kg of IGF-1 treatment. The second phase of IGF-1 treatment phase was the Open Label Extension (OLE) where RTT patients were given the same dose escalation as MAD phase initially and then maintained the maximum dose of 120 µg/Kg of IGF-1 for an extended period of 17 weeks, total time of OLE was 20 weeks.



2.2 Samples from Versilia Hospital, Italy

In collaboration with Pini and colleagues we received a number of whole blood samples stored in PAXgene blood RNA tubes (QIAGEN, cat. No. 762165) corresponding to before and after treatment from the clinical trial mecamermin to treat RTT patients, in the Tuscany Rett Center, Versilia Hospital, Italy (Giorgio Pini, Congiu, Benincasa, et al. 2016). 10 female, Caucasian patients with classic RTT with *MECP2* mutations between the ages of 2 and 12 years old were recruited from the Tuscany Rett Centre to participate in the trial and receive treatment. 10 untreated RTT patients that were matched in age and disease severity were used as a comparison for the treated trial participants. Families provided written consent for their child participation which was approved by Comitato Etico Locale Sperimentazioni approval number 328. Patients had to fulfil the following inclusion/exclusion criteria:

Inclusion

1. Participants must have clinical diagnosis of classical RTT
2. Participants must have confirmed *MECP2* mutations

Exclusion

1. Participants must not be younger than two years old
2. Participants must not be older than 13 years of age
3. Participants must be in stage III of RTT disorder (Hagberg 2002)
4. Participants must not display non-classic variants of RTT
5. Participants must not have presence of neoplasia
6. Participants must not have history of hypoglycaemia

The clinical endpoints for this study were to measure the severity of disease after treatment with mecamermin to further examine the efficacy of the treatment in human RTT patients. Trial participants were treated with twice daily subcutaneous mecamermin injections of 100 µg/kg for a period of 20 weeks. This dose and treatment period is comparable to the OLE phase of the BCH trial. For the first and final weeks of treatment trial participants were given half doses of mecamermin.

To measure disease severity of RTT patient's three different approaches were used. Clinicians assessed RTT patients with International Severity Score (ISS) according to Kerr and colleagues (Kerr et al. 2001). ISS has five different subscales where patients are rated zero no abnormality, one mild abnormality or two severe abnormality. The

subscales of the ISS are growth and development, muscular-skeletal appearance, movement, mental-cortical and autonomic subscales. Rett severity score (RSS) was assessed by two blinded independent observers where patients were scored on 10 positive features and 10 negative features of RTT summarised in table 2.2. RSS was measured on a scale from one to five where low scores correspond to most severe and highest scores correspond to no impairment. Thirdly RTT patient's brain activity was measured through electroencephalogram (EEG) where mean theta and delta wave amplitude and frequencies as well as alpha band power were calculated.

Table 2.2 - Criteria for the Rett severity scale (RSS). This scale was measured with 10 negative features of RTT (left) and 10 positive features of RTT (right). These features were all marked on a one to five scale where one = extreme impairment, two = severe impairment, three = moderate impairment, four = mild impairment, five = absence of impairment.

Negative Features	Positive Features
Hand wringing	Pointing
Biting	Manipulating
Rocking	Reaching for something
Superficial breathing	Ability to imitate
Moaning/shouting	Yes or no head movement
Tongue chewing	Reactivity to call
Vacant staring	Reactivity to an object
Bruxism	Smiling in response to a stimulus
Breathholding/apnoea	Deliberate vocalisation
Valsalva manoeuvre	Attention

Of the trial participants in Versilia Hospital we had access to number of whole blood samples but only four patients had at least one sample corresponding to before treatment and after treatment. These patients were given codes from RTT1 to RTT4. The time intervals between before and after treatment differed substantially across the four patients from four months to over two years. We categorised each blood samples into those taken before mecasermin administration (VH-t0), those taken at the end of mecasermin administration (VH-t1) and finally samples taken beyond end mecasermin treatment (VH-t2). All these data are displayed in table 2.3 below.

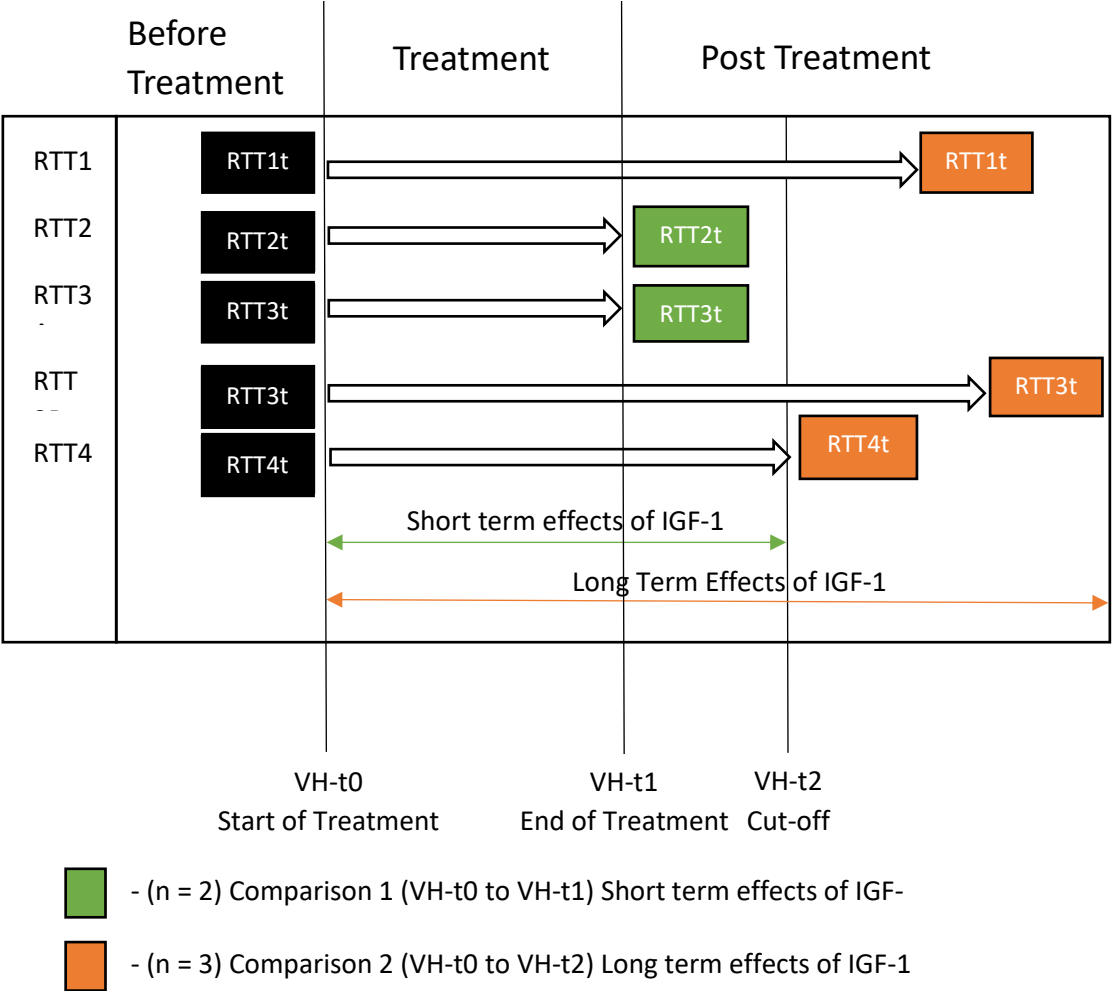
Table 2.3 - Blood sample data for mecasermin trial based in Versilia Hospital. Table includes sample codes, date of blood samples, are samples from before or after treatment, dates of start and end of mecasermin administration and time difference between before and after treatment.

Patient code	Sample date	Time point	Date of mecasermin Start	Date of mecasermin end	Time difference from VH-t0
RTT1	19/03/2013	VH-t0	19/03/2013	06/08/2013	
RTT1	05/08/2014	VH-t2	19/03/2013	06/08/2013	1y, 4m, 17 d
RTT2	02/10/2012	VH-t0	02/10/2012	12/02/2013	
RTT2	12/02/2013	VH-t1	02/10/2012	12/02/2013	4m, 10 d
RTT3	25/09/2012	VH-t0	25/09/2012	05/02/2013	
RTT3	18/11/2014	VH-t2	25/09/2012	05/02/2013	2y, 1m, 24 d
RTT3	05/02/2013	VH-t1	25/09/2012	05/02/2013	4m, 11 d
RTT4	25/02/2014	VH-t0	16/07/2014	03/03/2015	
RTT4	03/03/2015	VH-t2	16/07/2014	03/03/2015	1y, 6 d

Given that time interval from VH-t0 to VH-t1 corresponded to the duration of the mecasermin administration while VH-t0 to VH-t2 was a much longer period after the end of treatment and we considered these separately. Taking into account that the wash out period between MAD and OLE for the BCH trial was 12 to 30 weeks it was not reasonable to think that DGE from VH-t0 to VH-t2 was effected directly by mecasermin treatment. However the VH-t0 to VH-t1 was the time period where patients were receiving mecasermin treatment so this was more likely a direct effect of mecasermin treatment.

Figure 2.3 below displays a schematic of the timeline for each of the RTT samples and how they relate to the course of mecasermin treatment. We will refer to the comparison VH-t0 to VH-t1 as the short term effects IGF-1 action (Green coloured samples in figure 2.3) and VH-t0 to VH-t2 as the long term effects of IGF-1 action (Orange coloured samples in figure 2.3) for clarity.

Figure 2.3 - Structure of the whole blood samples from Versilia hospital mecasermin trial. Boxes represent blood samples with patient codes (RTT1, RTT2, RTT3 and RTT4) and time points (VH-t0, VH-t1 and VH-t2). Arrows go from untreated samples VH-t0 to samples at the end of treatment VH-t1 (green) and post treatment samples of VH-t2 (orange). These two different time intervals VH-t0 to VH-t1 and VH-t0 to VH-t2 will be used in the later DGE analysis and will be used to examine the short term effects of IGF-1 and the long term effects of IGF-1 treatment



A summary of the participant’s data from both the BCH and Versilia Hospital mecasermin trials including age, gender, mutation type and RTT stage are listed below in table 2.4.

Table 2.4 - Summary of participants of mecasermin trials from Versilia Hospital and Boston Children's Hospital data. Data includes sample codes gender age, mutation and RTT stage as per Hagberg criteria (Hagberg 2002)

Sample code	Gender	Age (Years)	MECP2 Mutation	RTT Stage
<i>Versilia Hospital Trial</i>				
RTT1	Female	4	c.1096-1184del98	III
RTT2	Female	2	T158M	III
RTT3	Female	10	R306C	III
RTT4	Female	4	c.157del46n	III
<i>Boston Children's Hospital Trial</i>				
RTT1	Female	3	c.538C>T	II
RTT2	Female	7	c.1135_1142delCCCGTG	NA*
RTT3	Female	7	c.1135_1142delCCCGTG	NA*
RTT4	Female	2	c.790_808del19	II
RTT5	Female	5	Deletion of E3 and partial E4 (6.0-7.1 kb)	III
RTT6	Female	4	c.1159_1273del	III
RTT7	Female	8	c.763C>T	III
RTT8	Female	4	c.763C>T	III
RTT9	Female	8	c.473C>T	III
RTT10	Female	3	c.916C>T	NA*
RTT11	Female	10	Deletion of E1 and E2	III
RTT12	Female	8	c.965C>T	III

* Participants with MECP2 related disorders (RTT2, RTT3 and RTT10 of BCH trial) that were not diagnosed with classic RTT were included in the MAD but not the extended OLE of the BCH mecasermin trial, therefore their RTT severity was NA.

2.3 Determining Response of Mecasermin Trial Participants

In these trials although there were significant improvements these varied with different patients. One important comparison that we considered was between participants that positively responded to mecasermin with those patients that did not respond to mecasermin. The response criteria for both studies is stated below and then described in further detail:

1. BCH mecasermin trial:

- i) Responders = Patients with decrease in apnoeic index from moderate severe (> five events per hour) to mild (< five events per hour) from beginning to end of trial.
- ii) Non-responders = no significant change in apnoeic index from beginning to end of trial.

2. Versilia hospital mecasermin trial:

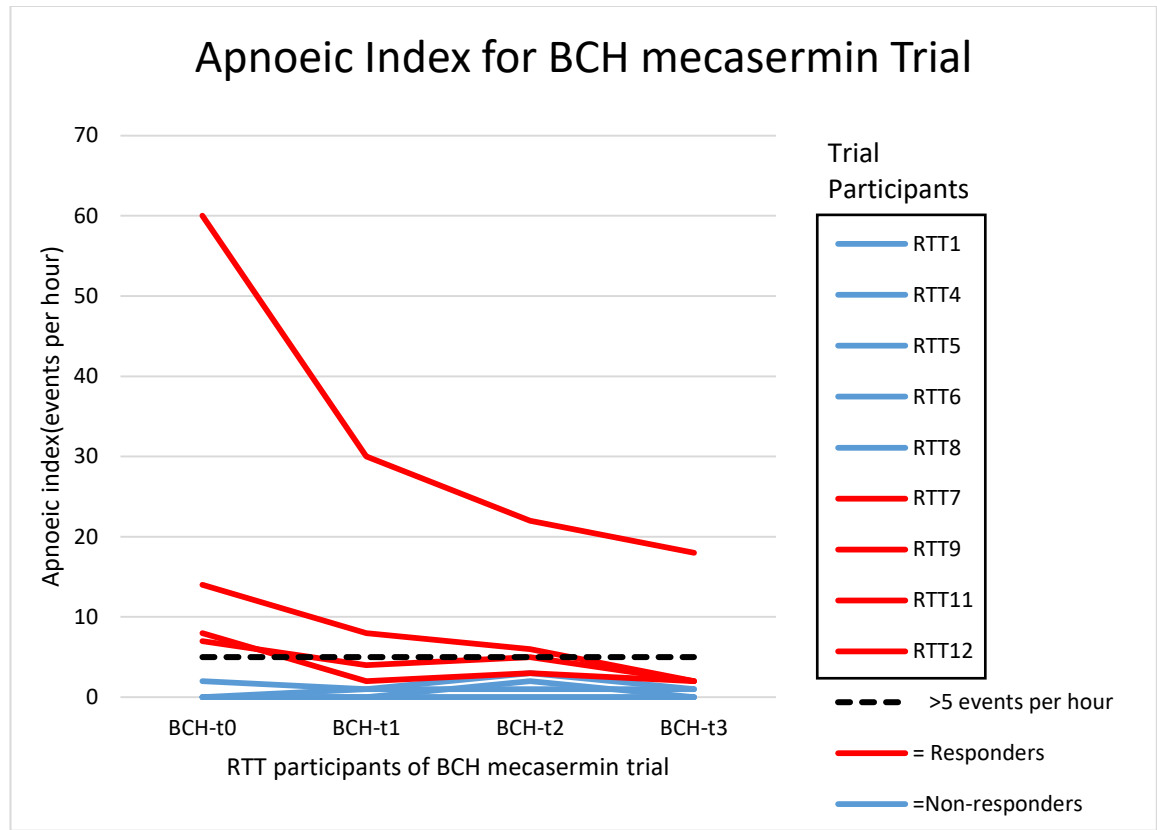
- i) Major responders = Patients with decrease in PBZ clinical severity score of three or greater from beginning to end of trial.
- ii) Minor responders = Patients with minimal decrease in PBZ clinical severity score of one or less from beginning to end of trial

In BCH trial the main improvement was found in apnoeic index. A graph charting the apnoeic index of all RTT trial participants across BCH trial is shown below in figure 2.4 and the data for this graph is shown in table 2.5. The apnoeic index was an adaption of the apnoea/hypopnoea index used in a previous clinical study in healthy individuals using dexmedetomidine and remifentanil therapy (Hsu et al. 2004).

Table 2.5 - Apnoeic index data for RTT trial participants over time points of the BCH mecasermin trial. These data correspond to the graph below in figure 2.4.

Apnoeic Index				
Time point	BCH-t0	BCH-t1	BCH-t2	BCH-t3
<i>RTT1</i>	0	1	1	1
<i>RTT4</i>	0	0	2	0
<i>RTT5</i>	0	0	0	0
<i>RTT6</i>	2	1	3	1
<i>RTT8</i>	0	0	0	0
<i>RTT7</i>	8	2	3	2
<i>RTT9</i>	7	4	5	2
<i>RTT11</i>	60	30	22	18
<i>RTT12</i>	14	8	6	2

Figure 2.4 - Graph showing the change in apnoeic index in BCH mecasermin trial. Trial participants were deemed responders (red) if they had improvement from moderate-severe to mild apnoea, trial participants that did not have this improvement or did not have moderate-severe apnoea to begin with were deemed non-responders (blue). The cut-off point for moderate-severe apnoea (>five events per hour) is shown by the dotted black line.



The purpose of differentiating by apnoeic index is to examine whether we can tell if there is a different pattern of DGE in the trial participants when they show this improvement to apnoeic index. We split the analysis into two separate groups responders (n = 4) and non-responders (n = 5) based on these apnoeic index improvements (see above figure 2.4 and table 2.5) and repeated the time point comparisons with edgeR and DESeq2 analysis.

Apnoeic index is the number of apnoeas per hour, where an apnoea is any event (cessation of breathing) over 10 seconds. A rate of five apnoeas or higher per hour was considered moderate-severe and less than five apnoeas per hour was considered mild. Responders were patients with improvement from moderate-severe apnoeic index to mild from BCH-t0 to BCH-t3, while non-responders were patients with no improvement to apnoeic index from BCH-t0 to BCH-t3.

In the Versilia Hospital trial significant improvements were found in ISS and RSS when comparing mecamermin treated and untreated participants. The difference between the trial analysis and our analysis is that we have samples from same patients while the trial compared different treated and untreated patients. To understand better which of our participants responded to treatment we examined the Pini, Bonuccelli, Zappella (PBZ) scale. PBZ scale is an adaption of the ISS scale designed by Pini, Bonuccelli, Zappella and colleagues hence its name (Giorgio Pini, Congiu, Zappella, et al. 2016) that has been shown to better capture the symptoms of RTT and its variant disorders. It showed the same trend in severity but had greater sensitivity to the severity of the more debilitating variants of the disorder. It captures broadly the features of the RSS scale and also further household abilities that patients might display. The basic structure of the PBZ scoring system is separated into three categories motor ability, communication and sensory ability and daily living. The 26 features of PBZ are displayed in the table 2.6 below:

Table 2.6 - Criteria for PBZ scale. The PBZ scale was assessed by clinicians marking each category presented below 1 to 4 with 1 being the mildest and 4 being the most severe.

Motor ability	Communication and sensory ability	Daily living
Mimicry	Language production	Going up stairs
Eye contact	Comprehension	Going down stairs
Dyspraxia	Auditory attention	Swallowing
Stereotypies	Visual attention	Independence
Inability of ambulation	Tactile attention	Falls (unrelated to other problems)
Sitting position	Breathing	
Muscle tone	Convulsions	
Nutrition/eating	Appetite	
Dystonia	Sleepiness during daytime	
	Night sleep	
	Quality of Sleep	

These 26 features are then ranked from one to four with one being mildest and four being the most severe. Based on the total RTT patients are rated mild 0-35, moderate 36-70 or severe 71-104. The application of the daily living component makes the PBZ scale more similar to the Rett syndrome behavioural questionnaire (RSBQ) normally used by

clinical trials (Mount et al. 2002) to assess carers perceived improvements to patients not just the negative symptoms. Unlike the RSBQ, PBZ were measured by clinicians which removes some of the subjectivity of carers or family can bring.

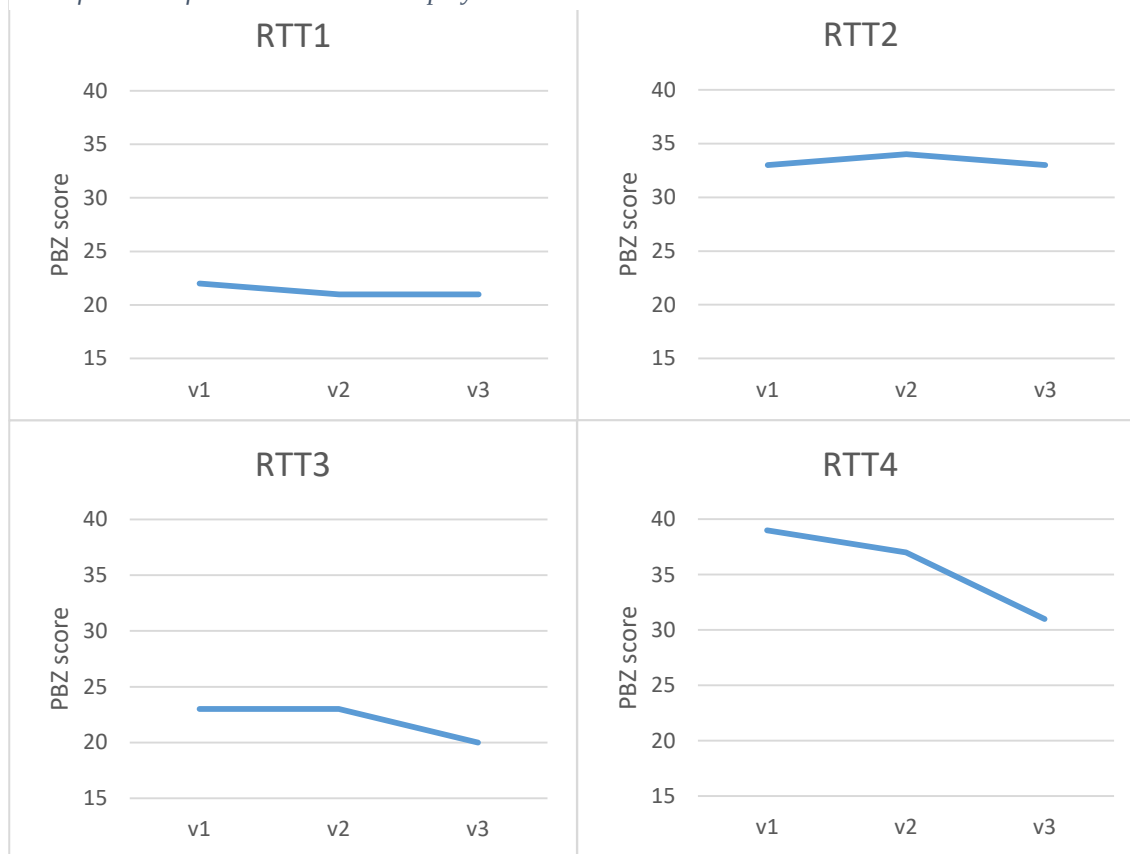
Considering the improvements in Versilia hospital trial were found in decreased severity scores we investigated the PBZ score of patients across the trial. When plotted the PBZ scores show two different profiles those with minimal improvement/no improvement and those with greater improvements. Figure 2.5 below shows the plot of PBZ scores each patient over three visits. These visits took place before, in the middle of and at the end of treatment. In table 2.7 below the dates for each of the visits where PBZ was assessed along with start/end dates for mecasermin administration.

Table 2.7 - Dates for PBZ assessments and mecasermin treatment for Versilia hospital trial.

Sample	Visit 1	Visit 2	Visit 3	Start of mecasermin	End of mecasermin
RTT1	19/03/2013	28/05/2013	06/08/2013	19/03/2013	06/08/2013
RTT2	02/10/2012	11/12/2012	12/02/2013	02/10/2012	12/02/2013
RTT3	25/09/2012	04/12/2012	05/02/2013	25/09/2012	05/02/2013
RTT4	01/07/2014	30/09/2014	19/11/2014	1/07/2014	19/11/2014

Using improvements seen in clinical severity based on PBZ score we determined response in patients of Versilia hospital mecasermin trial. We categorised RTT3 and RTT4 as major responders as they had greater improvements (decrease in PBZ of three or more) by the end of treatment and categorised RTT1 and RTT2 as minor responders as they had a more stable profiles with minimal change (decrease in PBZ of one or less) by end of treatment (see figure 2.5 below). These classes were major and minor responders as opposed to responders and non-responders because RTT1 did show a small improvement with a decrease of one between first and second visit.

Figure 2.5 - Plot of PBZ scores for RTT samples from Versilia hospital mecasermin trial. PBZ scores were recorded by clinicians at three visits at the beginning, middle and end of mecasermin treatment. Dates for each of the PBZ visits are displayed in table 2.6.



2.4 Laboratory and Analysis Methods

Sample Collection and Storage

Whole blood samples from Versilia Hospital mecasermin trial were taken by venepuncture, delivering the blood tissue into specialised PAXgene blood RNA tubes (QIAGEN, cat. no. 762165). Samples were stored at -80°C which stabilises the RNA and microRNA (miRNA) in the sample so RNA can be extracted at a later date. Samples were shipped at -80°C to Trinity Translational Medicine Institute in St. James's hospital Dublin where they were again kept at -80°C until ready for extraction of RNA.

Extraction of RNA from whole blood

RNA extraction from whole blood was conducted using the PAXgene blood RNA kit (QIAGEN, cat. no. 762164) and was completed as follows. Collected blood samples from the PAXgene blood RNA tubes (PBRT) were allowed to incubate at room temperature for two hours in the PBRT once removed from storage. Blood samples were transferred to sterile 15mL tubes to allow the samples to be centrifuged at $3000\times g$ using a swing-out motor. The supernatant was removed by decanting to leave a cell pellet

which was then resuspended in 4 ml of RNase-free water. Centrifugation was repeated and supernatant was removed leaving a cell pellet. 350µL of resuspension buffer (BR1) was added to the pellet, vortexing to dissolve. Sample was transferred to 1.5mL Eppendorf tubes and 300µL of binding buffer (BR2) and 40µL of proteinase K and vortexed for five seconds before incubating at 55°C for 10 minutes on a shaker incubator at 400RPM. Sample was then transferred into PAXgene shredder spin column over a 2mL tube and centrifuged at 19,000 xg for three minutes at room temperature. The supernatant was removed without disturbing the pellet and placed in a new 1.5mL tube. 350 µL of 100% ethanol (EtOH) was added to the sample and mixed by vortexing. 700 µL of the sample mixture was added at a time to the RNA spin column over a processing tube and centrifuged at 19,000xg for one minute at room temperature. The flow-through was discarded until all sample had been passed through the spin column. 10 µL of DNaseI stock solution and 70 µL of RDD per sample were mixed by inversion and then added directly to the spin column's membrane. This was allowed to incubate at room temperature for 15 minutes. 350 µL of wash buffer (BR3) was added to the spin column and centrifuged at 19,000xg for three minutes discarding flow-through. 500 µL of wash buffer (BR4) was added to the spin column and centrifuged at 15,000xg for three minutes and centrifuged a second time to ensure a dry membrane, this wash and dry with BR4 was repeated. The spin column was placed over a new 1.5mL Eppendorf tube and 40 µL of elution buffer (BR5) was placed directly on the membrane and centrifuged for one minute at 19,000xg to elute purified RNA. This elution step was repeated twice for a total volume of 80 µL of purified RNA. The sample was incubated at 65°C without shaking for five minutes and then immediately placed on Ice. The purified RNA sample was stored at -70C in 12 µL aliquots.

Determination of Concentration and RNA Integrity

The concentration and RNA integrity of the extracted RNA from the blood samples of the RTT patients was measured through the use of the RNA 6000 nano chip kit (Agilent Technologies, reorder number 5067-1511) and Agilent Bioanalyser 2100. Upon arrival of the kit the ladder was stored at -20°C before denaturing the ladder at 70°C for two minutes and then aliquoting into 3 µL of ladder into 1.5mL Eppendorf tubes, which were stored long term at -80°C. All other reagents in the kit were stored in a fridge at 2-8°C.

Before use all reagents were let equilibrate for 30 minutes at room temperature while the RNA samples themselves were left to thaw on ice. 550 µL of the agilent RNA 6000 nano gel matrix was placed into a spin filter and centrifuged for 10 minutes at 1500 ±

20%. 65 μ L of the filtered gel was then placed into microfuge tube provided, these can be stored at 2-8C for up to one month. RNA 6000 nano dye concentrate was vortexed for 10 second and spun down in a microcentrifuge before adding 1 μ L of the dye concentrate to the filtered gel matrix. This mixture was vortexed thoroughly and kept chilled and removed from light. The gel dye mix was centrifuged for 10 minutes at room temperature at 13,000 xg. The electrodes of the 2100 Bioanalyser were cleaned by putting 350 μ L RNase zap (Ambion, cat. No. 9780) and 350 μ L RNase free H₂O into separate chips. The RNase Zap chip was placed in the lid of the Bioanalyser for one minute, the RNase free H₂O chip was placed in the lid of the Bioanalyser for a further 30 seconds. The lid was then left open to air dry for 10 seconds. The RNA 6000 RNA nano chip was placed into the chip priming station (Agilent, reorder number 5065-4401) the clip set to the highest setting, the base was set to position C and the plunger of the syringe left at 1 mL. 9 μ L of gel-dye mix was pipetted (using reverse pipetting to prevent air bubbles) directly into the well marked with G of the RNA 6000 nano chip. The priming station was closed over the chip and the plunger depressed and held by the clip to pressurise for 30 seconds. The clip was released and the plunger allowed to depressurise until it has moved back to 0.3mL, wait five seconds and then gently pulling back to 1 mL. 9 μ L of gel-dye mix was then pipetted into the well marked g (the two wells directly above the G well). 5 μ L of RNA 6000 nano marker into the 12 sample wells and single ladder well. 1 μ L of ladder from the 3 μ L aliquot which was thawed on ice was added to the ladder well. 1 μ L of denatured sample (at 70C for two minutes) was added into each sample well, each sample well not used had 1 μ L of marker added to it. The RNA chip was then vortexed in the IKA vortex mixer, model MS-3 (IKA, Ident. No. 0003617000) at 2000rpm for 60 seconds. The loaded chip was then placed into underneath the electrodes in the chip holder, the lid was closed and the total eukaryotic RNA nano assay was run to assess the RNA concentration and RNA integrity.

Library Preparation using NEBNext Ultra Directional RNA and Poly(A) messenger RNA (mRNA) Magnetic Isolation Module

RNA samples were prepared with NEBNext Ultra Directional RNA (New England BioLabs, NEB #E7420S), NEBNext poly(A) mRNA magnetic isolation module (New England BioLabs, #E7490) and Agencourt AMPure XP beads (Becker Coulter, A63880). Based on the concentrations of the RNA samples measured using the Bioanalyser and RNA 6000 nano kit 100 ng of RNA samples + 10% to accommodate for pipetting error was used for library preparation.

All reagents from the NEBNext Ultra Directional RNA, NEBNext poly(A) mRNA magnetic isolation module were stored at -20°C and thawed and kept on ice during use. To prepare the first strand reaction buffer and random primer mix was prepared by mixing 8 µL of first strand synthesis reaction buffer (5X) and 2 µL of random primers and 10 µL of nuclease-free H₂O per sample which was kept on ice for later use. NEBNext oligo d(T)₂₅ beads were vortexed and 20 µL of the beads were transferred to a sterile 0.2 polymerase chain reaction (PCR) tube which was washed twice with 100 µL of RNA binding buffer (2X) which was mixed and removed from the beads using the magnetic rack. The beads were resuspended in 50 µL of RNA binding buffer and 110 ng (100ng + 10% for pipetting error) of sample RNA diluted with up to 50 µL using nuclease free H₂O. From this point on all incubations occurred at room temperature unless otherwise stated. The sample bead mix was heated for five minutes at 65°C and held at 4°C to denature RNA and bind poly-A mRNA to the beads. The beads were resuspended and allowed to incubate at for five minutes and this was repeated this was to allow the fragmented RNA to bind with oligo beads. To remove and discard the supernatant the PCR tube can be placed on the magnetic rack to separate the beads from the liquid and this waste supernatant was aspirated off with a pipette making sure not to disturb the magnetic beads that form a ring or a pellet shape on the edge of the tube. This technique was used throughout the protocol.

The binding buffer was removed from the oligo d(T)₂₅ beads by leaving the beads to separate for two minutes on the magnetic rack and then discarding the supernatant, leaving the beads with fragmented RNA bound. The beads were then washed with 200 µL of wash buffer to remove unbound RNA using the magnetic rack incubating for two minutes before discarding supernatant. This wash step was repeated and 50 µL of tris buffer was added to the sample bound beads pipetting up and down to mix thoroughly. The tris RNA bead mix was then heated at 80°C for two minutes in a thermocycler, holding at 25°C for elution of poly-A mRNA from the beads. 50 µL of RNA binding buffer (2X) was added to sample. The beads and RNA binding buffer mixed by pipetting and incubating for five minutes this was repeated a second time to allow the RNA and beads to re-bind, the tube was then placed on the magnetic rack for two minutes and the supernatant was discarded. The beads were removed from the rack and were washed again using the 200 µL of wash buffer mixing by pipetting and then removed after incubating on a magnetic rack for two minutes ensuring all supernatant was removed. The mRNA was eluted with 15.5 µL of the previously prepared first

strand synthesis mix to resuspend the beads and then incubated at 94°C for 15 minutes in a thermocycler. The bead first strand synthesis mix was immediately placed on the magnetic rack after incubation, 13.5 µL of the eluted RNA supernatant was collected and placed in a clean nuclease free PCR tube and placed on ice.

0.5 µL of murine RNase inhibitor, 1 µL of ProtoScript II reverse transcriptase and 5 µL of freshly prepared 0.1 µg/µL of actinomycin (Sigma, #A1410), which was protected from light, were all added to the fragmented mRNA and mixed by gentle pipetting. The tubes containing the sample were placed in a thermocycler and heated with the as follows; 10 minutes at 25°C, 15 minutes at 42°C, 15 minutes at 70°C and then held at 4°C. These temperature allowed for the generation of single stranded complementary deoxyribonucleic acid (cDNA).

48 µL of RNase-free H₂O, 8 µL of second strand synthesis reaction buffer (10X) and 4 µL of second strand synthesis enzyme mix were added to the single stranded cDNA and mixed by gentle pipetting, this mixture was incubated on thermocycler for one hour at 16°C.

The Agencourt AMPure XP beads were allowed to come to room temperature and were vortexed before use. 144 µL of AMPure XP beads were added to the second strand synthesis product, mixing well with a vortex and allowing to incubate at room temperature for five minutes. The tubes were quickly spun in a microcentrifuge before placing them on the magnetic rack for five minutes and the supernatant was discarded without disturbing the DNA bound magnetic beads. 200 µL of freshly prepared 80% EtOH was added to wash the beads without disturbing them, the EtOH was allowed to incubate for 30 seconds before discarding supernatant and repeating this ethanol wash again. The DNA bound magnetic beads were then left to air dry for five minutes by leaving the tube with an open lid on the magnetic rack. The DNA from the beads was eluted with 60 µL of 10mM tris-hydrochloride (tris-HCl) added and mixed by vortex and spun by microcentrifuge before incubating the tube for two minutes on a magnetic rack. 55.5 µL of the eluted DNA supernatant was transferred to a clean RNase-free PCR tube and stored at -20°C until needed for the next library preparation steps of the protocol.

6.5 µL of NEBNext end repair reaction buffer (10X) and 3 µL NEBNext end prep enzyme mix were both added to the purified cDNA from the previous steps, these were mixed by gentle pipetting. This mix then incubated as follows on the thermocycler (with a heated lid); 30 minutes at 20°C, 30 minutes at 65°C and then held at 4°C.

The cDNA was removed from thermocycler and 15 μ L of blunt/TA ligase master mix, 1 μ L diluted NEBNext adaptor (1:9 adaptor to 10mM tris-HCl) and 2.5 μ L RNase-free H₂O were all added to the cDNA and mixed by pipetting and spun down in a microcentrifuge. The sample was then incubated for 15 minutes at 20°C on a thermocycler with no heated lid.

16.5 μ L of RNase-free H₂O was added to the ligation reaction mixture, 100 μ L of resuspended AMPure XP beads were added and mixed into the solution with vortex and incubated minutes at room temperature and spun down by microcentrifuge. The tube was then incubated on magnetic rack for five minutes before discarding the supernatant with unbound cDNA while preserving the cDNA bound magnetic beads. The beads were then washed with 200 μ L of freshly prepared 80% EtOH for 30 seconds and discarded without removing the tube from the magnetic rack, this EtOH step was repeated again. All EtOH was aspirated from the tube and the beads were air dried for five minutes while remaining on the magnetic rack. The purified cDNA was resuspended with 10mM of tris-HCl, mixed by vortex and incubated for two minutes on the magnetic rack. This supernatant elution was transferred to a new RNase-free PCR tube.

50 μ L of AMPure XP beads were added to 50 μ L of supernatant, mixed using vortex, and left to incubate for five minutes. The mixture was placed on the magnetic rack for five minutes and the supernatant was discarded and the wash steps with 80% of EtOH twice to remove the unwanted unbound cDNA. The beads were left on the magnetic rack left to air-dry for five minutes with lid left open. 19 μ L of 10mM tris-HCl was used to resuspend the beads, this solution was mixed and left to incubate for two minutes. The tube was then placed on the magnetic rack and 17 μ L supernatant elution was transferred once the solution was cleared, and placed in a new RNase-free PCR tube.

3 μ L of NEBNext USER Enzyme, 25 μ L NEBNext HiFi PCR master mix, 2.5 μ L universal PCR primer and 2.5 μ L of index primer were added to each sample cDNA library from the previous section and mixed gently by pipetting. The universal and indexing primers were from NEBNext multiplex oligos for Illumina sets one and two (New England Biolabs, #E7335, #E7500). The following PCR cycle was set up on the thermocycler:

Table 2.8 - PCR cycles for library enrichment. Cycles used for PCR library enrichment as described by NEBNext Ultra Directional RNA library preparation Illumina instruction manual

Process	Temperature	Time	Cycle Number
Digestion	37°C	15 minutes	1
Denaturation	98°C	30 seconds	1
Denaturation	98°C	10 seconds	15
Annealing and Extension	65°C	75 seconds	
Final Extension	65°C	5 minutes	1
Holding	4°C	∞	

Agencourt AMPure XP beads were resuspended before use and 45 µL (0.9X) of beads was added to the PCR reaction product, mixing well by vortex and incubating for five minutes. The samples were spun down with a microcentrifuge and incubated on a magnetic rack for five minutes. The supernatant was discarded and 200 µL of freshly prepared 80% EtOH was added to wash the beads for 30 seconds before being discarded. This 80% EtOH wash step was repeated. The Beads were then air dried for five minutes while the tube remained open on the magnetic rack. 23 µL of tris-ethylenediaminetetraacetic acid (TE) buffer was added to elute the purified DNA, mixed well by vortexing and quickly spun down by microcentrifuge and incubated for two minutes at room temperature. The tube was placed on the magnetic rack and 20 µL of eluted DNA was removed from the tube without disturbing the magnetic beads. This supernatant was the purified index tagged cDNA libraries compatible with Illumina sequencing protocols.

Assessing library quality and removal of unwanted peaks

The Library quality of each sample was assessed by running a high sensitivity DNA assay on an Agilent Bioanalyser 2100 and determining whether the library peak which occurs roughly at 200 base-pair (bp) at least reaches a molarity compatible with the minimal required inputs for the miSeq protocol. The presents of non-specific peaks in the library would also require cleaning especially adapter dimer peaks (as seen in figure 2.6 below), which are usually a smaller peak around 127 bp. This corresponds to binding of adaptors from the adaptor ligation step which can dimerise with itself. This impurity would be amplified and sequenced in downstream processing. When these peaks were detected the samples were put through an additional cleaning step using the Agencourt AMPure XP beads consisting of a 1.0X repeat of the purification to achieve

this the sample was brought to a volume of 50 μL and an additional 50 μL of resuspended AMPure beads and the technique was repeated with two washes of freshly prepared 80% EtOH. An example of a library cleaned of dimer peaks can be seen below in figure 2.7.

Agilent high sensitivity DNA kit (Agilent, 5067-4626) was used for the high sensitivity DNA assay. All reagents were allowed equilibrate for 30 minutes before use at room temperature. Dye concentrate was mixed using a vortex for 10 seconds and 15 μL of dye were added to the high sensitivity DNA gel matrix vial and mixed by vortex for 10 seconds the gel-dye mix was filtered through a spin column for 10 minutes at $2240g \pm 20\%$. (Both dye concentrate and gel-dye mix were protected from light.) 9 μL of the filtered gel-dye mix was added to the bottom of the well marked G in the third row of the fourth column of the wells. The chip was then pressurised using the chip primer station for 60 seconds with the base set at the C mark and the release clip set at the bottom setting and pressing down the syringe on the chip until it is held. After the 60 seconds the release clip was released and after five seconds pulled the plunger of the syringe to 1mL and the chip was removed from the priming station. 9 μL of gel-dye mix was added to the remaining wells marked G (first, second and fourth row of the fourth column). 5 μL of high sensitivity DNA marker was added to the ladder and 11 sample wells. 1 μL of ladder was added to the ladder well and 1 μL of sample was added to each of the sample wells adding further 1 μL of the marker to any unused sample wells. The loaded chip was then vortexed using IKA vortex mixer model MS3 (IKA, Ident. No. 0003617000) at 2400 rpm for 60 seconds before placing the chip underneath the electrodes in the chip holder of the Agilent Bioanalyser 2100 and a high sensitivity DNA assay was run.

Figure 2.6 - Bioanalyser graph for contaminated library preparation. Graph illustrating the presence of contaminants in a sample library preparation Identified here as secondary peak between 50 and 100 bp. Samples with these characteristics were cleaned with additional steps. X-axis =fluorescence (FU), Y-axis = number of base pairs

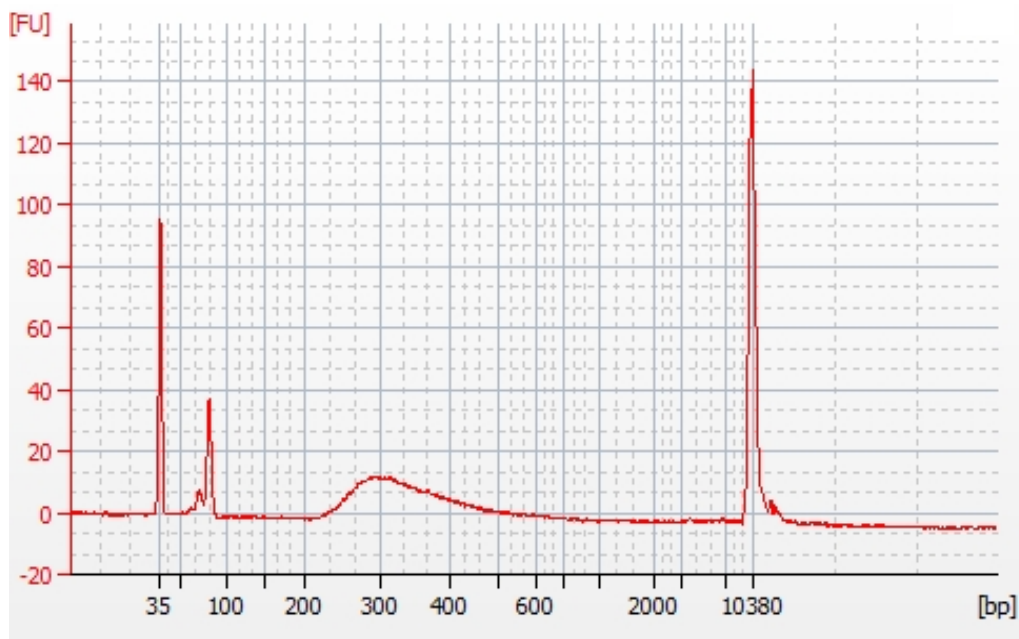
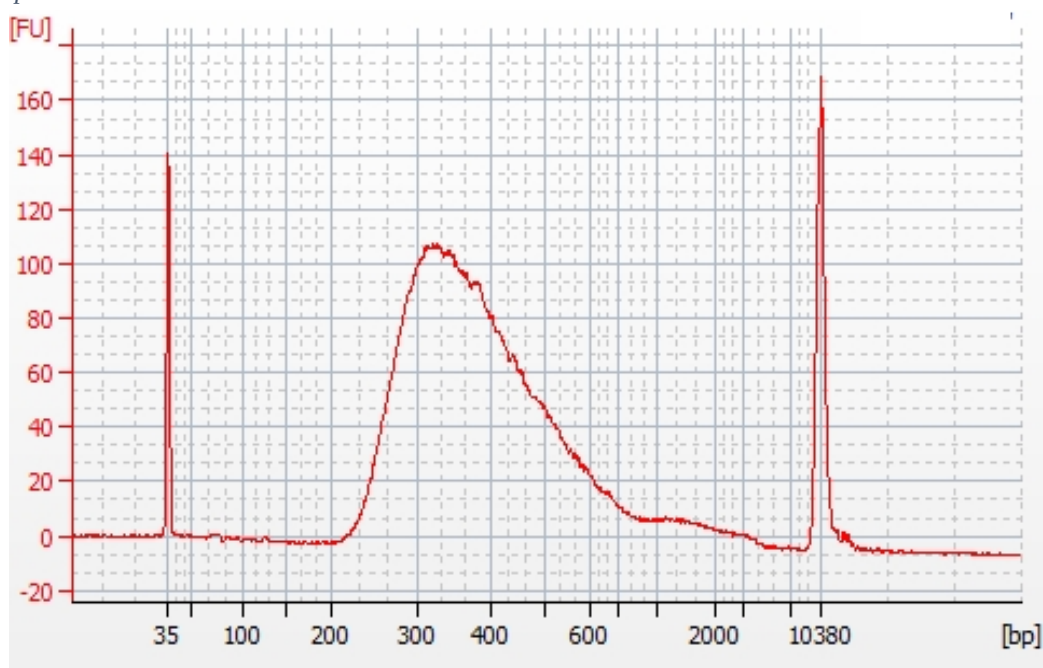


Figure 2.7 - Bioanalyser of graph clean pooled library preparations. Graph of pooled cleaned cDNA libraries of all samples before loading into miSeq. There was no detectable presence of any unspecific peaks that represent contaminants such as adapter dimers. X-axis = fluorescence (FU), Y-axis = number of base pairs



Sequencing by Illumina miSeq

To prepare the libraries for sequencing by Illumina MiSeq with a 50 cycle V2 kit (Illumina, # MS-102-2001), the libraries were pooled together so as to generate a pool of 5nM with a final pool volume of 60 μ L. The pooled library was then measured on the Agilent Bioanalyser 2100 to ensure no impurities were present and that the distribution of the cDNA library was roughly the same, this was done with the same method as before using the high sensitivity DNA chip.

The hybridisation buffer HT1 was thawed from -20°C by storing it at 2-8°C before chilling on ice, the V2 cartridge was thawed in a basing of room temperature water which was brought up to a water-level mark on the side of the cartridge. A freshly prepared 0.2nM of sodium hydroxide (NaOH) solution was prepared using 200 μ L of 1nM stock NaOH and 800 μ L of laboratory grade H₂O. A 20pM PhiX library was prepared using tris-HCl 8.5pH, with a 0.1% tween 20 and 0.2nM NaOH which was kept on ice. The pooled library was diluted down from 5nM to 4nM with tris-HCl and 5 μ L of the 0.2nM NaOH was added to library. This solution was mixed by vortexing briefly, spinning down briefly by microcentrifuge and incubated at room temperature for 5 minutes. 990 μ L of the chilled HT1 is added to the 10 μ L of pooled library bringing the concentration to 20pM. The PhiX library was also prepared using 10 μ L of denatured library with 990 μ L of pre-chilled HT1 buffer diluting it from 20nM to 20pM PhiX library. The library was further diluted from 20pM to 6pM by adding 180 μ L of 20pM library to 420 μ L of chilled HT1 giving a final library concentration of 6pM in 600 μ L of library. To make the 1% spike-in control with the PhiX and pooled library 6 μ L of the pooled library was removed leaving 594 μ L and then replaced with 6 μ L of the 20pM PhiX. The flow cell was removed from its container, buffers and salts were thoroughly washed away using a laboratory grade H₂O. An alcohol wipe and lint-free tissue was used to dry the glass and gasket ensuring that no streaks or tissue fibres were present. The flow cell was placed into the flow cell stage of the miSeq and the electrode pins aligned to the gaskets. The PR2 bottle was inserted in beside an empty wash bottle lowering the sippers into the bottles, the Radio-frequency identification (RFID) tracker that was present on the side of the bottle and the on the flow cell identified and registered the kit with the miseq.

600 μ L of the 6pM library mix was loaded into the loading port of the V2 cartridge piercing the seal with the pipette tip and pipetting the mix into the very bottom of the port. The cartridge was inserted into the cartridge holder of the chiller reagent. The

RFID tracker identifies that the cartridge is fully loaded and will match to the flow cell, PR2 bottle and wash bottle. The miSeq machine was logged onto a previously made BaseSpace account, and a sample sheet was created using a list of the sample names matched to the index IDs and primer sequences as shown in table 2.9.

Table 2.9 - miSeq sample sheet. Sample sheet showing index primers and the corresponding index ID sequences for RTT patient's library preparation

Sample Name	Index ID	Index Primer	Index ID Sequences
RTT1 VH-t2	A004	4	TGACCA
RTT1 VH-t0	A005	5	ACAGTG
RTT3 VH-t0	A001	1	ATCACG
RTT3 VH-t2	A006	6	GCCAAT
RTT3 VH-t1	A007	7	CAGATC
RTT4 VH-t2	A008	8	ACTTGA
RTT4 VH-t0	A009	9	GATCAG
RTT2 VH-t1	A013	13	AGTCAA
RTT2 VH-t0	A015	15	ATGTCA

Links to software packages used

Illumina BaseSpace <https://basespace.illumina.com/home/index>

edgeR <https://bioconductor.org/packages/release/bioc/html/edgeR.html>

DESeq2 <http://bioconductor.org/packages/release/bioc/html/DESeq2.html>

iPathway Guide <https://advaitabio.com/ipathwayguide/>

FASTQC <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

HISAT2 <http://ccb.jhu.edu/software/hisat2>

Stringtie <http://ccb.jhu.edu/software/stringtie>

SAMtools <http://samtools.sourceforge.net>

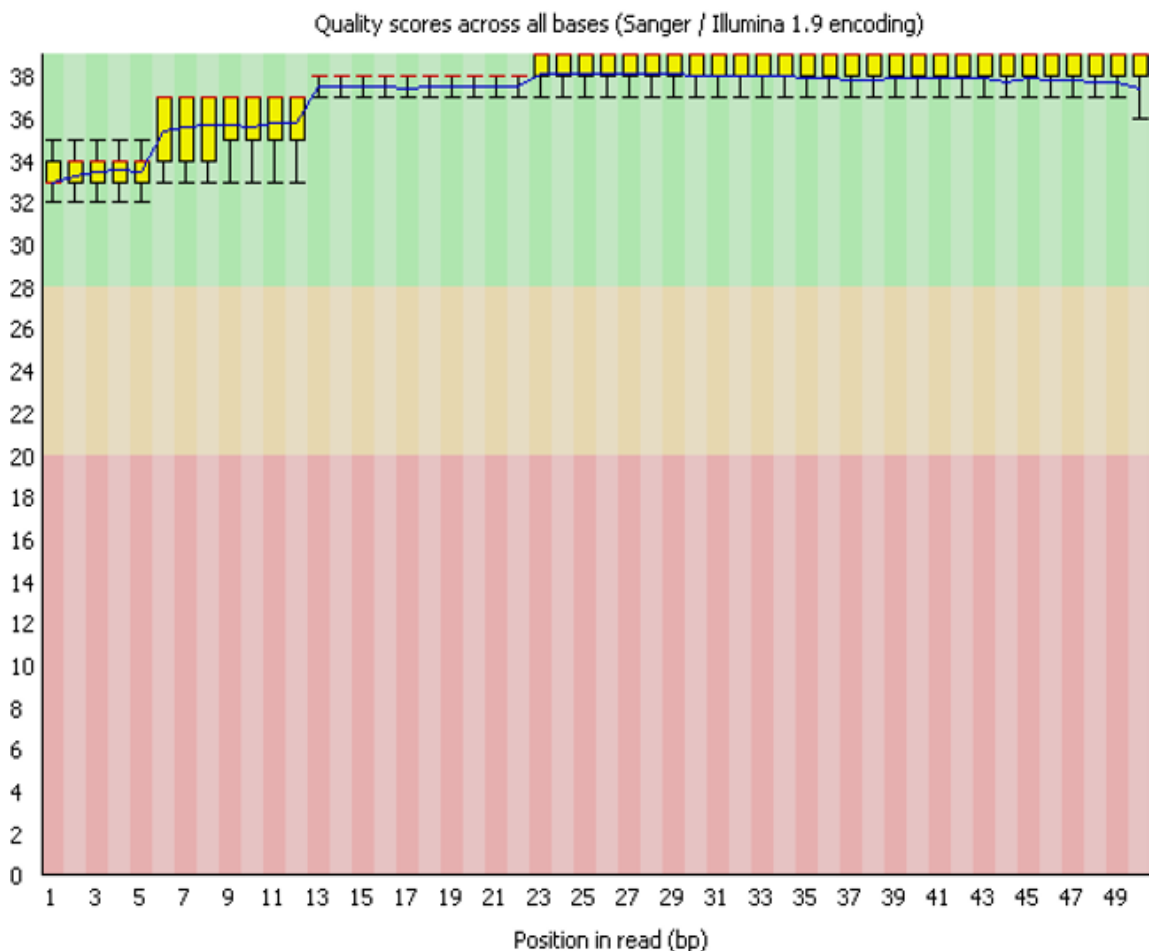
R Studio <https://www.rstudio.com/products/rstudio/download/>

Ballgown <https://bioconductor.org/packages/release/bioc/html/ballgown.html>

Data processing and generation of raw counts and FPKM tables

Data was produced on the miSeq and was uploaded to a BaseSpace account corresponding to the user previously created. The raw data was downloaded as FASTQ format files using BaseSpace downloader. These FASTQ files for each sample were assessed for quality using the FASTQC program to analyse the quality of the RNA-Seq data. This analysis of the samples revealed that all samples showed high percentage $\geq$ Q30 as illustrated from the example RTT2 VH-t1 in figure 2.8 below which was typical of our sequencing data.

Figure 2.8 - Example of FASTQC plot. An example of the Q score graph for sample RTT2 VH-t1, showing quality scores over all bases in a given read. These scores for RNA-Seq data were generated by inputting FASTQ files into the program FASTQC. X-axis represents the position of the read in base pairs, and Y-axis shows the Q scores.



This data was converted into raw counts per gene using the Stringtie -eB method of the “New Tuxedo protocol” devised by Pertea and colleagues (Pertea et al. 2016). This protocol used the programs Hisat2, samtools-1.5 and Stringtie to align the FASTQ files to a reference genome *Homo sapiens* Genome Reference Consortium Human Build 38 (GRch38). These aligned reads were in the sequence alignment/map format which were

large and cumbersome. The sequence alignment/map files were then converted into the more compact binary alignment/map (BAM) format using samtools-1.5. These aligned compressed BAM files were then used to generate read coverage tables in Stringtie.

The raw count tables and transcript count tables were then generated from Stringtie using the python script, PrepDE.py. This method of using Stringtie bypasses the de novo transcript assembly and merge functions in Stringtie, and instead uses the known transcript isoforms of the genes from the reference genome. PrepDE.py script was used to output two different files, one was a gene transfer format (GTF) file which stored all the transcripts that fully matched to the reference annotation transcripts. The other output was a gene abundance file that contained the fragments per kilobase of million mapped reads (FPKM) information. The GTF and gene abundance files were then read through Ballgown in order to input them into more manageable tables using RStudio. RStudio is a platform that can be used for the processing of packages from the Bioconductor Suite, a project that provides an open source of software for the purposes of bioinformatics. These packages are written in the R programming language which was specifically designed for statistical computing and generation of graphics.

Ballgown was used to match the assembled transcripts to annotations of the reference human genome, GRCh38. Ballgown matched the abundances generated in Stringtie and applied phenotypic information about experimental groups and specific samples. This produced a per sample table in a comma separated value (CSV) format, one table for gene counts and one table for FPKM values. The gene count table was to be used in the DGE packages edgeR and DESeq2 to test whether each gene was significantly differentiated comparing different control and experimental conditions.

Bulk Single Cell RNA Barcoding and Sequencing

Collaborators at Boston Children's Hospital (BCH) performed a single cell RNA barcoding and sequencing (Bulk SCRB-Seq) at Massachusetts institute of technology (MIT) MICRO Center on the whole blood samples from the BCH mecamermin trial. The blood samples described in table 2.1 for RTT1 to RTT12 were used for the bulk SCRB-Seq.

The bulk SCRB-Seq technique was first developed by Soumillon and colleagues and is a 3' digital expression RNA-Seq protocol. It uses a template switching reverse transcriptase method to isolate mRNA from single cells. Sample mRNA is then converted to cDNA attaching unique molecular identifiers, adaptors and barcodes that

enable specific identification of samples from a pooled source. The pooled library is then sequenced in multiplex. When this technique was applied to heterogeneous populations of cells the major gene expression patterns in sample can be measured. Bulk SCRB-Seq technique is much more cost efficient method to traditional single cell RNA-Seq techniques. Soumillon and colleague's findings showed that mean gene expression across single cell profiles and corresponding levels from bulk unsorted cell of the same time point were correlated with a p value $<10^{-100}$ and R of 0.8. Despite the strong correlation this method is less sensitive than deeper sequencing due to lower coverage. However sensitivity is strong enough to detect the more salient gene expression patterns (Soumillon et al. 2014). A recent comparison of four different single cell RNA-Seq methods showed that SCRB-Seq along with Smart Seq/C1 has the highest sensitivity and the second highest statistical ($R^2>0.8$) (Ziegenhain et al. 2016)

For extracting the gene expression information from the raw data our colleagues at MIT MICRO centre used end sequence analysis toolkit (ESAT). This program, designed by Derr and colleagues can analyse bulk RNA-Seq and is designed specifically for handling 3' DGE. ESAT addresses some of the common issues that 3' DGE can cause like poorer annotation and mappability of 3' transcripts due to greater fraction of untranslated regions (UTR) at the 3' end compared to 5' end or coding regions of the gene (Derr et al. 2016). The pipeline begins by demultiplexing the samples. Reads are then kept based on standard RNA-Seq quality control metrics. R1 and R2 reads are combined into a single end fastq file with barcode information under each read header. Single end reads with the same sequence and barcodes are retained for downstream analysis. Reads were then mapped on to GRCh37 and quantified per-gene using ESAT. For the count data generated counts per million are used to normalise for library size. The raw count tables were then ready for DGE analysis with edgeR and DESeq2 packages.

Differential Gene Expression Analysis

The BCH mecaermin trial samples were quantified with the bulk SCRB-Seq technique. The end product of this technique was a gene count table with rows representing genes detected and columns representing each sample. Similarly the Versilia Hospital mecaermin trial samples were quantified by new tuxedo protocol designed by Pertea and colleagues (Pertea et al. 2016). Its end product too was a gene count table with each gene represented on a separate row and each sample in a separate column. For both gene count tables, one count represents a single gene transcript quantified in the sample.

To analyse this quantification of genes we used two bioconductor packages designed for interpreting RNA-Seq data with low sample numbers. These were edgeR and DESeq2 (Robinson et al. 2009; Love et al. 2014). These software packages work in the R studio environment which uses R programming language calculations and graphical representation of data sets such as bioinformatics data (Venables et al. 1995). Both packages test for significance of DGE between two or more conditions, providing Log₂ fold change (Log₂FC) in the expression of a given gene along with p-value and corrected p-values for multiple testing using Benjamini-Hochberg (BH) (Benjamini & Hochberg 1995). These two packages provide the best performance as analysis tools in terms of statistical power for RNA-Seq with small sample number relative to other popular analysis packages (Ching et al. 2014).

Network and Pathway Analysis

To better understand the biological implication of the differential expression from the RNA-Seq experiments the results of the edgeR and DESeq2 analysis were run through another analysis program called iPathway guide. This tool can be used to provide more biologically meaningful interpretation of DGE.

This online resource uses impact analysis that was developed by Draghici and colleagues. It takes an overrepresentation approach and also measures absolute perturbation of a gene set in a given pathway and then using a statistical model function that takes into account biological factors gives a significance level in a given affected pathway. The impact analysis uses classical statistics but also considers biological factors such as magnitude of expression, direction, gene type, position in a given pathway and interactors, to apply a systems biology approach to interpreting DGE (Draghici et al. 2007). The input for the iPathway guide analysis was a custom file format of tab-delimited text containing gene symbol, log₂ fold change corrected p-value for each comparison. Cut off threshold was corrected p-value under 0.05 and log₂ fold change greater than 0.6.

Statistical methods

EdgeR uses empirical Bayes estimations based on a negative binomial model and uses exact testing to determine DGE (Robinson et al. 2009). DESeq2 uses the same empirical Bayes and negative binomial distribution but controls the level of shrinkage based on the amount of information in the data to give more stable and reproducible result. DESeq2 uses the Wald test to measure if shrunken estimates divided by standard error is

significant compared to standard normal distribution (Love et al. 2014). For detailed description of edgeR and DESeq2 statistical models and testing see the following publications “Small sample estimation of negative binomial dispersion, with applications to SAGE data.” by Robinson and Smyth and “Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2” by Love and colleagues respectively (Robinson & Smyth 2008; Love et al. 2014). For full description of statistical methods used in iPathway Guide network analysis see the impact analysis section of “A systems biology approach for pathway analysis” by Draghici and colleagues (Draghici et al. 2007).

Chapter Three: Results

3.1 Summary of Results

Below are three tables that show significant results we attained from our analyses of mecasermin trials in Versilia hospital and in Boston Children's Hospital. Firstly in accordance to our first hypothesis (see section 1.10) we compared samples before and after treatment with mecasermin. Table 3.1 below summarises the number of significant results from these time point comparisons:

Table 3.1 - Significant results from time point comparisons from mecasermin trials from Versilia hospital and BCH. Significant results were taken as expression changes that were <0.05 for corrected p -value. For edgeR and DESeq2 this was adjusted p -value and for network analysis it was Bonferroni corrected p -value.

Time Interval	EdgeR Analysis	DESeq2 Analysis	Network Analysis*
VH-t0 to VH-t1	11	1	1
VH-t0 to VH-t2	15	4	3
BCH-t0 to BCH-t1	0	0	0
BCH-t0 to BCH-t2	0	0	0
BCH-t0 to BCH-t3	0	0	0
BCH-t1 to BCH-t2	2	0	0
BCH-t1 to BCH-t3	0	0	0
BCH-t2 to BCH-t3	0	0	0

* Network analysis by iPathway Guide used the results from edgeR analysis and were considered significant if Bonferroni corrected p -value was <0.05 and only the top 5 pathways were considered.

The results above show very little change between the time point comparisons for the BCH trial. There was no significant changes in the DESeq2 analysis and only two significantly expressed genes in edgeR analysis. Given that there was an objective measure of improvement in BCH trial i.e. the decrease to apnoeic index we repeated the analysis by dividing the patients by response to determine if there was a difference between responders and non-responders. The results are displayed below in table 3.2:

Table 3.2 - Significant results from responders and non-responders from BCH mecasermin trial.
Significant results were taken as expression changes that were <0.05 for corrected p -value. For edgeR and DESeq2 this was adjusted p -value and for network analysis it was Bonferroni corrected p -value.

Response	Time interval	EdgeR Analysis	DESeq2 Analysis	Network Analysis*
Responders (n=4)	BCH-t0 to BCH-t1	0	0	0
	BCH-t0 to BCH-t2	3	0	1
	BCH-t0 to BCH-t3	0	0	0
	BCH-t1 to BCH-t2	0	0	0
	BCH-t1 to BCH-t3	0	0	0
	BCH-t2 to BCH-t3	1	0	5
Non-Responders (n=5)	BCH-t0 to BCH-t1	0	0	0
	BCH-t0 to BCH-t2	1	0	0
	BCH-t0 to BCH-t3	6	0	1
	BCH-t1 to BCH-t2	36	5	1
	BCH-t1 to BCH-t3	38	0	1
	BCH-t2 to BCH-t3	1	0	0

* Network analysis by iPathway Guide used the results from edgeR analysis and were considered significant if Bonferroni corrected p -value was <0.05 and only the top 5 pathways were considered.

When the BCH trial participants were categorised by response as per our second hypothesis, the numbers of significant genes expressed increased in the time point comparisons. This was mainly found in the non-responders although responders also showed four as opposed to two significant genes in the edgeR analysis. The edgeR results for non-responders found a total of 82 gene significantly expressed across the trial. The majority of these were at the intervals BCH-t1 to BCH-t2 and BCH-t1 to BCH-t3. The non-responder's time interval BCH-t1 to BCH-t2 showed the five significant genes for DESeq2 analysis in BCH trial.

Finally we analysed the patients based on their response to mecasermin treatment, as described in section 2.3. We divided the BCH participants into responders and non-responders and the Versilia Hospital participants into major responders and minor responders. Due to the different structure of these trials we only compared untreated samples at t0 (VH-t0 and BCH-t0). As the results of these comparisons were high, probably due to the different genetic backgrounds of patients, we only considered the overlapping genes between the two trials. These results are displayed below in table 3.3.

Table 3.3 - Significant results from response based comparisons from Versilia hospital and BCH mecaseimerin trial. Significant results were taken as expression changes that were <0.05 for corrected p -value. For edgeR and DESeq2 this was adjusted p -value and for network analysis it was Bonferroni corrected p -value.

Comparison	EdgeR Analysis	DESeq2 Analysis	Network Analysis*
Versilia Hospital: Major responders versus Minor Responders at VH-t0	85	10	0
BCH: Responders versus Non-Responders at BCH-t0	5028	2748	1
Overlap between Versilia Hospital and BCH	10	0	0

* Network analysis by iPathway Guide used the results from edgeR analysis and were considered significant if Bonferroni corrected p -value was <0.05 and only the top 5 pathways were considered.

3.2 Time-Point Comparisons

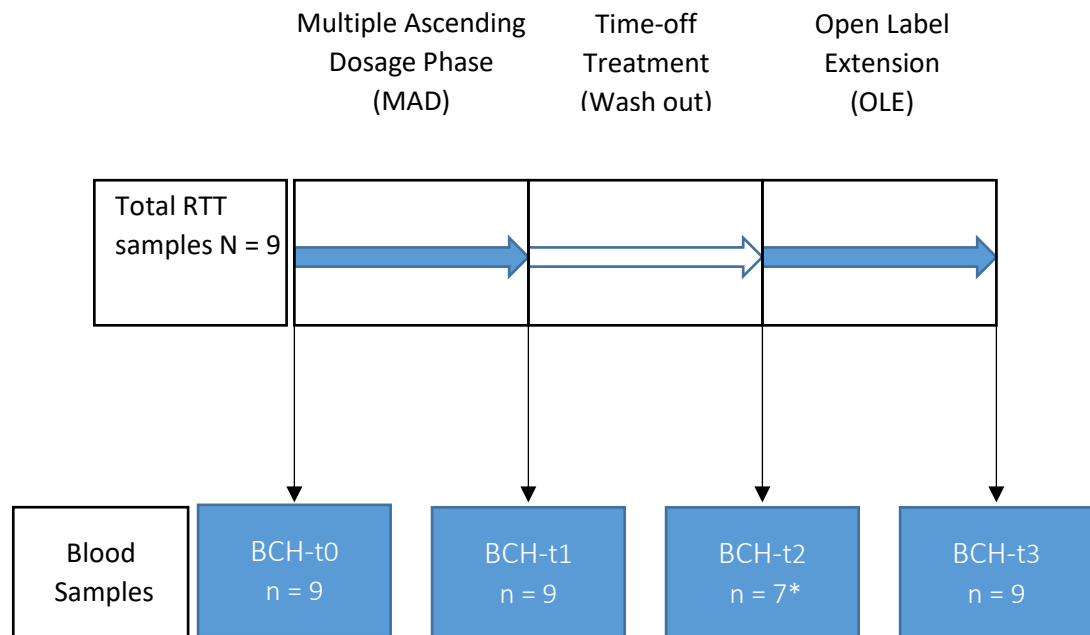
As previously described the limited blood samples available from Versilia Hospital mecaseimerin trial meant the DGE analysis had to be split into two different comparisons. The short term effects of IGF-1 action looked at samples at the end of 20 week treatment period and compared them to the untreated samples from the same patient. For the long term effects of IGF-1 action we compared samples that were taken over a year after treatment at varying time intervals and compared them to untreated samples from the same patient. For a clear illustration of the comparisons see figures 2.3 and table 2.3 in the previous chapter as they illustrate a schematic of the comparisons and a table with all the dates for each sample.

This division of samples lowers the statistical power of the experiment. However given the period of time between VH-t1 to VH-t2 is large and exceeds the wash out period used in the BCH study the two comparisons cannot be considered the same.

In the mecaseimerin trial in BCH we have a greater number of time points to compare. Here we have BCH-t0, BCH-t1, BCH-t2 and BCH-t3 and because of the structure of the trial there are a number of meaningful comparisons that we can make. Figure 3.1 below shows a schematic of the structure of the trial with the number of samples at each time point (not including the *MECP2* related disorder patients that were included in the MAD phase). In total the intervals we compared were BCH-t0 to BCH-t1, BCH-t0 to BCH-t2, BCH-t0 to BCH-t3, BCH-t1 to BCH-t2, BCH-t1 to BCH-t3 and BCH-t2 to BCH-t3 for all samples.

For both trials we used both edgeR analysis, DESeq2 analysis and network analysis for the time point comparisons. The network analysis results in all cases were based on the results from edgeR inputs though DESeq2 was also used as an input as well.

Figure 3.1 - Schematic of the blood samples analysed in BCH mecasermin trial. Shows the number of samples per time point and where they occur. The time points *BCH-t0*, *BCH-t1*, *BCH-t2* and *BCH-t3* correspond to screening (beginning) of MAD, day 29 (end) of MAD, baseline (beginning) of OLE and visit 5 (end) of OLE respectively.



**BCH-t2* had two less samples one of these was a responder and the other a non-responder. *BCH-t2* corresponds to baseline of OLE.

Due to low sample numbers we consider mainly the genes overlapping between edgeR and DESeq2 as they will be more robust, however this is not always possible especially in the BCH analysis. The full list of genes that were significantly differentially expressed in each analysis can be found in appendix A and B.

Mecasermin trial in Versilia Hospital

In the short term effects of IGF-1 comparison we found that one gene overlapped in both edgeR and DESeq2 analysis. *Mitochondrially encoded ATP Synthase 6 (MT-ATP6)* was found to be strongly down regulated in VH-t1 samples compared to VH-t0 and had the greatest significance with Benjamini-Hochberg (BH) adjusted p-value of 1.23×10^{-13} and 3.56×10^{-12} in edgeR and DESeq2 analysis respectively. *MT-ATP6* had the largest $\log_2$ fold change of -8.01 and -8.1 in edgeR and DESeq2 analysis respectively.

MT-ATP6 encodes a protein involved in the production of the F_1F_0 -ATP synthase subunit that enhances ATP production as well as ionic homeostasis of the mitochondria during stress stimuli. *MT-ATP6* has been found to be expressed in the nucleus and observed to translocate to the mitochondria to increase the production of F_1F_0 -ATP synthase subunit under stressful cellular conditions (Moghadam et al. 2013). This gene has been associated with several neurological pathologies including Leigh syndrome and neurogenic muscle weakness ataxia and retinitis pigmentosa (NARP) (Rahman et al. 1996; Oost 1993) characterised by progressive neurodevelopmental symptoms due to dysfunctions of mitochondrial ATP production (Sgarbi et al. 2006).

In the long term effects comparison of IGF-1 in the BCH trial we found four genes that were overlapping between the edgeR and DESeq2 analysis. These genes were *elongation factor 1 alpha 1 pseudogene 7 (EEF1A1P7)*, *Plasmalipin (PLLP)*, *ras oncogene family member 37(RAB37)* and *mitochondrially encoded cytochrome c oxidase subunit 2 (MT-CO2)*.

The *EEF1A1P7* gene was up-regulated in VH-t2 samples with respect to VH-t0 samples by $\log_2$ FC of 7.25 and 6.59 and a BH adjusted p-value of 1.44×10^{-14} and 1.4×10^{-2} in edgeR and DESeq2 analysis respectively. This gene which has similar homology to *EEF1A1* was a pseudogene and as such does not have any functions.

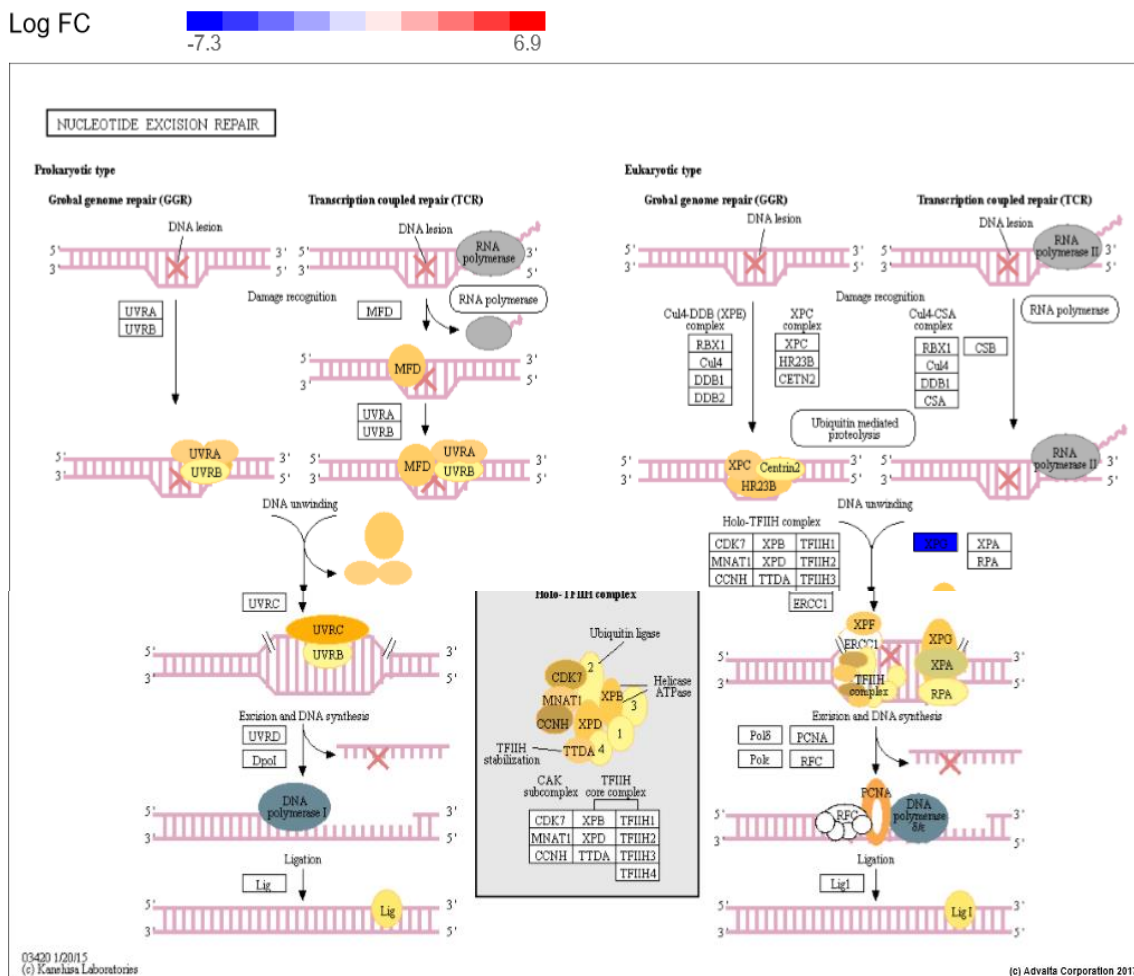
The *PLLP* gene was up-regulated in VH-t2 samples with respect to VH-t0 samples by $\log_2$ FC 7.12 and 6.47 and a BH adjusted p-value of 1.96×10^{-7} and 3.1×10^{-2} in edgeR and DESeq2 analysis respectively. *PLLP* has been found to be involved in the myelination process which occurs in Golgi complex (Yaffe et al. 2015). *PLLP* has also been found to localise to the putative Bardet-Biedl syndrome type 2 (BBS2) region. This disorder has been characterised by obesity, cognitive dysfunction, rod-cone dystrophy and renal abnormalities (Forsythe & Beales 2013).

The *RAB37* was up-regulated in VH-t2 sample with respect to VH-t0 samples by $\log_2$ FC of 2.88 and 2.91 and a BH adjusted p-value of 5.02×10^{-8} and 3.44×10^{-3} in edgeR and DESeq2 analysis respectively. *RAB37* encodes a small ras GTPase and has been implicated in signal transduction, vesicle trafficking and mast cell degranulation (Masuda et al. 2000; Higashio et al. 2016). *RAB37* has been shown to have specific functions in controlling insulin exocytosis during glucose induced secretion (Ljubicic et al. 2013).

The most significant of the differentially expressed genes in the long term effects of IGF-1 comparison was the *MT-CO2* gene. *MT-CO2* was decreased by a log₂ fold change of -10.85 and -10.36 and a BH adjusted p-value of 1.44×10^{-14} and 7.85×10^{-11} for edgeR and DESeq2 analysis respectively. This gene has an integral role in the respiratory chain where it encodes cytochrome c oxidase subunit 2 which is part of the complex IV. Cytochrome C oxidase functions as a redox coupled proton pump that establishes an electrochemical gradient enabling ATP synthase by generating ATP from adenosine diphosphate (ADP) (Yoshikawa 1998; Stock 1999). Mutations in *MT-CO2* has been associated with mitochondrial disorders and mitochondrial Encephalomyopathy, lactic acidosis and stroke-like episodes (MELAS) (Kytovuori et al. 2017; Wong et al. 2001; Rossmanith et al. 2008).

For the network analysis, the results obtained from edgeR analysis and DESeq2 analysis were both entered as inputs for iPathway Guide impact analysis. The results from the short term effects of IGF-1 comparison produced one significant result for the edgeR analysis but nothing for the DESeq2 analysis. The nucleotide excision repair (NER) pathway had a Bonferroni corrected p-value of 5.0×10^{-3} in over-representation analysis. This over-representation was found as a decrease in log fold change of the *ERCC excision repair 5 endonuclease (ERCC5)*.

Figure 3.2 - Schematic of nucleotide excision repair pathway (KEGG: 03420). Network analysis from short term effects of IGF-1 based on edgeR results. The schematic shows the decrease in log fold change of XPG or ERCC excision repair 5, endonuclease.



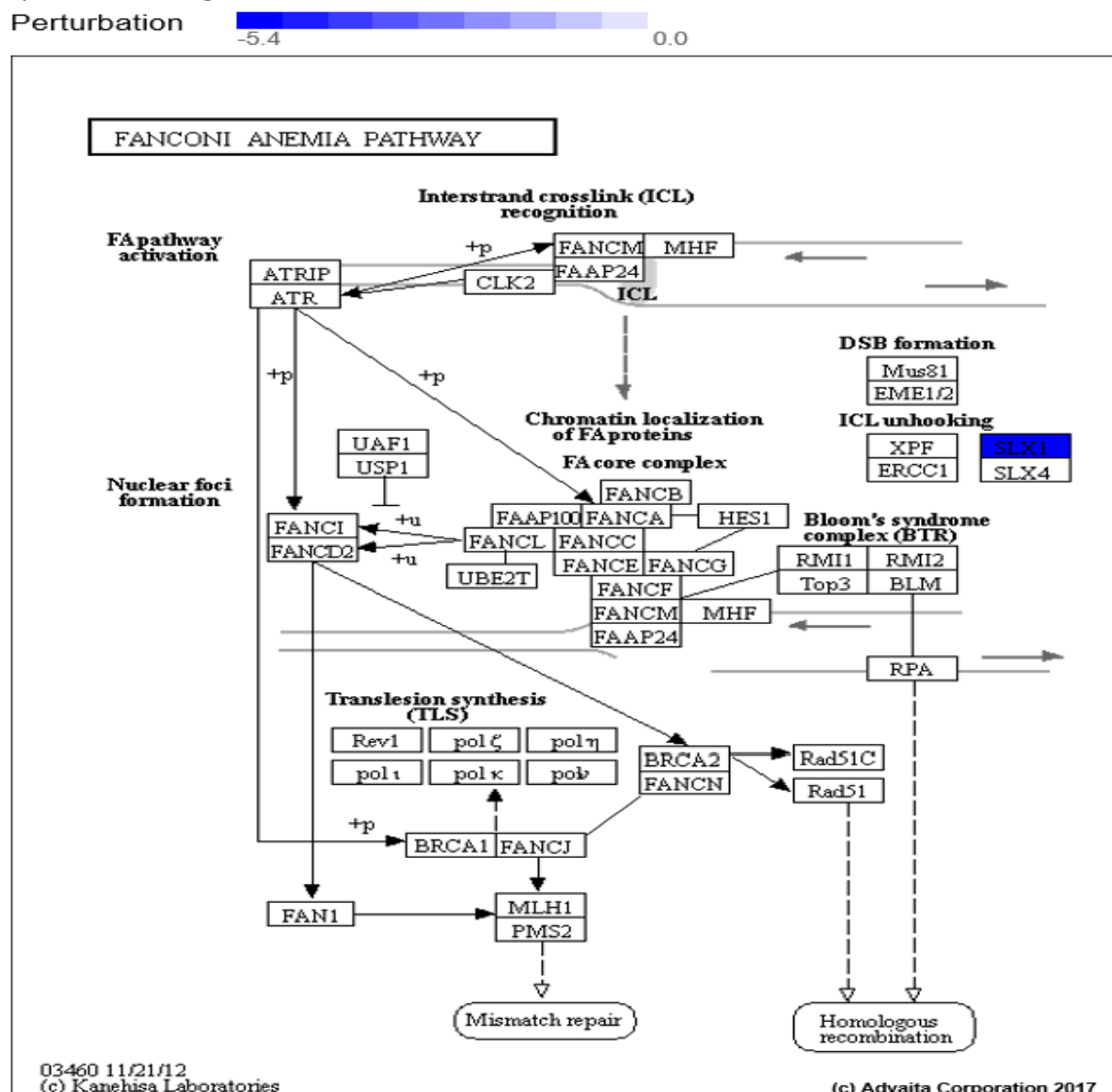
The NER pathway describes the mechanism of DNA repair in both prokaryotic and eukaryotic organisms. NER provides a versatile system that removes DNA that has been damaged from environmental sources such as ultraviolet damage by sunlight. Figure 3.2 highlights the *ERCC excision repair 5 endonuclease* (*ERCC5* or *XPG*) gene (in blue) that is shown to be down regulated in the NER pathway. The functional role of the *ERCC5* in the NER pathway is to catalyse the 3' incision and stabilise an open complex and is thought to play a role in catalysing the 3' incision at damaged DNA sites (de Laat et al. 1999; Schäfer 2013).

The network analysis based on the long term effects of IGF-1 results found three pathways significantly altered based on the input from edgeR analysis results. These pathways were Fanconi anemia pathway, oxidative phosphorylation (over-representation

analysis only) and Huntington's disease. Although these pathways were also found in the DESeq2 analysis they were not found significant when Bonferroni corrected.

Fanconi anemia is an autosomal recessive disorder syndrome that leads to predisposition to human cancer due to genomic instability. The clinical features of Fanconi anemia includes developmental abnormality, bone marrow failure and skeletal defects, hypopigmentation and predisposition to cancer (Moldovan & D'Andrea 2009). Interestingly this pathway has been implicated in the coordination of several functions in DNA crosslink regulation including translesion synthesis, homologous recombination and NER (Niedzwiedz et al. 2004). The effected Fanconi anemia pathway displayed below in figure 3.3 shows a decrease in perturbation of *SLX1* homolog A structure-specific endonuclease subunit (*SLX1*) gene. This pathway was perturbed with a Bonferroni corrected p-value of 1.6×10^{-2} .

Figure 3.3 - Schematic of the Fanconi anemia pathway (KEGG: 03460). Network analysis from the long term effects of IGF-1 based on the edgeR results. The pathway shows a decrease of $\log_2$ fold change of 5.4 in the *SLX1* gene.



The second pathway significantly affected in the network analysis was oxidative phosphorylation with a Bonferroni corrected p-value of 3.2×10^{-2} . This was based on the over-representation analysis only. The analysis identified the *MT-CO2* and *NADH: Ubiquinone Oxidoreductase Subunit A1 (NDUFA1)* as the over-represented genes in the pathway with *MT-CO2* decreased and *NDUFA1* increased as shown in figures 3.4 and 3.5.

Figure 3.4 - Schematic of the oxidative phosphorylation pathway (KEGG: 00190). Network analysis from the long term effects of IGF-1 based on the edgeR results. The pathway shows a decrease in log fold change of 10 in *MT-CO2* and an increase in log fold change of 6.8 in *NDUFA1* genes.

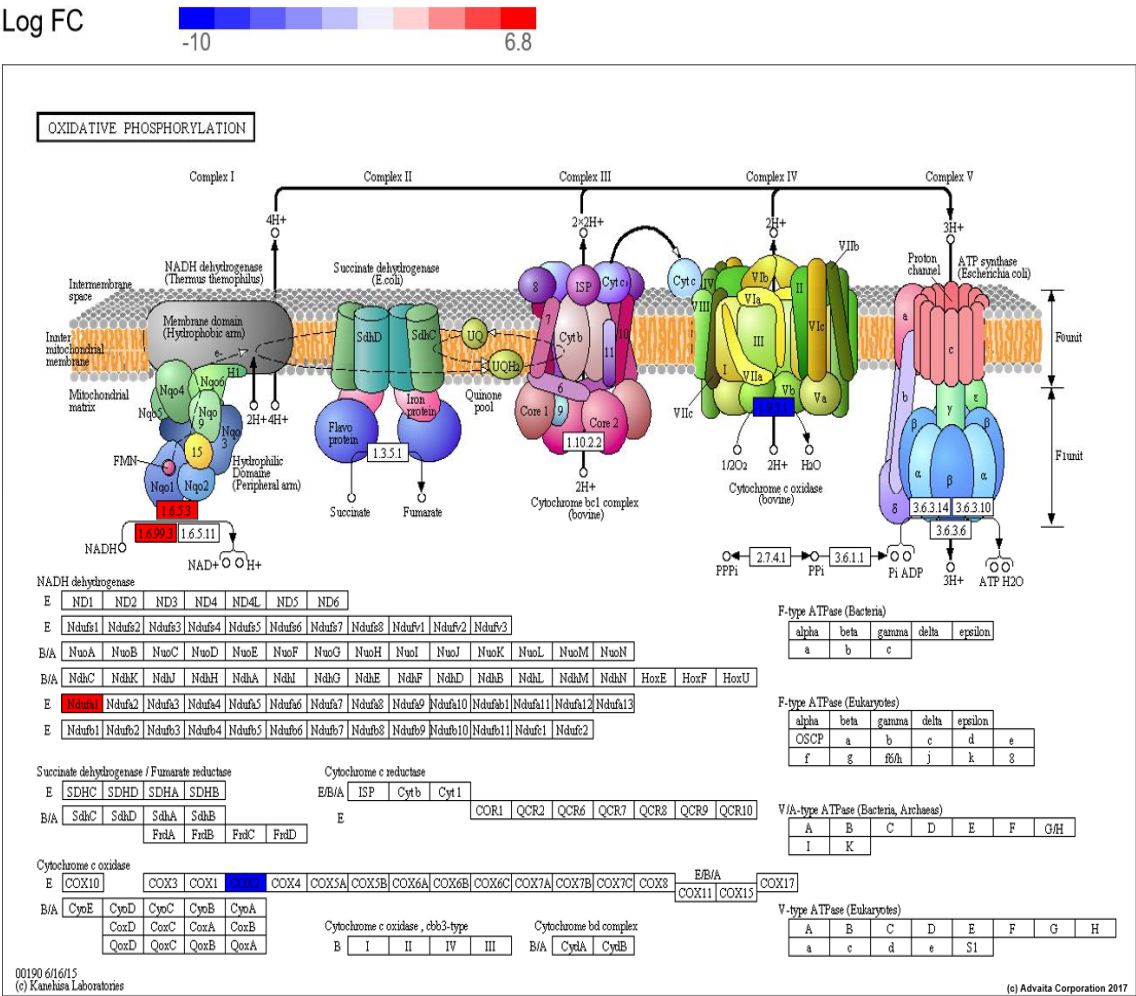
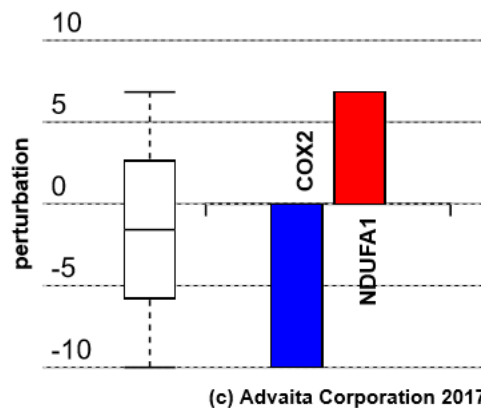


Figure 3.5 - Box plot of absolute perturbation values for over-represented genes in the oxidative phosphorylation pathway.

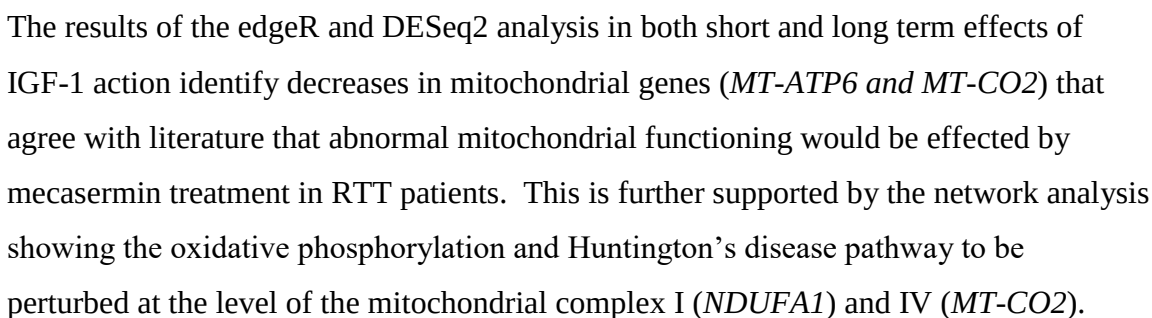


The *MT-CO2* gene had previously been identified by both edgeR and DESeq2 analysis however the *NDUFA1* gene was only observed in the edgeR. *NDUFA1* is a gene associated with complex I of the electron transport chain. Mutations in this gene are associated with mitochondrial encephalomyopathy and complex I deficiency (Loeffen et al. 1998; Fernandez-Moreira et al. 2007; Potluri et al. 2009).

The final pathway significantly effected in the long term effects of IGF-1 was the Huntington's disease pathway. This pathway was linked to the previous oxidative phosphorylation pathway shown in figures 3.4 and 3.5. Again the same decrease in *MT-CO2* and increase in *NDUFA1* that was seen in the oxidative phosphorylation pathway (figure 3.5) was found in Huntington's disease pathway this can be seen in figure 3.6 below.

Much evidence of mitochondrial dysfunction has been identified in Huntington's disease especially in mouse models that mimic the poly- cytosine-adenine-guanine (CAG) repeats that characterise the disease. Mitochondrial dysfunction has been observed in mutant *Huntingtin* (*Htt*) knock-in mice (Hdh^{Q111}) showing significant reduction of ATP in striatal neurons of mice embryos (Milakovic & Johnson 2005). Orr and colleagues showed that N-terminal fragments of *Htt* can effect ATP generation and mitochondrial trafficking even without forming aggregations (Orr et al. 2009). Striatal neuronal cultures of mutant *Htt* mouse models have shown mitochondria have a vulnerability to Ca^{2+} overloading in both neurons and astrocytes (Lim et al. 2008; Oliveira & Gonçalves 2009). These studies show illustrate the link between oxidative phosphorylation pathway and the Huntington's disease pathway.

Perturbation



For the BCH trial we first examined each time point including all the RTT patients. Only two genes were found to be significant in the edgeR analysis and no genes were found significant in the DESeq2 or network analysis. The two significant genes were

Transmembrane Protein 176A (TMEM176A) and *Transmembrane Protein 176B (TMEM176B)* both of which were decreased by log₂ fold change -2.72 and -2.52 and BH adjusted p-value of 1.3×10^{-4} and 5.69×10^{-4} respectively. The decrease in both these genes was at time point BCH-t1 to BCH-t2 and raise the possibility of the some sort of immunological reaction as these genes are associated with dendritic cell immaturity (Condamine et al. 2010). This reaction occurs at the interval BCH-t1 to BCH-t2 during the wash-out period of trial when participants finished the MAD phase and were not currently being treated with mecasermin.

3.3 Time-point Comparison in Responders and Non-Responders of BCH

In the BCH mecasermin trial apnoeic index significantly decreased in a number of patients with clinically significant apnoea. Clinically significant events were considered any apnoeic episode over 10 seconds in length, while moderate-severe apnoea was five or more events per hour. Five RTT trial participants had moderate-severe apnoea and four of these RTT trial participant's condition improved to mild apnoea by the end of treatment. In this way we determined the subgroups responders and non-responders (see section 2.3)

For edgeR analysis we found a total of four significant genes in the responders and 82 significant genes in the non-responders time point comparisons. In the DESeq2 analysis there was no significant genes found in the responder's time points and five significant genes in the non-responders at the interval BCH-t1 to BCH-t2. These DESeq2 genes all overlapped with findings in the edgeR analysis for the same time point and were *TMEM176A*, *TMEM176B*, *BR serine/threonine kinase 1 (BRSK1)*, *Joining Chain of Multimeric IgA and IgM (JCHAIN)* and *Dematin Actin Binding Protein (DMTN)*. For the network analysis there was no overlap between the results produced from the edgeR and the DESeq2.

For the significant genes in comparisons in responders (n = 4), three genes were found at time point from BCH-t0 to BCH-t2 and one gene was found at time point from BCH-t2 to BCH-t3.

Androgen dependent TFPI regulating protein (ADTRP) was found to up regulated from BCH-t0 to BCH-t2 by a log₂ fold change of 6.02 and BH adjusted p-value of 0.017. *ADTRP* has been shown to up-regulate the *tissue factor pathway inhibitor (TFPI)* expression and activation in an androgen-dependent manner. This has been proposed

with some supporting evidence to be a mechanism for anticoagulation protection in endothelium (Lupu et al. 2013).

Major histocompatibility complex class II DR beta 5 (HLA-DRB5) was found to be down-regulated from BCH-t0 to BCH-t2 by a $\log_2$ fold change of -2.33 and BH adjusted p-value of 1.24×10^{-3} . Interestingly in the subsequent interval from BCH-t2 to BCH-t3 *HLA-DRB5* was significantly up-regulated with a $\log_2$ fold change of 3.06 and a BH adjusted p-value of 1.51×10^{-5} . This gene plays an important role in the immune system as a major histocompatibility complex (MHC) class II cell receptor present on antigen presenting cells which recognise specific antigens to mediate immune response through CD4 T-cells. CD4 T-cells or helper T-cells activate dendritic cells to recruit killer T cells (Robinson & Delvig 2002; Bodmer et al. 1994).

Smith-Magenis syndrome chromosome region candidate 5 (SMCR5) was found to be down-regulated from BCH-t0 to BCH-t2 by a $\log_2$ fold change of -2.97 and BH adjusted p-value of 2.0×10^{-2} . *SMCR5* a gene located in chromosome 17p11.2 has been proposed as a candidate gene responsible for Smith-Magenis Syndrome along with a number of other genes identified in a Smith-Magenis syndrome mouse model (Bi et al. 2002). Smith-Magenis syndrome patients have intellectual disabilities, head growth abnormalities, brachycephaly and sleep disturbance (Greenberg et al. 1991)

As already mentioned *HLA-DRB5* was initially down regulated in BCH-t0 to BCH-t2 and then up regulated in BCH-t2 to BCH-t3. Given the function of *HLA-DRB5* to mediate immune functioning and the dramatic swing in fold change from decreased to increased suggests some sort of change in immune response. The decrease occurred when patients were initially treated in MAD phase and went off treatment (BCH-t0 to BCH-t2) and then when placed back on treatment during the OLE phase *HLA-DRB5* increased (BCH-t2 to BCH-t3).

There was a much greater amount of significant genes found in the non-responders compared to responders in the BCH trial. The edgeR and DESeq2 analysis even overlapped at BCH-t1 to BCH-t2 for non-responders with five significant genes which we will now discuss. This interval corresponds with wash-out period between MAD and OLE. Again *TMEM176A* and *TMEM176B* genes were found to be decreased in DESeq2 analysis with a $\log_2$ fold change of -3.67 and -3.57 and BH adjusted p-value of 7.79×10^{-9} and 7.93×10^{-11} respectively. A similar result was found in the total participant comparison in edgeR analysis.

Brain specific serine/threonine kinase 1 (BRSK1) was found to be down-regulated from BCH-t1 to BCH-t2 by a $\log_2$ fold change of -1.62 and BH adjusted p-value of 2.27×10^{-3} . This gene functions as a kinase for serine/threonine and promotes axon-dendrite polarity and synaptic development (Sample et al. 2015).

Joining chain of multimeric IgA and IgM (JCHAIN) was found to be down-regulated from BCH-t1 to BCH-t2 by a $\log_2$ fold change of 1.22 and BH adjusted p-value of 1.2×10^{-2} . This gene acts to link to monomer units IgM and IgA immunoglobulins to make IgM pentamers and IgA polymers respectively (Johansen et al. 2001).

Dematin actin binding protein (DMTN) was found to be down-regulated from BCH-t1 to BCH-t2 by a $\log_2$ fold change of -1.32 with an adjusted p-value of 2.6×10^{-2} . *DMTN* gene encodes a protein that acts as an actin binding bundling protein structurally functions in the erythrocyte, KO mouse models of *DMTN* causes haemolytic anemia (Lu et al. 2016).

The decrease in *TMEM176A* and *TMEM176B* at the wash-out period of non-responders and the swing from decreased to increased expression of *HLA-DRB5* from in responders are interesting results as they point towards different immunological functions being modulated depending on response. Considering the glial cell reactivity mechanism in the literature immune functioning genes may also contribute centrally to glial reactivity.

3.4 Response Based Comparisons

We wanted to understand the DGE between patients with positive clinical outcome compared to those without these improvements after mecasermin treatment. To do this patients were categorised by clinical improvements. This was based on decreases in PBZ score in Versilia Hospital trial (major responder's decreased by $\geq$ three, minor responder's decreased by $\leq$ one) and decrease in apnoeic index in BCH trial (responder's decreased in apnoeic index from moderate-severe to mild, $<$ five events per hour. Non-responders did not significantly change) and we compared samples before treatment.

In this comparison samples are not matched and genetic backgrounds of these samples are not controlled for. Therefore this comparison is much more likely to identify false positives. This was reflected in the results (displayed in table 3.3) showing that 85 and 10 genes significantly expressed in edgeR and DESeq2 analysis respectively in Versilia hospital and 5028 and 2748 genes significantly expressed in edgeR and DESeq2 analysis respectively of BCH trial. Interestingly network analysis only identified one significant

pathway, this was an over-representation of ribosome pathway in the BCH edgeR analysis which had 5028 significant genes.

To decrease the probability of finding false positives we considered only genes that overlapped in analyses of both trials. We identified 10 genes that were significant in edgeR analysis. These genes were *interleukin 7 receptor (IL7R)*, *S100 calcium binding protein A12 (S100A12)*, *asparaginyl-tRNA synthetase 2, mitochondrial (NARS2)*, *proline rich nuclear receptor coactivator 2 (PNRC2)*, *mast cell expressed membrane protein 1 (MCEMP1)*, *adenosine deaminase (ADA)*, *S100 calcium binding protein A8 (S100A8)*, *BCL2 related protein A1 (BCL2A1)*, *ribosomal protein L27 (RPL27)* and *SEC11 homolog C signal peptidase complex subunit (SEC11C)*. The results of these analyses are displayed in the table 3.4 below.

Table 3.4 - Overlapping significant genes from edgeR analysis based response comparisons in BCH and Versilia hospital trials. The table shows the genes with log₂FC and corrected p-values.

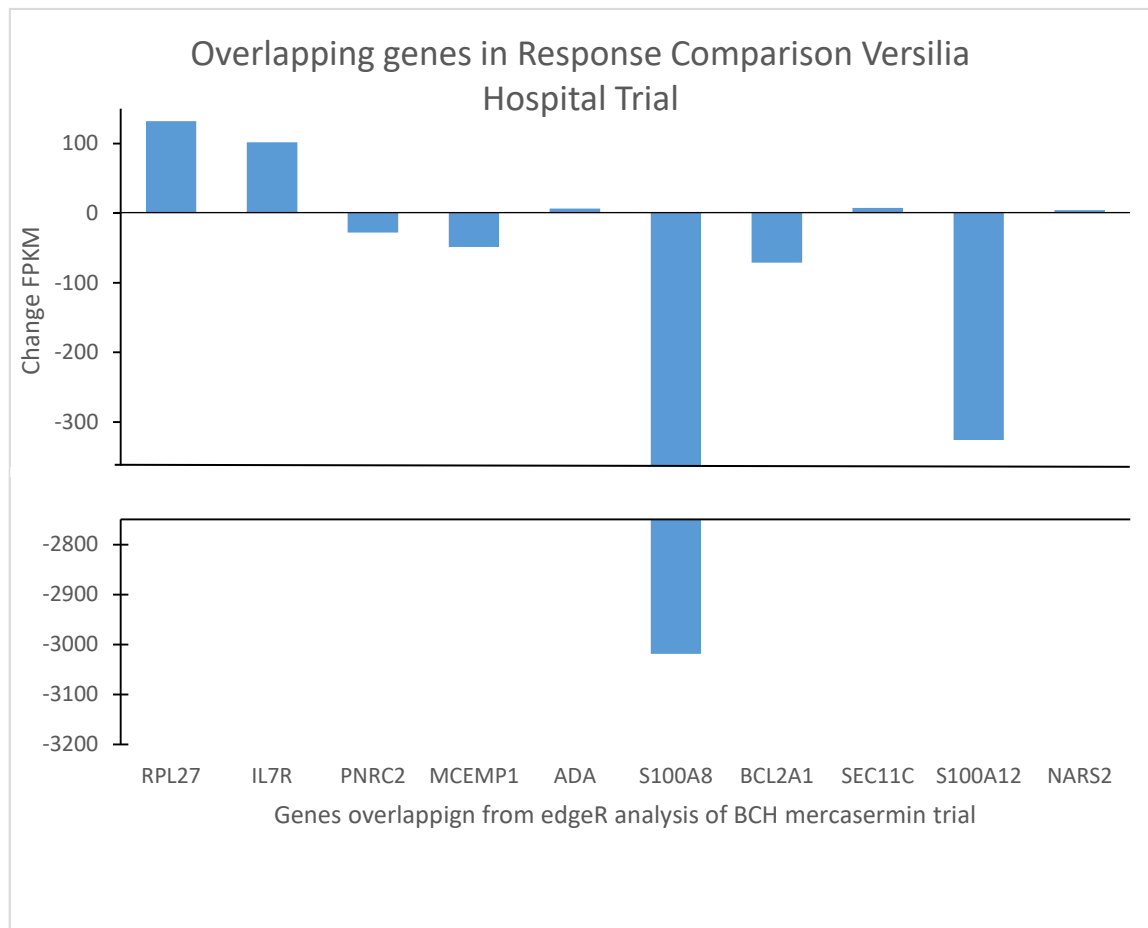
Gene Name	BCH Log ₂ FC	Versilia Hospital Log ₂ FC	BCH p-value	Versilia Hospital p-value
IL7R*	-1.20	3.47	3.68x10 ⁻²	5.65 x10 ⁻⁸
S100A12	-2.26	-2.18	4.77 x10 ⁻³	3.29x10 ⁻²
NARS2	1.29	5.46	2.81x10 ⁻²	3.95x10 ⁻²
PNRC2	-1.02	-3.37	4.61x10 ⁻²	1.62 x10 ⁻⁵
MCEMP1*	-1.59	-4.16	2.99x10 ⁻²	1.04x10 ⁻²
ADA*	-1.78	5.81	2.78 x10 ⁻³	1.04x10 ⁻²
S100A8	-2.13	-3.00	5.75 x10 ⁻³	1.71x10 ⁻²
BCL2A1	-1.62	-1.91	6.88 x10 ⁻³	1.91x10 ⁻²
RPL27*	-1.23	9.44	3.45x10 ⁻²	2.03 x10 ⁻¹⁵
SEC11C*	-1.47	2.52	9.34 x10 ⁻³	3.26x10 ⁻²

* Genes with different direction of expression in the two trials (see figure 3.7 and figure 3.8 for plots)

The relative expression of these overlapping genes is depicted below in figure 3.7 and figure 3.8 by plotting the changes in FPKM values for Versilia Hospital trial and log₂

count values for the BCH trial. Both $\log_2$ count and FPKM values are normalised for library size and should give a clearer picture of the data.

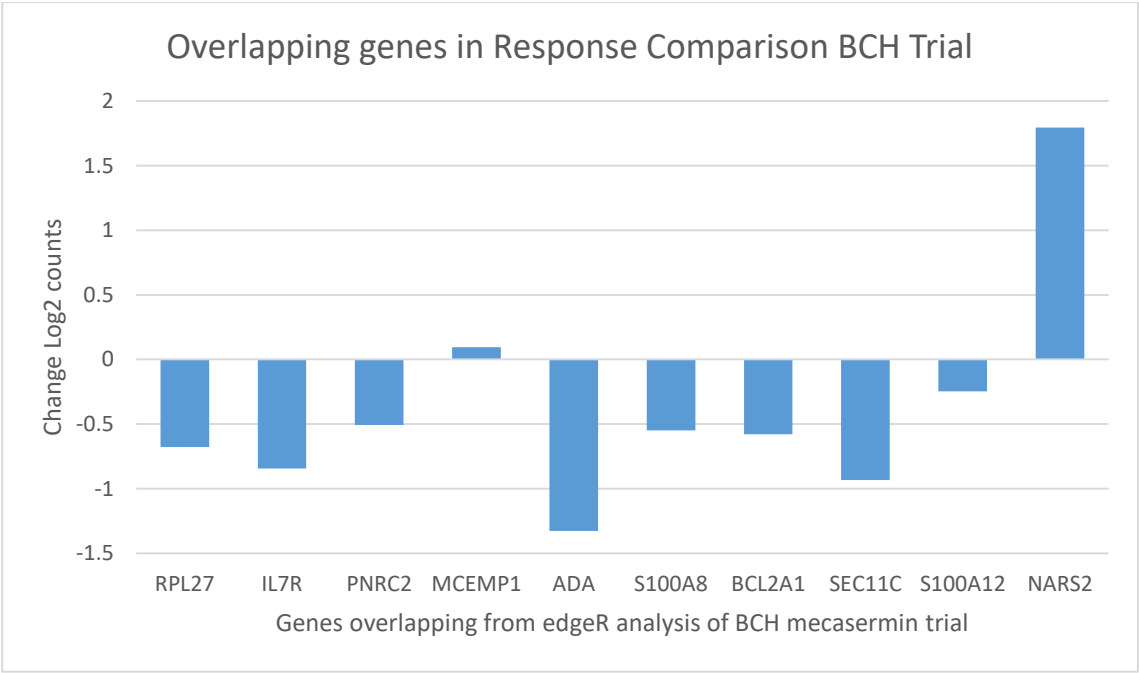
Figure 3.7 - Expression of overlapping genes from response comparison based on FPKM values of Versilia Hospital mecaseimerin trial. Positive values on this graph are genes with higher expression in major responders and lower expression in minor responder (vice versa for negative values). Data is distorted from S100A8 which is clearly a large outlier in expression.



In the case of the expression data from Versilia Hospital trial (figure 3.7 above) all FPKM data matched the direction of fold change described in table 3.3. However the direction of $\log_2$ fold change does not match for five genes in the BCH trial namely *MCEMP1*, *IL7R*, *ADA*, *RPL27* and *SEC11* (see figure 3.8 below). These genes were excluded as the data did agree and therefore were more likely false positives.

These results provide us with interesting candidates associated with positive clinical response to treatment across both trials. As we have small sample numbers further study in clinical trials, animal or in vitro models are required to determine whether these genes could have predictive relationship with positive response to IGF-1 treatment.

Figure 3.8 - Expression of overlapping genes from response comparisons based on log₂ counts values of BCH mecasermin trial. Positive values on this graph are genes with higher expression in major responders and lower expression in minor responder (Vice versa for negative values). Importantly though MCEMP1 was identified to be downregulated in edgeR analysis in this graph we see that it is upregulated.



Chapter Four: Discussion

4.1 Main findings

We summarise the main findings from our analysis in whole blood samples of these two clinical trials below and each point is then discussed in more detail:

1. The strongest expression changes of the Versilia Hospital trial were the large decreases in expression of mitochondrially functioning genes *MT-ATP6* and *MT-CO2* when comparing samples before and after mecasermin treatment.
 2. When considering total RTT patients in the BCH trial, there was no change in expression profiles (DGE) during mecasermin treatment phases (MAD and OLE).
 3. When stratifying RTT patients based on beneficial response to mecasermin treatment there were a number of immune functioning genes differentially expressed in responders (decreased in apnoeic index to <five events per hour) and non-responders (no significant change to apnoeic index).
 - i) In responders *HLA-DRB5* was initially decreased at start of MAD phase to the end of wash out period and then increased in the OLE phase.
 - ii) In non-responders *TMEM176A* and *TMEM176B* were both decreased in the wash out period between treatment phases.
 4. When comparing the expression profiles (DGE) in response based comparisons we found five genes expressed in both studies in the same direction. These comparisons were responders (decreased in apnoeic index to <five events per hour) versus non-responders (no significant change in apnoeic index) in BCH trial and major responders (decrease in PBZ score $\geq$ three from beginning to end of treatment) versus minor responders (decrease in PBZ score $\leq$ one from beginning to end of treatment) in Versilia hospital. These genes were *S100A8*, *S100A12*, *NARS2*, *PNRC2* and *BCL2A1*.
1. As part of our first hypothesis (see section 1.11) we compared the DGE of whole blood samples before and after mecasermin treatment in both trials. *MT-ATP6* and *MT-CO2* genes were strongly decreased in the short and long term effects of mecasermin respectively. They were found significantly expressed with both edgeR and DESeq2 indicating that these findings were particularly reliable. The fact that both short and long

term comparisons found strong decreases to these genes imply a sustained effect on mitochondrial functioning gene expression resulted from treatment.

MT-ATP6 has been associated with Leigh Syndrome and neurogenic muscle weakness, ataxia, retinitis pigmentosa (NARP). In the context of RTT there are an intriguing set of parallels across these conditions. Leigh syndrome is a developmental disorder characterised by early postnatal onset between three and 12 months, motor deterioration and regression through hypotonia, spasticity, cerebellar ataxia and or neuropathies. The severity Leigh syndrome results in the 50% mortality rate in patients under three (Thorburn et al. 2017). NARP is relatively more stable condition compared with Leigh Syndrome (Rojo et al. 2006) and therefore patients have longer survive rates. However they suffer proximal muscle weakness and ataxia as well as vision problems and learning difficulties (Thorburn et al. 2017.). Both RTT and Leigh syndrome have early postnatal onset, loss of acquired skills and progressive decline due to cerebellar ataxia and neuropathy. The also applies to RTT and NARP except for early postnatal onset.

MT-CO2 was found to be downregulated in the long term effects of IGF-1 comparison. This gene encodes a subunit of the cytochrome C oxidase protein. *MT-CO2* mutations are rare but have been associated with MELAS a disorder characterised by vomiting and seizure episodes, recurrent stroke-like symptoms such as hemiparesis and cortical blindness (Kobayashi et al. 1991; Kytovuori et al. 2017).

2. In the BCH trial comparisons using total patients (n = 9) we found there was no change in expression profile of patients during either treatment phases (MAD or OLE) or when comparing across the whole trial from BCH-t0 to BCH-t3. The main difference between the two trials was the additional MAD and wash out phases in BCH where patients were treated with increasing doses of treatment to assess pharmacokinetics. The gradual increase in dosage (see figure 2.2) was used in both MAD and OLE phases. In contrast Versilia hospital trial patients received half dosages in the first and last weeks of treatment with full dosage over 18 weeks. This sharper dosage produced greater changes in expression profile (DGE). A possible explanation for this lack of DGE in BCH trial is that prior exposure to mecasermin in the MAD phase and/or the slow increase to full dosage (see figure 2.2) has a diluting effect on the expression changes induced by mecasermin. We conclude that the dosage scheme of mecasermin in Versilia hospital trial has a greater effect on the molecular mechanisms causing DGE, than dosage scheme of mecasermin in BCH trial.

The only two genes found significantly expressed in comparison using all patients (n=9) in the BCH trial were found in the wash out period (BCH-t1 to BCH-t2) when patients were not treated. *Transmembrane protein 176A (TMEM176A)* and *transmembrane protein 176B (TMEM176B)* were both decreased in this time interval in edgeR analysis.

3. When we stratified BCH patients into responders and non-responders as per hypothesis two (see section 1.11), we measured a stronger decrease in *TMEM176A* and *TMEM176B* in non-responding patients (n = 5) in both edgeR and DESeq2 analysis. The stronger decrease and smaller corrected p-values suggest that *TMEM176* changes were more likely a mechanism of non-responders although it was observed in the total comparison too. Comparison in responders (n = 4) did not see a decrease to *TMEM176* expression but did detect significant difference in another immune functioning gene *HLA-DRB5*. *HLA-DRB5* was initially decreased from start of MAD phase to end of wash out phase, and then increased in expression in the OLE phase. *HLA-DRB5* is a MHC II cell receptor whose expression has been associated with attenuation of multiple sclerosis severity (Caillier et al. 2008). Given that we do not have immunological data for the BCH trial it is not possible to rule out intercurrent infection or hypersensitivity reaction from treatment. However we did have blood tests from the Versilia hospital and they did not show any clear sign of immune infection or hypersensitivity reaction.

4. Finally -and perhaps most importantly- we found a large number of genes to be significantly expressed in patients with positive response compared to non-responding patients before mecasermin treatment. As this was not a paired comparison the number of significant genes was high. Therefore we considered only those genes overlapping in both studies. These genes were *S100A12*, *S100A8*, *NARS2*, *PNRC2* and *BCL2A1* and were differentially expressed in positively responding patients (responders and major responders) relative to non-responding patients (non-responders and minor responders). *S100A8* and *S100A12* both encode S100 calcium binding proteins and were decreased in mecasermin responding patients relative to the minor/non-responding patients. Both of these genes have proinflammatory functions and are associated with neurological inflammation (Cho et al. 2016; Wakisaka et al. 2014). *S100A8* has been shown to be increased in microglia three days post traumatic brain injury (TBI) In the human brain (Beschorner et al. 2000). While *S100A12* has been found to be elevated in patients of acute ischemic stroke with poor functional recovery (Wakisaka et al. 2014). These studies indicate that S100 calcium binding proteins could have utility as a biomarker for

neuropathologies. Another overlapping gene *NARS2* had increased expression in responders compared to non-responders. It is also worth mentioning that mutations to *NARS2* have also been shown to cause Leigh syndrome (Simon et al. 2015) like mutations to *MT-ATP6*.

A table summarising the main results from our analyses are shown below in table 4.1.

Table 4.1 - Summary of the main findings from analyses.

Comparison	Gene	Corrected p-value edgeR (DESeq2)	Log ₂ Fold Change edgeR (DESeq2)
VH-t0 to VH-t1 (n = 2)	<i>MT-ATP6</i>	1.24x10 ⁻¹³ (3.56x10 ⁻¹²)	-8.01 (-8.11)
VH-t0 to VH-t2 (n = 3)	<i>MT-CO2</i>	1.44x10 ⁻¹⁴ (7.85x10 ⁻¹¹)	-10.85 (-10.36)
	<i>RAB37</i>	5.02x10 ⁻⁸ (3.44x10 ⁻³)	2.88 (2.91)
	<i>EEF1A1P7</i>	1.35x10 ⁻¹⁰ (1.45x10 ⁻²)	7.25 (6.59)
	<i>PLLP</i>	1.97x10 ⁻⁷ (3.12x10 ⁻²)	7.11 (6.47)
BCH-t1 to BCH-t2 Non-responders (n=5)	<i>TMEM176A</i>	1.07x10 ⁻¹³ (7.79x10 ⁻⁹)	-3.64 (-3.67)
	<i>TMEM176B</i>	1.6 x10 ⁻¹⁶ (7.93x10 ⁻¹¹)	-3.56 (-3.57)
BCH-t1 to BCH-t2 Total (n=9)	<i>TMEM176A</i>	1.3 x10 ⁻⁴	-2.72
	<i>TMEM176B</i>	5.69 x10 ⁻⁴	-2.52
BCH-t0 to BCH-t2 Responders (n=4)	<i>HLA-DRB5</i>	1.24x10 ⁻³	-2.33
BCH-t2 to BCH-t3 Responders (n=4)	<i>HLA-DRB5</i>	1.51x10 ⁻⁵	3.06

Overlapping Responders vs Non-responders*	<i>S100A12</i>	1.89X10 ⁻²	-2.22
	<i>NARS2</i>	3.4X10 ⁻²	3.38
	<i>PNRC2</i>	2.3X10 ⁻²	-2.2
	<i>S100A8</i>	1.14X10 ⁻²	-2.57
	<i>BCL2A1</i>	1.3X10 ⁻²	-1.77

*For *Versilia hospital* comparison this was major responders versus minor responders see section 2.3. Corrected p-values and Log₂fold changes were average values between BCH and *Versilia hospital* trials.

4.2 Context of Literature

This is a novel study examining the differential gene expression of RTT blood in two mecamermin clinical trials and there is not an extensive literature to contextualise our results. Previous work in *Mecp2* KO mice model of RTT showed partial reversal of many RTT symptoms including improvements to spine density, brain weight, survival, motor activity, breathing variability and heart rate (Tropea et al. 2009; Castro et al. 2014). As set out in the literature review there are common mechanisms of RTT pathology identified in transcriptomic analysis of RTT tissues that may be informative to our analysis. These common mechanisms were abnormal mitochondrial functioning, abnormal synaptic functioning and glial cell activation. We will examine how these mechanisms and other evidence in the literature apply to our findings.

Even before finding the association between *MECP2* and RTT researchers noted mitochondrial dysfunction in cases of RTT (Cornford et al. 1994; Eeg-Olofsson et al. 1989; Dotti et al. 1993). More significant evidence has been gathered on the link between oxidative stress (OS) and mitochondrial dysfunction in RTT pathogenesis using OS markers and gene expression analysis (Leoncini et al. 2011; Lintas et al. 2012; Müller & Can 2014). De Felice and colleagues demonstrated that OS biomarker (F₄-neuroprostanes) was increased by hypoxic events associated with typical RTT patient (De Felice et al. 2009; Signorini et al. 2011).

Numerous findings in RTT mice models have also suggested mitochondrial dysfunctions including increases to complex III gene *Ubiquinol-cytochrome c reductase core protein 1* (*Uqcrc1*) and vulnerability to ROS in the hippocampus (Kriaucionis et al. 2006; Großer et al. 2012). De Filippis and colleagues found that mitochondrial over production

of ROS resulted from dysfunctions in mitochondrial complex II in *MECP2*-308 heterozygous female. These mice had reduced levels of ATP in the cerebellum (De Filippis et al. 2015). These studies demonstrate how individual deficiencies in mitochondrial genes can impact RTT at a functional level. Interestingly IGF-1 has also been found to reverse antioxidant deficiency and increased OS in rats (García-Fernández et al. 2008).

In case control experiments on RTT blood lymphomonocytes increased mitochondrial functioning genes such as COX complex genes and *NDUFA1* were increased (Pecorelli et al. 2013). In our analysis *MT-CO2* (a subunit of COX) and *NDUFA1* were significantly expressed in long term effects comparison of Versilia hospital. *MT-CO2* is decreased by mecasermin treatment however *NDUFA1* was increased so we cannot say simply mecasermin mechanisms correct abnormally high mitochondrially functioning genes. Clearly though mecasermin appears to have a regulatory effect on mitochondrial functioning genes. This relationship is further complicated by the observation that *MT-CO1* (subunit 1 of COX) was increased in RTT frontal cortex which was biologically validated in a SH-SY5Y cell lines. Indeed this deficiency was strong enough to effect COX enzyme activity (Gibson et al. 2010). Understanding the exact role of these decreases in mitochondrial genes (*MT-ATP* and *MT-CO2*) in attenuating RTT phenotype requires to be further investigation.

The role of immune response in RTT has grown recently (De Felice et al. 2016) and it could be a promising area for new therapeutics. Mutations in *MECP2* have been shown to decrease mitogenic stimulation in T cell cultures from RTT patients (Balmer et al. 2002). Although not thought to adversely affect immune response later findings indicate subclinical inflammation in RTT blood which was related to mutation severity (Cortelazzo, De Felice, Pecorelli, et al. 2014). In *MECP2* deficient mice cultured media from microglia effected the wild type neurons producing an aberrant dendritic morphology (Maezawa & Jin 2010). Glial cells are key components of both innate and adaptive response in the CNS and therefore the involvement of glial cells in RTT links immune response directly to the central pathology.

TMEM176A and *TMEM176B* were both decreased in the wash-out period between treatment phases of total and non-responder comparisons. These genes are associated with maintenance of immature dendritic cells (Condamine et al. 2010) and have even been found to have roles in schizophrenia and Alzheimer's disease (Melchior et al. 2010;

Kos et al. 2017). In relation to the treatment, IGF-1 has been shown to modulate dendritic cell function in human cord blood by promoting maturation and survival through MAPK and PI3K pathways (Liu et al. 2003). While IGF-2 KO mice were found to increase the proportion of immature dendritic cells and decrease proliferation of T-cells (Hansenne et al. 2006). From these studies we can postulate that treatment may be increasing maturity of dendritic cells like with IGF-1 administration in human cord blood. More experiments are required to confirm this and to distinguish whether this mechanism contributes to treatment response in RTT or not.

The swing of expression in *HLA-DRB5* from decreased (MAD and wash out phase) to increased (OLE) in responders represents a distinct immune modulation in RTT patients that had significant improvements from treatment. From the swing in expression we could hypothesise that these changes are dose dependent. Ethnicity can also be a factor in MHC II diversity however this is largely controlled for in our analysis. All of the Versilia Hospital patients are Caucasian and all but two in the BCH trial are Caucasian, the remaining two are Asian (approximately 85% Caucasian and 15% Asian across both studies). To understand the full mechanism of how MHC II function relates to mecamermin further investigation is needed.

The overlapping genes *S100A8*, *S100A12*, *NARS2*, *PNRC2* and *BCL2A1* found in response based comparison for both trials provide us with preliminary associative genes to test for predicative relationship with response to mecamermin in RTT. As we only assessed the association of genes with positive response an independent clinical trial would need to establish if these genes can actually predict mecamermin response in RTT patients assessing specificity and sensitivity. To do this an adequately powered trial would need to be designed and analysed using predictive models with a clear set of clinical parameters and end points. Such a study should include desired response outcomes and thresholds for expressions. This means that we are some way off confirming these genes as biomarkers for RTT but these are interesting preliminary data identified in two independent studies that warrant further study.

It is also worth noting that S100 calcium binding proteins are encoded by *S100A8* and *S100A12*. These proteins/genes have already been found to mark the neurological damage induced from TBI in the case of *S100A8* and acute ischemic stroke in the case of *S100A12*. *S100A8* expression in peripheral blood was found in to be significantly increased in older patients suffering TBI 48hr post injury. *S100A12* protein plasma

levels in blood were higher in patients with poorer recovery from ischemic stroke (Cho et al. 2016; Wakisaka et al. 2014). In fact a partially deficient IGF-1 mouse model has shown that both *S100A9* and *S100A8* expression was increased with treatment by IGF-1 which reverses oxidative brain damage and cognitive functioning (Puche et al. 2016). Although our results show *S100A8* and *S100A12* were decreased in positively responding patients these studies support the idea that genes may have utility in marking patients more likely to have benefit from mecasermin treatment. However these associations need to be further validated in a predictive model

4.3 Limitations

It is important to consider the very strong limiting factors of this analysis. These limitations either lower statistical power or make the results more difficult to interpret. Statistical power is inversely proportional to the probability of making a type II error in other word finding false negatives. The strongest influencer of statistical power in this analysis was sample size. Though a power calculation was not done for these studies they were certainly underpowered for standard statistics tests considering low sample numbers. However it is important to recognise that the software used to analyse DGE (edgeR and DESeq2) have been specifically designed to work reliably even with very small sample sizes. In fact Robinson and Smyth state “Not only is our strategy applicable even with the smallest number of libraries, but it also proves to be more powerful than previous strategies when more libraries are available.” (Robinson & Smyth 2007; Love et al. 2014).

The data from the miSeq experiment in Versilia Hospital trial was primarily used to determine the quality of the library preparation from whole blood RNA. Although the library was of high quality, its purpose was to assess if samples were worth sequencing using the higher specification Illumina HiSeq. As we cannot change sample numbers an increase in the sequencing depth would theoretically increase power of future experiments but not to the same extent as having greater sample numbers (Ching et al. 2014). This higher sequencing depth could be achieved with HiSeq sequencing. Generally real time-PCR techniques are also required to confirm differential gene expression results in RNA sequencing experiments. This has not been done in our analysis but should be considered for future validation of all genes.

As with most studies of rare diseases, RTT populations are small and heterogeneous. This makes it difficult to design and recruit clinical trials (Katz et al. 2016). The fact that

RTT patients cannot communicate themselves and most clinical endpoints are poorly defined further complicates interpretation of trial results. This problem of poorly defined endpoints is illustrated in our analysis with the use of different severity scoring systems ISS, RSS and PBZ to detect improvement. The PBZ score is based on similar criteria as ISS but with an assessment of daily living tasks included. ISS and PBZ were both used in the Versilia hospital trial. Although showing similar reliability and sensitivity as ISS, PBZ score requires a proper clinical trial validation to be able to use it internationally (Giorgio Pini, Congiu, Zappella, et al. 2016). Arbitrary clinical endpoints can be hard to avoid in neuropsychiatric severity scoring as assessment will be generally be done by a clinician, carer or family member in trials.

Whole blood gene expression in general has implicit problems too. Heterogeneity of sample composition and relatively high expression of non-informative globin mRNA (Min et al. 2010) are unavoidable. In this analysis the non-informative globin mRNA genes were removed from raw counts before DGE analysis was performed. This removed the need for globin removal steps to RNA processing which can often effect the concentration of RNA output. Although this method does limit the detection of genes it can be potentially overcome with better coverage available in higher through-put sequencing (Shin et al. 2014).

One specific aspect of the Versilia Hospital trial to consider was the long term effects of IGF-1. This comparison had time points after treatment at three varied time intervals (see table 2.3). Although IGF-1 has many long term effects including development, as outlined in section 1.5, the length of time since treatment makes it less likely DGE was caused by mecaseprin treatment. That being said both short and long term comparisons found mitochondrial genes to be strongly decreased indicating a prolonged effect on mitochondrial functioning. IGF-1 has long term effects on growth, survival and neurodevelopment (Chernausek et al. 2007; Laron 2008; Wrigley et al. 2017) so a long term comparison is still clinically relevant to treatment.

Study design for both trials was heterogeneous but with many similar aspects. We will now discuss those similarities and differences. BCH trial had a preliminary MAD phase, wash out period and open label extension, while Versilia Hospital trial structure was analogous to the open label extension. Both trial had similar full dosage 120 µg/kg in BCH and 100 µg/kg in Versilia hospital trials. However BCH used a gradual stepped increase over three weeks until maximum dose (see figure 2.2) and Versilia hospital trial had one week at half dose before reaching maximum dose.

Versilia Hospital trial cohort was made up of entirely Caucasian patients while BCH trial contained two Asian patients with the rest being Caucasian. This bias towards Caucasian trial participants is another limitation of our analysis. An epidemiological study of RTT with a large database showed no bias in RTT prevalence in white, black or Hispanic populations (Kozinetz et al. 1993) yet our cohort was for the majority Caucasian with no black or Hispanic representation. The patient cohorts had similar mean age, 5 years old in Versilia Hospital trial and 5.78 years in BCH trial. Two patients in the BCH were stage II (developmental regression) while all others RTT patients in the trials were stage III (pseudostationary phase). Even though the age profiles are similar the non-linear progression of RTT introduces further complications because the pseudostationary stage can see some return of communication and or apparent slowing of the motor decline (Hagberg 2002). So even though we were able to remove the confounding factor of trial structure by only comparing untreated samples at VH-t0 and BCH-t0 in response based comparison, there were still key differences in the patient cohorts that could potentially effect these results.

Despite the differences there were many similarities including diagnosis of patients, gender, presence of *MECP2* mutation, similar maximum dose of mecamsermin, similar treatment time period (20 weeks total in OLE and Versilia hospital trial), with similar age profiles and both trials showing mixed clinical response to treatment. All things considered we conclude that a comparative analysis could have genuine merit in understanding the relationship between mecamsermin treatment and positive clinical response, but further experimentation and validation, such as RT-PCR, are also required.

One of the most difficult challenges for RTT research is finding non-arbitrary biomarkers for measuring clinical endpoints in patients. More integrative ways of assessing RTT patients based on meaningful measures and biomarkers will be essential for translating preclinical findings to approved treatments. We will discuss an example of such an approach in different disorder in the subsequent section.

4.4 Future Directions

Recently a phase II clinical trial for mecamsermin has been completed and found no significant improvement in severity scales such as ADAMS, RSBQ, parent targeted symptom visual analogue scale (PTSVAS), clinical global impression (CGI), parent global impression and Kerr severity score. The trial did however find significant improvements to stereotypic behaviour and social communication subscales (O'Leary et

al. 2018). O’Leary and colleagues identified a number of possible contributing factors to the mainly negative results. These factors were low statistical power, inadequate dosage, placebo effects and carry forward effects. Due to the limitations of the trial and evidence of significant secondary improvements they conclude that there still may be value to mecasermin treatment in RTT. Interestingly they used the same dosing scheme as was used in the OLE of BCH trial for treatment. Accessing blood samples from this study would provide an ideal setting for further validation of our results. However access to such resources requires a huge amount of work and cooperation.

The compound Trofinetide (NNZ-2566) which is analogue of the (1-3) IGF-1 tripeptide is also under assessment in a phase II clinical trial with RTT patients showing tolerability of treatment and clinically meaningful improvements in a number of features of the disorder including communication, behaviour breathing abnormality motor function and seizures activity (Glaze et al. 2017). Analysis of these samples and data too could help strengthen our findings.

Biomarkers other than genetic expression factors could also be beneficial in understanding RTT pathophysiology. Recently Keogh and colleagues used EEG to differentiate between RTT clinical subgrouping (classic, Hanefeld and PSV) based on interelectrode coherence and network architecture (Keogh et al. 2018). Eye tracking and visual paired-comparison has been used to investigate the cognitive phenotype of RTT patients (Djukic et al. 2012; Rose et al. 2013). The use of multiple biomarker paradigms in RTT especially non-invasive forms like genetic blood based tests, EEG or eye-tracking could have great utility.

A central problem in RTT research appears to be scarcity of samples that leads to underpowered research. Efforts need to be made to pool resources, limit confounding factors and increase statistical power in neurodevelopmental studies in general. This type of analyses requires huge collaboration and cooperation across institutes. The longitudinal European autism project (LEAP) supported by the innovative medicines initiative (IMI) is an example of such a study in ASD. LEAP takes a number of different biomarker paradigms (including functional magnetic resonance imaging, diffusion tensor imaging, EEG, eye-tracking, cognitive tests, biochemical and genomic analysis of blood/saliva and hair samples) and analyses a large enough study cohort (ASD n = 437, control n = 300) to detect significant changes with adequate statistical power (Loth et al. 2017). To use the LEAP model for RTT would be desirable, however we should take into account the disease prevalence and variation in presentation. LEAP does provide a

frame-work to work towards for exploring multiple paradigms at a statistically adequate level.

4.5 Conclusion

Rett Syndrome has been a model system for developing strategies to treat neuropsychiatric disorder. Its monogenetic nature has allowed the disorder to be mimicked in murine models and has led to the development of many promising therapeutic drugs. Despite promising findings in preclinical models researchers have yet to translate any of these to effective human treatments.

MECP2, the gene strongly associated with Rett, is a global transcriptional regulator and therefore gives rise to a nuanced pathophysiology (Ehrhart et al. 2016). The involvement of features like metabolic imbalance, mitochondrial dysfunction and oxidative stress as well as the primary neuronal pathologies have led to alternative therapeutic strategies for RTT.

In this project we measured DGE in two independent open label clinical trials testing safety and preliminary efficacy of human recombinant IGF-1 (mecasermin). We hypothesised that there would be differences to patient expression profiles (DGE) after treating with mecasermin. To measure this DGE we performed sequencing analysis techniques with two software packages, edgeR and DESeq2 specifically designed for experiments with small sample size. The analysis identified strong decreases in expression of *MT-ATP6* and *MT-CO2* in Versilia Hospital trial. The *TMEM176A* and *TMEM176B* genes function in immature dendritic cells were significantly decreased in the BCH trial between treatment phases when using total patients (n = 9) and non-responding patients (n = 5). Responder comparisons (n = 4) showed an initial decrease in *HLA-DRB5* in MAD and wash out period and then an increase during OLE. These genes seem to suggest immune response playing an important role in response to mecasermin.

Secondly we hypothesised that expression profiles in patients that responded to mecasermin treatment would be different from expression profile of patients who did not respond to treatment in blood samples before administration of the drug. We found a great many genes significantly expressed in this comparison. When overlapping the results in both studies we identified five genes significantly expressed, *S100A8*, *S100A12*, *NARS2*, *PNRC2* and *BCL2A1* in the same direction. Of these genes two had

previously been identified as potential biomarkers in traumatic brain injury and stroke, *S100A8* and *S100A12*.

Our analysis has further highlighted the role of mitochondrial functioning and immunological functioning genes in RTT patients treated with mecamermin. Additionally it has identified a number of genes potentially involved in positive clinical response to mecamermin in RTT patients. These genes warrant further biological validation possibly through quantitative PCR or high through-put sequencing.

References:

- Ager, S. et al., 2006. Predictors of Scoliosis in Rett Syndrome. *Journal of Child Neurology*, 21(9), pp.809–813. Available at: <http://journals.sagepub.com/doi/10.1177/08830738060210091501>.
- Amir, R.E. et al., 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, 23(october), pp.185–188. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10508514>.
- Ananiev, G. et al., 2011. Isogenic pairs of wild type and mutant induced pluripotent stem cell (iPSC) lines from rett syndrome patients as In Vitro disease model. *PLoS ONE*, 6(9).
- Anderson, A. et al., 2014. Twenty years of surveillance in Rett syndrome: what does this tell us? *Orphanet Journal of Rare Diseases*, 9(1), p.87. Available at: <http://ojrd.biomedcentral.com/articles/10.1186/1750-1172-9-87>.
- Anderson, M.F. et al., 2002. Insulin-like growth factor-I and neurogenesis in the adult mammalian brain. *Brain Res Dev Brain Res*, 134(1–2), pp.115–122. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11947942.
- Archer, H. et al., 2007. Correlation between clinical severity in patients with Rett syndrome with a p.R168X or p.T158M MECP2 mutation, and the direction and degree of skewing of X-chromosome inactivation. *Journal of medical genetics*, 44(2), pp.148–152.
- Archer, H.L. et al., 2006. CDKL5 mutations cause infantile spasms, early onset seizures, and severe mental retardation in female patients. *Journal of Medical Genetics*, 43(9), pp.729–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16611748%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/16611748?ordinalpos=4&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum.
- Asmita, K., 2008. Analysis of protein domains and RTT mutations indicate that multiple regions influence chromatin binding dynamics of the chromatin associated protein MECP2 in vivo. *Cell Journal*, 6(5), pp.760–769.
- Ballas, N. et al., 2009. Non-cell autonomous influence of MeCP2-deficient glia on

- neuronal dendritic morphology. *Nature Neuroscience*, 12(3), pp.311–317. Available at: <http://www.nature.com/doifinder/10.1038/nn.2275>.
- Ballestar, E., Yusufzai, T.M. & Wolffe, A.P., 2000. Effects of rett syndrome mutations of the Methyl-CpG binding domain of the transcriptional repressor MeCP2 on selectivity for association with methylated DNA. *Biochemistry*, 39(24), pp.7100–7106.
- Balmer, D. et al., 2002. MECP2 mutations in Rett syndrome adversely affect lymphocyte growth, but do not affect imprinted gene expression in blood or brain. *Human Genetics*, 110(6), pp.545–552.
- Baron-Van Evercooren, A. et al., 1991. Expression of IGF-1 and insulin receptor genes in the rat central nervous system: a developmental, regional and cellular analysis. *Journal of Neuroscience Research*, 28, pp.244–253.
- Barres, B.A. et al., 1992. Cell death and control of cell survival in the oligodendrocyte lineage. *Cell*, 70(1), pp.31–46.
- Bebbington, A. et al., 2012. The phenotype associated with a large deletion on MECP2. *European Journal of Human Genetics*, 20(9), pp.921–927. Available at: <http://www.nature.com/doifinder/10.1038/ejhg.2012.34>.
- Ben-Shachar, S. et al., 2009. Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Human Molecular Genetics*, 18(13), pp.2431–2442.
- Benjamini, Y. & Hochberg, Y., 1995. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. , 57(1), pp.289–300.
- Beschorner, R. et al., 2000. Differential regulation of the monocytic calcium-binding peptides macrophage-inhibiting factor related protein-8 (MRP8/S100A8) and allograft inflammatory factor-1 (AIF-1) following human traumatic brain injury. *Acta Neuropathologica*, 100(6), pp.627–634.
- Bi, W. et al., 2002. Genes in a refined Smith-Magenis syndrome critical deletion interval on chromosome 17p11.2 and the syntenic region of the mouse. *Genome Research*, 12(5), pp.713–728.
- Bodmer, J.G. et al., 1994. Nomenclature for factors of the HLA system, 1994. *Tissue Antigens*, 44, pp.1–18.

- Bondy, C. & Lee, W.H., 1993. Correlation between insulin-like growth factor (IGF)-binding protein 5 and IGF-I gene expression during brain development. *J Neurosci*, 13(12), pp.5092–5104. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7504720.
- Bondy, C.A. & Cheng, C.M., 2004. Signaling by insulin-like growth factor 1 in brain. *European Journal of Pharmacology*, 490(1–3), pp.25–31.
- Braunschweig, D. et al., 2004. X-chromosome inactivation ratios affect wild-type MeCP2 expression within mosaic Rett syndrome and Mecp2-/+ mouse brain. *Human Molecular Genetics*, 13(12), pp.1275–1286.
- Caillier, S.J. et al., 2008. Uncoupling the roles of HLA-DRB1 and HLA-DRB5 genes in multiple sclerosis. *Journal of immunology (Baltimore, Md. : 1950)*, 181(8), pp.5473–80. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4346327&tool=pmcentrez&rendertype=abstract>.
- Cancedda, L., 2004. Acceleration of Visual System Development by Environmental Enrichment. *Journal of Neuroscience*, 24(20), pp.4840–4848. Available at: <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0845-04.2004>.
- Carney, R.M. et al., 2003. Identification of MeCP2 mutations in a series of females with autistic disorder. *Pediatric Neurology*, 28(3), pp.205–211.
- Carson, M.J. et al., 1993. Insulin-like growth factor I increases brain growth and central nervous system myelination in tTransgenic mice. *Neuron*, 10(4), pp.729–740. Available at: <http://www.sciencedirect.com/science/article/pii/0896627393901730>.
- Castro, J. et al., 2014. Functional recovery with recombinant human IGF1 treatment in a mouse model of Rett Syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 111(27), pp.9941–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24958891>.
- Chahrour, M. et al., 2008. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science (New York, N.Y.)*, 320(5880), pp.1224–9. Available at: <http://www.sciencemag.org/content/320/5880/1224.short>.
- Chen, L. et al., 2015. MeCP2 binds to non-CG methylated DNA as neurons mature,

- influencing transcription and the timing of onset for Rett syndrome. *Proceedings of the National Academy of Sciences*, 112(17), p.201505909. Available at: <http://www.pnas.org/content/112/17/5509.full>.
- Chen, R.Z. et al., 2001. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nature genetics*, 27(3), pp.327–331.
- Chernausk, S.D. et al., 2007. Long-term treatment with recombinant insulin-like growth factor (IGF)-I in children with severe IGF-I deficiency due to growth hormone insensitivity. *Journal of Clinical Endocrinology and Metabolism*, 92(3), pp.902–910.
- Cheung, A.Y.L. et al., 2011. Isolation of MECP2-null Rett Syndrome patient hiPS cells and isogenic controls through X-chromosome inactivation. *Human Molecular Genetics*, 20(11), pp.2103–2115.
- Cheung, A.Y.L. et al., 2012. X-chromosome inactivation in Rett syndrome human induced pluripotent stem cells. *Frontiers in Psychiatry*, 3(MAR), pp.1–16.
- Ching, T., Huang, S. & Garmire, L.X., 2014. Power analysis and sample size estimation for RNA-Seq differential expression. *Rna*, pp.0–13.
- Cho, Y.E. et al., 2016. Older age results in differential gene expression after mild traumatic brain injury and is linked to imaging differences at acute follow-up. *Frontiers in Aging Neuroscience*, 8(JUN), pp.1–11.
- Cifu, D. et al., 2009. Clinical practice guideline: Management of Concussion/Mild Traumatic Brain Injury. *The Journal of Rehabilitation Research and Development*, 46(6), p.CP1. Available at: <http://www.rehab.research.va.gov/jour/09/46/6/pdf/cpg.pdf>.
- Ciucci, F. et al., 2007. Insulin-like growth factor 1 (IGF-1) mediates the effects of enriched environment (EE) on visual cortical development. *PLoS ONE*, 2(5).
- Colak, D. et al., 2011. Genomic and transcriptomic analyses distinguish classic Rett and Rett-like syndrome and reveals shared altered pathways. *Genomics*, 97(1), pp.19–28. Available at: <http://dx.doi.org/10.1016/j.ygeno.2010.09.004>.
- Colantuoni, C. et al., 2001. Gene expression profiling in postmortem Rett Syndrome brain: differential gene expression and patient classification. *Neurobiology of disease*, 8(5), pp.847–65. Available at:

<http://www.sciencedirect.com/science/article/pii/S0969996101904280>.

- Colvin, L. et al., 2004. Refining the phenotype of common mutations in Rett syndrome. *Journal of medical genetics*, 41(1), pp.25–30.
- Condamine, T. et al., 2010. Tmem176B and Tmem176A are associated with the immature state of dendritic cells. *Journal of leukocyte biology*, 88(3), pp.507–515.
- Cornford, M.E. et al., 1994. Neuropathology of Rett Syndrome: Case Report With Neuronal and Mitochondrial Abnormalities in the Brain. *Journal of Child Neurology*, 9(4), pp.424–431. Available at: <http://jcn.sagepub.com/cgi/doi/10.1177/088307389400900419>.
- Cortelazzo, A., De Felice, C., Pecorelli, A., et al., 2014. Beta-actin deficiency with oxidative posttranslational modifications in rett syndrome erythrocytes: Insights into an altered cytoskeletal organization. *PLoS ONE*, 9(3).
- Cortelazzo, A., De Felice, C., Guerranti, R., et al., 2014. Subclinical inflammatory status in Rett syndrome. *Mediators of Inflammation*, 2014.
- Costa, V. et al., 2012. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. *European Journal of Human Genetics*, 21(10), pp.134–142. Available at: <http://www.nature.com/doi/10.1038/ejhg.2012.129>.
- Coy, J.F. et al., 1999. A complex pattern of evolutionary conservation and alternative polyadenylation within the long 3′-untranslated region of the methyl-CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene expression. *Human molecular genetics*, 8(7), pp.1253–62. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10369871>.
- Crosson, J. et al., 2017. Evaluation of QTc in Rett syndrome: Correlation with age, severity, and genotype. *American Journal of Medical Genetics, Part A*, 173(6), pp.1495–1501.
- Cuddapah, V.A. et al., 2014. Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in rett syndrome. *Journal of Medical Genetics*, 51(3), pp.152–158.
- D’Ercole, J.A., Ye, P. & O’Kusky, J.R., 2002. Mutant mouse models of insulin-like growth factor actions in the central nervous system. *Neuropeptides*, 36(2–3), pp.209–220. Available at:

<http://linkinghub.elsevier.com/retrieve/pii/S0143417902908930>.

- Deng, V. et al., 2007. FXRD1 is an MecP2 target gene overexpressed in the brains of Rett syndrome patients and Mecp2-null mice. *Human Molecular Genetics*, 16(6), pp.640–650.
- Derr, A. et al., 2016. End Sequence Analysis ToolKit (ESAT) expands the extractable from single cell RNA-seq experiments. *Genome research*, pp.1397–1410. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27470110>.
- Dhindsa, R.S. et al., 2015. Epileptic encephalopathy-causing mutations in *DNM1* impair synaptic vesicle endocytosis. *Neurology Genetics*, 1(1), p.e4. Available at: <http://ng.neurology.org/lookup/doi/10.1212/01.NXG.0000464295.65736.da>.
- Djukic, A. et al., 2012. Rett syndrome: Basic features of visual processing - A pilot study of eye-tracking. *Pediatric Neurology*, 47(1), pp.25–29. Available at: <http://dx.doi.org/10.1016/j.pediatrneurol.2012.04.009>.
- Dolce, A. et al., 2013. Rett syndrome and epilepsy: An update for child neurologists. *Pediatric Neurology*, 48(5), pp.337–345. Available at: <http://dx.doi.org/10.1016/j.pediatrneurol.2012.11.001>.
- Dotti, M.T. et al., 1993. Mitochondrial dysfunction in Rett syndrome. An ultrastructural and biochemical study. *Brain & development*, 15(2), pp.103–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8214327>.
- Downs, J. et al., 2016. How can clinical ethics guide the management of comorbidities in the child with Rett syndrome? *Journal of Paediatrics and Child Health*, 52(8), pp.809–813.
- Downs, J. et al., 2015. The Natural History of Scoliosis in Females with Rett Syndrome. *Spine*, 41(10), pp.856–863.
- Draghici, S. et al., 2007. A systems biology approach for pathway level analysis. *Genome Research*, pp.1537–1545.
- Eeg-Olofsson, O. et al., 1989. Rett syndrome: Genetic clues based on mitochondrial changes in muscle. *American Journal of Medical Genetics*, 32(1), pp.142–144. Available at: <http://doi.wiley.com/10.1002/ajmg.1320320131>.
- Ehrhart, F. et al., 2016. Rett syndrome - biological pathways leading from MECP2 to disorder phenotypes. *Rare Diseases*, pp.1–21. Available at:

<http://www.biorxiv.org/content/early/2016/07/08/062786.abstract>.

- Evans, J.C. et al., 2005. Early onset seizures and Rett-like features associated with mutations in CDKL5. *European journal of human genetics : EJHG*, 13(10), pp.1113–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16015284>.
- Fehr, S. et al., 2011. Trends in the diagnosis of rett syndrome in Australia. *Pediatric Research*, 70(3), pp.313–319.
- De Felice, C. et al., 2014. Oxidative brain damage in Mecp2-mutant murine models of Rett syndrome. *Neurobiology of Disease*, 68, pp.66–77. Available at: <http://dx.doi.org/10.1016/j.nbd.2014.04.006>.
- De Felice, C. et al., 2016. Rett syndrome: An autoimmune disease? *Autoimmunity Reviews*, 15(4), pp.411–416. Available at: <http://dx.doi.org/10.1016/j.autrev.2016.01.011>.
- De Felice, C. et al., 2009. Systemic oxidative stress in classic Rett syndrome. *Free Radical Biology and Medicine*, 47(4), pp.440–448. Available at: <http://dx.doi.org/10.1016/j.freeradbiomed.2009.05.016>.
- Fernandez-Moreira, D. et al., 2007. X-linked NDUFA1 gene mutations associated with mitochondrial encephalomyopathy. *Annals of Neurology*, 61(1), pp.73–83.
- De Filippis, B. et al., 2015. Mitochondrial free radical overproduction due to respiratory chain impairment in the brain of a mouse model of Rett syndrome: Protective effect of CNF1. *Free Radical Biology and Medicine*, 83, pp.167–177. Available at: <http://dx.doi.org/10.1016/j.freeradbiomed.2015.02.014>.
- De Filippis, B., Ricceri, L. & Laviola, G., 2010. Early postnatal behavioral changes in the Mecp2-308 truncation mouse model of Rett syndrome. *Genes, Brain and Behavior*, 9(2), pp.213–223.
- Forsythe, E. & Beales, P.L., 2013. Bardet–Biedl syndrome. *European Journal of Human Genetics*, 21(1), pp.8–13. Available at: <http://www.nature.com/doifinder/10.1038/ejhg.2012.115>.
- Fromer, M. et al., 2016. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience*, 19(11), pp.1442–1453. Available at: <http://www.nature.com/doifinder/10.1038/nn.4399>.
- García-Fernández, M. et al., 2008. Low doses of insulin-like growth factor I improve

- insulin resistance, lipid metabolism, and oxidative damage in aging rats. *Endocrinology*, 149(5), pp.2433–2442.
- Gibson, J.H. et al., 2010. Downstream targets of methyl CpG binding protein 2 and their abnormal expression in the frontal cortex of the human Rett syndrome brain. *BMC neuroscience*, 11, p.53.
- Glaze, D.G. et al., 2017. A Double-Blind, Randomized, Placebo-Controlled Clinical Study of Trofinetide in the Treatment of Rett Syndrome. *Pediatric Neurology*. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0887899417304058>.
- Glaze, D.G. et al., 2010. Epilepsy and the natural history of Rett syndrome. *Neurology*, 74(11), pp.909–912.
- Greenberg, F. et al., 1991. Molecular Analysis of the Smith-Magenis Syndrome: A Possible Contiguous-Gene Syndrome Associated with del(17)(p11.2). *American Journal of Human Genetics*, 49, pp.1207–1218.
- Großer, E. et al., 2012. Oxidative burden and mitochondrial dysfunction in a mouse model of Rett syndrome. *Neurobiology of Disease*, 48(1), pp.102–114. Available at: <http://dx.doi.org/10.1016/j.nbd.2012.06.007>.
- Guo, J.U. et al., 2014. Distribution, recognition and regulation of non-CpG methylation in adult mammalian brain. , 17(2), pp.215–222.
- Guy, J. et al., 2001. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nature genetics*, 27(3), pp.322–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11242117>.
- Hagberg, B., 2002. Clinical manifestations and stages of Rett syndrome. *Mental Retardation and Developmental Disabilities Research Reviews*, 8(2), pp.61–65.
- Hagberg, B. & Rasmussen, P., 1986. “FORME FRUSTE” OF RETT SYNDROME - A CASE REPORT. *American Journal of Medical Genetics*, 181, pp.175–181.
- Hansenne, I. et al., 2006. Dendritic Cell Differentiation and Immune Tolerance to Insulin-Related Peptides in Igf2-Deficient Mice. *The Journal of Immunology*, 176(8), pp.4651–4657. Available at: <http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.176.8.4651>.
- Hardy, J.J. et al., 2017. Assessing the accuracy of blood RNA profiles to identify patients with post-concussion syndrome: A pilot study in a military patient population. *Plos*

One, 12(9), p.e0183113. Available at:
<http://dx.plos.org/10.1371/journal.pone.0183113>.

Higashio, H., Satoh, Y. & Saino, T., 2016. Mast cell degranulation is negatively regulated by the Munc13-4-binding small-guanosine triphosphatase Rab37. *Scientific reports*, 6(February), p.22539. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/26931073><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4773767>.

Holzenberger, M. et al., 2003. IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature*, 421(6919), pp.182–187. Available at:
<http://www.nature.com/doifinder/10.1038/nature01298>.

Hsu, Y.W. et al., 2004. Dexmedetomidine pharmacodynamics: Part I - Crossover comparison of the respiratory effects of dexmedetomidine and remifentanyl in healthy volunteers. *Anesthesiology*, 101(5), pp.1066–1076.

Johansen, F.E., Braathen, R. & Brandtzaeg, P., 2001. The J chain is essential for polymeric Ig receptor-mediated epithelial transport of IgA. *Journal of immunology (Baltimore, Md. : 1950)*, 167(9), pp.5185–5192.

Jones, P.L. et al., 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature genetics*, 19(june), pp.187–191. Available at:
<http://www.nature.com/doifinder/10.1038/561>.

Julu, 2008. Cardiorespiratory challenges in Rett ' s syndrome. , pp.1981–1983.

Julu, P.O. et al., 1997. Immaturity of medullary cardiorespiratory neurones leading to inappropriate autonomic reactions as a likely cause of sudden death in Rett's syndrome. *Archives of disease in childhood*, 77(5), pp.464–5. Available at:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1717371&tool=pmcentrez&rendertype=abstract>.

Julu, P.O.O. et al., 2001. Characterisation of breathing and associated central autonomic dysfunction in the Rett disorder. , pp.29–37.

Katz, D.M. et al., 2016. Rett Syndrome: Crossing the Threshold to Clinical Translation. *Trends in Neurosciences*, 39(2), pp.100–113.

Kempermann, G., 2008. The neurogenic reserve hypothesis: what is adult hippocampal neurogenesis good for? *Trends in Neurosciences*, 31(4), pp.163–169.

- Keogh, C. et al., 2018. Clinical and genetic Rett syndrome variants are defined by stable electrophysiological profiles. *BMC Pediatrics*, pp.1–15. Available at: <https://bmcpediatr.biomedcentral.com/articles/10.1186/s12887-018-1304-7>.
- Kerr, A.M. et al., 2001. Guidelines for reporting clinical features in cases with MECP2 mutations. *Brain and Development*, 23(4), pp.208–211.
- Khwaja, O.S. et al., 2014a. Safety, pharmacokinetics, and preliminary assessment of efficacy of mecasermin (recombinant human IGF-1) for the treatment of Rett syndrome. *Proceedings of the National Academy of Sciences*, 111(12), pp.4596–4601. Available at: <http://www.pnas.org/cgi/doi/10.1073/pnas.1311141111>.
- Khwaja, O.S. et al., 2014b. Supporting Information: Safety, pharmacokinetics, and preliminary assessment of efficacy of mecasermin (recombinant human IGF-1) for the treatment of Rett syndrome. , (727265), pp.1–8.
- Kim, K. et al., 2010. Epigenetic memory in induced pluripotent stem cells. *Nature*, 467(7313), pp.285–290. Available at: [papers3://publication/doi/10.1038/nature09342](https://doi.org/10.1038/nature09342).
- Kline, D.D. et al., 2010. Exogenous Brain-Derived Neurotrophic Factor Rescues Synaptic Dysfunction in Mecp2-Null Mice. *Journal of Neuroscience*, 30(15), pp.5303–5310. Available at: <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5503-09.2010>.
- Kobayashi, Y. et al., 1991. Respiration-deficient Cells Are Caused by a Single Point Mutation in the Mitochondrial tRNA-Leu (UUR) Gene in Mitochondrial Myopathy , Encephalopathy , Lactic Acidosis , and Strokeliike Episodes (MELAS). , pp.590–599.
- Kos, M.Z. et al., 2017. Exome sequences of multiplex, multigenerational families reveal schizophrenia risk loci with potential implications for neurocognitive performance. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, (March).
- Kozinetz, C.A. et al., 1993. Epidemiology of Rett Syndrome: A Population-Based Registry. *Pediatrics*, 91(2), pp.445–450.
- Kriaucionis, S. et al., 2006. Gene Expression Analysis Exposes Mitochondrial Abnormalities in a Mouse Model of Rett Syndrome. *Molecular and Cellular*

Biology, 26(13), pp.5033–5042. Available at:

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1489175/%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC1489175/pdf/1665-05.pdf>.

Krishnaraj, R., Ho, G. & Christodoulou, J., 2017. RettBASE: Rett Syndrome Database Update. *Human Mutation*, (July 2016), pp.1–10. Available at: <http://doi.wiley.com/10.1002/humu.23263>.

Kron, M. et al., 2012. Brain Activity Mapping in Mecp2 Mutant Mice Reveals Functional Deficits in Forebrain Circuits, Including Key Nodes in the Default Mode Network, that are Reversed with Ketamine Treatment. *Journal of Neuroscience*, 32(40), pp.13860–13872.

Kytovuori, L. et al., 2017. Case report: A novel frameshift mutation in the mitochondrial cytochrome c oxidase II gene causing mitochondrial disorder. *BMC Neurology*, 17(1), p.no pagination. Available at: <http://www.biomedcentral.com/bmcneurol/%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emex&NEWS=N&AN=616235257>.

de Laat, W., Jaspers, N. & Hoeijmakers, J., 1999. Molecular mechanism of nucleotide excision repair. *Gene. Dev*, 13(7), pp.768–785. Available at: <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?dbfrom=pubmed&id=10197977&retmode=ref&cmd=prlinks%5Cnpapers2://publication/uuid/997141EC-928B-413E-B9B7-636F231106CC>.

Lagger, S. et al., 2017. MeCP2 recognizes cytosine methylated tri-nucleotide and dinucleotide sequences to tune transcription in the mammalian brain. *PLoS Genetics*, 13(5), pp.1–26.

Laron, Z., 2005. Do deficiencies in growth hormone and insulin-like growth factor-1 (IGF-1) shorten or prolong longevity? *Mechanisms of Ageing and Development*, 126(2), pp.305–307.

Laron, Z., 2001. Insulin-like growth factor 1 (IGF-1): a growth hormone. *Molecular pathology : MP*, 54(5), pp.311–6. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1187088&tool=pmcentrez&rendertype=abstract>.

Laron, Z., 2008. The GH-IGF1 axis and longevity. The paradigm of IGF1 deficiency. *Hormones*, 7(1), pp.24–27.

- Laurvick, C.L., 2006. Rett syndrome in Australia: a review of the epidemiology. *J Pediatr*, 148(3), pp.347–52.
- Leoncini, S. et al., 2011. Oxidative stress in Rett syndrome: natural history, genotype, and variants. *Redox report : communications in free radical research*, 16(4), pp.145–53. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21888765>.
- Li, G. et al., 2017. Effect of simvastatin on CSF Alzheimer disease biomarkers in cognitively normal adults. *Neurology*, 89(12), pp.1251–1255.
- Li, Y. et al., 2013. Global transcriptional and translational repression in human-embryonic- stem-cell-derived rett syndrome neurons. *Cell Stem Cell*, 13(4), pp.446–458.
- Lim, D. et al., 2008. Calcium homeostasis and mitochondrial dysfunction in striatal neurons of Huntington disease. *Journal of Biological Chemistry*, 283(9), pp.5780–5789.
- Lin, P. et al., 2016. Transcriptome analysis of human brain tissue identifies reduced expression of complement complex C1Q Genes in Rett syndrome. *BMC Genomics*, pp.1–11. Available at: <http://dx.doi.org/10.1186/s12864-016-2746-7>.
- Lintas, C., Sacco, R. & Persico, A.M., 2012. Genome-wide expression studies in Autism spectrum disorder, Rett syndrome, and Down syndrome. *Neurobiology of Disease*, 45(1), pp.57–68. Available at: <http://dx.doi.org/10.1016/j.nbd.2010.11.010>.
- Liu, E., Law, H.K.W. & Lau, Y.L., 2003. Insulin-Like Growth Factor I Promotes Maturation and Inhibits Apoptosis of Immature Cord Blood Monocyte-Derived Dendritic Cells through MEK and PI 3-Kinase Pathways. *Pediatric Research*, 54(6), pp.919–925.
- Liu, J.P. et al., 1993. Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r). *Cell*, 75(1), pp.59–72.
- Ljubicic, S. et al., 2013. The GTPase Rab37 Participates in the Control of Insulin Exocytosis. *PLoS ONE*, 8(6), pp.1–9.
- Llorens-Martín, M., Torres-Alemán, I. & Trejo, J.L., 2009. Mechanisms Mediating Brain Plasticity: IGF1 and Adult Hippocampal Neurogenesis. *The Neuroscientist*, 15(2), pp.134–148. Available at: <http://nro.sagepub.com>.
- Loeffen, J. et al., 1998. The X-chromosomal NDUFA1 gene of complex I in

mitochondrial encephalomyopathies: Tissue expression and mutation detection. *Journal of Inherited Metabolic Disease*, 21(3), pp.210–215.

- Lopez, A.M. et al., 2016. Suppression of brain cholesterol synthesis in male Mecp2-deficient mice is age dependent and not accompanied by a concurrent change in the rate of fatty acid synthesis. *Brain Research*, 1654(May 2016), pp.77–84. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0006899316307247>.
- Loth, E. et al., 2017. The EU-AIMS Longitudinal European Autism Project (LEAP): Design and methodologies to identify and validate stratification biomarkers for autism spectrum disorders. *Molecular Autism*, 8(1), pp.1–19.
- Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), p.550. Available at: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8>.
- Lu, Y. et al., 2016. Gene disruption of dematin causes precipitous loss of erythrocyte membrane stability and severe hemolytic anemia. *Blood*, 128(1), pp.93–103.
- Lupu, C. et al., 2013. endothelial cells in normal conditions and in response to androgen Novel protein ADTRP regulates TFPI expression and function in human endothelial cells in normal conditions and in response to androgen. , 118(16), pp.4463–4471.
- Lyst, M.J. et al., 2013. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nature Neuroscience*, 16(7), pp.898–902. Available at: <http://www.nature.com/doifinder/10.1038/nn.3434>.
- Maezawa, I. & Jin, L.W., 2010. Rett Syndrome Microglia Damage Dendrites and Synapses by the Elevated Release of Glutamate. *Journal of Neuroscience*, 30(15), pp.5346–5356. Available at: <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5966-09.2010>.
- Mangatt, M. et al., 2016. Prevalence and onset of comorbidities in the CDKL5 disorder differ from Rett syndrome. *Orphanet journal of rare diseases*, 11, pp.1–16. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27080038><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4832563>.
- Marchetto, M.C.N.M. et al., 2010. A model for neural development and treatment of Rett

- syndrome using human induced pluripotent stem cells. *Cell*, 143(4), pp.527–39.
Available at:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3003590&tool=pmcentrez&rendertype=abstract><http://www.sciencedirect.com/science/article/pii/S0092867410011864><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3003590&tool=pmcentrez&>.
- Markowska, A.L., Mooney, M. & Sonntag, W.E., 1998. Insulin-like growth factor-1 ameliorates age-related behavioral deficits. *Neuroscience*, 87(3), pp.559–569.
- Masuda, E.S. et al., 2000. Rab37 is a novel mast cell specific GTPase localized to secretory granules. *FEBS Letters*, 470(1), pp.61–64.
- Maxwell, S.S. et al., 2013. Chromatin context and ncRNA highlight targets of MeCP2 in brain. *RNA Biology*, 10(11), pp.1741–1757. Available at:
<http://www.tandfonline.com/doi/abs/10.4161/rna.26921>.
- Meffert, M.K. et al., 2003. NF- κ B functions in synaptic signaling and behavior. *Nature Neuroscience*, 6(10), pp.1072–1078. Available at:
<http://www.nature.com/doi/abs/10.1038/nn1110>.
- Melchior, B. et al., 2010. Dual Induction of TREM2 and Tolerance-Related Transcript, Tmem176b, in Amyloid Transgenic Mice: Implications for Vaccine-Based Therapies for Alzheimer's Disease. *ASN Neuro*, 2(3), p.AN20100010. Available at:
<http://journals.sagepub.com/doi/10.1042/AN20100010>.
- Milakovic, T. & Johnson, G.V.W., 2005. Mitochondrial respiration and ATP production are significantly impaired in striatal cells expressing mutant huntingtin. *Journal of Biological Chemistry*, 280(35), pp.30773–30782.
- Milani, D. et al., 2005. Another patient with MECP2 mutation without classic rett syndrome phenotype. *Pediatric Neurology*, 32(5), pp.355–357.
- Min, J.L. et al., 2010. Variability of gene expression profiles in human blood and lymphoblastoid cell lines. *BMC Genomics*, 11(1), p.96. Available at:
<http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-96>.
- Moghadam, A.A. et al., 2013. How the nucleus and mitochondria communicate in energy production during stress: Nuclear MtATP6, an early-stress responsive gene, regulates the mitochondrial F1F0-ATP synthase complex. *Molecular*

Biotechnology, 54(3), pp.756–769.

- Moldovan, G.-L. & D'Andrea, A.D., 2009. How the Fanconi Anemia Pathway Guards the Genome. *Annual Review of Genetics*, 43(1), pp.223–249. Available at: <http://www.annualreviews.org/doi/10.1146/annurev-genet-102108-134222>.
- Molina, J.R. & Adjei, A.A., 2006. The Ras/Raf/MAPK Pathway. *Journal of Thoracic Oncology*, 1(1), pp.7–9. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1556086415315069>.
- Motil, K.J. et al., 2012. Gastrointestinal and nutritional problems occur frequently throughout life in Girls and women with Rett Syndrome. *Journal of Pediatric Gastroenterology and nutrition*, 55(3), pp.292–298.
- Mount, R.H. et al., 2002. The Rett Syndrome Behaviour Questionnaire (RSBQ): Refining the behavioural phenotype of Rett syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 43(8), pp.1099–1110.
- Müller, M. & Can, K., 2014. Aberrant redox homeostasis and mitochondrial dysfunction in Rett syndrome. *Biochemical Society transactions*, 42(4), pp.959–64. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25109986>.
- Nan, X. et al., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683), pp.386–389. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9620804<http://www.nature.com/nature/journal/v393/n6683/pdf/393386a0.pdf>.
- Nan, X., Campoy, F.J. & Bird, A., 1997. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, 88(4), pp.471–481.
- Neul, J.L. et al., 2010. Rett Syndrome: Revised Diagnostic Criteria and Nomenclature. *Annual Neurology*, 68(6), pp.944–950. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3058521/pdf/nihms214759.pdf>.
- Neul, J.L. et al., 2008. Specific mutations in Methyl-CpG-Binding Protein 2 confer different severity in Rett syndrome. *Neurology*, 70(16), pp.1313–1321.
- Niedzwiedz, W. et al., 2004. The Fanconi anaemia gene FANCC promotes homologous recombination and error-prone DNA repair. *Molecular Cell*, 15(4), pp.607–620.

- Nuutinen, T. et al., 2005. Induction of clusterin/apoJ expression by histone deacetylase inhibitors in neural cells. *Neurochemistry International*, 47(8), pp.528–538.
- O’Kusky, J.R., Ye, P. & D’Ercole, a J., 2000. Insulin-like growth factor-I promotes neurogenesis and synaptogenesis in the hippocampal dentate gyrus during postnatal development. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20(22), pp.8435–42. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11069951>.
- O’Leary, H.M. et al., 2018. Placebo-controlled crossover assessment of mecamermin for the treatment of Rett syndrome. *Annals of Clinical and Translational Neurology*, pp.1–10. Available at: <http://doi.wiley.com/10.1002/acn3.533>.
- Oliveira, J.M.A. & Gonçalves, J., 2009. In situ mitochondrial Ca²⁺ buffering differences of intact neurons and astrocytes from cortex and striatum. *Journal of Biological Chemistry*, 284(8), pp.5010–5020.
- Olson, C.O. et al., 2014. Brain region-specific expression of MeCP2 isoforms correlates with DNA methylation within Mecp2 regulatory elements. *PLoS ONE*, 9(3).
- Orr, A.L. et al., 2009. N-Terminal Mutant Huntingtin Associates with Mitochondria and Impairs Mitochondrial Trafficking. , 28(11), pp.2783–2792.
- Park, I.-H. et al., 2008. Generation of human-induced pluripotent stem cells. *Nature Protocols*, 3(7), pp.1180–1186. Available at: <http://www.nature.com/doifinder/10.1038/nprot.2008.92>.
- Park, M.J. et al., 2014. Anaplerotic triheptanoin diet enhances mitochondrial substrate use to remodel the metabolome and improve lifespan, motor function, and sociability in MeCP2-null mice. *PLoS ONE*, 9(10).
- Pecorelli, A. et al., 2013. Genes related to mitochondrial functions, protein degradation, and chromatin folding are differentially expressed in lymphomonocytes of rett syndrome patients. *Mediators of Inflammation*, 2013.
- Percy, A.K. et al., 2010. Profiling Scoiosis in Rett Sydnrome. *Pediatric Research*, 67(4), pp.435–439.
- Pertea, M. et al., 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protocols*, 11(9), pp.1650–1667. Available at:

<http://dx.doi.org/10.1038/nprot.2016.095%5Cnhttp://10.1038/nprot.2016.095%5Cnhttp://www.nature.com/nprot/journal/v11/n9/abs/nprot.2016.095.html#supplementary-information>.

- Pini, G. et al., 2012. IGF1 as a Potential Treatment for Rett Syndrome: Safety Assessment in Six Rett Patients. *Autism Research and Treatment*, 2012, pp.1–14.
- Pini, G., Congiu, L., Benincasa, A., et al., 2016. Illness Severity, Social and Cognitive Ability, and EEG Analysis of Ten Patients with Rett Syndrome Treated with Mecasermin (Recombinant Human IGF-1). *Autism Research and Treatment*, 2016, pp.1–9. Available at: <http://www.hindawi.com/journals/aurt/2016/5073078/>.
- Pini, G., Congiu, L., Zappella, M., et al., 2016. PBZ (Pini, Bonuccelli, Zappella): Scala italiana per la valutazione clinica nella Sindrome di Rett. *AUTISMO e disturbi dello sviluppo*, 14(1).
- Pini, G. et al., 2016. Rett syndrome: a wide clinical and autonomic picture. *Orphanet Journal of Rare Diseases*, 11(1), p.132. Available at: <http://ojrd.biomedcentral.com/articles/10.1186/s13023-016-0499-7>.
- Piper, M.D.W. et al., 2002. Reproducibility of oligonucleotide microarray transcriptome analyses. An interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 277(40), pp.37001–37008.
- Pohodich, A.E. & Zoghbi, H.Y., 2015. Rett syndrome: Disruption of epigenetic control of postnatal neurological functions. *Human Molecular Genetics*, 24(R1), pp.R10–R16.
- Polo, J.J.M. et al., 2010. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature biotechnology*, 28(8), pp.848–855.
- Pomp, O. et al., 2011. Unexpected X chromosome skewing during culture and reprogramming of human somatic cells can be alleviated by exogenous telomerase. *Cell Stem Cell*, 9(2), pp.156–165. Available at: <http://dx.doi.org/10.1016/j.stem.2011.06.004>.
- Popken, G.J. et al., 2004. In vivo effects of insulin-like growth factor-I (IGF-I) on prenatal and early postnatal development of the central nervous system.

- Neuroscience*, 19, pp.2056–2068.
- Potluri, P. et al., 2009. A novel NDUFA1 mutation leads to a progressive mitochondrial complex I-specific neurodegenerative disease. *Molecular Genetics and Metabolism*, 96(4), pp.189–195. Available at: <http://dx.doi.org/10.1016/j.ymgme.2008.12.004>.
- Puche, J.E. et al., 2016. Partial IGF-1 deficiency induces brain oxidative damage and edema, which are ameliorated by replacement therapy. *BioFactors*, 42(1), pp.60–79.
- Rahman, S. et al., 1996. Leigh syndrome: Clinical features and biochemical and DNA abnormalities. *Annals of Neurology*, 39(3), pp.343–351.
- Renieri, A. et al., 2009. Diagnostic criteria for the Zappella variant of Rett syndrome (the preserved speech variant). *Brain and Development*, 31(3), pp.208–216. Available at: <http://dx.doi.org/10.1016/j.braindev.2008.04.007>.
- Rett, A., 2016. On a remarkable syndrome of cerebral atrophy associated with hyperammonaemia in childhood. *Wiener Medizinische Wochenschrift*, 166(11–12), pp.322–324.
- Ricceri, L., De Filippis, B. & Laviola, G., 2008. Mouse models of Rett syndrome: from behavioural phenotyping to preclinical evaluation of new therapeutic approaches. *Behavioural pharmacology*, 19(5–6), pp.501–517.
- Robinson, J.H. & Delvig, A.A., 2002. Diversity in mhc class ii antigen presentation. *Immunology*, 105(3), pp.252–262.
- Robinson, M.D., McCarthy, D.J. & Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139–140.
- Robinson, M.D. & Smyth, G.K., 2007. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, 23(21), pp.2881–2887.
- Robinson, M.D. & Smyth, G.K., 2008. Small-sample estimation of negative binomial dispersion with applications to SAGE data. *Biostatistics*, 9(2), pp.321–332.
- Rojo, A. et al., 2006. NARP-MILS syndrome caused by 8993 T > G mitochondrial DNA mutation: A clinical, genetic and neuropathological study. *Acta Neuropathologica*, 111(6), pp.610–616.
- Rose, S.A. et al., 2013. Rett syndrome: An eye-tracking study of attention and

recognition memory. *Developmental Medicine and Child Neurology*, 55(4), pp.364–371.

Ross, P.D. et al., 2016. Exclusive expression of MeCP2 in the nervous system distinguishes between brain and peripheral Rett syndrome-like phenotypes HMG Advance Access. *Human Molecular Genetics*, 0(0), pp.1–16. Available at: <http://hmg.oxfordjournals.org/>.

Rossmannith, W. et al., 2008. Isolated cytochrome c oxidase deficiency as a cause of MELAS. *Journal of medical genetics*, 45(2), pp.117–21. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18245391>.

Rumbak, D.M. et al., 2016. Spinal Fusion for Scoliosis in Rett Syndrome with an Emphasis on Respiratory Failure and Opioid Usage. *Journal of Child Neurology*, 31(2), pp.153–158. Available at: <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L607730609%5Cnhttp://dx.doi.org/10.1177/0883073815585352>.

Sample, V. et al., 2015. Polarized activities of AMPK and BRSK in primary hippocampal neurons. *Molecular Biology of the Cell*, 26(10), pp.1935–1946. Available at: <http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E14-02-0764>.

Saywell, V. et al., 2006. Brain magnetic resonance study of Mecp2 deletion effects on anatomy and metabolism. *Biochemical and Biophysical Research Communications*, 340(3), pp.776–783.

Schärer, O.D., 2013. Nucleotide excision repair in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 5(10), pp.1–19.

Segatto, M. et al., 2014. Cholesterol metabolism is altered in Rett syndrome: A study on plasma and primary cultured fibroblasts derived from patients. *PLoS ONE*, 9(8), pp.1–6.

Sgarbi, G. et al., 2006. Inefficient coupling between proton transport and ATP synthesis may be the pathogenic mechanism for NARP and Leigh syndrome resulting from the T8993G mutation in mtDNA. *Biochem. J*, 395, pp.493–500.

Shah, R.R. & Bird, A.P., 2017. MeCP2 mutations: progress towards understanding and treating Rett syndrome. *Genome Medicine*, 9(1), p.17. Available at: <http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-017-0411-7>.

- Shahbazian, M.D. et al., 2002. Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. *Human Molecular Genetics*, 11(2), pp.115–124.
- Shahbazian, M.D. et al., 2002. Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron*, 35(2), pp.243–254.
- Shin, H. et al., 2014. Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. *PLoS ONE*, 9(3), pp.1–11.
- Signorini, C. et al., 2011. F4-neuroprostanes mediate neurological severity in Rett syndrome. *Clinica Chimica Acta*, 412(15–16), pp.1399–1406. Available at: <http://dx.doi.org/10.1016/j.cca.2011.04.016>.
- Simon, M. et al., 2015. Mutations of Human NARS2, Encoding the Mitochondrial Asparaginyl-tRNA Synthetase, Cause Nonsyndromic Deafness and Leigh Syndrome. *PLoS Genetics*, 11(3), pp.1–26.
- Skene, P.J. et al., 2010. Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. *Molecular Cell*, 37(4), pp.457–468. Available at: <http://dx.doi.org/10.1016/j.molcel.2010.01.030>.
- Sonntag, W.E. et al., 2000. The effects of growth hormone and IGF-1 deficiency on cerebrovascular and brain ageing. *Journal of anatomy*, 197 Pt 4(2000), pp.575–585.
- Soumillon, M. et al., 2014. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv*, p.003236. Available at: <http://biorxiv.org/content/early/2014/03/05/003236.abstract>.
- Steffenburg, U., Hagberg, G. & Hagberg, B., 2001. Epilepsy in a representative series of Rett syndrome. *Acta paediatrica*, 90, pp.34–39.
- Stock, D., 1999. Molecular Architecture of the Rotary Motor in ATP Synthase. *Science*, 286(5445), pp.1700–1705. Available at: <http://www.sciencemag.org/cgi/doi/10.1126/science.286.5445.1700>.
- Sugino, K. et al., 2014. Cell-type-specific repression by methyl-CpG-binding protein 2 is biased toward long genes. *J Neurosci*, 34(38), pp.12877–12883. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25232122>.

- Suter, B. et al., 2014. Brief report: MECP2 mutations in people without rett syndrome. *Journal of Autism and Developmental Disorders*, 44(3), pp.703–711.
- Tanaka, Y. et al., 2014. Transcriptional regulation in pluripotent stem cells by methyl CpG-binding protein 2 (MeCP2). *Human Molecular Genetics*, 23(4), pp.1045–1055.
- Tatar, M., 2001. A Mutant Drosophila Insulin Receptor Homolog That Extends Life-Span and Impairs Neuroendocrine Function. *Science*, 292(5514), pp.107–110. Available at: <http://www.sciencemag.org/cgi/doi/10.1126/science.1057987>.
- Tatebe, H. et al., 2017. Quantification of plasma phosphorylated tau to use as a biomarker for brain Alzheimer pathology: pilot case-control studies including patients with Alzheimer's disease and down syndrome. *Molecular Neurodegeneration*, 12, pp.1–11. Available at: <https://molecularneurodegeneration.biomedcentral.com/track/pdf/10.1186/s13024-017-0206-8?site=molecularneurodegeneration.biomedcentral.com>.
- Thorburn, D.R., Rahman, J. & Rahman, S., Mitochondrial DNA-Associated Leigh Syndrome and NARP. *GeneReviews® [Internet]*. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK1173/>.
- Trappe, R. et al., 2001. MECP2 Mutations in Sporadic Cases of Rett Syndrome Are Almost Exclusively of Paternal Origin. *The American Journal of Human Genetics*, 68(5), pp.1093–1101.
- Tropea, D. et al., 2009. Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice. *Proceedings of the National Academy of Sciences of the United States of America*, 106(6), pp.2029–2034. Available at: <http://www.pnas.org/content/106/6/2029.long%5Cnpapers3://publication/doi/10.1073/pnas.0812394106>.
- Venables, P.J.W. et al., 1995. Abundance of an endogenous retroviral envelope protein in trophoblasts suggests a biological function. *Virology*, 211(2), pp.589–592.
- de Vries, D.D. et al., 1993. A Second missense Mutation in Mitochondrial ATPase 6 gene in Leigh's Syndrome. *Annals of Neurology*, 34(3), pp.410–412.
- Wakisaka, Y. et al., 2014. Plasma S100A12 is associated with functional outcome after ischemic stroke: Research for Biomarkers in Ischemic Stroke. *Journal of the*

- Neurological Sciences*, 340(1–2), pp.75–79. Available at:
<http://dx.doi.org/10.1016/j.jns.2014.02.031>.
- Weese-Mayer, D.E. et al., 2006. Autonomic nervous system dysregulation: Breathing and heart rate perturbation during wakefulness in young girls with rett syndrome. *Pediatric Research*, 60(4), pp.443–449.
- Wong, L.J.C. et al., 2001. Severe lactic acidosis caused by a novel frame-shift mutation in mitochondrial-encoded cytochrome c oxidase subunit II. *American Journal of Medical Genetics*, 102(1), pp.95–99.
- Wrigley, S., Arafa, D. & Tropea, D., 2017. Insulin-Like Growth Factor 1: At the Crossroads of Brain Development and Aging. *Frontiers in Cellular Neuroscience*, 11(February), p.14. Available at:
<http://journal.frontiersin.org/article/10.3389/fncel.2017.00014/full>.
- Xu, X., Miller, E.C. & Pozzo-Miller, L., 2014. Dendritic spine dysgenesis in Rett syndrome. *Frontiers in neuroanatomy*, 8(September), p.97. Available at:
<http://journal.frontiersin.org/article/10.3389/fnana.2014.00097/abstract>.
- Yaffe, Y. et al., 2015. The myelin proteolipid plasmalogen forms oligomers and induces liquid-ordered membranes in the Golgi complex. *Journal of cell science*, 128(13), pp.2293–302. Available at:
<http://jcs.biologists.org/content/early/2015/05/20/jcs.166249>.
- Ye, P. et al., 2010. β -Catenin mediates insulin-like growth factor-1 (IGF-1) actions to promote cyclin D1 mRNA expression, cell proliferation and survival in oligodendroglial cultures. *Glia*, 58(9), pp.1031–1041.
- Yoshikawa, S., 1998. Redox-Coupled Crystal Structural Changes in Bovine Heart Cytochrome c Oxidase. *Science*, 280(5370), pp.1723–1729. Available at:
<http://www.sciencemag.org/cgi/doi/10.1126/science.280.5370.1723>.
- Zhao, S. et al., 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*, 9(1).
- Zhao, Y. et al., 2013. Neurobiology of Disease Loss of MeCP2 function is associated with distinct gene expression changes in the striatum. , 59, pp.257–266.
- Ziegenhain, C. et al., 2016. Comparative analysis of single-cell RNA-sequencing methods. , 49(0).

Zoghbi, H.Y., 2016. Rett Syndrome and the Ongoing Legacy of Close Clinical Observation. *Cell*, 167(2), pp.293–297. Available at: <http://dx.doi.org/10.1016/j.cell.2016.09.039>.

Appendices

Appendix A – Full List of Significant DGE Results from edgeR Analysis

Boston Children's Hospital Results:				
Comparison	Time Interval	Gene Symbol	Log₂ Fold Change	Corrected P-value
Total (n = 9)	BCH-t1 to BCh-t2	TMEM176A	-2.72	1.3 x10 ⁻⁴
		TMEM176B	-2.52	5.69 x10 ⁻⁴
Non-Responders (n = 5)	BCH-t0 to BCH-t3	DEFA3	3.89	2.17 x10 ⁻²
		DEFA1B	3.69	2.17 x10 ⁻²
		DEFA1	3.69	2.17 x10 ⁻²
		CD177	-4.05	3.43 x10 ⁻²
		ZFHX2	-3.12	3.43 x10 ⁻²
		DEFA4	3.34	3.43x10 ⁻²
	BCH-t0 to BCH-t2	ERVMER34-1	-4.32	2.98 x10 ⁻⁶
	BCH-t1 to BCH-t3	DEFA3	3.66	2.05 x10 ⁻³
		DEFA1B	3.56	2.05 x10 ⁻³
		DEFA1	3.56	2.05x10 ⁻³
		DEFA4	3.5	2.05 x10 ⁻³
		ERVMER34-1	2.85	1.98 x10 ⁻²
	BCH-t1 to BCH-t2	TMEM176B	-3.56	1.60 x10 ⁻¹⁶
		TMEM176A	-3.64	1.07 x10 ⁻¹³
		SLC4A1	-2.04	1.29 x10 ⁻⁷
		KRT1	-2.14	1.70 x10 ⁻⁷
		AHSP	-1.7	1.75 x10 ⁻⁵
		PAX8-AS1	-3.16	4.69 x10 ⁻⁴
		BRSK1	-1.62	5.11 x10 ⁻⁴
		RSAD2	-2.27	1.23 x10 ⁻³
		TRIM58	-1.52	1.83 x10 ⁻³
		SMIM1	-1.71	4.58 x10 ⁻³
		FAM210B	-1.41	4.58 x10 ⁻³
		HOPX	1.66	4.76 x10 ⁻³
		BLVRB	-1.37	4.76 x10 ⁻³

		HEMGN	-1.51	6.02 x10 ⁻³
		SLC25A39	-1.36	6.52 x10 ⁻³
		SNCA	-1.33	7.04 x10 ⁻³
		DMTN	-1.32	7.71 x10 ⁻³
		UBXN6	-1.33	7.71 x10 ⁻³
		RBM38	-1.36	7.87 x10 ⁻³
		E2F2	-1.75	1.37 x10 ⁻²
		BCL2L1	-1.31	1.48 x10 ⁻²
		SELENBP1	-1.31	1.68 x10 ⁻²
		GYPB	-1.49	2.77 x10 ⁻²
		HBM	-1.23	2.77 x10 ⁻²
		STRADB	-1.25	3.08 x10 ⁻²
		CSE1L-AS1	-5.14	3.13 x10 ⁻²
		JCHAIN	1.19	3.14 x10 ⁻²
		GSPT1	-1.21	3.14 x10 ⁻²
		ALAS2	-1.25	4.42 x10 ⁻²
		LOC100288152	5.15	4.57 x10 ⁻²
		LOC101927697	-1.18	4.58 x10 ⁻²
		PDHX	2.34	4.58 x10 ⁻²
		RUNDC3A	-1.56	4.58 x10 ⁻²
		TRIM2	5.12	4.58 x10 ⁻²
		EPB42	-1.28	4.73 x10 ⁻²
		DCAF12	-1.17	4.73 x10 ⁻²
Responders (n = 4)	BCH-t2 to BCH-t3	ERVMER34-1	4.08	1.11 x10 ⁻⁵
	BCH-t0 to BCH-t2	HLA-DRB5	-2.33	1.24 x10 ⁻³
		ADTRP	6.02	1.71 x10 ⁻²
		SMCR5	-2.97	2.09 x10 ⁻²
Overlapping Responders vs Non- Responders	BCH-t2 to BCH-t3	HLA-DRB5	3.06	1.51 x10 ⁻⁵
	BCH-t0	RPL27	-1.23	3.45 x10 ⁻²
		IL7R	-1.19	3.69 x10 ⁻²
		PNRC2	-1.01	4.61 x10 ⁻²
		MCEMP1	-1.59	2.99 x10 ⁻²
		ADA	-1.78	2.78 x10 ⁻³
		S100A8	-2.13	5.75 x10 ⁻³

		BCL2A1	-1.63	6.88 x10 ⁻³
		SEC11C	-1.41	9.34 x10 ⁻³
		S100A12	-2.26	4.77 x10 ⁻³
		NARS2	1.3	2.81 x10 ⁻²
Versilia Hospital Italy				
Comparison	Time Interval	Gene Symbol	Log₂ Fold Change	Corrected p=value
Short Term Effects of IGF-1 (n=2)	VH-t0 to VH-t1	MT-ATP6	-8.01	1.24 x10 ⁻¹³
		HLA-G	-7.92	2.19 x10 ⁻⁷
		ERCC5	-7.31	6.84 x10 ⁻⁵
		BIVM-ERCC5	6.91	2.08 x10 ⁻⁴
		CCDC85C	7.09	3.79 x10 ⁻⁴
		SNX14	-6.82	1.17 x10 ⁻³
		GPATCH3	-6.59	8.13 x10 ⁻³
		RAB37	2.79	1.19 x10 ⁻²
		INO80B-WBP1	6.62	1.83 x10 ⁻²
		SLC7A6	2.60	1.91 x10 ⁻²
		AC024909.3	4.51	3.18 x10 ⁻²
Long Term Effects of IGF-1 (n=3)	VH-t0 to VH-t2	MT-CO2	-10.85	1.44 x10 ⁻¹⁴
		EEF1A1P7	7.25	1.35 x10 ⁻¹⁰
		NDUFA1	6.84	1.25 x10 ⁻⁹
		RAB37	2.88	5.02 x10 ⁻⁸
		PLLP	7.12	1.96 x10 ⁻⁷
		ARL2BP	-3.27	1.39 x10 ⁻⁴
		DNAJC14	6.00	1.39 x10 ⁻⁴
		WASHC3	6.07	2.42 x10 ⁻⁴
		LGALS2	6.24	7.59 x10 ⁻⁴
		AC023055.1	-5.61	4.23 x10 ⁻³
		SLX1B	-5.44	4.53 x10 ⁻³
		SLX1A	-5.34	9.09 x10 ⁻³
		AC020915.2	-5.68	1.57 x10 ⁻²
		UBE2I	-1.43	2.04 x10 ⁻²
		GTF2IRD2B	5.28	2.04 x10 ⁻²

Overlapping Major Responders vs Minor Responders (n=2)	VH-t0	RPL27	9.44	2.03×10^{-15}
		IL7R	3.47	5.65×10^{-8}
		PNRC2	-3.37	1.62×10^{-5}
		MCEMP1	-4.16	1.04×10^{-2}
		ADA	5.81	1.04×10^{-2}
		S100A8	-3.00	1.71×10^{-2}
		SEC11C	2.52	3.26×10^{-2}
		S100A12	-2.18	3.30×10^{-2}
		BCL2A1	-1.91	1.91×10^{-2}
		NARS2	5.46	3.95×10^{-2}

Appendix B – Full List of Significant DGE Results from DESeq2 Analysis

Boston Children’s Hospital Results:				
Comparison	Time Interval	Gene Symbol	Log ₂ Fold Change	Corrected p-value
Non-Responders (n = 5)	BCH-t1 to BCH-t2	TMEM176B	-3.57	7.93 x10 ⁻¹¹
		TMEM176A	-3.67	7.79 x10 ⁻⁹
		BRSK1	-1.62	2.27 x10 ⁻³
		JCHAIN	1.22	1.24 x10 ⁻²
		DMTN	-1.32	2.63 x10 ⁻²
Versilia Hospital Italy Results				
Comparison	Time Interval	Gene Symbol	Log ₂ Fold Change	Corrected p=value
Short term effects of IGF-1	VH-t0 to VH-t1	MT-ATP6	-8.11	3.56 x10 ⁻¹²
Long term effects of IGF-1	VH-t0 to VH-t2	MT-CO2	-10.36	7.85 x10 ⁻¹¹
		RAB37	2.91	3.44 x10 ⁻³
		EEF1A1P7	6.59	1.44 x10 ⁻²
		PLLP	8.50	3.13 x10 ⁻²