UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università di Padova
Dipartimento di Pediatria

SCUOLA DI DOTTORATO DI RICERCA IN:

MEDICINA DELLO SVILUPPO E SCIENZE DELLA PROGRAMMAZIONE

INDIRIZZO: MALATTIE RARE

CICLO: XXIII

GENETIC VARIATIONS IN THE *MECP2* 3'UTR AND IN THE *ARX* GENE INFLUENCE THE PATHOGENESIS OF NEURODEVELOPMENTAL DISORDERS

Direttore della Scuola: Ch.mo Prof. Giuseppe Basso
Coordinatore di Indirizzo: Ch.mo Prof. Giorgio Perilongo
Supervisore: Dr.ssa Alessandra Murgia

Dottorando : Elisa Bettella

Luglio 2011

Publications and Abstracts arising from this thesis

Sartori S, Polli R, Bettella E, Rossato S, Andreoli W, Vecchi M, Giordano L, Accorsi P, Di Rosa G, Toldo I, Zamponi N, Darra F, Dalla Bernardina B, Perilongo G, Boniver C, Murgia A. - Pathogenic Role of the X-Linked Cyclin-Dependent Kinase-Like 5 and Aristaless-Related Homeobox Genes in Epileptic Encephalopathy of Unknown Etiology With Onset in the First Year of Life. *J Child Neurol. 2011 Apr 11*

Giordano L, Sartori S, Russo S, Accorsi P, Galli J, Tiberti A, Bettella E, Marchi M, Vignoli A, Darra F, Murgia A, Bernardina BD - Familial Ohtahara syndrome due to a novel *ARX* gene mutation. *Am J Med Genet A. 2010 Dec;152A(12):3133-7*

Bettella E, Ho G, Murgia A, Christodoulou J - Non coding elements that regulate MeCP2 expression. *Abstract and poster at the 5th Australian Health & Medical Research Congress 2010, November 2010 Melbourne Australia*

Bettella E, Sartori S, Giordano L, Darra F, Polli R, Russo S, Dalla Bernardina B, Murgia A - *ARX* gene and Ohtahara syndrome: beyond polyalanine tract expansions. *Abstract and poster at the 58th Annual Meeting The American Society of Human Genetics, November 2008 Philadelphia USA*

Sartori S, Polli R, Di Rosa G, Bettella E, Tricomi G, Tortorella G, Murgia A - Early-onset seizure variant of RETT Syndrome in a 47, XXY boy with a novel *CDKL5* mutation. *Abstract and poster at the 58th Annual Meeting The American Society of Human Genetics, November 2008 Philadelphia USA*

Index

List of Figures

List of Tables

Abbreviations

AD	Autistic disorder
AS	Angelman Syndrome
AREs	AU-rich elements
ARX	Aristaless related homeobox
BDNF	Brain-Derived Neuratrophic Factor
CDKL5	Cyclin-Dependent Kinase Like 5
CPSF	Cleavage and polyadenylation specific factors
CREB1	CAMP Responsive Element Binding Protein 1
CS	Polyadenylation cleavage site
DNA	Deoxyribonucleic Acid
DNMT1	DNA (sytosine-5)-Methyltransferase 1
EEG	Electroencephalography
FHD	Fork Head Domain
FOXG1	Fork Head Box G1
FXS	Fragile X syndrome
Groucho/TLE1	Trasducin like enhancer of split 1
GTBD	Groucho-Binding Domain
HDAC	Histone deacetylase
MeCP2	Methyl-CpG Binding Protein 2
miRNA	microRNA
MLPA	Multiplex Ligation Dependent Probe Amplification
mRNA	messenger RNA
ncRNA	Non-coding RNA molecules
NLS	Nuclear Localization signal
NS-XLMR	Non-Syndromic X-Linked Mental Retardation
OAR	Aristaless domain
OMIM	Online Mendelian Inheritance in Man
OS	Othahara syndrome
poly(A)	Polyadenylation signal
Q-PCR	Real Time Quantitative PCR
RNA	Ribonucleic Acid
RTT	Rett Syndrome
ssRNA	single strand RNAs
TRD	Transcription Repression Domain
UBE3A	Ubiquitin Protein Ligase E3A
XCI	X-Chromosome Inactivation
XMEID	X-linked myoclonic epilepsy with severe intellectual disability
XLAG	X-linked lissencephaly with abnormal genitalia

Abstract

Neurodevelopmental disorders are a large group of common and complex disorders seen in paediatric, child-neurological and clinical genetic practice. They include a wide set of clinically diverse and genetically heterogeneous disorders of varying severity. Even when the clinical and behavioural phenotypes of these disorders are known, the complete mechanisms of these pathologies are still not completely elucidated. The *MECP2* and *ARX* genes are two fundamental genes for a proper development of the brain, pathogenic sequence variations in these genes are known to be responsible respectively of Rett syndrome (RTT) and early epileptic encephalopathies (EIEE).

Patogenic variations in the *MECP2* coding region are responsible for 90-95% of classical RTT and for the 40-60% of atypical RTT. The effect and the impact of sequence variations in the very long 3'UTR *MECP2* non-coding region in RTT are still not known. This project investigated the possible role of the *MECP2* 3'UTR in the pathogenesis of RTT through the analysis of the expression levels of *MECP2* isoforms, the sequencing of the *MECP2* 3'UTR, *in silico* analysis and i*n vitro* functional analysis of 3'UTR sequence alterations. Our results showed altered expression levels of both isoforms among a cohort of 22 RTT patients, without sequence variations in the *MECP2* coding region. Two SNPs, one known and one unreported sequence variations were identified as a consequence of screening of the proximal part of the *MECP2* 3'UTR. The functional analysis of a few variations in the 3'UTR showed that luciferase reporter protein expression levels were increased in neuronal and in non-neuronal cell lines.

Pathogenic variations in the *ARX* gene are responsible of a broad spectrum of neurodevelopmental disorders with and without malformations. Initially only expansions of the polyalanine tracts in this gene were associated with Ohtahara syndrome. As a result of the *ARX* gene analysis, two previously unreported pathogenic point mutations, in a well-conserved domain, were detected in three boys with Ohtahara syndrome futher confirming the pathogenic role of this gene.

The complexity of the etiopathogenesis of the neurodevelopmental disorders remains to be further elucidated and other studies will be necessary to expand our understanding of these disorders to provide accurate diagnosis and genetic counselling for the patients

.

1. Introduction

Neurodevelopmental disorders are a large group of common and complex disorders seen in paediatric, child-neurological and clinical genetic practice. They include a wide set of clinically diverse and genetically heterogeneous disorders of varying severity. They are mostly associated with mild to severe intellectual disability and impact on normal brain functioning, affecting behaviour and cognitive abilities. These disorders include a large number of diseases that share similar features involving brain dysfunction (abnormalities in motor and sensory systems, language and speech problems) and cognitive impairment. The wide spectrum of neurodevelopmental disorders includes Tourette syndrome, dysmorphic syndromes, epilepsy, attention-deficit hyperactivity disorder, autistic spectrum disorder, Down syndrome, Fragile X syndrome, Prader-Willi syndrome, Angelman syndrome and Rett syndrome (Bishop, 2010).

The clinical and behavioural phenotypes of these disorders are known, but the complete mechanisms of these pathologies are still not completely elucidated, as these disorders can have environmental influences as well as genetic causes. Linkage and association studies have been performed over the years to test possible associations between genetic variation and these neurodevelopmental disorders, but inconsistent and sometimes controversial results have been obtained (Wetmore and Garner, 2010). Advanced molecular genetic technologies have improved knowledge about the mechanisms underlying these disorders. An abnormal number of chromosomes, single gene defects, the gene-dosage imbalances, genetic background and epigenetic factors play important roles in the pathogenesis of the central nervous system disorders (Gropman and Batshaw, 2010).

Emerging evidence links sequence variations in the coding/non-coding regions, epigenetic dysfunction (such as abnormalities in DNA methylation, and histone modification) or sequence variation in microRNAs with the progression of neurodevelopmental disorders (Urdinguio et al., 2009).

1.1 Neural development

Neural development comprises the processes that generate and shape the nervous system during the embryogenesis. The central nervous system (CNS) is the most complex structure. In fact, the CNS is characterized by ten trillion cells that form finely regulated and complex networks (Ottens, 2009). The neural plate, a specialized region of the ectoderm, transforms into the neural tube (*Figure 1.1*). The neural tube is the precursor of the brain (neurons and glia) and spinal cord. The neuronal crest cells (neuroepithelial cells) migrate to form the elements of the peripheral nervous system, e.g. sensory, autonomic and enteric neurons, plus Schwann and satellite cells, as well as several non-neural structures, e.g. melanocytes, craniofacial mesenchyme and some cardiac septa.

The generation of the brain is produced during a time window that closes before birth. Correct spatial/temporal order and generation of neurons during development are a prerequisite for the establishment of the correct neuronal connectivity. It is then clear that minimal changes during this period can have a great impact at neurogenesis levels and overall later in life leading to neurodevelopmental disorders (Purves et al., 2001).

The *MECP2* and *ARX* genes are two fundamental genes for proper development of the brain. It is known that pathogenic sequence variations in the *MECP2* gene and in the *ARX* gene are responsible respectively for two important neurodevelopmental disorders such as Rett syndrome (RTT) and early epileptic encephalopathies (EIEE).

In fact, the *MECP2* gene is important for the CNS development, neuronal maturation, dendritic morphology and synaptic transmission (Na and Monteggia, 2011). Shahbazian *et al* showed that MeCP2 rodent protein localizes in the mouse spinal cord and brainstem at embryonic day 12 (E12),subsequently it is present in thalamus, cerebellum, hypothalamus, and hippocampus. MeCP2 protein is highly expressed in neurons throughout the brain from early postnatal development into adulthood (Shahbazian et al., 2002a). MeCP2 loss may be therefore interfere with neuronal maturation and synaptogenesis, culminating in abnormal development of the CNS (Na and Monteggia, 2011). RTT patients and MeCP2 null mice neurons are characterized by fewer dendritic spines and reduced arborization. Moreover, MeCP2 over expression results in significant reduction in the number of spines and in dendritic branching of the hippocampal neurons (Chapleau et al., 2009; Na and Monteggia, 2011).

The *ARX* gene is essential for the development of the central and peripheral nervous systems during embryogenesis. Arx protein is localized in the mouse anterior neural plate; in addition, Arx shows a different pattern of expression in the ventral and dorsal forebrain (Friocourt et al., 2006) and interestingly, newborn mutant mice were characterized by a smaller brain and smaller olfactory bulbs. Experiments in *Xenopus laevis* show that xArx is important in the development of the brain and confirmed the essential role of this gene in forebrain specification and development.

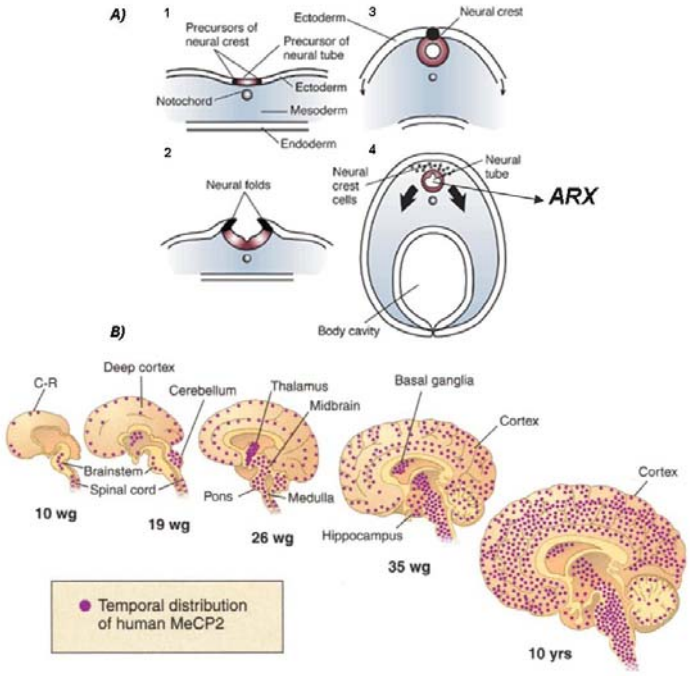


Figure1.1: Neural development

The neuronal differentiation is represented in two cartoon representations. A) The notochord induces the cells of the ectoderm directly above it to become the primitive nervous system (1) that wrinkles and folds over (2). In this way, the neural tube is formed (3-4) that is the precursor of the brain and spinal cord. The neural crest cells migrate to various locations throughout the embryo, where they will initiate the development of the elements of the peripheral nervous system. Picture from (Goodlett and Horn, 2001).
B) Spatial-temporal distribution of the MeCP2 that is abundant in mature neurons. MeCP2 is expressed in Cajal-Retzius (C-R) midbrain, thalamus, cerebellum and deep cortical neurons ganglia, hypothalamus, hippocampus, and superficial cortical layers appears later, and the number of MeCP2-positive neurons in the cerebral cortex continues to increase until 10 years of age. Wg, weeks of gestation. Picture from (Zoghbi, 2003).

1.2 Clinical presentation of Rett syndrome

In 1966 Dr Andreas Rett first described a disorder in German medical literature as a syndrome of “cerebral atrophy and hyperammonemia which was only observed in girls and characterized by autistic behaviour, dementia, apraxia of gait, loss of facial expression and stereotyped use of the hands” (Rett, 1966). Seventeen years later, Hagberg and colleagues described a cohort of 35 French, Portuguese and Swedish

female patients with a very similar phenotype and dubbed the disorder Rett syndrome (Hagberg et al., 1983). In 1999, the discovery of the causative X-linked gene *MECP2,* Methyl-CpG-Binding Protein 2, improved recognition of this syndrome and enabled the development of tests for early diagnosis and prenatal detection (Amir et al., 1999), and led to an expansion of the clinical phenotypes associated with *MECP2* mutations (Weaving et al., 2005)

Rett syndrome (RTT, OMIM #312750) is a postnatal progressive neurodevelopmental disorder that predominantly affects girls. The incidence varies from 0.43-0.71:10000 females in France (Bienvenu et al., 2006) to 1.3:10000 female births by the age of 15 years in Australia (Laurvick et al., 2006) and it is considered the most common cause of severe mental retardation in girls after Down syndrome.

In 1988 diagnostic criteria were set up by the Rett syndrome Diagnostic Work Group (Group, 1988). Subsequently, these criteria have been slightly modified following the European Paediatric Neurology Society Meeting (Hagberg et al., 2002). In 2010 the RettSearch Consortium revisited the diagnostic criteria to simplify the diagnosis of RTT. The diagnostic criteria involve 4 main, 2 exclusion and a variety of supportive criteria (*Table 1.1*). Classical RTT (also known as typical RTT) is defined by a) the presence of a period of regression followed by recovery or stabilization and b) all four main criteria as well as both exclusion criteria. While supportive criteria are not required they are often present in typical RTT. Atypical RTT is diagnosed when a) a period of regression followed by recovery or stabilization is observed in patients as well as b) at least two of the four main criteria and five of the eleven supportive criteria (Neul et al., 2010).

Table 1.1: The revised diagnostic criteria for Rett syndrome from Neul *et al* 2010

Main criteria	
1. Partial or complete loss of acquired purposeful hand skills 2. Partial or complete loss of acquired spoken language 3. Gait abnormalities: impaired (dyspraxia) or absence of ability 4. Stereotypic hand movements such as hand wringing/squeezing, clapping/tapping, mouthing and washing/rubbing automatisms	
Exclusion criteria	
1. Brain injury secondary to trauma (pre- or postnatal), neurometabolic disease or severe infection that causes neurological problems 2. Grossly abnormal psychomotor development in first 6 months of life	
Supportive criteria	
1. Breathing disturbance 2. Bruxism when awake 3. Impaired sleep pattern 4. Abnormal muscle tone 5. Peripheral vasomotor disturbances 6. Scoliosis/kyphosis	7. Growth retardation 8. Small cold hands and feet 9. Inappropriate laughing/screaming spells 10. Diminished response to pain 11. Intense eye communication – “eye pointing”

1.2.1 Classical Rett syndrome

The clinical phenotype for RTT generally consists of a normal pre- and perinatal period, followed by a progressive loss of intellectual functioning, motor skills and communicative abilities, gait abnormalities and hand stereotypies. In 1986, Hagberg and Witt-Engerström proposed four stages in the progression of Rett syndrome (Hagberg and Witt-Engerstrom, 1987) providing a valid and practical classification of the clinical progression (*Figure 1.2*). Although the progression of the clinical events is common to all the patients, the duration and the age onset at each stage varies from patient to patient.

Stage 1 or early infancy stage:

RTT patients appear to have a normal development up to 6-18 months of age achieving the ability to walk and say a few words. Patients may develop deceleration of head growth, leading to microcephaly, disinterest or diminished play interest and hypotonia. Einspieler *et al*. observed abnormal general movements of 14 RTT infants during the first 4 months of life, demonstrating that the disorder may manifest within the first weeks of life (Einspieler et al., 2005a, b).

Stage 2 or rapid regression stage

RTT patients enter a period of developmental stagnation between the ages of 1 and 3 years. This stage is characterized by general growth retardation, loss of weight and a poor posture resulting from muscle hypotonia. As the disorder progresses, patients lose the ability to use their hands in a purposeful way and involuntary stereotypic hand activities become more evident. These stereotypies are the hallmark of Rett syndrome and consist of hand washing or hand wringing, clapping, flapping as well as hand-mouthing or hair twirling (Hagberg, 1995). In this stage a wide range of features may manifest. These include social withdrawal, loss of language, irritability and self-abusive behaviour, as well as the manifestation of autistic features such as an expressionless face, hypersensitivity to sound, lack of eye contact and disinterest to the surrounding environment. Most of the girls suffer from breathing abnormalities, apnea and seizures, which range from easily controlled to intractable epilepsy (Jian et al., 2006; Nomura and Segawa, 2005).

Stage 3 or stationary stage

Symptoms stabilize in the period spanning from childhood to adulthood. Autistic-like features begin to ameliorate and the child again begins to interact with the environment between 5 to 10 years old. However, intellectual abilities are severely affected. Motor

debility slowly progresses to muscle hypertonus, joint contractures and scoliosis, which can lead to wheelchair dependency. In addition, breathing and cardiac abnormalities become evident during this third stage. The phenotypic severity increases until 15 years old and then plateaus at 25 years of age, leading to a stable condition in adulthood (Chahrour and Zoghbi, 2007; Nomura and Segawa, 2005).

Stage 4 or late motor deterioration stage

In adulthood, although epileptic episodes are less frequent and less severe, patients can live in a severely debilitated physical condition with muscle rigidity and parkinsonism. The degree of intellectual disability remains severe and virtually no meaningful speech is retained (Chahrour and Zoghbi, 2007; Nomura and Segawa, 2005).

Patients with RTT can live into the fourth and fifth decades of life, however many die earlier due to various complicating factors. Causes of death include: sudden unexpected death in epilepsy (SUDEP), brainstem autonomic failure with or without cardiac arrhythmias, pneumonia/aspiration, acute gastric dilatation and rupture, inflicted injury and medication-related problems (Freilinger et al., 2010).

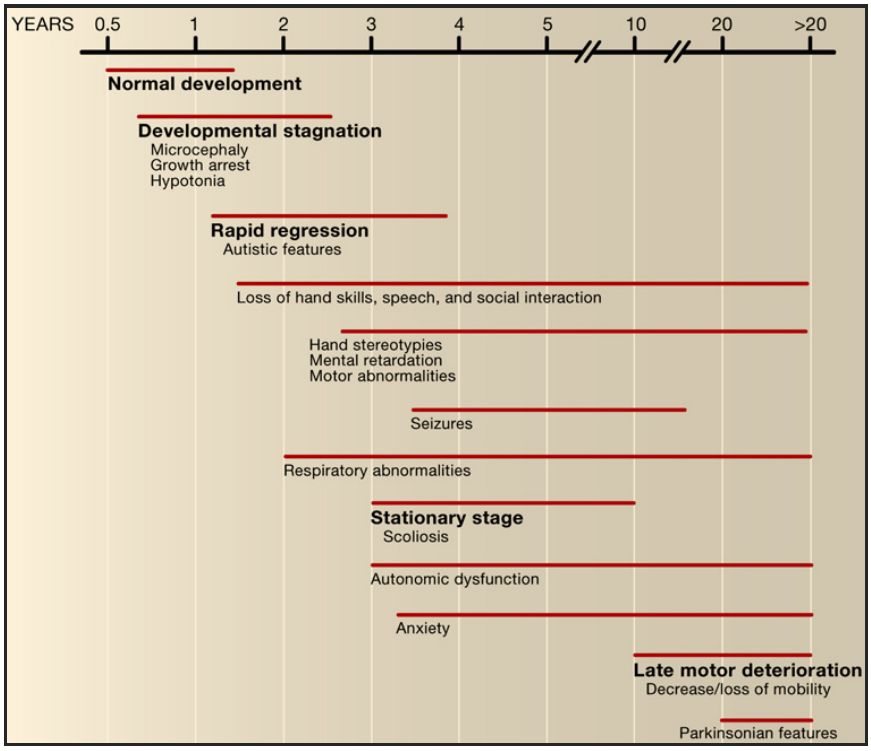


Figure 1.2: The stage progression of classical Rett syndrome

A normal prenatal and early psychomotor development (stage 1) is followed by development deterioration, loss of acquired speech and onset of stereotypic hand movements (stage 2). Stage 3 can last for years with improvement of the communicative skills and degeneration of neuromotor problems. Progression ends with devastating motor deterioration. Picture from Weaving *et al*., 2005.

1.2.1 Atypical Rett syndrome

Hagberg and Gillberg described the presence of 5 “Rettoid” variants, called also Rett syndrome variants or atypical forms, that may have more or less severe features than classical RTT (Hagberg and Gillberg, 1993). Some of these atypical forms have been recognized in only a small number of cases, making it difficult to create any clear statement concerning the clinical features. However, three distinct atypical forms have been reported in a number of patients the so-called *Zappella, Hanefeld* and *Rolando* variants (*Table 1.2*) (Neul et al., 2010).

The mildest of the 5 “Rettoid” variants, ***forme fruste***, has a late age of onset, typically between 1 to 3 years, compared to the classical variant. The second mildest variant known as the ***preserved speech variant*** (*PSV or Zappella variant*) is characterized by the preservation of some degree of speech and mobility where some patients are able to articulate a few words, even if not in the right context, and use their hands to a larger degree. (Zappella et al., 1998; Zappella et al., 2001). The ***early onset seizure variant*** (*Hanefeld variant)* is characterized by seizure onset between the first week and 5 months of age, and have a very severe reduction of psychomotor development in association with severe hypotonia (Hanefeld, 1985). The ***late regression variant*** is characterized by late onset seizures which usually occur in the third stage and lesser overall clinical severity (Moser et al., 2007; Steffenburg et al., 2001). The most severe variant is the ***congenital form*** (*Rolando form*) which manifests from the first months of life and where girls show psychomotor delay, hypotonia, electroencephalogram abnormalities and severe microcephaly (Rolando, 1985).

Rett syndrome was initially assumed to predominantly occur only in girls and to be lethal in males due to the location of *MECP2* on the X-chromosome. However, few reports of males affected with RTT phenotype and mutations in MECP2 gene are described in literature. Affected males with pathogenic *MECP2* mutations, already described in RTT females, are characterized by neonatal encephalopathy, leading to early death after a severe developmental delay, hypotonia, seizures and respiratory abnormalities (Schanen et al., 1998; Villard, 2007). However, mosaic patients for these mutations or 47, XXY Klinefelter patients manifest the classical Rett phenotype (Hoffbuhr et al., 2001; Leonard et al., 2001; Schwartzman et al., 2001). Patients with *MECP2* mutations, that are not found in RTT females, manifest moderate to profound mental retardation or psychiatric disorders (Khajuria et al., 2011; Lynch et al., 2003; Masuyama et al., 2005; Wan et al., 1999). Male patients with duplication of the whole *MECP2* gene are characterized by a

infantile hypotonia, recurrent respiratory infection, severe MR, absence of speech development, seizures and spasticity (Del Gaudio et al., 2006; Friez et al., 2006; Meins et al., 2005; Smyk et al., 2008).

Table 1.2: The variant forms of atypical Rett syndrome

Preserved Speech Variant (Zappella Variant)	Early onset Variant (Hanefeld Variant)	Congenital Variant (Rolando Variant)
Clinical features: • Regression at 1-3 yrs, prolonged plateau phase • Milder reduction of hand skills • Recovery of language after regression - mean age of recovery is 5 yrs. - single words or phrases • Milder intellectual disability (IQ up to 50) • Autistic behaviour common • Decreased frequency of typical RTT features - rare epilepsy - rare autonomic dysfunction - milder scoliosis and kyphosis - normal head circumference	*Clinical features:* • Early onset of seizures - before 5 months of life - infantile spasm - refractory myoclonic epilepsy - seizure onset before regression • Decreased frequency of typical RTT features - rare epilepsy - rare autonomic dysfunction - milder scoliosis and kyphosis - normal head circumference	*Clinical features:* • Grossly abnormal initial development - severe psychomotor delay - inability to walk • Severe postnatal microcephaly before 4 months • Regression in first 5 months • Lack of typical intense "RTT" eye gaze • Typical RTT autonomic abnormalities present • Specific movements abnormalities

The diagram shows the three atypical variants forms of Rett syndrome after the clinical revision. The diagram is adapted from Neul *et al* 2010.

1.3 The Neuropathology of Rett syndrome

Rett syndrome is a disorder of the central nervous system that affects the cerebral cortex, basal ganglia, limbic system, cerebellum, brainstem, caudate nucleus and the autonomic system. RTT in not considered a disorder of neurodegeneration, but rather a neurodevelopmental disorder (Armstrong, 2001; Shahbazian et al., 2002a).

Neuroanatomical examination has shown that the RTT brain weighs less than the brain of controls when matched for both age and height (*Figure 1.3*) (Armstrong et al., 1999). The weight of other organs such as the heart, liver, kidney and spleen are within the normal range until 8-12 years of age, at which stage they start to deviate from normal. These findings support the idea that RTT is not a progressive degenerative disorder, but it is more likely the result of incomplete development or deceleration of the growth rate (Armstrong, 2001; Armstrong et al., 1999). Morphological studies in RTT post-mortem brain samples show a characteristic neuropathology, including decreased neuronal size (*Figure 1.3B*) and increased neuronal density in the cerebral cortex, hypothalamus and the hippocampal formation (Bauman et al., 1995). Other studies have reported

abnormalities in dendritic growth and spine dysgenesis in the frontal cortex and motor cortices pyramidal neurons (Belichenko et al., 1997).

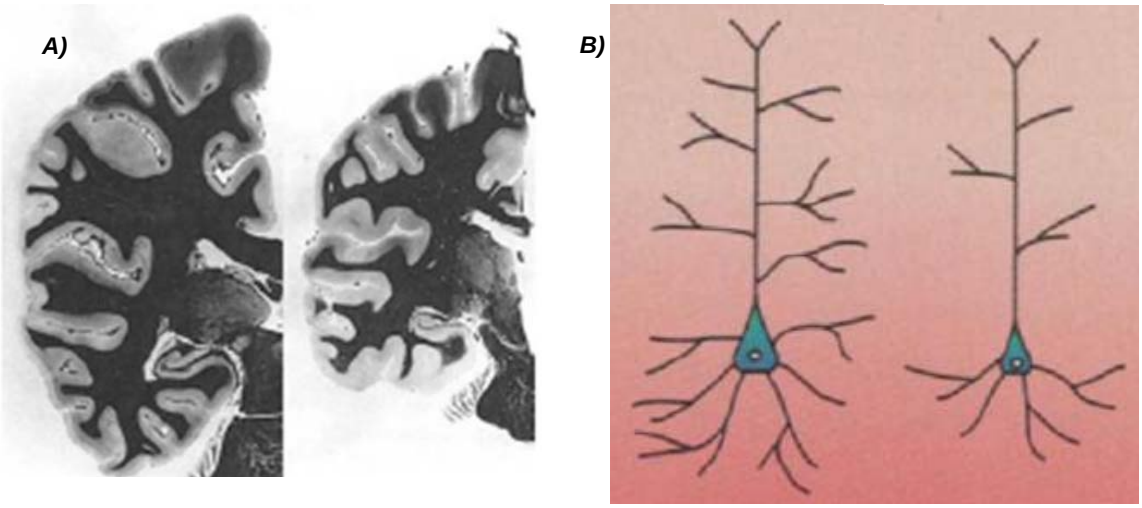


Figure 1.3 : Brain and dendritic abnormalities in RTT patients

A) A brain coronal section of a normal control (left) versus age matched (21 years old) RTT patients (right) showing overall reduction in size of RTT brain (Jellinger, 2003).
B) Schematic representation of pyramidal neurons from control and RTT brains. Reduction in the cell body and in dendritic branching are showed (Zoghbi, 2003).

1.4 The *MECP2* gene

1.4.1 The Methyl Binding Protein Family

Methyl-CpG binding proteins belong to a family of proteins conserved in mammals with a common conserved 70 amino acid sequence that binds single methylated CpGs: the methyl binding domain (MBD) (Hendrich and Bird, 1998).

In vertebrates, proteins containing the MBD constitute a large and well-studied family of proteins which includes MeCP2, MBD1, MBD2, MBD3 and MBD4 (Cross et al., 1997; Filion et al., 2006; Hendrich and Bird, 1998; Jorgensen and Bird, 2002; Prokhortchouk et al., 2001). The MBD is not the only protein domain that can permit recognition of methylated DNA; for instance, the protein Kaiso uses a three-zinc-finger motif to bind DNA and includes Kaiso (ZBTB33), ZBTB4 and ZBTB38 (Filion et al., 2006; Prokhortchouk et al., 2001). Six other proteins have a MBD related domain: BAZ2A (TIP5), BAZ2B, KMT1E (ESET/SETDB1), KMT1F (CLLD8/SETDB2) and two uncharacterized proteins, KIAA1461 and KIAA1887, that were renamed MBD5 and MBD6 based on their homology to the MBD domain (Laget et al., 2010).

MeCP2, MBD1 and MBD2 proteins can bind DNA through their MBD domains. MeCP2, MBD1, MBD2 and MBD3 all have a transcriptional repressor domain although they are members of different transcription complexes (*Figure 1.4a*). In contrast to these functions, MBD4 acts as a DNA repair protein (Hendrich et al., 1999). Additionally, a glutamate repeat (E-repeat) and a DNA glycosylase domain have been found in MBD3 in MBD4 respectively, with no specific function identified. Other proteins such as Kaiso, ZBTB4 and ZBTB38, are able to bind methylated DNA through their zinc-finger domain. ZBTB4 and ZBTB38 require only one methylated CpG for binding, while Kaiso requires two.

The molecular structure of the MBD domain of MBD1 and MeCP2 has been characterized through X-ray and nuclear magnetic resonance spectroscopy studies. The domain folds into an α/β sandwich structure comprising a layer of a twisted β sheet, backed by another layer formed by the α helix and a hairpin loop at the C-terminus. The β sheet is composed of two long inner strands (β2 and β3), and of two shorter outer strands (β1 and β4). The loop L1, that links β2 and β3, and the N-terminus of the α helix, form a hydrophobic pocket that is able to insert into the DNA major groove and interact with the methyl CpG groups (*Figure 1.4b/c*) (Ho et al., 2008; Ohki et al., 2001)

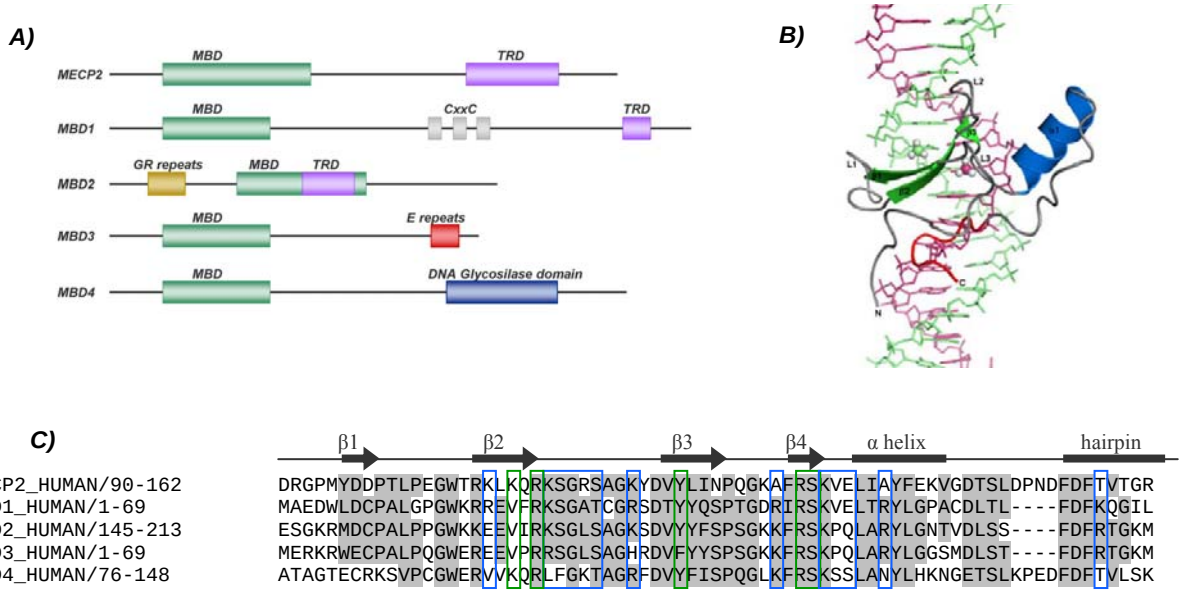


Figure 1.4: The molecular structure of the Methyl CpG Binding Protein family

A) A schematic representation of the methyl CpG binding (MBD) protein family is shown. MBD = methyl CpG binding domain, TRD = transcriptional repressor domain, CxxC = cystein rich domain, GR repeats = glycine and arginine repeats, E repeat = glutamate repeat. Adapted from Dhasarathy *et al*, 2008.
B) The ribbon diagram shows the molecular structure of the *MECP2* MBD domain in complex with the DNA. The methyl groups of the mCpG pair are shown as spheres. The b strands (green) and a helix (blue) are connected by the loops L1, L2 and L3.(Ho et al., 2008) C) Alignment of the MBD domain. Secondary structure elements from the X-ray structure of MeCP2 are denoted by arrows (β strand) and bars (α helix- hairpin). Conserved residues are shaded in grey, residues important for DNA binding are boxed in blue, while residues contacting the cytosine methyl group are boxed in light green.

1.4.2 The Methyl-CpG binding Protein 2 gene

The *MECP2* gene (OMIM *300005) maps on the human X chromosome (Xq28) between the L1 cell adhesion molecule gene (*L1CAM*, OMIM #308840) and the RCP/GCP vision loci (Adler et al., 1995; Amir et al., 1999; D'Esposito et al., 1996). The gene has 4 exons spanning a region of more than 75 kb in length (Moretti and Zoghbi, 2006) and encodes a ubiquitous DNA binding protein with a high affinity for methylated CpG dinucleotides.

The *MECP2* gene encodes two different protein isoforms: MeCP2_e1 (also known as MeCP2B or MeCP2α) which is 498 amino acids in length and is encoded by exons 1, 3 and 4, and MeCP2_e2 (also known as MeCP2A or MeCP2β) which is 486 amino acids in length and is encoded by exons 2, 3 and 4 (*Figure1.5*). It was demonstrated that *MECP2_e1* expression levels are 10 times higher in the human adult brain and cerebellum than the *e2* isoform (Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004).

The *MECP2* 3'UTR is one of the longest UTRs documented. There are 4 different transcripts associated with multiple polyadenylation sites with different lengths ~1.8, ~5.4, ~7.2 and ~10.2 kb (Pelka et al., 2005) (*refer to Section: 1.9.1 The long 3'UTR of MECP2*).

1.4.3 The Methyl-CpG binding Protein 2 protein

The MeCP2 protein (UniProtKB/Swiss-Prot P51608) is a 53 kDa nuclear protein, and a member of the Methyl-CpG binding protein family (*Figure 1.5*). It is composed of four functional domains: the highly conserved Methyl-CpG binging domain (MBD; amino acid 78-163) which is necessary to bind DNA; the central transcriptional binding domain (TRD; amino acid 205-310) required for transcriptional silencing, two nuclear localization signals (NLS; amino acid 174-190 and 255-271) (Kudo, 1998; Nan et al., 1996) and the protein-protein interaction domain (WW; amino acid 384-387) characterized by two tryptophan (W) residues, spaced by 20-22 amino acid that interact with proline-rich motifs (Buschdorf and Stratling, 2004).

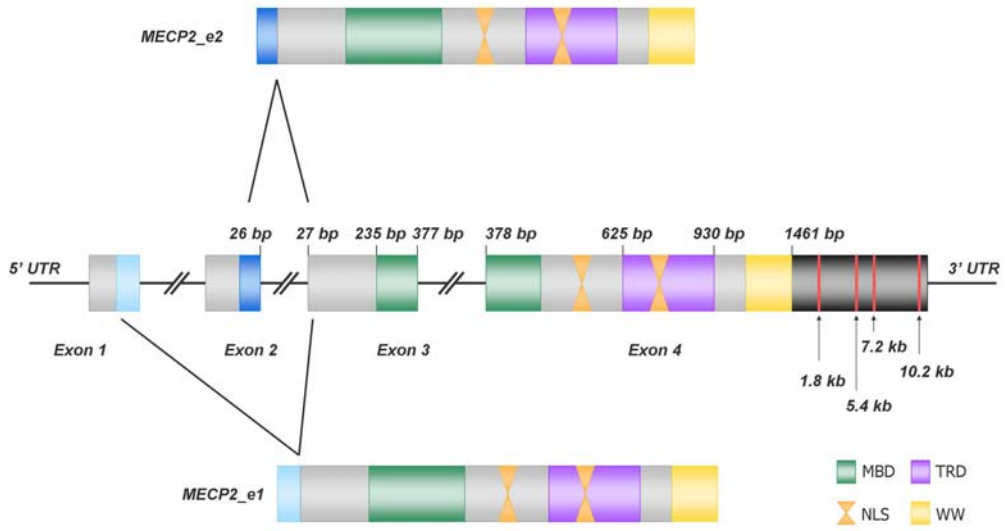


Figure 1.5: The structure of *MECP2* gene and its proteins

The gene consists of 4 exons and there are 2 mRNA splice variants. The two isoforms differ in their N-terminal region; MeCP2_e1 has 486 amino acids and MeCP2_e2 has 498 amino acids. There are different domains: the methyl-CpG binding domain (MDB) is represented in green, the transcriptional binding domain (TRD) is represented in purple, the nuclear localization signals (NLS) are represented in orange and the WW domain is represented in yellow. The long 3'UTR (not drawn to scale) is represented in black and the poly(A) tracts are represented in pink.

1.5 The functions of MeCP2

Initially MeCP2 was shown to bind to methylated CpG dinucleotides. Epigenetic studies have changed the historical concept that MeCP2 is only a transcriptional repressor of methylated gene promoters and has demonstrated new roles for MeCP2 in chromatin remodelling and as a modulator of alternative splicing of pre-mRNA. In addition, MeCP2 is also able to bind to the unmethylated promoters of activated genes and to associate with the transcriptional factor CREB1 (Chahrour and Zoghbi, 2007; Lasalle and Yasui, 2009).

1.5.1 MeCP2 as a DNA binding protein

MeCP2 binds specifically to the 5-methylcytosine residues in single, symmetrical CpG dinucleotides at unmethylated (Yasui et al., 2007), hemimethylated or fully methylated DNA sites (Nan et al., 1996). Recent studies have shown that a sequence motif of (A/T)≥4, adjacent to the CpG sites, improves MeCP2 binding to methylated and unmethylated DNA (Ghosh et al., 2010; Klose et al., 2005). MeCP2 has also been shown to be able to bind to DNA packed in nucleosomes. Indeed the Histone H1 is displaced

from chromosome by MeCP2 suggesting that MeCP2 does not require exposed DNA (Nan et al., 1997).

1.4.2 MeCP2 as a transcriptional repressor

Nan *et al* in 1998 suggested a possible link between MeCP2 binding to methylated DNA and the transcriptional repression machinery. Co-immunoprecipitation experiments in HeLa cells (Nan et al., 1998) and *Xenopus* embryos (Jones et al., 1998) have shown that the TRD of MeCP2 interacts with the co-repressor complex containing Sin3A (paired amphipathic helix protein Sin3A) and HDAC1/2 (histone deacetylases) to form a transcriptional repressor complex. This complex mediates the alteration of the chromatin architecture by promoting nucleosome clustering which consequently results in the silencing of genes with methylated CpG sites (*Figure 1.6a*) (Jones et al., 1998; Wade, 2005).

Trichostatin A, an HDAC inhibitor, has been demonstrated to partially reverse the MeCP2 transcriptional activity. Moreover, *in vitro* experiments have shown MeCP2 to repress the transcription of methylated promoters and SV40- or GAL4-containing promoters (Bird and Wolffe, 1999; Nan et al., 1997; Yu et al., 2000). In addition, it has been demonstrated that MeCP2 interaction with Sin3A is not stable in extracts from rodent tissues, cultured cells, or *Xenopus laevis* oocytes. These data suggest that other factors may be involved in the co-repressor complex to modulate gene expression (Klose and Bird, 2004). MBD2 and MBD3 have been found to interact with Mi2/NuRD, a member of the SWI/SNF superfamily of ATPases that repress transcription (Zhang et al., 1999). Harikrishnan *et al.* has demonstrated that MeCP2 is able to associate with Brahma (BRM), a subunit of the human SWI/SNF ATPase-dependent remodelling complex. These complexes alter nucleosome structure and DNA accessibility to transcriptional machinery (Harikrishnan et al., 2005).

It has also been proposed that the presence of a transcriptional repressor mechanism does not involve the presence of HDACs. It was known that DNMT1 (DNA methyltrasferase) preserves DNA methylation patterns after DNA replication and was able to interact with HDACs (Fuks et al., 2000). A possible interaction between MeCP2 and DNMT1 was then suggested and a close interaction between these two proteins was confirmed in the absence of HDAC proteins (Kimura and Shiota, 2003). This discovery implies that HDACs, DNMT1 and MeCP2 may interact in related pathways in which DNA

methylation and histone acetylation affect chromatin organization by maintaining a stable epigenetic state (*Figure 1.6b*).

Besides DNA methylation and histone deacetylation, histone methylation is emerging as an important epigenetic mechanism linked to the organization of chromatin structure and the regulation of gene expression. In 2003, it was shown that MeCP2 facilitates methylation of Histone H3 at Lysine 9 (H3K9) (Fuks et al., 2003). These results indicate that MeCP2 acts as a fundamental link between DNA methylation, histone deacetylation and histone methylation, and thus reinforces the repressive function of these forms of epigenetic modifications.

1.4.3 MeCP2 as a transcriptional activator

MeCP2 interacts with the Y-box-binding protein 1 (YB-1, also known as p50 and EF1A), which is the major component of messenger ribonucleoprotein particles (mRNPs), mRNA packaging proteins and mRNA template activity in protein synthesis. Moreover, the MeCP2/YB-1 complex appears to function as a regulator of alternative splicing of many genes (Young et al., 2005).

Further analysis has revealed that there is a correlation between the binding of MeCP2 to the promoters of specific genes and their expression. In fact, MeCP2 binds to the *JUNB* promoter (proto-oncogene component of the AP1 transcript complex) when it is transcriptionally active (Yasui et al., 2007).

Furthermore, in 2008 it was shown that MeCP2 plays a role in the transcriptional activation of neurons (*Figure 1.6c*). Chahrour *et al*. performed gene-expression microarray analysis of the hypothalamus from Mecp2-null and Mecp2-duplication mice. Changes in the gene expression levels were found in both of the mouse models, and showed that approximately 85% of the genes were activated by MeCP2 in the hypothalamus. Subsequently, it has been demonstrated that MeCP2 binds to the promoter regions of activated genes (*Sst, Opk1, Gamt and Gprin1*), as well as repressed genes (*Mef2c* and *A2bp1*). Moreover, it was shown that a physical and functional cooperation exists between MeCP2 and the transcription factor, CREB1 (OMIM *123810), at the promoter region of *Sst* (gene upregulated by MeCP2), while no MeCP2 and CREB1 cooperation was detected at the promoter region of *Mef2c* (gene repressed by MeCP2) (Chahrour et al., 2008) supporting the role for MeCP2 as transcriptional activator.

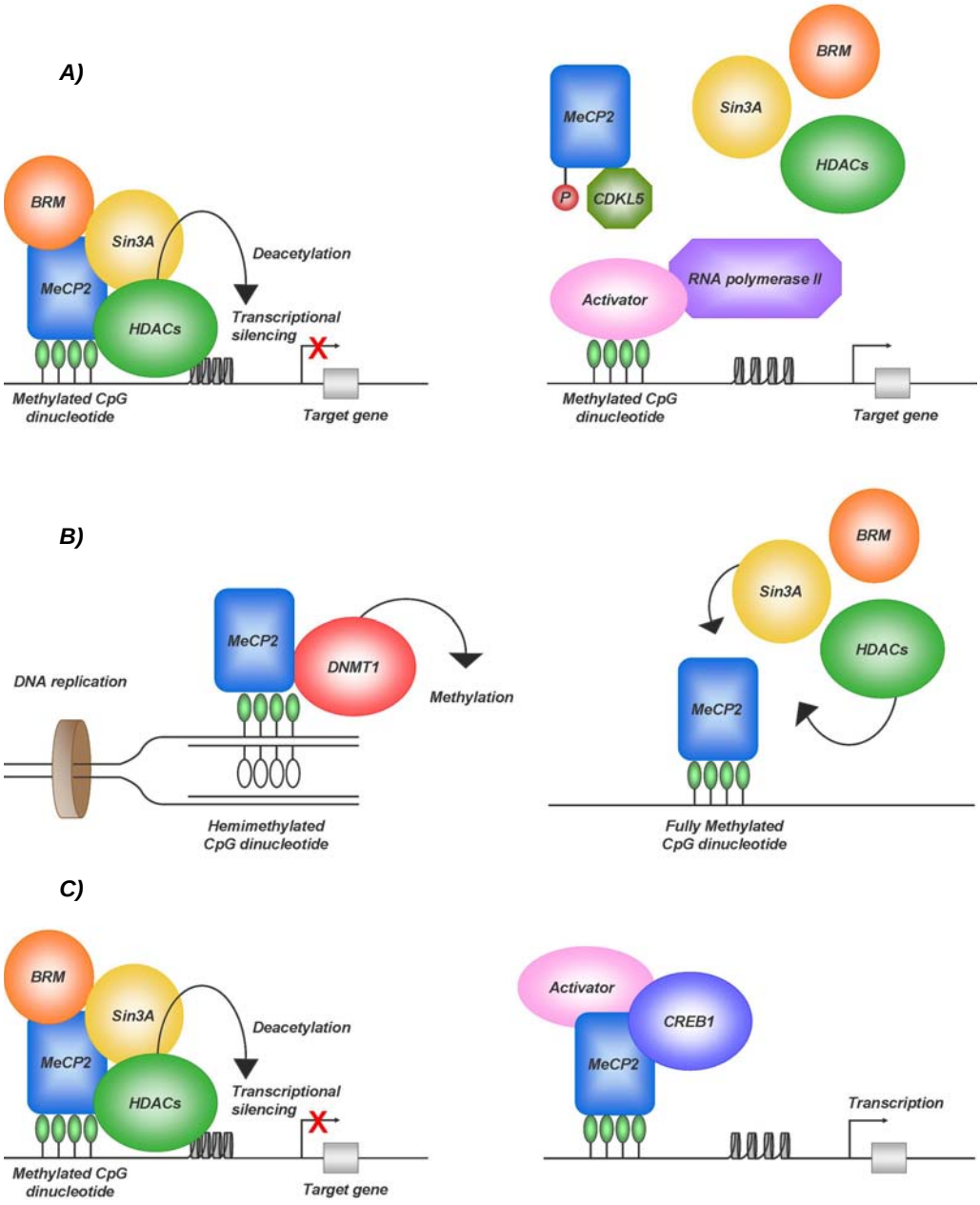


Figure 1.6: Schematic representation of MeCP2 regulation of transcription, methylation and chromatin remodelling

A) MeCP2 binds to methylated DNA sites and recruits chromatin remodelling complexes such as transcriptional co-repressor Sin3A, histone deacetylases (HDACs) and a SWI/SNF chromatin remodelling protein BRM. This complex leads to modification of chromatin integrity (condensation) with consequent inaccessibility of the transcriptional machinery to promoter regions. When MeCP2 is not bound, the chromatin remodelling complexes are not required. The lack of MeCP2 binding to CpG sites could be due to the activity of CDKL5 (cyclin-dependent kinase like 5), which may bind to and phosphorylate MeCP2 resulting in its inactivation. Picture from Bienvenu & Chelly (Bienvenu and Chelly, 2006).
B) During DNA replication DNMT1 binds to hemimethylated CpG sites in double stranded DNA and interacts with MeCP2. The complex then methylates the cytosine in the newly synthesized strand. MeCP2 could change conformation and recognize fully methylated double-stranded DNA leading to the possible formation of the MeCP2 / Sin3A / HDAC complex. Picture adapted from Bienvenu & Chelly 2006 and Kimura & Shiota 2003.
C) MeCP2 binds to methylated DNA sites and recruits chromatin remodelling complexes as transcriptional co-repressor Sin3A, histone deacetylases (HDACs) and a SWI/SNF chromatin remodelling protein BRM. MeCP2 can act as transcriptional activator, recruiting CREB1 protein, thereby activating transcription complexes. Picture adapted from Chahrour *et al*. 2008.

1.5 MeCP2 has specific target genes

Gene expression changes in the brain as well as neuronal and non-neuronal tissues from mice lacking MeCP2 and/or from patients with pathogenic *MECP2* mutations have identified putative MeCP2 target genes (*Table 1.3*) that may contribute to neuronal development and be relevant in the pathogenesis of Rett syndrome.

Table 1.3: MeCP2 target genes

Gene	Function	References
BDNF/Bdnf	Neuronal development and neuronal plasticity, learning and memory	(Chen and Shyu, 1995; Martinowich et al., 2003)
UBE3A/Ube3A	Proteolysis	(Makedonski et al., 2005; Samaco et al., 2005)
DLX5/Dlx5, DLX6/Dlx6	Transcription factor	(Horike et al., 2005)
xHairy2α	Neuronal repressor	(Stancheva et al., 2003)
IGF2	Cell proliferation	(Ballestar et al., 2005)
MPP1	Signal transduction	(Ballestar et al., 2005)
GABRB3/Gabrb3	GABA receptor subunit	(Samaco et al., 2005)
Fkbp5	Hormone signalling	(Nuber et al., 2005)
Sgk1	Ion channel activation	(Nuber et al., 2005)
ID1/Id1, ID2/Id2, ID3/Id3	Transcriptional regulation	(Peddada et al., 2006)
Uqcrc1	Mitochondrial respiratory complex subunit	(Kriaucionis and Bird, 2004)
Crh	Anxiety and stress response	(McGill et al., 2006)
IGFBP3/Igfbp3	Hormone signalling	(Itoh et al., 2007)
FXYD1/Fxyd1	Na+ / K+ - ATPase activity	(Deng et al., 2007; Jordan et al., 2007)
Reln	Cell signalling	(Jordan et al., 2007)
Gtl2/Meg3	Non-coding RNA	(Jordan et al., 2007)
JUNB	Early response gene, oncogene	(Yasui et al., 2007)
RNASEH2A	Ribonucleotide cleavage from RNA-DNA complex	(Yasui et al., 2007)
Sst	Hormone signalling	(Chahrour et al., 2008)
Oprk1	Opioid receptor	(Chahrour et al., 2008)
Gamt	Methyltrasferase	(Chahrour et al., 2008)
Gprin1	Neurite formation	(Chahrour et al., 2008)
Mef2c	Myogenesis	(Chahrour et al., 2008)
A2bp1	Splicing	(Chahrour et al., 2008)
Creb1	Transcriptional co-activator	(Chahrour et al., 2008)
DNM1	Synaptic vesicle endocytosis	(Gibson et al., 2010)
CLU	Chaperone function	(Gibson et al., 2010)
CO1	Cytochrome c oxidase subunit 1	(Gibson et al., 2010)

The target genes with altered expression in the presence of dysfunctional MeCP2 protein are listed.

One of the most studied targets of MeCP2 is a neuronal activity-dependent gene called brain derived neurotrophic factor (*BDNF*, OMIM *113503). BDNF plays an important role in normal brain development, in learning and memory. It has been implicated in several diseases such as susceptibility to memory impairment, bipolar disorder (OMIM #125480), schizophrenia (OMIM #181500), obsessive-compulsive disorder (OCD, OMIM #164230), Huntington disease (HD, OMIM #143100) and Alzheimer disease (AD, OMIM #104300).

It has been shown that BDNF levels are reduced in Mecp2 loss-of-function mutant mice (Chang et al., 2006) and restoring BDNF in the brains of Mecp2 mutant mice ameliorates

many of their RTT-like physiological and behavioural deficits (Larimore et al., 2009). Recently, the role of a BDNF functional polymorphism (p.V66M) on the severity of Rett syndrome has been investigated. Recent clinical reports have studied the correlation between the *BDNF* genotype and RTT phenotype and they have given contradictory results. Nectoux *et al*. speculated that the presence of the p.V66M polymorphism correlates with later age of onset of epilepsy (protective effect against early onset seizure) in RTT patients (Nectoux et al., 2008). Conversely, Ben Zeev *et al*. showed that p.V66M polymorphism increases the risk of seizure onset, particularly in those patients with the p.R168X mutation (Ben Zeev et al., 2009). This latter study showed that the p.V66M variation is associated with an earlier age of epilepsy onset, whereas the p.V66 protein has a protective effect (Nissenkorn et al., 2010).

Another putative target is the ubiquitin ligase E3A gene (*UBE3A,* OMIM *601623) which is located on chromosome 15. Mutations in this gene have been reported in patients with Angelman syndrome (AS), a neurodevelopmental disorder that shows overlapping clinical features with Rett syndrome (Watson et al., 2001). *MECP2* mutations and expression defects have been demonstrated in AS brain samples corroborating the idea that MeCP2 may regulate the expression levels of *UBE3A*. The complexity of the *UBE3A* locus and the differences in assays, reagents, and developmental time points, used by the different studies, can explain some discrepancies in the results between groups that had identified *UBE3A* as a MeCP2 target gene.

Both Samaco and Makedonski demonstrated a reduction of *UBE3A/Ube3a* and *GABRB3/Gabrb3* in adult Mecp2-deficient mice (Samaco et al., 2005). Makedonski reported a reduction of *UBE3A* expression in RTT patients brains and neonatal Mecp2-deficient mice and proposed an aberrant imprinting model of the 15q11–q13 chromosome (Makedonski et al., 2005). On the other hand, Jordan *et al*. showed that *Ube3a* expression was not altered in *Mecp2*-deficient brain mice (Jordan and Francke, 2006).

MeCP2 has also been shown to increase the expression of *Dlx5* and *Dlx6* (Distal-less Dll 5, OMIM*600028, and Dll6, OMIM *600030), a homeobox gene family involved in various developmental processes, including limb formation and neurogenesis (Horike et al., 2005). Chromatin immunoprecipitation and chromatin conformation capture have shown that MeCP2 mediates the formation of an 11 kb chromatin loop bringing together the *DLX5* and *DLX6* sequences, thus leading to a transcriptionally silent acetylated and methylated chromatin loop configuration (Horike et al., 2005). This loop was found to be

absent in chromatin of MeCP2-null mice brains, suggesting that the formation of a silent chromatin loop is possible mechanism of gene regulation by MeCP2.

1.6 Mutations in the *MECP2* gene

Most of the mutations in *MECP2* are *de novo* and have been reported in more than 90-95% of classical Rett patients (Weaving et al., 2005). More than 270 unique pathogenic *MECP2* mutations have been identified, which include missense mutations that change a single amino acid, nonsense mutations that result in a truncated incomplete or unstable protein, small deletions or insertions leading to frame-shifts, as well as the deletion of whole exons or the deletion of one copy of the entire *MECP2* gene (*Figure1.7)*.

The most common and recurrent mutations (p.R106W, p.R133C, p.T158M, p.R168X, p.R255X, p.R270X, p.R294X and p.R306C) are thought to arise as a result of spontaneous de-amination of methylated cytosine residue that causes a C to T transition at CpG dinucleotide in exons 3 and 4. These alterations give rise to missense and nonsense mutations that constitute approximately 55% of all the pathogenic *MECP2* mutations. (Bebbington et al., 2010; Christodoulou et al., 2003)

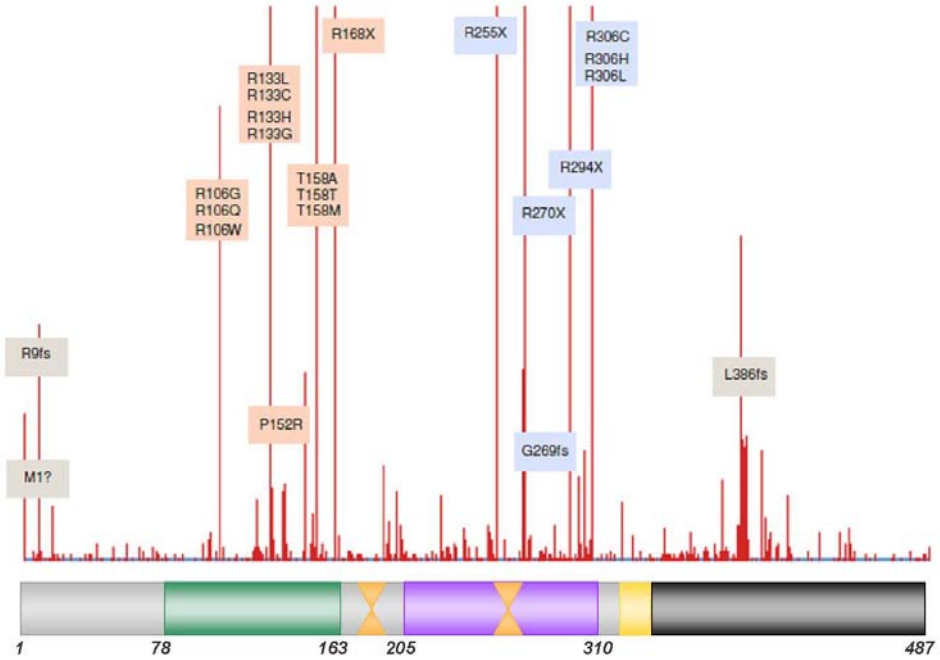


Figure 1.7: Distribution of pathogenic *MECP2* mutations

Pathogenic mutations are distributed along the protein and they are represented by red bars. Mutations that affect the methyl binding domain (MBD) are labelled in red, while mutations affecting the transcriptional repressor domain (TRD) are represented in blue. This image is an adaptation of an image from RettBASE (Christodoulou et al., 2003).

Interestingly, small deletions frequently occur in hotspot regions in exon 4 whilst large deletions involve both exons 3 and 4, which make up approximately a further 15% of all mutations. An 11 bp intragenic deletion prone region (DPR) at the 3' end of exon 4, characterized by short direct repeat elements, has been considered the principal reason for the genomic instability of this region (Laccone et al., 2004).

Different studies have described deletions in *MECP2* using Southern blot, Multiplex Ligation dependent Probe Amplification (MLPA) or Real Time Quantitative PCR (Q-PCR) (Archer et al., 2006b; Ariani et al., 2004; Bourdon et al., 2001a; Erlandson et al., 2003; Laccone et al., 2004; Pan et al., 2006; Ravn et al., 2005b; Scala et al., 2007; Schollen et al., 2003; Yaron et al., 2002). All these studies confirm that large deletions of *MECP2* are a common cause of classical RTT and that gene dosage analysis is an important investigation step for a complete diagnostic strategy.

Until recently, exon 1 was considered non-coding and therefore mutation analysis of only exons 2, 3 and 4 of the gene were conducted. Mutations in exon 1 are indeed rare and present in only 3% of cases. Data from multiple studies have shown the importance of mutations in this exon. As a matter of fact, an 11 bp common deletion, localized in a region with an AAG repeat, creates a frame-shift and consequently a premature stop codon in the MeCP2_e1 protein isoform, with apparent consequences also in MeCP2_e2 isoform translation (Amir et al., 2005; Mnatzakanian et al., 2004; Ravn et al., 2005a; Saxena et al., 2006). Saxena *et al.* detected the 11 bp exon 1 deletion in a subject with a milder RTT phenotype, while Scala *et al*. identified it in a patient with a classical RTT phenotype (Saxena et al., 2006; Scala et al., 2007).

Four public databases have been established to catalogue data on reported mutations in *MECP2* and are good sources to aid laboratories and clinicians who perform diagnostic testing of the *MECP2* gene: RettBASE (http://mecp2.chw.edu.au), MeCP2.org.uk (http://www.mecp2.org.uk/), the Italian Rett syndrome database (www.biobank.unisi.it) and EuroRETT (http://www.eurorett.eu).

1.7.1 Genotype - phenotype correlations

Genotype-phenotype correlation studies have been conducted by different groups to determine whether different types or specific mutations can explain the variability of phenotypes found in Rett syndrome. Information about mutation type and position, X-

inactivation status and severity of clinical manifestations (regression period, epilepsy, scoliosis, hand use, ambulation and language) has been examined.

Ariani *et al*. found no correlation between genotype and phenotype, but patients with truncating mutations had a higher incidence of awake respiratory dysfunction and lower levels of cerebrospinal fluid homovanillic acid (Amir et al., 2000). Conversely, it has been shown that truncating mutations led to a more severe phenotype than missense mutations and in particular early truncations caused a more severe clinical presentation in comparison with C-terminal deletions (Cheadle et al., 2000; Hoffbuhr et al., 2002). Chae *et al*. have reported no significant difference in overall severity between missense and truncating mutations (Chae et al., 2002; Weaving et al., 2003) while Weaving *et al*. have reported that truncation mutations are associated with and overall increased severity of the phenotype (Weaving et al., 2003). Schanen *et al*. demonstrated that patients with missense mutations in the TRD had a best overall scores and better preservation of head growth and language skills (Schanen et al., 2004).

A German and an Australian study showed that mutations in the nuclear localization signal region were associated with a more severe phenotype (Colvin et al., 2004; Huppke et al., 2002). Huppke and Smeets showed that patients with deletions in the transcriptional repressor domain (TRD) and in the C-terminal region caused an overall milder phenotype (Huppke et al., 2002; Smeets et al., 2005).

A better association was found between specific mutations and certain functions such as ambulation, hand use, language and onset of regression. Generally, patients harbouring the p.R133C, p.R294X, p.R306C variations or C-terminal deletions were less severely affected (Bebbington et al., 2008; Colvin et al., 2004; Kerr and Prescott, 2005; Neul et al., 2008; Schanen et al., 2004). In particular, patients with these milder alterations showed a delayed onset of regression, delayed onset of stereotypies and relative preservation of hand skills (Bebbington et al., 2008). The p.R133C was associated with conserved language skills and the p.R270X and p.R255X with “developed” speech (Bebbington et al., 2008). Patients with the p.R168X mutation were not able to walk or use words, but they had retained hand use, whilst those with the p.R306C mutation had improved language skills (Neul et al., 2008). Bebbington and Kerr found the p.R168X and p.T158M variants to be the more severe mutations (Bebbington et al., 2008; Kerr and Prescott, 2005).

Genotype - phenotype correlation analysis studies have produced conflicting data as the pathogenic sequence variations have variable phenotype expressivity. The usage of different clinical severity scoring systems, including the Kerr (Kerr et al., 2001), Percy (Schanen et al., 2004) and Pineda (Monros et al., 2001) scales, varying diagnostic criteria, the effect of age and disease progression, as well the possible effect of skewing of X chromosome inactivation (XCI), would be contributing factors.

Skewed X chromosome inactivation

X chromosome inactivation (XCI) is an important dosage compensation mechanism present in female mammals. XCI normally occurs randomly in differentiating embryonic cells in females resulting in approximately half of the cells expressing one normal allele and the other half expressing the other allele. If the process is not random (skewed), then the X-chromosome with either the normal, or the mutated gene, may be preferentially active in the majority of the cells. Therefore, skewed XCI potentially can contribute to the variability of the Rett syndrome. Skewed XCI may cause the silencing of the mutant *MECP2* allele thereby attenuating the severity of the disorder or may accentuate the severity of the disorder if the wildtype (normal) *MECP2* is preferentially silenced.

Skewed patterns of XCI have been reported in females who were asymptomatic or had a less severe RTT clinical phenotype. In the latter cases, non random XCI attenuates the effects of the *MECP2* mutations by preferential inactivation of the mutant allele (Amir et al., 2000; Hoffbuhr et al., 2002; Knudsen et al., 2006; Van den Veyver and Zoghbi, 2001). However other studies of XCI and phenotype correlation have given discordant results. Nielsen and colleagues did not find an association between skewed X inactivation and phenotype (Nielsen et al., 2001). Moreover, skewed XCI did not provide a valid explanation for the normal phenotype of a mother carrying the same *MECP2* mutation as her RTT daughter (Takahashi et al., 2008). Interestingly, Ishii and colleagues reported a case of monozygotic RTT twins with very different clinical features. The twin with mild symptoms had skewed XCI while the twin with the more severe symptoms presented random XCI (Ishii et al., 2001).

All the above studies were conducted using peripheral blood from patients. However other studies have examined the XCI pattern in the brains of RTT patients. Zoghbi *et al*. found balanced XCI in cortical tissues and moderately skewed XCI in liver, leukocytes and fibroblasts in three RTT patients (Zoghbi et al., 1990). Anvret and Wahlström also showed balanced XCI in brain samples from two RTT patients, but did not specify the

regions tested (Anvret and Wahlstrom, 1994). La Salle and colleagues found balanced XCI in the cerebral cortex and cerebellum of two RTT patients (LaSalle et al., 2001) and Shahbazian *et al*. found random XCI in both cerebral cortex and cerebellum of 10 RTT patients (Shahbazian et al., 2002b). Gibson and colleagues characterized the XCI pattern in neuronal and non-neuronal regions of a cohort of RTT patients, showing no significant differences in frontal and occipital cortex between RTT and controls brains. Moreover, they observed variability of XCI across different neuronal regions and also in non-neuronal tissues such as heart, musucle and kidney (Gibson et al., 2005).

A possible explanation for the differences between all these studies may be attributable to a number of factors such as grouping mutations that have different effects on MeCP2 function, different inclusion criteria (age, small cohorts of patients) and variability in the clinical classification. Thus, it would seem that the variability in the clinical phenotypes of Rett syndrome is probably the result of complex interactions between the type of mutation, functional domain affected, X-chromosome inactivation and other factors, including enviromental factors, yet to be identified.

1.7.2 Sequence variations in the 3'Untranslated regions and Human diseases

Traditionally only mutations in the coding regions of genes have been associated with pathological phenotypes. Variants in the 5' untranslated region (5'UTR) or in the gene promoter region may interfere with transcription or translation and variants located in the coding sequence and in introns may affect mRNA or protein expression and the regulation of splicing. A growing number of studies are revealing that the 3' untranslated region (3'UTR) of mRNA is implicated in the regulation of gene expression. Indeed, a recent study based on the Human Gene Mutation Databases (HGMD, http://www.hgmd.org) estimated that around 0.2% of the disease-associated mutations reside in the 3'UTR regulatory region of genes (Chen et al., 2006).

Sequence variations in the 3'UTR have been associated with a broad range of pathologies, including: myotonic dystrophy (DM, OMIM #160900), a neuromuscular disease characterized by dystonia, mental retardation and muscular development defects (Timchenko, 1999); immunodysregulation, polyendocrinopathy and enteropathy X-linked syndrome (IPEX, OMIM #304790) (Bennett et al., 2001) and congenital heart disease (CHD, OMIM #600576). In addition, defects in the 3'UTR region have been linked with cancer (Lopez de Silanes et al., 2007).

mRNAs are not simply intermediate molecules between DNA and protein, but they have important cellular functions in 3' end formation and polyadenylation, nucleo/cytoplasm export, mRNA subcellular localization, stabilization, degradation and translational efficiency (Chabanon et al., 2004; Conne et al., 2000; Mignone et al., 2002). The 3'UTRs contain *cis*-regulatory elements such as AU-rich elements (AREs), iron response elements (IREs) and secondary structures, that govern the activity of *trans*-acting factors, which in turn influences processes such as translation, storage, turnover and transport (Lopez de Silanes et al., 2007). In fact, mutations affecting the termination codon, polyadenylation signals or secondary structure of the 3'UTR mRNA region may produce translational deregulation, may affect protein expression or may affect the binding of *trans*-acting factors and regulatory proteins. The growing number of 3'UTR-mediated human diseases suggests that this region of the gene may be a "molecular hotspot" for pathology and further investigation may be informative in understanding the role of this region and how alterations can lead to human diseases (Conne et al., 2000).

1.8 MicroRNA biology

MicroRNAs (miRNAs) are a large class of non-coding RNA molecules (ncRNAs) that were first described in *Caenorhabditis elegans* (Lee et al., 1993; Wightman et al., 1993). Since then, miRNAs have been identified in numerous species, tissues and cell lineages.

Recent genome-wide studies have shown that miRNAs are potent regulators of gene expression, affecting a wide range of biological functions such as cell differentiation, cell proliferation, organ development and cell apoptosis (Bartel, 2004). They perform these functions by repression of their target genes. A recent estimation predicted that more than 60% of all human genes are regulated by microRNAs and that a single miRNA can potentially target the 3'UTR of several hundred different transcripts (Silahtaroglu and Stenvang, 2010). Evidence accumulated over recent years indicates that a vast number of normal and pathological pathways are controlled by miRNA (Taft et al., 2010).

Microarray analyses have indicated that miRNAs are particularly abundant in the brain, suggesting a role in neuronal development or plasticity. Dysfunction in miRNA biogenesis pathways and also single nucleotide polymorphisms in the miRNA sequence or in the 3'UTR binding site (disrupting/creating the putative miRNA target binding sites) may potentially contribute to the pathogenesis of disorders. A fine regulation of these processes is essential for correct cellular neuronal development and function. Indeed,

alterations in miRNA regulation have been implicated in the etiology of neurodegenerative diseases such as Alzheimer disease (AD), Parkinson disease (PD), Tourette syndrome (GTS), Huntington disease (HD), Fragile X syndrome (FXS), and Rett syndrome (Chang et al., 2009; De Smaele et al., 2010; Saba and Schratt, 2010).

1.8.1 MicroRNA biogenesis pathway

miRNA genes are transcribed in the nucleus by RNA polymerase II or III into primary miRNA (pri-miRNA) transcripts with a 5' cap and a 3' polyadenylated tail (*Figure 1.8)*. The pri-miRNA contains a hairpin stem-loop of 33 nucleotides, a terminal loop and two single stranded regions flanking the hairpin (Cai et al., 2004).

Following transcription, the nuclear RNase III endonuclease Drosha forms a complex with its cofactor DGCR8 protein (DiGeorge critical region 8, also known as Pasha in *Drosophila melanogaster* and *Caenorhabditis elegans*). The Drosha/DGCR8/Pasha complex cleaves the pri-miRNA transcript, that can be several kilobases long, into a 60-70 nucleotides long precursor miRNA (pre-miRNA). It is believed that the DGCR8 protein stably interacts with the pri-miRNA, while Drosha cleaves the 5' and 3' end of the pri-miRNA (Winter et al., 2009).

After nuclear processing, the pre-miRNA is then exported to the cytoplasm through the nuclear export factor, Exportin-5 (EXP5), which forms a complex with cofactor Ran-GTP. The RNase III endonuclease Dicer forms a complex in association with RNA binding domain proteins (TRBP, Tar RNA binding protein) and the Argonaute (AGO 1-4) protein family. This complex is called the RISC complex (RNA induced silencing complex), and it binds and cleaves the pre-miRNA at the loop level, creating a miRNA duplex about 22 nucleotide long characterized by protruding overhangs at each 3' end. The miRNA duplex contains a mature miRNA guide strand and the passenger strand called miRNA*. The RISC complex is then responsible for the miRNA* passenger strand degradation. At this point the miRNA guide strand binds to the complementary region located in the 3'UTR of the transcript.

The exact aspects of the miRNA action are still unknown and further studies are necessary. Emerging evidence suggests that the biosynthetic pathway is not universal to all miRNAs, but steps can be replaced or omitted. MicroRNAs can be intragenic and they can have their own promoter (monocistronic) or share the same promoter (polycistronic). Moreover, miRNAs can be intronic with their own promoter or they can use the promoter

of their host gene (Ozsolak et al., 2008). In this case, some miRNAs can be processed directly as pri-miRNAs or as mirtrons bypassing the Drosha cleavage (Berezikov et al., 2007). Furthermore, based on studies on small interfering RNA (siRNAs), it is currently thought that the selection of one strand from the miRNA/miRNA* duplex is based on the thermodynamics of the strand's ends: a strand with lower stability at the 5'-end is preferentially selected (Erson and Petty, 2008; Kim, 2005). Ro *et al*. showed that the relative expression levels of the two strands vary among tissues and that both the miRNA strands can be functional (Ro et al., 2007). It seems that correct control of miRNA levels is essential to maintain normal cellular function because deregulation of miRNAs is often associated with human diseases.

1.8.2 MicroRNA and Human diseases

As mentioned before, miRNAs are involved in a wide range of biological processes. Alterations in miRNA processing may affect their functions and consequently may have a pathological role in human diseases. A large number of studies have demonstrated a tight association between miRNA expression and cancer development, but only a few recent studies have established a correlation between miRNAs and human genetic disorders (Erson and Petty, 2008). Genetic alterations in miRNA processing may occur at different levels such as i) point mutations affecting miRNA biogenesis, *ii)* point mutations in miRNA mature sequences and *iii)* point mutations in miRNA binding target sites.

i) Point mutations affecting miRNA biogenesis

Different proteins are involved in the miRNA biogenesis pathway, so loss-of-function mutations in these basic components may be responsible for human diseases. DiGeorge syndrome (DGS, OMIM #188400) is characterized by cardiovascular and craniofacial defects, immunodeficiency and neurobehavioural alterations; the DiGeorge syndrome Critical Region gene *8* (*DGCR8*, OMIM *609030) is commonly deleted in DGS patients. *DGCR8* is an essential member for miRNA biogenesis as it is an essential component of the Drosha complex (Wang et al., 2007). Fragile X syndrome (FXS, OMIM #300624) characterized by moderate to severe mental retardation, macro-orchidism, and distinct facial features, is caused by trinucleotide (CGG) repeat expansions in the *FMR1* gene (OMIM *309550). FMRP1 protein is an RNA-binding protein implicated in mRNA transport at synapses and in recent years FMRP has also been implicated in the recruitment of miRNA-RISC complex (Li et al., 2008). It is possible to speculate that gene haploinsufficiency and alterations in miRNA biogenesis may have a potential impact in synaptic development. However these hypotheses require more investigations before to

support a direct role between altered miRNA process and human genetic diseases could be complicated.

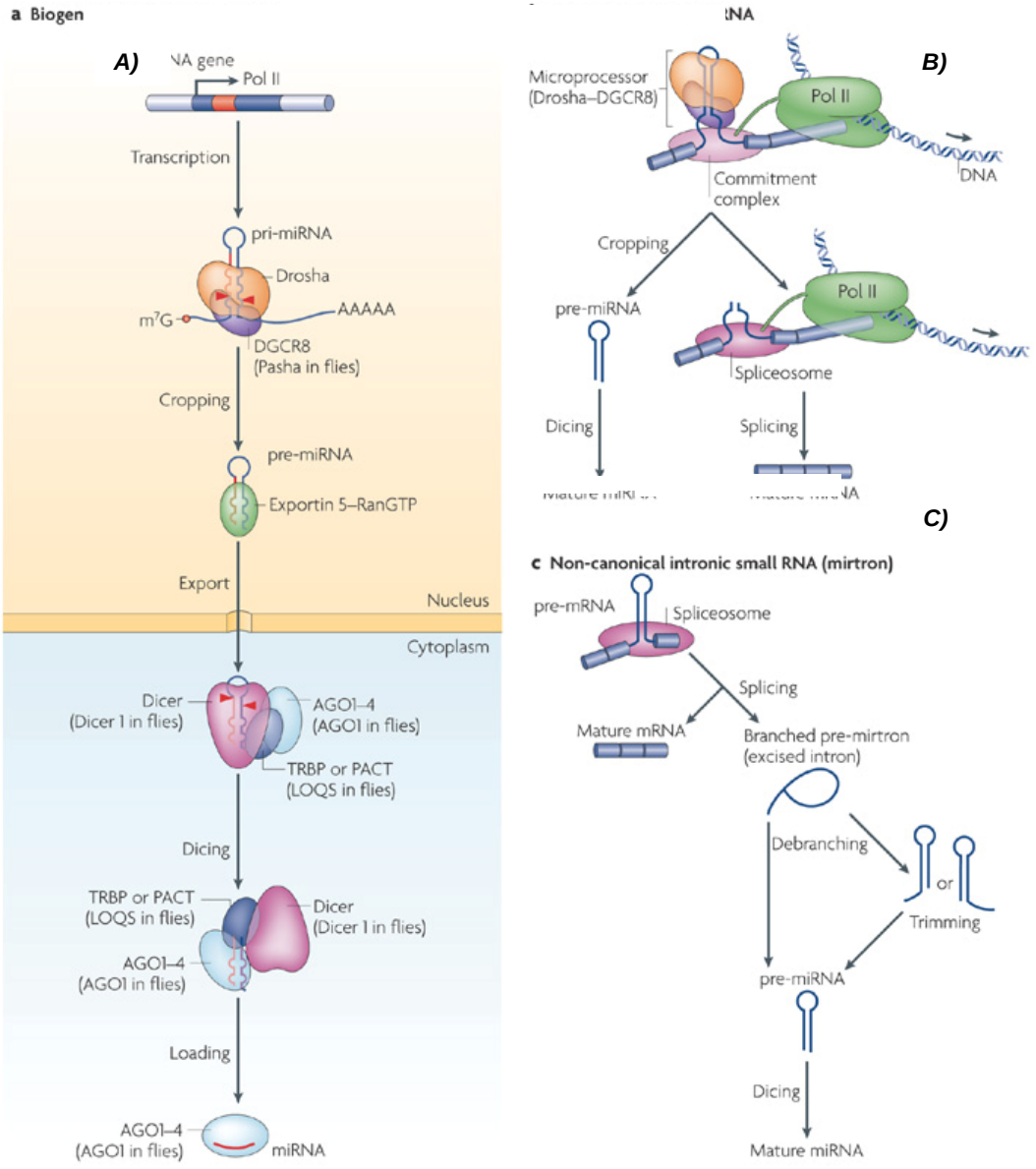


Figure 1.8: MicroRNAs biogenesis in mammals

A) A long primary microRNA (pri-miRNA) is generally transcribed by RNA polymerase II and it is 5' capped and polyadenylated. A nuclear complex Drosha/DGCR8/Pasha (DiGeorge syndrome critical region gene 8, Pasha in Drosophila melanogaster and Caenorhabditis elegans) cleaves the pri-miRNA into a 60-70 nucleotides stem-loop structure (pre-miRNA). Exportin 5 recognises the pre-miRNA and exports it into the cytoplasm after binding to Ran-GTP. The cytoplasmic RNAase Dicer forms a RISC complex (RNA inducing silencing complex) with the RNA-binding partner TRBP and Argonaute. Incorporation of mature miRNA duplex into the complex leads to the degradation of the miRNA* passenger strand. (In brackets the fly components).
B) Canonical intronic miRNAs are co-transcribed before splicing. The splicing complex is thought to bind the introns while Drosha cleaves the miRNA hairpin. It is only then that the pre-miRNA undergoes the miRNA biogenesis pathway.
C) A non-canonical pathway is generated when miRNA expression is controlled from host gene promoter; mirtrons are then generated and they bypass the Drosha-processing step. Picture from Kim *et al.* 2009.

ii) Point mutations in miRNA mature sequences

Point mutations in the microRNA-96 gene (*MIRN96*, OMIM *611606, miR-96+13G>A and +14C>A) were reported for the first time by Mencìa *et al*. in two unrelated families affected by non-syndromic progressive hearing loss (DFNA50, OMIM #613074) (Mencia et al., 2009). The involvement of miR-96 in hearing loss was corroborated by the characterization of the *Mirn96* mouse (Lewis et al., 2009).

iii) Point mutations in miRNA binding target sites

Mutations in the 3'UTR region of mRNA transcripts may create or destroy putative miRNA target sites. This possibility was demonstrated by a study on the Texel sheep, where a mutation (g.+6723G>A) in the 3'UTR of the myostatin gene (*MSTN, Ovis aries* NM_001009428) created a *de novo* target site for two miRNAs (miR-1 and miR-206) that are highly expressed in skeletal muscle. As a result, MSTN is decreased causing muscular hypertrophy in these sheep (OMIA ID: 2813) (Clop et al., 2006).

In the same year, a paper linked 3'UTR mutations with a human genetic disorder. Two 3'UTR changes (c.*43G>T and c.*50G>A) in the receptor expression-enhancer protein 1 gene (*REEP1*, OMIM *609139) have been associated with hereditary spastic paraplegia type 31 (*HSPG31*, OMIM #610250) which affects upper motor neurons and their axonal projections. These variations alter a conserved binding site for miR-140, which is expressed in rat and monkey cortex and also in rat cultured primary cortical neurons (Zuchner et al., 2006). Beetz *et al*. identified the previously reported c.*43 G>T variation and a new change (c.*14C>T) affected miR-691 target sites in two unrelated HSPG31 families (Beetz et al., 2008).

A mutation (c.*281A>T) in the 3'UTR of the histone deacetylase 6 gene (*HDAC6*, OMIM *606543) was found in all affected females and in one male in a family characterized by dominant X-linked chondrodysplasia. This alteration is responsible for the disruption of the miR-433 target site. Consequently the lack of recognition leads to overexpression of HDAC6 protein and to the pathological phenotype (Simon et al., 2010).

Common polymorphisms (SNPs) in the 3'UTR have been demonstrated to modulate microRNA binding ability and to be responsible for some human genetic diseases. An example, the polymorphism rs5186 (c.*1166A/C) in the angiotensin II type 1 receptor gene (*AGTR1*, OMIM 106165) is associated with cardiovascular disease. The C allele variant decreased the ability of miR-155 to bind to its target sequence and consequently *AGTR1* expression was inhibited in fibroblasts. Clinical features such as hypertension,

cardiac hypertrophy, aortic stiffness and myocardial infarction have been associated with the *AGTR1* c.*1166C allele variant, providing the first example of association between a mutation in a miRNA binding site and cardiovascular disease (Martin et al., 2007). In the same year, an independent study confirmed that the *AGTR1* c.*1166C allele is associated with elevated levels of AGTR1 protein due to the abolishment of the miR-155 target site (Sethupathy et al., 2007).

Wang *et al.* identified that the C allele of the polymorphism rs12720208 (c.*166 C/T) disrupts the miR-433 target sites in the 3'UTR of the fibroblast growth factor gene (*FGF20*, OMIM 605558) increasing FGF20 translation. It is known that dysregulation in the *α-synuclein* gene (*SNCA*, OMIM *163890) causes Parkinson disease (PD), so high levels of FGF20 protein can upregulate α-synuclein expression and lead to PD (Wang et al., 2008).

Irritable bowel syndrome, diarrhea-predominant (IBS-D) is a complex disorder associated with dysfunction in the serotonergic pathway. Kapeller *et al*. reported that the A allele variant of rs62625044 (c.*76 G/A) in the serotonin 3E receptor gene (*HTR3E*, OMIM *610123) affected miR-510 binding to the *HTR3E* 3'UTR. The association between the A allele and the diarrhea phenotype of IBS patients was shown in both a UK and German female cohort of IBS-D patients (Kapeller et al., 2008). Subsequently, another gene involved in the serotonergic pathway has been investigated. In 2009, Jensen *et al*. focused their attention on the serotonin receptor 1B gene (*HTR1B*, OMIM *182131) as a deletion in this gene had been previously associated with aggressive behaviour in mice (Saudou et al., 1994). They focused their attention on the rs13212041 (A/G) polymorphism highlighting that the ancestral A allele repressed the expression of the *HTR1B* gene by miR-96. Then, a genotyping study was carried out in 359 college students demonstrating that aggression-related/conduct-disorder phenotype was significantly enriched in individuals with the A allele than individuals with the G allele (Jensen et al., 2009).

A cohort of Tourette Syndrome (TS) patients was analyzed for candidate gene screening. Abelson *et al*. selected among them SLIT and NTRK-like family member 1 gene (*SLITRK1*, OMIM 8609678) as a candidate gene. They detected a 3'UTR mutation (var321) in the target site sequence for miR-189 in two unrelated TS patients. This variation was not detected in 4296 Caucasian and African-American controls demonstrating a significant association with TS (Abelson et al., 2005). However, the involvement of *SLITRK1* var321 in TS has been subsequently questioned by other papers. An Italian and American large family-based studies did not find association

between the variant 321 and Tourette syndrome (Fabbrini et al., 2007; Pasquini et al., 2008; Scharf et al., 2008).

An association between the major histocompatibility complex, class I, G (*HLA-G*, OMIM *142871) and asthma in childhood (OMIM #600807) was observed in an American study conducted in 2007. They reported in particular that the G allele variant of rs1063320 (c.*3142 C/G) affected miR-148 binding to the *HLA-G* 3'UTR. They found that the GG genotype is protective against childhood asthma when the mother is affected, but it is associated with modest risk when the mother is not affected (Tan et al., 2007). However, Moffatt *et al*. did not find an association between rs1063320 or any other HLA-G polymorphisms and childhood asthma (Moffatt et al., 2007).

Alterations in the miRNA-mRNA target site recognition have also been associated with the onset of cancer. The A allele of the rs17084733 (G/A) was shown to affect the miR-221 and -222 binding in the 3'UTR of the v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog gene (KIT, OMIM*164920). This SNP deregulated the expression of the KIT protein and consequently contributed to papillary thyroid carcinoma (PTC, OMIM#188550) (He et al., 2005). Adams *et al*. reported that the polymorphism rs93410170 (C/T) in the 3'UTR of the estrogen receptor 1 gene (*ESR1*, OMIM +133430) had the capacity to modulate the risk of breast cancer. The T allele of the polymorphism rs93410170 (C/T) seemed to enhance miR-206 binding, thereby lowering *ESR1* expression. The authors assumed therefore that this T allele offers some protection against breast cancer, as there is higher frequency of the T allele and lower incidence of breast cancer in Hispanic and European populations (Adams et al., 2007). More recently, Brendle *et al*. analyzed the effect of integrin genes on breast cancer risk. In particular, they focused their attention on the integrin beta-4 gene (*ITGB4*, OMIM *147557), demonstrating that the A allele of the SNP rs743554 (G/A) caused a loss of the miR-34a binding site. A worse survival outcome was associated with this polymorphism, which may influence breast tumour aggressiveness and survival (Brendle et al., 2008). Despite this conclusion more genome-wide association studies are necessary to confirm these associations.

Landi *et al*. reported a positive association between a SNP and colorectal cancer risk. They identified that the C allele of rs17281995 (G/C) in the 3'UTR of the B lymphocyte activator antigen gene (*CD86*, OMIM *601020), was associated with colorectal cancer. This SNP involved the targeted site of five different miRNAs, resulting in a reduced

binding for miR-337, -582 and -200a and in an increased matching for miR-184 and -212 (Landi et al., 2008).

A goal of pharmacogenomic research is the identification of polymorphisms in drug candidate genes (drug metabolizing enzymes, transporters and targets), but recent attention has focused on miRNAs as gene expression regulatory elements. In fact, a SNP rs34764978 (C/T), in the 3'UTR of the dihydrofolate reductase gene (*DHFR*, OMIM +126060), causes miR-24 loss of function and consequently resistance to methotrexate (MTX), an important chemotherapeutic agent commonly used in the treatment of many types of cancer (Mishra et al., 2007).

1.9 The 3'UTR of the *MECP2* gene

1.9.1 The long 3'UTR of the *MECP2* gene

The discovery of a cDNA fragment without an open reading frame between *MECP2* and *IRAK* (interleukin 1 receptor associated kinase 1, OMIM *300283) in humans and mice has led to the identification of an exceptionally long 3'UTR of *MECP2,* which can generate a 10.2-kb alternative transcript observed on Northern blots (Coy et al., 1999; D'Esposito et al., 1996; Reichwald et al., 2000). It became clear that alternative polyadenylation [poly (A)] signals in *MECP2* give rise to four different transcripts ~1.8 kb, ~5.4 kb, ~7.2 kb and ~10.2 kb in length, which are differentially expressed in various tissues (Pelka et al., 2005).

In 1996, Esposito *et al*. showed that the ~1.8kb *MECP2* transcript was ubiquitously expressed in different human adult tissues such as the heart, brain, lung, liver, skeletal muscle, kidney and pancreas (D'Esposito et al., 1996). Except for liver, lung and placenta, the other tissues analysed also displayed a >9.5kb and a >7.5kb transcript (D'Esposito et al., 1996). Subsequently, Coy *et al*. have analysed the ~10.2kb human *MECP2* transcript and demonstrated that this transcript was detectable in fetal heart, brain, lung and kidney and in adult tissues such as heart, brain, placenta, lung, liver, muscle, kidney and pancreas. They demonstrated that the ~10.2kb human *MECP2* transcript is more abundant than the ~5.4kb in fetal brain, but the ~5.4kb transcript is more abundant in the adult brain (Coy et al., 1999). Reichwald and colleagues demonstrated that the ~10.2kb transcript is barely detectable in lung and liver, absent in ovary, but relatively abundant in skeletal muscle, kidney, and pancreas, while the shorter ~1.8kb

transcript is barely detectable in brain and lung, but very abundant in heart, skeletal muscle and spleen. In addition, they identified a weak transcript ~7.5kb in size in heart, brain, skeletal muscle and pancreas (Reichwald et al., 2000).

In 2003, a correlation between higher MeCP2 expression levels and increasing age was discovered in comparison with normal post-mortem human tissues (from fetal to young adult age); moreover, an inverse correlation between the usage of the ~10.2kb long *MECP2* 3'UTR and increasing age was shown, suggesting that a "switch" is present in polyadenylation and it may be responsible for the elevated levels of MeCP2 protein (Balmer et al., 2003). Samaco *et al*. compared post-mortem brain of autistic individuals and age-matched controls; they observed altered levels of *MECP2* expression in four out of five autistic patients without *MECP2* mutations (Samaco et al., 2004). Pelka and colleagues showed using time point quantitative PCR that the levels of the *Mecp2* transcripts in mouse are high in specific areas of the brain (substantia nigra, basal ganglia and occipital cortex) during fetal development and decrease in the young animal and rise again in older animals. (Pelka et al., 2005).

1.9.2 A well conserved 3'UTR

The great length of the *MECP2* 3'UTR raised the question of conservation between different species. The sequence comparison between human and mouse *MECP2* /*Mecp2* 3'UTR sequences exhibit an average of 74% identity (Reichwald et al., 2000). The analysis showed the presence of at least eight blocks of strong sequence similarity distributed along the sequence, interrupted by regions of low homology (Coy et al., 1999; Santos et al., 2008). The functional implications of these conserved blocks is currently unknown, but one may speculate that these sequences are under evolutionary pressure and therefore functionally important (Coy et al., 1999). To see the conserved blocks and the localization of polyalanine tracts refer to *Appendix 1*.

Computational analysis of the human and mouse *MECP2*/*Mecp2* 3'UTR show the presence of a canonical highly conserved in human/mouse 5'-AAUAAA-3' hexamer in the 3'UTR of the ~1.8kb *MECP2* transcript, while the highly conserved in human/mouse, but less efficient, 5'-UAUAAA-3' hexamer is detectable in the 3'UTR of the ~10.2kb *MECP2* transcript (*Figure 1.9)*. Moreover, prediction analysis showed that the polyadenylation sequence of the ~5.4kb transcript is the 5'-AAUAAG-3' hexamer and the sequence for the ~7.2kb transcript is the highly conserved 5'-AAUAUU-3' hexamer (Pelka et al., 2005;

Singh et al., 2008). These polyadenylation signals, also known as upstream core polyadenylation signals, are usually located 10-30 nucleotides upstream of the polyadenylation cleavage site (CS), where *trans*-acting factors such as cleavage and polyadenylation specific factors (CPSF) specifically bind. The polyadenylation cleavage site (CS) is usually followed by a U/GU-rich element region (downstream core polyadenylation signal) which is known to bind cleavage stimulating factors (CstF). The presence of these motifs is important for correct pre-mRNA endonucleolytic cleavage and polyadenylation (Chen et al., 2006; Coy et al., 1999).

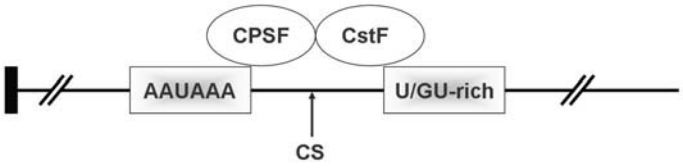


Figure 1.9: Schematic representation of the 3'regulatory motifs in human

The most frequent polyadenylation signal is an AAUAAA hexamer located downstream of the translational terminator (black box). A U/GU-rich sequence follows the cleavage site (CS), which is indicated with an arrow. Specific factors such as the trans-acting factors, cleavage and polyadenylation specific factor (CPSF) and cleavage stimulating factor (CstF) bind to the conserved hexamer and to the U/GU-rich element respectively. CPST and CstF are represented by ovals. Figure adapted from Chen *et al*. 2006.

1.9.3 Point mutations in the *MECP2* 3'UTR

It is possible that mutations in the *MECP2* 3'UTR could disrupt the interaction between the miRNAs and their target sites in the *MECP2* 3'UTR, leading to loss of appropriate gene regulation and causing the RTT phenotype. Large screening studies have been carried out and the number of variations in the 3'UTR has grown in recent years, highlighting the possible important role of this type of sequence variation in RTT.

MECP2 3'UTR molecular analysis was carried out in a cohort of 37 Israeli female patients with classical RTT (n=17) and with non-specific neurodevelopmental delay (n=20). In this study, a girl with psychomotor retardation, progressive microcephaly and spasticity was found to be a carrier for a missense mutation as well as an insertion of 96 bp after the stop codon (c.*96 dupA), which was also detected in her normal mother (*Table 1.4*). The same insertion was detected in a non-classical RTT girl and in her healthy father. The authors therefore inferred that this variation was a non-pathogenic mutation (Yaron et al., 2002).

In 2003, Lobo-Menendez and colleagues screened, in a cohort of 175 autistic patients, the 3'UTR of *MECP2* and they identified two non-pathogenic disease polymorphisms in two

unrelated males. Both the c.*9G>A and c.*55C>G variants were inherited from phenotypically normal mothers and they were also seen in normal control males (Lobo-Menendez et al., 2003). Moreover, the c.*9G>A variation had also been reported in patients with autism or atypical RTT with autism features (Bourdon et al., 2001b; Coutinho et al., 2007; Lam et al., 2000; Santos et al., 2008).

A c.*36G>C polymorphism was identified in a female patient with Angelman syndrome and in her unaffected mother during a molecular screening in negative patients for Fragile-X syndrome, Angelman syndrome or Prader-Willi syndrome (Kleefstra et al., 2004).

Shibayama *et al*. screened *MECP2* coding regions, associated splice junctions and conserved regions of the 3'UTR in 214 patients with schizophrenia (n=106), autism (n=24), attention-deficit hyperactivity disorder (ADHD) (n=25), bipolar illness (BPI) (n=24), alcoholism (n=17), puerperal psychosis (n=15) and phobia (n=3). Three sequence variations were identified: two alterations in Caucasian patients with autism, c.*177G>C in a female and c.*5348T>C in a male, and a c.*98dupA in an African-American male with ADHD. The putative variations were present in the heterozygous mother of the two autism cases establishing their likely non-pathogenic effect (Shibayama et al., 2004a). The c.*98dupA variation was also reported in Portuguese and Spanish papers (Santos et al., 2008; Tejada et al., 2006); and a Tunisian paper reported this alteration in association with a missense mutation in a classical RTT patient (Fendri-Kriaa et al., 2010).

In addition, a Finnish study analyzed the *MECP2* untranslated region in a cohort of 118 mentally retarded individuals with no RTT features, infantile autism, cerebral palsy, fetal alcohol syndrome, or other known etiology. A c.*93G>A variation was detected in two unrelated males and in a male with a second intronic variation. The variation was already listed in the IRSF *MECP2* Database as a polymorphism not causing disease (Ylisaukko-Oja et al., 2005).

In 2007, a Slavic study identified four novel sequence variation in the 3'UTR in 87 female classical RTT patients. A c.*328G>A variant had a parental origin (2 patients out of 87), while the parental origin of the other 3 variations (c.*92C>T, c.*359G>C and c.*363G>C) was unknown as the DNA from parents were not available (Zahorakova et al., 2007). On the other hand, Fedri-Kriaa *et al*. reported a female RTT patient with a novel sequence variation c.*92C>G that was not detected in 100 Tunisian control chromosomes (Fendri-Kriaa et al., 2010).

A heterogeneous cohort of 66 Portuguese female patients with classical (n=35) and atypical (n=16) RTT, mental retardation with autism (n=13) and Angelman syndrome-like clinical presentation (n=2), were studied by Santos and colleagues. They identified a variation already described in the literature as a polymorphism (c*8503delC), and two new sequence variants (c*1134G>A and c.*8500C>G) not found in 218 Portuguese controls. The c.*8500C>G and c*8503delC variants were detected in a patient with mental retardation, ataxia and epilepsy and in her unaffected father (c.*8500C>G) and in her unaffected mother (c*8503delC). The c*1134G>A variation was detected in a patient with congenital RTT, however this girl had a large rearrangement of the *MECP2* gene that more likely explained the RTT features (Santos et al., 2008).

To date, there is only one study that has screened the entire length of the 3'UTR. Coutinho and colleagues screened the entire *MECP2* 3'UTR in 172 Caucasian Portuguese patients with autism (141 males and 31 females). They identified 21 novel sequence variations in 46 individuals, but only 12 variations were present only in autistic patients. The authors then quantified the *MECP2* mRNA levels in four male patients carrying the c.*371G>C, c.*554G>A, c.*2556T>A and c.*2956G>A variants. The levels of *MECP2* mRNA were lower in these patients when compared with patients of a similar phenotype, but without any variations in *MECP2*. The authors postulated that the reduced amount of transcripts was correlated with mRNA instability resulting in degradation (Coutinho et al., 2007).

Table 1.4: Identified variations in the 3'UTR of the *MECP2* gene

Nucleotide change	Type of variation	N	Phenotype	References
c.*8C>T	3'UTR	2	Unaffected family member	Directly submitted
c.* 9 G>A	3'UTR	18	Autism, atypical RTT, control, unaffected family member	Bourdon *et al*., 2001, Coutinho *et al.,* 2007, Lobo-Menendez *et al.*, 2003, Santos *et al.,* 2008
c.*36 G>C	3'UTR	1	Angelman syndrome	Kleefstra *et al*., 2004
c.*55 C>G	3'UTR	3	Autism, unaffected family member, control	Lobo-Menendez *et al*., 2003
c.*92 C>G	3'UTR	1	Classical RTT	Fendri-Kriaa *et al*., 2010,
c.*92 C>T	3'UTR	1	Classical RTT	Zahorakova *et al.,* 2007
c.*93G>A	3'UTR	3	Autism, sporadic mental retardation (Not Rett syndrome), control	Directly submitted, Coutinho *et al.,* 2007, Ylisaukko-oja *et al*., 2005
c.*96dupA	3'UTR	3	Classical RTT, unaffected family member	Yaron *et al*., 2002
c.*98dupA	3'UTR	13	ADHD, unaffected family member, sporadic mental retardation (Not Rett synd)rome, mental retardation and autism, control,	Directly submitted, , Shibayama *et al*., 2004, Santos *et al.,* 2008, Tejada *et al*., 2006, Cardff in RETTBASE, Khajuria in RETTBASE
c.*122 delT	3'UTR	1	Not certain RTT	Directly submitted
c.*139 G>A	3'UTR	2	Autism, unaffected family member	(Xi et al., 2007)
c.*177 G>C	3'UTR	2	Autism, unaffected family member	Shibayama *et al*., 2004
c.*204 G>A	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*328 G>A	3'UTR	4	Classical RTT, unaffected family member	Zahorakova *et al.,* 2007

Nucleotide change	Type of variation	N	Phenotype	References
c.*359 G>C	3'UTR	1	Classical RTT	Zahorakova *et al.,* 2007
c.*363 G>C	3'UTR	1	Classical RTT	Zahorakova *et al.,* 2007
c.*371 G>C	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*393G>A	3'UTR	1	Non-Rett syndrome control	Coutinho *et al.,* 2007
c.*487 G>C	3'UTR	1	Unaffected family member	(Hoffbuhr et al., 2001)
c.*489 G>C	3'UTR	1	Control	Coutinho *et al.,* 2007
c.*529 G>T	3'UTR	1	Control	Coutinho *et al.,* 2007
c.*544 G>A	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*554 G>A	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*767 G>T	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*806 G>A	3'UTR	1	Control	Coutinho *et al.,* 2007
c.*831G>C	3'UTR	1	Control	Coutinho *et al.,* 2007
c.*861 T>G	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*875 dupA	3'UTR	1	Control	Coutinho *et al.,* 2007
c.*878 C>G	3'UTR	23	Autism, control	Coutinho *et al.,* 2007
c.*1134 G>A	3'UTR	1	Congenital RTT	Santos *et al.,* 2008
c.*1237 T>C	3'UTR	1	Control	Coutinho *et al.,* 2007
c.*1368 C>A	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*1737 G>A	3'UTR	13	Autism, control	Coutinho *et al.,* 2007
c.*2556 T>A	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*2657 G>A	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*2706 G>A	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*2956 G>A	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*3477 G>A	3'UTR	5	Autism, control	Coutinho *et al.,* 2007
c.*3658 C>T	3'UTR	1	Autism	Coutinho *et al.,* 2007
c.*3662 A>G	3'UTR	1	Control	Coutinho *et al.,* 2007
c.*3878 G>C	3'UTR	2	Autism, control	Coutinho *et al.,* 2007
c.*4086_*4087delGT	3'UTR	1	Control	Coutinho *et al.,* 2007
c.*4576 A>C	3'UTR	3	Autism, control	Coutinho *et al.,* 2007
c.*5348T>C	3'UTR	2	Autism, unaffected family member	Shibayama *et al*., 2004
c.*5486_*5487dupAT	3'UTR	2	Autism, control	Coutinho *et al.,* 2007
c.*5839 C>T	3'UTR	1	Control	Coutinho *et al.,* 2007
c.*7748 C>T	3'UTR	2	Autism, control	Coutinho *et al.,* 2007
c.*7856 A>C	3'UTR	18	Autism, control	Coutinho *et al.,* 2007
c.*8500 C>G	3'UTR	1	Unaffected family member	Santos *et al.,* 2008
c.*8503 delC	3'UTR	11	Autism, unaffected family member, control	Coutinho *et al.,* 2007, Santos *et al.,* 2008
c.*8503 dupC	3'UTR	2	Control	Coutinho *et al.,* 2007
c.[*8500C>G]+ [*8503delC]	3'UTR	1	Mental retardation (not Rett synd)	Santos et al., 2008
c.[1189G>A] +[*55C>G]	missense, 3'UTR	1	Not known	Directly submitted
c.[1451G>C] +[*98dupA]	missense, 3'UTR	1	Classical RTT	Yaron et al., 2002
c.[378-61C>G] +[*93G>A]	intronic, 3'UTR	1	Sporadic mental retardation (Not Rett syndrome)	Ylisaukko-oja et al., 2005
c.[880C>T] +[*9G>A]	nonsense, 3'UTR	1	Not known	Directly submitted
c.[916C>T] +[*98dupA]	missense, 3'UTR	1	Classical RTT	Fendri-Kriaa et al., 2010

This table is an adaptation from the RETTBASE table (http://mecp2.chw.edu.au) and it shows the nucleotide change, the type of variation, the number of patients carrying that mutation (n), the phenotype and the references.

1.10 Ovelapping phenotypes between RTT and other neurodevelopmental disorders

Although mutations in *MECP2* are responsible for almost all known cases of classical Rett syndrome, 5-10% of typical Rett patients and 40-60% of atypical cases do not appear to have *MECP2* mutations (Weaving et al., 2005). So, the existence of other genes responsible for the atypical Rett phenotype is also a possibility. (*Figure 1.10*)

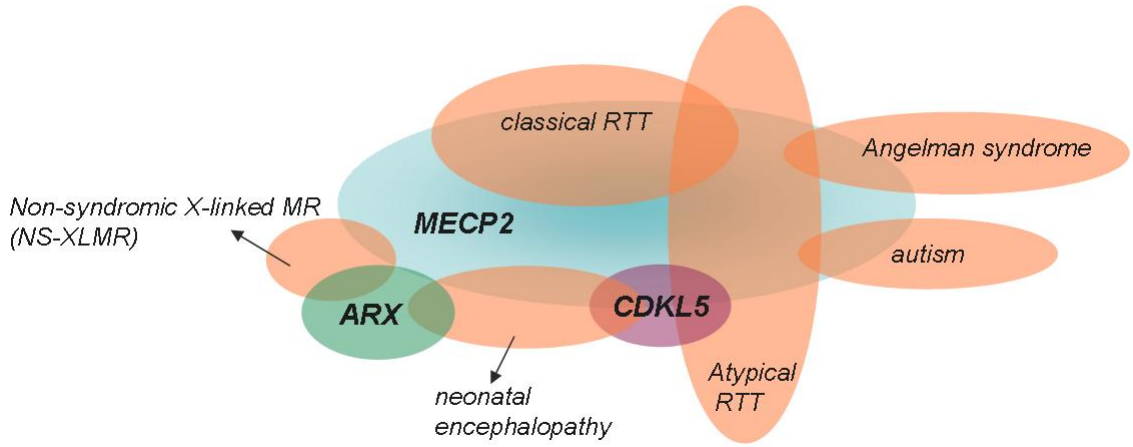


Figure 1.10: Overlapping phenotypes

It is well recognised that Rett syndrome and Angelman syndrome (AS, OMIM #105830) have overlapping clinical features, with global developmental delay, severe speech and communication impairment, progressive microcephaly, movement disorders, seizures, stereotypic hand movements, and autistic behaviours (Scheffer et al., 1990).

One specific mutation in the *MECP2,* p.A140V is the genetic cause of X linked mental retardation syndrome with psychosis, pyramidal signal and macro-orchidism (PPM-X syndrome, OMIM #300055) (Klauck et al., 2002).

MECP2 mutations are responsible for a wide spectrum of neurological disorders, ranging from mild mental retardation to severe neonatal encephalopathy in males patients. Males, who carry a mutation which is also found in females with Rett syndrome, develop severe neonatal encephalopathy and usually die in their first year of life (Schanen et al., 1998). However, males with symptoms similar to classical Rett syndrome result from somatic mosaicism (Clayton-Smith et al., 2000) or a 47, XXY karyotype (Hoffbuhr et al., 2001; Leonard et al., 2001; Schwartzman et al., 2001). The neurological presentation of males who carry mutations in the *MECP2* which are not found in females with RTT, range from severe to mild non syndromic X-linked mental retardation (NS-XLMR) (Couvert et al.,

2001; Gomot et al., 2003; Orrico et al., 2000; Villard, 2007). Moreover, duplications of the whole *MECP2* gene (and sometimes genes in its vicinity) have been described in male patients with severe intellectual disability, delayed milestones, absence of language, hypotonia replaced by spasticity and contractures and often severe infections (Belligni et al., 2010; Campos et al., 2010; Echenne et al., 2009; Prescott et al., 2009; Van Esch et al., 2005; Villard, 2007).

Mutations in the *MECP2* have been identified in autistic females (Carney et al., 2003). On the other hand, other studies have suggested that mutations in the *MECP2* are unlikely to play a major role in the aetiology of autism (Beyer et al., 2002; Lobo-Menendez et al., 2003; Ylisaukko-Oja et al., 2005).

Mutations in the cyclin dependent kinase like 5 gene *(CDKL5,* OMIM *300203) have been associated with a broad spectrum of phenotypes such as early-onset untreatable seizures, infantile spasms, severe mental retardation and RTT-like features and in patients with the atypical RTT early onset seizures variant or Hanefeld variant (Artuso et al., 2010; Mari et al., 2005; Nemos et al., 2009; Russo et al., 2009; Sartori et al., 2009; Scala et al., 2005; Tao et al., 2004; Weaving et al., 2004; White et al., 2010). It was suggested that *CDKL5* belongs in the same molecular pathway as MeCP2 (Tao et al., 2004). The two proteins showed a similar expression pattern in brain and they are similarly activated during neuronal maturation and synaptogenesis (Mari et al., 2005; Rusconi et al., 2008). However, Lin *et al*. have suggested that MeCP2 is not a direct substrate of CDKL5 (Lin et al., 2005), and so the debate regarding this issue persists.

The netrin G1 gene (*NTNG1,* OMIM *608818) was initially considered a possible novel candidate disease gene for atypical Rett syndrome (Borg et al., 2005). Additional studies suggested that *NTNG1* was not involved in atypical Rett syndrome and it should not be considered a priority in unexplained epileptic encephalopathy and atypical Rett syndrome (Archer et al., 2006a; Nectoux et al., 2007). The myocyte enhancer factor 2 gene (*MEF2C*, OMIM *600662) have been found to be associated with severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations that overlap to atypical Rett syndrome (Le Meur et al., 2010; Zweier et al., 2010). Interestilngly, sequence variations were not detected in a large cohort of RTT patients in a recent study (Armani et al., 2011).

The phenotype of RTT and Angelman syndrome overlap also with the Pitt-Hopkins syndrome (PHS, OMIM #610954) (Zweier et al., 2007). A pathogenic frame-shift variation

was detected in the autosomal transcription factor 4 gene (*TCF4*, OMIM *602272) in one female with atypical RTT, in whom facial dysmorphism suggested the possible diagnosis of Pitt-Hopkins syndrome (Armani et al., 2011).

Due to the overlapping phenotypic spectrum of *CDKL5* and *ARX* genes, screening of the *Aristaless*-related homeobox gene (*ARX,* OMIM *300382) should also be considered in order to investigate the contribution of this gene in patients with infantile spasms or autistic spectrum disorder with intellectual disability and early onset seizures.

1.11 Epileptic Encephalopathy with onset in the first year

Epileptic encephalopathies, with onset in the first year of life, are a group of conditions in which frequent seizures and/or major interictal paroxysmal activity are believed to contribute to the deterioration of cognitive, sensory and motor functions. The clinical and electroencephalography (EEG) characteristics of these conditions depend on the age of onset and may change over time (Dulac, 2001). The International League against Epilepsy (ILAE) has proposed a first classification scheme of epileptic seizures that is widely and universally accepted. Some revisions in the definitions have been made, but six specific epileptic encephalopathies in infancy are recognized: Ohtahara syndrome, early myoclonic encephalopathy, West syndrome, Dravet syndrome, myoclonic status in nonprogressive encephalopathies and migrating focal seizures in infancy (Engel, 2006). The etiology of these conditions often remains unknown and many cases do not always fulfil the electro-clinical definition.

The *ARX* gene have been associated recently with refractory early onset epilepsy with mental retardation (Archer et al., 2006; Bahi-Buisson et al., 2008; Buoni et al., 2006; Scheffer et al., 2002).

1.11.1 The *Aristaless*-related homeobox gene

The *ARX* gene is located on the chromosome X (Xp22) near the *POLA* gene (OMIM *312040), and is composed of 5 coding exons that encode a protein of 562 amino acids (*Figure 1.11*) (Stromme et al., 2002b). The ARX protein (UniProtKB/Swiss-Prot Q96QS3) belongs to large family of homeodomain transcription factors with crucial roles in

development during embryogenesis (Meijlink et al., 1999), and head development in particular (Galliot et al., 1999).

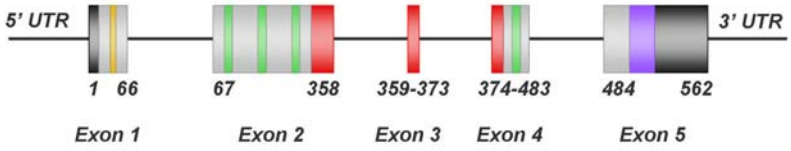


Figure1.11: Schematic representation of the *ARX* gene structure

The *ARX* gene consists of 5 exons. The known domains are represented: the octapepide is shown in yellow, polyalanine tracts are shown in green, the homeodomain is represented in red and the *Aristaless* domain is in purple. The 5' and 3'UTRs regions are represented in black. The numbers below the figure indicate the amino acids. Adapted from Gecz *et al.*, (Gecz et al., 2006).

The protein has important functional domains (*Figure 1.11*). A conserved 60 amino acid DNA binding motif called Paired-like or paired/Q50 homeodomain is located in exons 3 and 4 (amino acids 328-387), but its functional role is still poorly understood. The *Aristaless*, also known as OAR, C-peptide or *Paired*-tail domain (amino acids 527-542), is highly conserved. The function of this domain has not been fully determined. It has been suggested that it acts as an activator domain (Seufert et al., 2005), and performs an inhibitory function, as deletion of this domain results in increased DNA binding or transactivation of target promoters (Brouwer et al., 2003; Furukawa et al., 1997). The octapeptide YCIDSILG (amino acids 27-34) has a transcriptional repressor activity, and it is important in the interaction with the Groucho/TLE1 co-repressor (McKenzie et al., 2007). ARX has four polyalanine tracts: PA1 (amino acid 100-115) and PA2 (amino acid 144-155) at the N-terminal position of the protein, PA3 (amino acid 275-281) in the middle and PA4 (amino acid 432-440) between the homeodomain and the *Aristaless* domains. The function of these polyalanine tracts is still poorly understood, but it has been suggested that they play role in transcription regulation (Lavoie et al., 2003). The *ARX* gene also contains three nuclear localization signal motifs (NLS1, amino acids 82-89; NLS2, 325-332; and NLS3, 379-386).

1.11.2 Phenotype heterogeneity associated with ARX mutations

Mutations in the *ARX* gene have been associated with a large number of human disorders (*Table 1.5*) including non syndromic X-linked intellectual disability (NS-XLID, OMIM #300419) (Bienvenu et al., 2002), Partington syndrome (PRTS, OMIM #309510), consisting of intellectual disability with dystonic hand movements, ataxia and seizures (Partington et al., 2004); X-linked myoclonic epilepsy with severe intellectual disability (XMEID, OMIM # 300432) and X-linked infantile spasms or West syndrome (ISSX/WE,

OMIM #308350), consisting of infantile spasms, an electroencephalographic pattern of hypsarrhythmia and subsequent mental retardation (Stromme et al., 2002b). Moreover, mutations in the *ARX* are responsible for some cases of Ohtahara syndrome (EEG, OMIM #308350) with early infantile epileptic encephalopathy with a burst-suppression pattern, and the most severe outcome, Proud syndrome (PRTS/ACC/AG, OMIM #309510), which consists of X-linked mental retardation, agenesis of the corpus callosum and abnormal genitalia (Kato et al., 2004); hydranencephaly with abnormal genitalia (HYD/AG, OMIM #300215) and X-linked lissencephaly with abnormal genitalia (XLAG, OMIM #300215). XLAG typically includes severe congenital or postnatal microcephaly, lissencephaly, agenesis of the corpus callosum, hypothalamic dysfunction (principally manifested by hypothermia), midbrain malformations, neonatal-onset intractable epilepsy, ambiguous or underdeveloped genitalia and severely shortened lifespan (Kitamura et al., 2002; Uyanik et al., 2003).

A broad spectrum of mutations in the *ARX* gene is responsible for those phenotypes. In particular, insertion, nonsense, deletion, missense mutations in the evolutionarily conserved residues such as the homeodomain and *aristaless* domains lead to a more severe phenotype, suggesting the important functions of these domains. The polyalanine expansion mutations and some missense mutations have less severe consequences, but generally the severity of the disorder increases with the length of the polyalanine tract expansion (Shoubridge et al., 2010).

Table 1.5: Clinical phenotypes associated with *ARX* mutations

Pathogenic features	Phenotype
Non malformation group, possible features: epilepsy, infantile spasms, autism, dystonia and dysarthria	• Non syndromic X-linked intellectual disability - NS-XLID • Partington syndrome - PRT • X-linked myoclonic epilepsy with severe intellectual disability XMESID • X-linked infantile spasm or West syndrome - ISSX/WE • Ohtahara syndrome - OS
Malformation group, possible features: lissencephaly, hydranencephaly, abnormal genitalia and agenesis of the corpus callosum	• Proud syndrome - PRTS/ACC/AG • Hydranencephaly with abnormal genitalia - HYD/AG • X-linked lissencephaly with abnormal genitalia - XLAG

The most recurrent mutations that give rise to a non-malformation phenotype are often polyalanine expansions. The most frequent alteration is the c.429_452dup that occurs in the second polyalanine tract in exon 2. This change leads to a non-syndromic intellectual disability, Partington syndrome and West syndrome (Bienvenu et al., 2002; Gronskov et al., 2004; Kato et al., 2003; Laperuta et al., 2007; Nawara et al., 2006; Partington et al., 2004; Stromme et al., 2002a). The second most frequent mutation is the c.304ins$(GCG)_7$ in polyalanine tract 2, which gives rise to intellectual disability with variable features such

as infantile spasms, generalized dystonia and a hypsarrhythmic EEG pattern (Guerrini et al., 2007; Poirier et al., 2008; Stromme et al., 2002a; Wallerstein et al., 2008). A clinical phenotype of Ohtahara syndrome has been reported in two unrelated males carrying the c.304ins$(GCG)_7$ (Absoud et al., 2010) and the c.298_330dup GCGGCA$(GCG)_9$ duplication (Kato et al., 2007). Other polyalanine expansions have been identified, and particularly interesting is the description of three related males with a duplication of 27 bp (c.430_456dup), characterized by early onset infantile spasms and early death (Reish et al., 2009), while a 33 bp duplication (c.423_455dup [33dup]) gave rise to a less severe phenotype, possibly due to the presence of a glycine and its possible "positive" effect of the severity of the phenotype (Demos et al., 2009). A total of 9 other mutations occur in the octapeptide domain, homeodomain and *aristaless* domain, and lead to a non-malformation phenotype. Insertions, deletions, nonsense and splice site mutations in the homeodomain (13 of 27), polyalanine tract 1 (7 of 27), octapeptide domain (3 of 27), *aristaless* domain (1 of 27) and three frame-shift lead to malformation phenotypes (Shoubridge et al., 2010).

Aim of this thesis

Despite advances in molecular diagnosis, the molecular basis of many neurodevelopmental disorders are still uknown. This study has focused the attention on two severe disorders such as Rett syndrome and early infantile epileptic encephalopathy to expand the knowledge on the pathogenesis of these common and devastating diseases.

It is well known the important role of the *MECP2* gene in the RTT syndrome. Patogenic sequence variations in the *MECP2* coding region are responsible for 90-95% of classical RTT and for the 20-40% of atypical RTT. The effect and the impact of sequence variations in the 3'UTR *MECP2* non-coding region in the RTT pathogenesis are still not known in literuature. Recent studies revealed that sequence alterations in the 3'UTR of other genes may affect protein expression levels and that these variations are associated with a broad spectrum of diseases. The 3'UTR contains binding target sites for important elements called non-coding RNA molecules, or microRNA, which are involved in an extended range of biological processes. Sequence alterations in these binding target sites have been already correlated with altered expression levels due to a missing interaction between microRNA:3'UTR binding target sites. This project investigated the possible role of the *MECP2* 3'UTR in the pathogenesis of RTT syndrome through the analysis of the expression levels of *MECP2_e1* and *MECP2_e2* isoforms and the sequencing of the *MECP2* 3'UTR sequence. Moreover, an *in silico* prediction studies were carried to predict which microRNA binds to known variations in the target binding sites in the *MECP2* 3'UTR. Finally, an *in vitro* functional analysis of *MECP2* 3'UTR sequence alterations was performed with a reporter gene to detect if alterations in the microRNA:3'UTR binding target site interaction might dysregulate protein expression levels and be a causative factor of RTT syndrome.

Pathogenic sequence variations in the *ARX* gene are responsible of a large spectrum of neurodevelopmental disorders without malformation (epilepsy, infantile spasm, autism, dystonia and dysarthria) and with malformation (lissencephaly, hydranencephaly, abnormal genitalia and agenesis of the corpus callosum). Only polyalanine expansions were associate with EIEE and in particular with the Ohtahara syndrome. The clarify the role of this gene in this syndrome, its coding sequence was analysed in patients with early infantile epileptic enchephalopaty.

2. Materials and Methods

2.1 Materials

2.1.1 Bacterial Strain and Mammalian Cell lines

Host:	*Escherichia coli*
Strain:	One Shot® TOP10 Chemically Competent cells
Genotype:	F-mcrA, Δ(mrr-hsdRMS-mcrBC) ψ80lacZΔM15, ΔlacX74, recA1, araD139, Δ(araleu)7697, galU, galK, rpsL (Str^R), endA1, nupG
COS-7:	African green monkey SV40-transformed kidney fibroblast cell line
PC12:	Rat adrenal pheochromocytoma cell line
SH-SY5Y:	Human derived neuroblastoma cell line

2.1.2 Vectors Description

pCR®2.1-TOPO

pCR®2.1-TOPO vector contains a TOPO® cloning sites for rapid and efficient cloning of amplified PCR products. The vector is characterized by single, overhanging 3' deoxythymidine (T) residues for TA Cloning®. The vector has Ampicillin/Kanamycin resistance for easy selection in *E. coli* and a pUC origin for high copy replication of the plasmid in *E. coli*. The pCR®2.1-TOPO vector (*Appendix 2, Figure 1*) was used to enable the cloning of the *MECP2* 3' UTR fragments into the Chroma-Luc™ Reporter Vectors.

Chroma-Luc™ Reporter Vectors

Chroma-Luc™ Reporter Vectors are a new series of luciferase technology based on a native yellow-green luciferase gene originally cloned from the large Caribbean click beetle (CB) *Pyrophorus plagiophthalamus* (*Appendix 3*). All the Chroma-Luc™ Reporter Vectors include a SV40 promoter and enhancer which provide strong, constitutive expression in a variety of cell types. Both vectors contain an ampicillin resistance gene in *E. coli*. The vectors used are:

i) pCBR-Control vector, 5234 bp in length, contains a CBR*luc* synthetic cDNA sequence encoding the red luciferase enzyme and an SV40 polyadenylation signal (*Appendix 3, Figure 1)*.

ii) pCBG99-Control vector, 4793 bp in length, contains a CBG99*luc* synthetic cDNA sequence encoding the green 99 luciferase enzyme and an SV40 polyadenilation signal. A fragment of 1270 bp of the *MECP2* promoter/5'UTR was previously cloned into the

pCBG99-Control vector upstream the luciferase gene. Different 3'UTR fragments of *MECP2* were cloned downstream of the luciferase gene (*Appendix 3, Figure 2)*.

2.2 Patients recruitment

Two different cohort of patients were investigated in this study. Parental informed consent for genetic testing was obtained for each subject included in the study.

To analyse *MECP2* expression levels of isoform *e1* (*MECP2_e1*) and *e2* (*MECP2_e2*) through Quantitative Real-Time PCR with TaqMan probes, a cohort of 22 female RTT patients were investigated in this analysis. To analyse the *MECP2* 3'UTR with direct sequencing a group of 17 individuals was analysed. RTT patients samples were kindly provided by the Child Neurology Units of Padova and by Professor David Ravine, Western Australian Institute of Medical Research, University of Western Australia, WA. All patients were classified according to clinical diagnostic criteria as classical or atypical RTT. All patients selected were previously screened for point mutations in the coding region and exon-intron boundaries of *MECP2* and *CDKL5* genes by direct sequencing and DHPLC. All patients, who have previously been tested negative for point mutations, were also screened for deletions and duplications using MLPA or Quantitative Real Time PCR.

A cohort of 40 male and 33 female children with a diagnosis of epileptic encephalopathy and onset of seizures in the first year of life were recruited for *ARX* gene analysis by the Child Neurology Units of Padua, Verona, Messina, Brescia, and Ancona. The key criteria for patient selection were based on clinical evaluation, neurophysiological and neuro-radiological assessment and negative standard cytogenetic analysis. A comprehensive set of clinical data including seizure history, electroencephalographic recordings and magnetic resonance imaging (MRI) findings, was collected for each enrolled patient and retrospectively evaluated (Giordano et al., 2010; Sartori et al., 2011).

2.3 DNA extraction from blood samples

Genomic DNA was extracted from 3-5 ml of peripheral blood samples anti-coagulated with EDTA. The blood sample was transferred to a 50 ml Falcon tube and 3 volumes of 1X Red Blood Cell Lysis Buffer solution (RBC, 40,12 g NH_4Cl, 3,1 g $NaHCO_3$, 3,72 g Na_2EDTA, H_2O to 500 ml) were added. The tube was inverted and incubated at room

temperature for 10 minutes. The tube was centrifuged at 3600 rpm for 10 minutes at room temperature, the supernatant was decanted and 3 volumes of 1X RBC were added to resuspend the cell pellet. This step was repeated three times. A volume of 3 ml of cell lysis buffer 1X (CLB 10X, 1.21 gr Tris 100 mM, 7.44 gr Na_2 EDTA 200 mM, 10 gr SDS 10%, pH 7.45 H_2O to 100 ml) were added to the cell pellet to disrupt the cellular structure of the white blood cells and the cell nuclei. A volume of 0.9-1.6 ml of protein precipitation solution (46 gr CH_3COONH_4 6M, H_2O to 100 ml) were added to the cell lysate. The sample was then vortexed and centrifuged at 3600 rpm for 30 seconds. The supernatant was transferred to a clean 50 ml Falcon tube avoiding the collection of the precipitated protein. Five ml of isopropanol were added to the tube which it was gently inverted until the appearance of the translucent white threads of DNA.

The DNA was visible as a small white pellet after a centrifuge of 3600 rpm for 5 minutes. DNA was air dried and rehydrated by adding 100 to 400 µl of TE buffer (10 mM Tris buffer pH 8.0, 1mM EDTA) depending on the size of the pellet. The sample was centrifuged briefly and then transferred to a 1.5 ml tube.

2.4 Measurement of the quantity and purity of total DNA and RNA

The concentration of DNA or RNA was determined by measuring the absorbance at 260 nm (A260) in the spectrophotometer NanoDrop® (Thermo Scientific). This Spectrophotometer enables highly accurate analyses of extremely small volumes of samples with remarkable reproducibility. 1 unit corresponds to 50 µg/ml for DNA samples and 40 µg/ml for RNA samples at 260 nm. The concentration in µg/ml can be calculated from the absorbance value with the following formula:

A260 x dilution factor x 50 µg/ml (DNA)
A260 x dilution factor x 40 µg/ml (RNA)

The purity of DNA and RNA were determined by calculating the ratio of 260 nm and 280 nm (A260/A280) with respect to contaminants that absorb in the UV spectrum, such as proteins. DNA and RNA samples with an A260/A280 ratio between 1.8 and 2 were considered to be sufficiently pure for further experiments.

2.5 Polymerase Chain Reaction (PCR)

A standard DNA fragment amplification reaction includes: DNA polymerase, forward and reverse primers, deoxyribonucleotide triphosphates (dNTPs), $MgCl_2$ and genomic DNA or cDNA. The cycling conditions vary depending on the size of the product, the length of the oligonucleotide primers and the degree of mismatch calculated between the primer and template. Primers were designed using Primer3 online software (http://frodo.wi.mit.edu/primer3/) (Rozen and Skaletsky, 2000). The Primer3 conditions were set to default setting, with maximum self complementary set to 5 and maximum 3' self complementarity set to 1. Primers were ordered from Invitrogen™ or Sigma-Aldrich® and diluted to 100 µM using TE buffer (10 mM Tris buffer pH 8.0, 1mM EDTA).

2.5.1 PCR Amplification of the *ARX* gene

The *ARX* gene consists of 5 exons located at Xp22 chromosome band. Exon 2 is 877 bp long and has a CG content of 75%; for this reason, this exon was divided into three smaller overlapping fragments. The PCR reactions were carried out using the AmpliTaq Gold® DNA polymerase (Applied Biosystems) and PCR conditions are outlined in *Tables 2.2* and *2.3*.

2.5.2 PCR Amplification of the *MECP2* 3'UTR

The 3'UTR of the *MECP2* gene is ~10.2 kb in length. The mutational screening was conducted on the proximal part of the 3'UTR of about ~4.5 kb. The PCR reactions were carried out using the *PfuUltra* High-Fidelity DNA Polymerase AD (Statagene) and PCR conditions are outlined in *Tables 2.5* and *2.6*. The fragment was sequenced in a unidirectional forward walk and the primers used are listed in *Table 2.4*.

2.5.3 PCR Amplification of *MECP2* 3'UTR fragments for Cloning

Amplification of the *MECP2* 1.8 kb and 5.4 kb 3'UTR fragments was carried out using Expand Long Template PCR System (Roche). For each 3'UTR two sets of primers were required: the common forward primer was designed to contain a *PspXI* restriction enzyme site, while the two specific reverse primers contained a *SalI* restriction enzyme site to allow the cloning of the *MECP2* promoter/5'UTR-pCBG99 Luciferase vector. The PCR amplification and conditions are outlined in *Tables 2.7, 2.8* and *2.9.*

Table 2.1: Primer sequences used for the screening of the *ARX* gene

Construct name	Primer Forward 5'-3'	Primer Reverse 5'-3'	Size (bps)
ARX exon 1	ccaacacacacccatccat	ccgaacaccaaacatccaa	363
ARX exon 2.1	agtggagaaagaggccaagg	ctcgcggtacgacttgct	493
ARX exon 2.2	cctgggacacgctcaagat	cctccttgggtgacagctc	417
ARX exon 2.3	cgctgctcaaggagccc	cacagagtccaggagccaag	420
ARX exon 3	tggagtaggcctgccataga	ccaacccatctctctctctcc	215
ARX exon 4	caagggaagggacgggta	gtttgactcctgcctcctc	707
ARX exon 5	cctcgggaatatctggact	ttgagtggtgctgagtgagg	544

Table 2.2: PCR conditions for the specific *ARX* fragments

	Final Concentration						
Reagents	**exon 1**	**exon 2.1**	**exon 2.2**	**exon 2.3**	**exon 3**	**exon 4**	**exon 5**
10X Buffer	1X	1X	1X	1X	1X	1X	1X
25 $MgCl_2$	1.5 mM	1.5 mM	1.5 mM	1.5 mM	1.5 mM	1.5 mM	1.5 mM
2 mM dNTPs	0.2 mM	0.2 mM	0.2 mM	0.2 mM	0.2 mM	0.2 mM	0.2 mM
5 M Betaine	----	----	0.5 mM	----	----	0.5 mM	0.5 mM
100% DMSO	10%	10%	10%	10%	10%	10%	10%
5 µM Primer Forward	0.6 µM	0.6 µM	0.6 µM	0.6 µM	1 µM	0.6 µM	0.6 µM
5 µM Primer Reverse	0.6 µM	0.6 µM	0.6 µM	0.6 µM	1 µM	0.6 µM	0.6 µM
AmpliTaqGold 1U/µl	2.5 U/µl	2.5 U/µl	2.5 U/µl	2.5 U/µl	2.5 U/µl	2.5 U/µl	5 U/µl
DNA	200 ng	200 ng	200 ng	200 ng	200 ng	200 ng	200 ng
milliQ-final volume of	50 µl	50 µl	50 µl	50 µl	50 µl	50 µl	50 µl

Table 2.3: PCR cycling conditions

Step	Temperature	Time	Cycle
Initial denaturation	95°C	10 min	1X
Denaturation Annealing Extension	94°C 64°C 72°C	45 sec 30 sec 1 min	10X -0.5°C/cycle
Denaturation Annealing Extension	94°C 64°C 72°C	45 sec 30 sec 1 min	20X
Final extension	72°C	10 min	
Cooling	4°C	∞	1X

Table 2.4: Primer sequences used for the *MECP2* 3'UTR screening

Construct name	Primer Forward 5'-3'	Primer Reverse 5'-3'	Size (bps)
MECP2 4.5kb	ctttacacggagcggattgc	accctgggggaggcagtatcc	4512bp

Sequencing primers			
PspXI Forward	ctttacacggagcggattgc		
2nd	cccccatgtggtcgttagag		
3rd	tctcaagggacgtcctccaa		
4th	gagagcaaactggcttgaattgg		
5th	ggcctgtctgtgaagagcaa		
6th	cttatgcttccagaaacacccacagg		
7th	catagtggcctcgaactcct		
8th	acaggaggttggagggaaag		
9th	gctgtgacagtgatggcgaa		

Table 2.5: PCR reactions for the *MECP2* 3'UTR fragment

Reagents	Final Concentration
10X Buffer	1X
102 mM dNTPs	0.4 mM
5 M Betaine	0.5 M
5 µM Primer Forward	0.5 µM
5 µM Primer Reverse	0.5 µM
PfuUltraAD 1U/µl	0.4 U
DNA	200 ng
milliQ-final volume of	20

Table 2.6: Cycling conditions for the *MECP2* 3'UTR fragments

Step	Temperature	Time	Cycle
Initial denaturation	95°C	4 min	1X
Denaturation Annealing Extension	95°C 60°C 68°C	30 sec 45 sec 10 min	10X
Denaturation Annealing Extension	95°C 60°C 68°C	30 sec 45 sec 10 min +10sec/cycle	20X
Final extension	68°C	10 min	
Cooling	4°C	∞	1X

Table 2.7: Primer sequences used for the *MECP2* 3'UTR cloning

Construct name	Primer Forward 5'-3'	Primer Reverse 5'-3'	Size (bps)
1.8kb	agcagacctcgagaagcgtgactttacacgga gcggatt	gcgaagtcgacacgcaacaagtgtggttgggc accgcg	1400
5.4kb	agcagacctcgagaagcgtgactttacacgga gcggatt	gcgaagtcgacacgcaaaccctgggggaggc agtatcc	4500

Table 2.8: PCR conditions for the *MECP2* 3'UTR fragments: 1.8kb fragment on the left and 5.4 kb fragment on the right

Reagents	Final concentration
10X Buffer 1	1X
10 mM dNTPs	0.35 mM
5 µM Primer Forward	0.3 µM
5 µM Primer Forward	0.3 µM
Expand Long Template enzymes mix 5U/ µl	0.2 µl
DNA	200 ng
milliQ-final volume of	10 µl

Reagents	Final concentration
10X Buffer 2	1X
10 mM dNTPs	0.5 mM
5 µM Primer Forward	0.3 µM
5 µM Primer Forward	0.3 µM
Expand Long Template enzymes mix 5U/ µl	0.2 µl
DNA	200 ng
milliQ-final volume of	10 µl

Table 2.9: PCR cycling conditions for the *MECP2* 3'UTR fragments

Step	Temperature	Time	Cycle
Initial denaturation	94°C	4 min	1X
Denaturation Annealing Extension	92°C 55°C 68°C	10 sec 30 sec 6 min	10X
Denaturation Annealing Extension	92°C 55°C 68°C	10 sec 30 sec 6 min+20 sec/cycle	25X
Final extension	68°C	7 min	1X
Cooling	4°C	∞	1X

2.6 Agarose Gel Electrophoresis

Electrophoresis through agarose is used to separate, identify and purify nucleic acid fragments. The technique is simple, rapid to perform and capable of resolving fragments of DNA and RNA. Agarose gel electrophoresis was used to analyse DNA fragments. The electrophoretic mobility of DNA fragments mainly depends on the size of the fragment and on the conformation of the DNA, the type and concentration of agarose used, as well as the applied voltage and electrophoresis buffer used. Agarose gels have greater range of separation and can resolve DNA fragments from 50 bp to 20 kb in length.

A 1% agarose gel was used for large fragment separations (1-20 kb), a 2% agarose gel was used for smaller sized PCR fragment separations. All agarose gels contained 1 µg/ml ethidium bromide for visualization of the DNA bands under UV light. The samples were mixed with a glycerol-based loading dye (3 µl /5 µl sample) (*Table 2.10*) prior to gel loading. Gels were submersed in a 1X TBE buffer (*Table 2.10*) and electrophoresed at 80-120 V. A 1 kb Plus Ladder (New England BioLab™) or TriDye 100 bp Ladder (New England BioLab™) size markers were co-electrophoresed with DNA samples.

Table 2.10: Recipe for buffers used in agarose gel electrophoresis

Reagents	Composition
10x TBE buffer	108 g Tris base 55 g Boric Acid 40 ml 0.5 M EDTA pH 8.0
10 X DNA loading Buffer	0.25 g Xilene cyano 0.25 g Bromophenol Blue 1.25 ml 10% SDS 12.5 Glycerol

2.7 Automated Sequencing

PCR reactions were prepared for DNA sequencing by shrimp alkaline phosphatase (*SAP*, (Promega) and exonuclease I (*ExoI*, Epicentre) digestion to remove unincorporated dNTPs and excess single stranded primer oligonucleotides from a reaction mixture. To each PCR reaction 1U of Shrimp Alkaline Phosphatase (Roche Diagnostics) and 5U of exonuclease I (Epicentre) were added to 10-15 µl of PCR reaction and incubated at 37°C for 30 minutes followed by incubation at 80°C for 20 minutes.

Sequencing was performed using the Sanger method with Big Dye®Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems): to each sequence reaction 1 µl of Big Dye® Terminator v3.1, 2 µl of Sequencing Buffer 5X, 1 µl of 3.5 pmol/µl primers and 3-10ng of amplicon were added to a 10 µl final volume reaction and incubated as follow: 96°C 1 minute, 25 cycles at 96°C 10 seconds, 50°C 5 seconds and 60°C 4 minutes. An excess dye terminators can interfere with sequence basecalling in sequencing reactions, the sequence product was purified using MicroSpin™ columns containing Sephadex™ G-50 (Illustra AutoSeq G-50, Amersham Biosciences) following the manufacturer's instructions. Automated sequencing was performed by ABI PRISM 3100 Genetic Analyzer® (Applied Biosystems) sequencing apparatus. Some sequencing reactions, containing 50-250 ng of purified PCR template or 200-500 ng of plasmid DNA and 5 µM of primer, were sent to the Macrogen (Korea) to be analysed. Sequencing was performed using the Sanger method with 3730XL DNA sequencer machine (Macrogen, Korea).

2.8 General RNA methods

To avoid RNAse contamination, disposable plasticware and RNase-free water were used whenever possible. RNase-free water was prepared by stirring double distilled water with 0.01% (v/v) DEPC overnight, followed by autoclaving to remove the DEPC.

2.8.1 cDNA synthesis

Total RNA was reverse transcribed by SuperScript™ III Reverse Transcriptase (Invitrogen), an enzyme engineered from Moloney Murine Leukaemia Virus. The reverse-transcription was carried out in a volume of 13 µl of DEPC-treated water containing 5 ng of total RNA of each sample, 1 µl random primers (250 ng) and 1 µl of 10mM dNTPs. Each sample was incubated at 65°C for 5 minutes for primer hybridisation and incubated on ice at least 1 minute. Samples were briefly centrifuged and 4 µl of 5X First-Strand Buffer, 1 µl of 0.1M DTT, 1 µl of RNase OUT™ Recombinant RNase Inhibitor (40 units/µl, Invitrogen) and 1 µl of SuperScript™ III Reverse Transcriptase (200 units/µl) were added into the samples. The samples were incubated at 25°C for 5 minutes followed by 50°C for 60 minutes. The reaction was inactivated by heating at 70°C for 15 minutes and the cDNA was stored at -20°C.

2.8.2 PCR amplification of cDNA

Two µl of reverse transcribed cDNA was used for PCR amplification with specific primer pairs for the housekeeping Phosphoglycerate Kinase 1 (*PGK1,* OMIM *311800) and the *MECP2_e1* and *_e2* isoforms. The reaction was carried out using AmpliTaq Gold® DNA polymerase (Applied Biosystems) as outlined in the following tables. Amplified DNA samples and 100kb DNA ladders were resolved using a 2% agarose gel at 80 volts to confirm the presence of PCR products.

Table 2.11: Primer sequences for amplification of the *PGK1* and *MECP2* genes

Construct name	Primer Forward Sequence 5'-3'	Primer Reverse Sequence 5'-3'
PGK1	gcttggacaatggagccaag	tctgcaactttagctccgcc
MECP2_e1/e2	gctccctcctctcggagaga	gcttaagcttccgtgtccag

Table 2.12 PCR amplification of cDNA

Reagents	Final concentration
10X Buffer	1X
2 mM dNTPs	0.3 mM
100% DMSO	10 %
25 mM $MgCl_2$	2.5 mM
5 µM Primer Forward *PGK1*	0.5 µM
5 µM Primer Reverse *PGK1*	0.5 µM
5 µM Primer Forward *MECP2*	0.5 µM
5 µM Primer Reverse *MECP2*	0.5 µM
AmpliTaqGold 5U/µl	1 U/µl
MilliQ-final volume of	20 µl

Table 2.13: PCR cycling conditions

Step	Temperature	Time	Cycle
Initial denaturation	95°C	10 min	1X
Denaturation Annealing Extension	95°C 59°C 68°C	30 sec 30 sec 45 sec	35X
Final extension	68°C	10 min	1X
Cooling	4°C	∞	1X

2.9 Quantitative Polymerase Chain Reaction (Q-PCR)

Real-time PCR is a highly sensitive technique that allow amplification and quantification of a specific nucleic acid sequence through detection of the fluorescence emitted during the PCR reaction. There are three main fluorescence-monitoring systems for amplification: 1) hydrolysis probes, 2) hybridizing probes (TaqMan probe, molecular beacons and scorpions) and 3) DNA-binding agents (non-sequence specific fluorescent intercalating agent as SYBR-green).

2.9.1 TaqMan assay chemistry

The TaqMan assay consists of a pair of unlabeled PCR primers and a TaqMan probe, longer than the primers, with a reporter dye (FAM™ dye, 6-Carboxyfluorescein, or VIC® dye) linked to the 5′ end of the probe. A minor groove binder (MGB) and a non-fluorescent quencher (NFQ) are present at the 3'end of the probe. The MGB folds into the DNA minor groove enhancing the probe melting temperature (Tm) and increases the stability of DNA duplex/TaqMan MGB probes (*Figure 2.1*) (Afonina et al., 1997; Kutyavin et al., 2000; Lukhtanov et al., 1997). A non-fluorescent quencher, NFQ, is a dye that absorbs fluorescent light of a particular wavelength, and does not itself emit a fluorescent wavelength. The vicinity of the reporter dye to the quencher suppresses the reporter

fluorescence by Förster or Fluorescence Resonance Energy Transfer (FRET) (Forster, 1948).

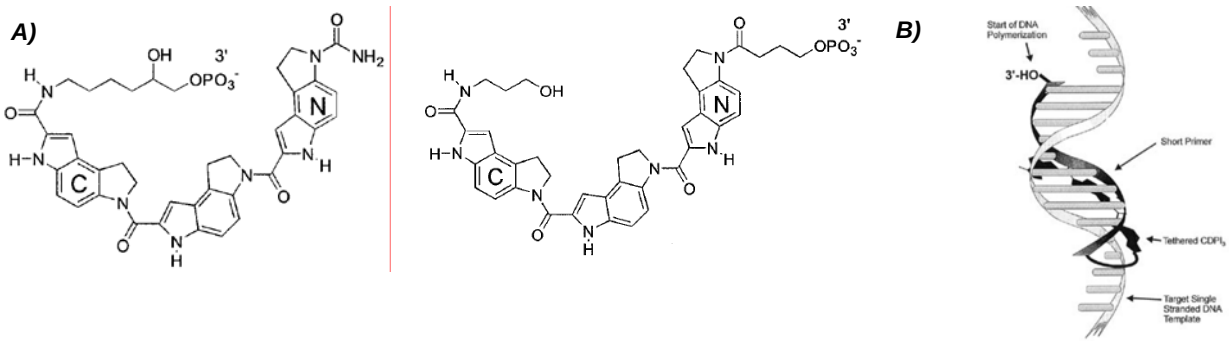


Figure 2.1: Structure of a minor groove binder
A) The figure shows the molecular structure of a minor grow binders called dihydrocyclopyrroloindole tripeptide with carbamyl group (CDPI3, left) and without (DPI3, right). The N- or C-terminus of the MGBs are linked to the 3' end of the TaqMan probe. Picture from Kutyavin *et al* 2000. B) The PCR template with the the MGB that folds into the minor grove of a DNA helix, diagram from Afonina *et al*., (Afonina et al., 1997).

During the amplification reaction, primers and TaqMan probe bind to the complementary target sequence; the Taq DNA polymerase cleaves the probe releasing the reporter dye with consequent emission of fluorescent light from the reporter dye.

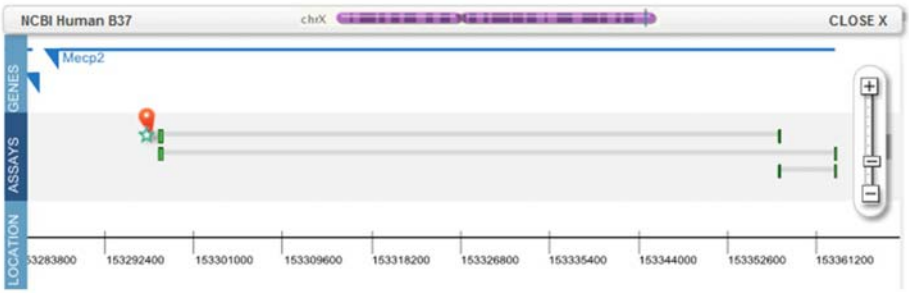
The TaqMan Assay was carried out using 2X TaqMan® Universal PCR Master Mix and the 1X TaqMan® Gene Expression Assay (Applied Biosystem) to detect the *MECP2_e1*, *MECP2_e2* isoforms and the housekeeping *PGK1* in a final volume of 10 µl. The thermal cycle conditions were performed according to the manufacturer's instructions using Rotor-Gene™6000 Corbett Life Science.

The TaqMan Gene Expression Assay is a sensitive and specific technique for the quantification of DNA and RNA based on a set of PCR primers and a dye-labelled probe. As *MECP2* isoforms are generated by the usage of two alterative start codons, the TaqMan probe approach seems optimal to detect the two different isoforms. In fact, *MECP2_e1* skips exon 2 and has an alternative N-terminus, translated from exon 1; while *MECP2_e2* is translated from a start codon in exon 2.

It is known that the TaqMan®Gene Expression assay has the ability to detect mRNA expression at low levels. Due to the very low amount of RNA previously collected from peripheral blood of RTT patients and due to the larger variability in the Ct values, it was necessary to introduce an extra PCR amplification step, before performing the Real-Time PCR. A multiplex PCR was performed using specific primer pairs to enrich only the selected products. As the reverse-transcription was performed with random decametres,

the multiplex PCR allowed the amplification of only the transcripts of interest such as *MECP2* and *PGK1*. The resulting products are enriched in product and in primers, deoxynuclotide triphospahte (dNTPs), Taq polymerase, magnesium chloride ($MgCl_2$). Consequently, a dilution of the enriched product was necessary to avoid the possible influence of these reagents in the following step of Real-Time PCR.

A)



B)

```
MECP2_e1        CCGGCGTCGGCGGCGCGCGCGCTCCCTCCTCTCGGAGAGAGGGCTGTGGTAAAAGCCGTC 60
MECP2_e2        CCGGCGTCGGCGGCGCGCGCGCTCCCTCCTCTCGGAGAGAGGGCTGTGGTAAAAGCCGTC 60
                ************************************************************

MECP2_e1        CGGAAAATGGCCGCCGCCGCCGCCGCCGCGCCGAGCGGAGGAGGAGGAGGAGGCGAGGAG 120
MECP2_e2        CGGAAAATGGCCGCCGCCGCCGCCGCCGCGCCGAGCGGAGGAGGAGGAGGAGGCGAGGAG 120
                ************************************************************

MECP2_e1        GAGAGACTG------------------------------------------------------ 129
MECP2_e2        GAGAGACTGCTCCATAAAAATACAGACTCACCAGTTCCTGCTTTGATGTGACATGTGACT 180
                *********

MECP2_e1        ------------------------------------------------------------
MECP2_e2        CCCCAGAATACACCTTGCTTCTGTAGACCAGCTCCAACAGGATTCCATGGTAGCTGGGAT 240

MECP2_e1        ------------GAAGAAAAGTCAGAAGACCAGGACCTCCAGGGCCTCAAGGACAAACC 176
MECP2_e2        GTTAGGGCTCAGGGAAGAAAAGTCAGAAGACCAGGACCTCCAGGGCCTCAAGGACAAACC 300
                            ********************************************
```

Figure 2.2: *MECP2*_e1 and *MECP2_e2* isoform sequences

A) The map represent the localization of the gene and the two transcripts. Picture from Applied Biosystem website (www.appliedbiosystems.com).
B) The two *MECP2* isoforms were aligned with ClustalW website. The *MECP2_e1* (NM_001110792) is encoded by exons 1 (green), 3 (blue) and 4 (not represented). The *MECP2_e2* (NM_004992) is encoded by exons 2 (pink), 3 (blue) and 4 (not represented). The two ATG start codons are represented in red.

2.9.2 Gene expression analyses

Quantitative analysis was performed by analysing the levels of *MECP2_e1* and *MECP2_e2* isoforms to the housekeeping phosphoglycerate kinase 1 (*PGK1*). The expression level of each isoform was determined using the Comparative $2^{(-\Delta\Delta Ct)}$ method (Livak and Schmittgen, 2001) which assumes that the amplification kinetics of the target gene and reference gene are approximately equal. The analysis was performed on triplicate samples using the following equations (1, 2 and 3), where cycle threshold (Ct) represents the cycle number of the PCR reaction at which the fluorescence signal of the

reaction crosses a threshold value. The first equation (1) is computed for each patients to normalise the amount of input DNA/RNA in each reaction by subtracting the Ct value of the reference gene/transcript (*PGK1*) from the Ct value of gene/transcript of interest, in this case *MECP2* isoforms The second equation (2) is used to calculate the difference in normalised expression between the patient and the controls. The last equation (3) expresses the linear fold change in expression between the experimental and calibrator sample.

(1) $\Delta Ct = Ct_{(MECP2)} - Ct_{(\text{reference } PGK1)}$

(2) $\Delta\Delta Ct = \Delta Ct_{(\text{patient})} - \Delta Ct_{(\text{normal control})}$

(3) Fold change = $2^{(-\Delta\Delta Ct)}$

2.9.3 Housekeeping expression stability analysis

Expression stability analysis of candidate reference genes was carried out using three publicly available software tools: geNorm (http://medgen.ugent.be/~jvdesomp/genorm/), NormFinder (http://www.mdl.dk/publicationsnormfinder.htm) and Bestkeeper (http://www.gene-quantification.de/bestkeeper.html#download) according to authors' instructions. *GeNorm* is a Visual Basic Application for Microsoft Excel (VBA) that compares each pair of housekeeping genes and calculates the gene stability value (M-value). The M value is the mean pair-wise variation between any gene and a set of candidate reference genes. The gene with the highest M-value is considered to have the less stable expression and it is sequentially removed from the analysis and a new M-value is then re-calculated for the remaining genes. The calculation proceeds until there is left only a pair of stable genes with the lowest M-value (Vandesompele et al., 2002).

NormFinder is another Visual Basic Application for Microsoft Excel (VBA) that uses a "model-based variation approach" to identify the reference gene with the least inter- and intra- variability. The program returns an arbitrary stability value and standard error value The gene with the lowest stability value is supposed to be the most stable gene out of the tested set of candidates (Andersen et al., 2004).

Bestkeeper defines the expression variability of a pool of genes by pair-wise correlation based on the raw experimental threshold cycle (Ct) values generated by real-time PCR. The software calculates the standard variation (SD), the Pearson coefficient of correlation (r) and the probability (p). All genes with SD values greater than 1.0 are considered to be unsuitable reference genes. A correlation coefficient r value close to 1.0, with a p-value

statistically significant ($p<0.05$), is an indicator of stable gene expression (Pfaffl et al., 2004).

2.10 General Bacteria and Cloning Methods

2.10.1 Bacterial culture media

The bacteria were cultured in Luria-Bertani broth (Lysogeny Broth, LB) and in its solid form. Agar and broth powder were dissolved in deionized water and sterilized by autoclaving. Bacterial growth media were autoclaved and cooled to 60°C prior to the addition of filter sterilized antibiotic.

Table 2.14: Recipe for Luria-Bertani broth and its solid form

Reagents	Composition		Reagents	Composition
LB-broth	10 g/L Bacto Tryptone 5 g/L Bacto Yeast extract 10 g/L $NaCl_2$		LB-Agar	10 g/L Bacto Tryptone 5 g/L Bacto Yeast extract 10 g/L $NaCl_2$ 15 g/L of Davis agar
Antibiotic-Ampicillin	final concentration of 100 µg/ml			

2.10.2 Plasmid DNA Mini- and Midi-preparation

LB-broth with bacteria containing the plasmid of interest was incubated overnight at 37°C with shaking. The overnight culture was centrifuged at 13,000 x g and the cell pellet was used to isolate plasmid DNA. Two protocols were used to isolate plasmid DNA. The PureLink™ Quick Plasmid Miniprep Kit (Invitrogen) yielded DNA of sufficient purity for automated sequencing. The HiSpeed Plasmid Midi Kit (QIAGEN) yielded DNA of sufficient purity for automated sequencing and for sub-cloning procedures. Cells were lysed following the manufacturer's instructions;

2.10.3 Transformation of Chemically Competent Cells

An aliquot of 2 µl of the required vectors was added into 10 µl of One Shot® Match1™ T1® Competent Cells (Invitrogen). The ligation mixture was incubated on ice for 30 minutes so the vectors could adhere to the bacteria cell wall. Cells were heat-shocked for 30 seconds at 42°C, transferred on ice and 250 µl of S.O.C. medium were added. Bacteria were then grown horizontally at 37°C for 1 hour at 225 rmp. Ten to fifty µl of competent cells were spread on a pre-warmed LB-agar and ampicillin selective plate (100 mg/ml). Plates were incubated at 37°C overnight.

2.10.4 DNA cloning

Generating 'A-tailing' to blunt-ended PCR fragments

Once the desired insert is produced by PCR, single 3'adeninde overhangs were created to allow the following cloning in TOPO-TA® vectors (Invitrogen). The reaction was carried out in a volume of 10 µl containing 0.2 µl of HotMaster Taq DNA polymerase (5-PRIME™) (5 U/µl), 1 µl of 10X Master Buffer, 1 µl of 2mM dATPs. The mix was vortexed, briefly centrifuged and incubated at 70°C for 15 minutes.

Cloning in TOPO-TA vectors

The TOPO-TA® cloning reaction were carried out using 1 µl of pCR® 2.1-TOPO vector, 4µl of *MECP2* 3'UTR fragments with single 3'adeninde overhangs, 1 µl of salt solution and milliQ to a final volume of 6 µl. The reaction was incubated 1 hour at room temperature.

Site Directed Mutagenesis

This protocol involved the use of the QuickChange® II Site-Directed Mutagenesis Kit (Stratagene) following the manufacturer's instructions. This protocol was used to remove errors in the *MECP2* 3'UTR fragments formed by the DNA polymerase during replication and to create alterations in the same fragments after cloning in pCR® 2.1-TOPO vector. The same protocol was used to create a new restriction enzyme recognition site in the *MECP2* promoter-pCBG99 to allow the cloning of the *MECP2* 3'UTR fragments (*Appendix 3, Figure 2).*

One hundred nanograms of the template plasmid was amplified using 1 unit of the PfuUltra™High-Fidelity DNA polymerase (2.5 U/µl, Stratagene), the PCR reaction was carried out using 200 ng/µl of each primers, 2 µl of 2 mM dNTPs and milliQ water to a final volume of 10µl. The cycle conditions were 95°C for 30 seconds: 18 cycles of 95°C for 30 seconds, 55°C for 1 minute, 68°C for 1minute/kb. The DNA amplified was digested with 1 µl of *DpnI* restriction enzyme (10U/ µl, New England BioLabs Inc) and 1X NEBuffer 4 at 37°C for 2 hours, followed by 20 minutes of enzyme inactivation step at 80°C. Two µl of the digested PCR reaction was then used to transform One Shot® Match1™ T1® Competent Cells as outlined in *Section 2.10.3 Transformation of Chemically Competent Cells)*. Selected colonies were sequenced using the universal primers M13F/M13R (Macrogen) for checking the base change.

Digestion with Restriction Endonucleases

Restriction endonucleases enzymes were used to cleave DNA and plasmids for promote the ligation and to check the ligation. NEBcutter V2.0 website (http://tools.neb.com/NEBcutter2/) was used to display the plasmid restriction map and to select specific restriction endonuclease sites. The restriction mix for cloning the *MECP2* 3'UTR into the backbone vector (*MECP2* promoter/5'UTR-pCBG99) was incubated at 37°C overnight according to *Table 2.15*. The restriction endonucleases were not suitable for a simultaneously double digestion, so sequential digestions were performed using each enzyme in the appropriate NEBuffer (New England BioLabs Inc).

Table 2.15: Restriction cutting mix for restriction analysis

Reagents	Final Concentration	
10X NEBuffer 3	1X	
10X NEBuffer 4	----	1X
10X BSA	1X	----
Enzyme SalI (20U/ µl)	1 µl	----
Enzyme PspXI (5U/ µl)	----	1 µl
Plasmid DNA	1-5 µg	1-5 µg
milliQ-final volume of	20 µl	20 µl

Rapid purification and concentration of DNA fragments and plasmid

The PureLink™PCR Micro Kit (Invitrogen) and the QIAquick® Spin (Qiagen) Kit were used to isolate, purify and extract high concentration DNA following the manufacturer's instructions. The elution steps were modified by incubating the columns for at least 30 minutes at room temperature.

Plasmid de-phosphorylation

The backbone vector *MECP2* promoter/5'UTR-pCBG99-Chroma-Luc™ Reporter vector (Promega) was linearized with *SalI* and *PspXI* restriction enzyme as previously described. Antarctic Phosphatase (New England BioLabs Inc) is an enzyme that catalyzes the removal of 5´phosphate groups from DNA, so the fragments lacking in the 5´ phosphoryl termini cannot self-ligate. The de-phosphorylation reaction was carried out in a volume of 30 µl containing 3 µl of 10XBuffer, 10 Units of enzyme and cut vector. The mixture was vortexed, briefly centrifuged and incubated at 37°C for 30 minutes and inactivated at 65°C for 10 minutes.

Ligation reactions

Ligation reactions were carried out using the T4DNA ligase (New England BioLabs Inc) derived from the T4 bacteriophage. T4DNA ligase catalyses the joining the 5´-phosphate and the 3´-hydroxyl groups of cohesive-ended adjacent nucleotides with the concomitant

hydrolysis of ATP to AMP and inorganic phosphate. The optimal molar ratio insert:vector was calculated using the following formula:

$$\text{ng of insert} = \frac{\text{ng of vector x kb size of insert}}{\text{kb size of vector}} \times \text{molar ratio of } \frac{\text{insert}}{\text{vector}}$$

Ligation reactions were conducted in 20 µl including 1 µl of T4DNA Ligase (2000U/µl) and 2 µl of 10X T4DNA Ligase Buffer and 5:1 or 10:1 DNA molar ratios of the insert:vector. The mixture was vortexed, briefly centrifuged and incubated at room temperature for 1 hour.

2.10.5 Bacteria Storage

For long term storage, 250 µl of 60% filter sterile glycerol was added to 750 µl of bacterial culture on ice. The culture and glycerol were mixed thoroughly to ensure even distribution of cells, and the cell suspension was stored at -80°C. For short term storage the bacteria were maintained on solid media at 4°C with appropriate selection.

2.11 General Tissue Culture Methods

2.11.1 Maintenance of Cell Cultures

COS-7 and SH-SY5Y cell lines were cultured in DMEM/F12 (Invitrogen) supplemented with 10% foetal bovine serum (FBS) (Invitrogen), 0.1 U/ml of penicillin (Invitrogen) and 100µg/ml of streptomycin (Invitrogen) and incubated in a 37°C incubator with 5% CO_2.
The PC12 cell line was cultured in DMEM-High Glucose (Invitrogen) supplemented with 5% Foetal Bovine Serum (FBS) (Invitrogen), 10% Horse Serum (Invitrogen), 0.1 U/ml of penicillin (Invitrogen) and 100µg/ml of streptomycin (Invitrogen) and incubated at 37°C incubator with 5% CO_2.

2.11.2 Splitting of Tissue Culture Cells

Cells were split 1:4-1:10 every 3-5 days depending on the speed of growth of individual cell line and the density required. Cells were washed with 10 ml of Phosphate Buffered Saline (PBS, Invitrogen) and then incubated at 37°C with TrypLE™ Express (Invitrogen) until the cells detached from the flask. The cells were then diluted 3 times the volume of TrypLE™ Express in culture media, transferred to a 50 ml falcon tube and pelleted by

centrifugation at 300 x g for 3 minutes at room temperature. The resulting cell pellet was resuspended in an appropriate volume of culture media and aliquoted into new culture flasks/plates. The appropriate volumes of media/TrypLE™ Express for various culture flasks are summarized in the *Table 2.16*.

Table 2.16: Media volumes for Tissue Culture Flasks

Culture vessel	Culture volume	Volume of TrypLE™ Express
T175 cm^2 flask	25-30 ml	10 ml
T75 cm^2 flask	10-15 ml	2 ml
T25 cm^2 flask	5-7 ml	500 µl
10 cm culture dish	10-15 ml	2 ml
6 well plate	2 ml	200 µl
12 well plate	1 ml	100 µl
24 well plate	500 µl	100 µl

2.11.3 Freezing Down of Cell Lines

Cell pellets from a T175 cm^2 culture flasks were collected as described in *section 2.11.2 Splitting of Tissue Culture Cells* and resuspended in 3 ml of DMEM/F12, 10% FBS, 10% DMSO for COS-7 cells and in DMEM-High glucose, 5% FBS, 10% Horse Serum, 10% DMSO for PC12 cells. Then 1 ml aliquots were transferred into 2 ml cryogenic vials. The cells were immediately transferred into a cryogenic freezing container (Nalgene) and transferred to a -80°C freezer, and when frozen they were placed in liquid nitrogen for long term storage.

2.11.4 Recovery of Frozen Cell Lines

Cell were thawed in a 37°C water bath and then transferred immediately into 10 ml of growth media and transferred into an appropriate culture flask containing fresh media.

2.11.5 Transfection of Cells Using Liposomes

The day before transfection, cells were trypsinized, counted and plated down into fresh culture dishes at an appropriate cell density. Plasmid DNA and PLUS™ Reagent (Invitrogen) were diluted in Opti-MEM® I Reduced Serum Medium (Invitrogen) and incubated at room temperature for 10 minutes.

Lipofectamine™ LTX (Invitrogen) was added into the above diluted DNA solution, mixed gently and incubated for 25 minutes at room temperature to form the DNA-Lipofectamine™ LTX complexes. One hundred µl of the DNA-Lipofectamine™ LTX complexes was then added drop wise to the cells and the 24 well-plate was returned to

the incubator at 37°C, 5% CO_2. The media was changed 4 hours post-transfection with fresh growth media if necessary. Volumes of the reagents used are shown in *Table 2.17.*

Table 2.17: Media and Transfection Reagents used in Tissue Culture

Cell line	Culture vessel	Volume plating media	Lipofectamine™ LTX	Plasmid DNA	PLUS™ Reagent
COS-7	24-well plate	500 µl	4 µl	750 ng	0.75 µl l
PC12	24-well plate	500 µl	4 µl	750 ng	0.75 µl l
SH-SY5Y:	24-well plate	500 µl	3 µl	750 ng	0.75 µl l

2.12 Luciferase Assay System

The Chroma-Glo™ Luciferase Assay System (Promega) was used to perform luciferase assays in order to study the role of two *MECP2* 3'UTR fragments in mammalian cell lines. This assay is designed to generate luminescence from a click beetle Chroma-Luc™ luciferases that emit green and red colours of light (maximum 357 nm and 613 nm respectively) after the addition of a single substrate.

The Chroma-Luc™ luciferases catalyse the oxidation of luciferin (*Figure 2.4*). Luciferin is a common term defined as an organic compound that exists in a luminous organism and provides the energy for light emission by being oxidized, normally in the presence of a specific luciferase. Generally, the luciferase enzyme catalyses the conversion of D-luciferin into luciferyl adenylate (LH2-AMP) using adenosine tri-phosphate (ATP) and magnesium (Mg^{2+}) with the release of inorganic pyrophosphate (PP_i). Luciferase acts as an oxygenase on adenyl luciferin to produce an electron excited state oxyluciferin molecule and carbon dioxide (CO_2). Then, the decay of the resulting oxyluciferin results in the emission of a photon of red or green light.

Glo Lysis Buffer, Chroma-Glo™ Luciferase substrate and the 24 well-plate were equilibrated at room temperature, as luciferase activity is temperature dependent. The 24 well-plate was washed twice with 1X PBS and 100 µl of Glo Lysis Buffer was added to allow the lysis of the cell cultures. After 5 minutes of incubation, the lysate cells were transferred into a white and black matrix isoplate (Wallac Isoplate™, Perkin Elmer™) and 100 µl of Chroma-Glo™ Luciferase substrate was added before measuring the luminescence. The quantitative luminescence analysis was carried out on the VICTOR™ X Light Luminescence Plate Reader (Perkin Elmer™) using two different filters to separate the red and green Chroma-Luc signals.

A 510/60 nm and a 610 nm long-pass filters were used to separate the red and green luminescence signals (*Figure 2.3*). The two signals have different kinetics. The red emitting luciferase signal (pCBR vector) has an half-life of approximately 30 minutes, while the green emitting luciferase signal (pCBG99 vector) has an half-life of more than 5 hours, so the red signal was measured before the green signal.

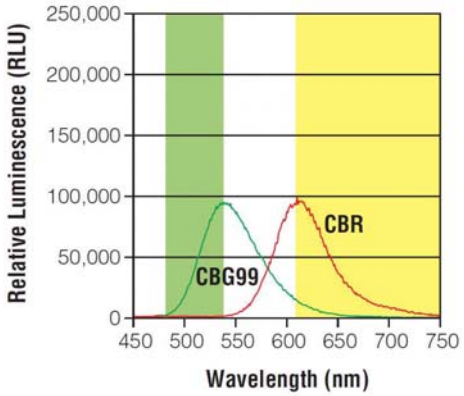


Figure 2.3: Emission spectra of the Chroma-Luc™ enzymes

The picture represents the emission spectra of the Chroma-Luc™ Luciferase vectors. The vectors used are pCBR-Control vector represented in red and pCBG99-Control vector represented in green. After addition of the Chroma-Glo™Luciferase substrate, the pCBR-Control vector emits red light at 613 nm, while the pCBG99-Control vector emits green light at 537 nm. The yellow shaded area on the right indicates the maximal transmittance range of a 610 long-pass filter, while the green shaded area on the left indicates the maximal transmittance range of a 510/60nm filter. Picture adapted from Chroma-Glo™ Luciferase Assay System-Technical Manual, Promega

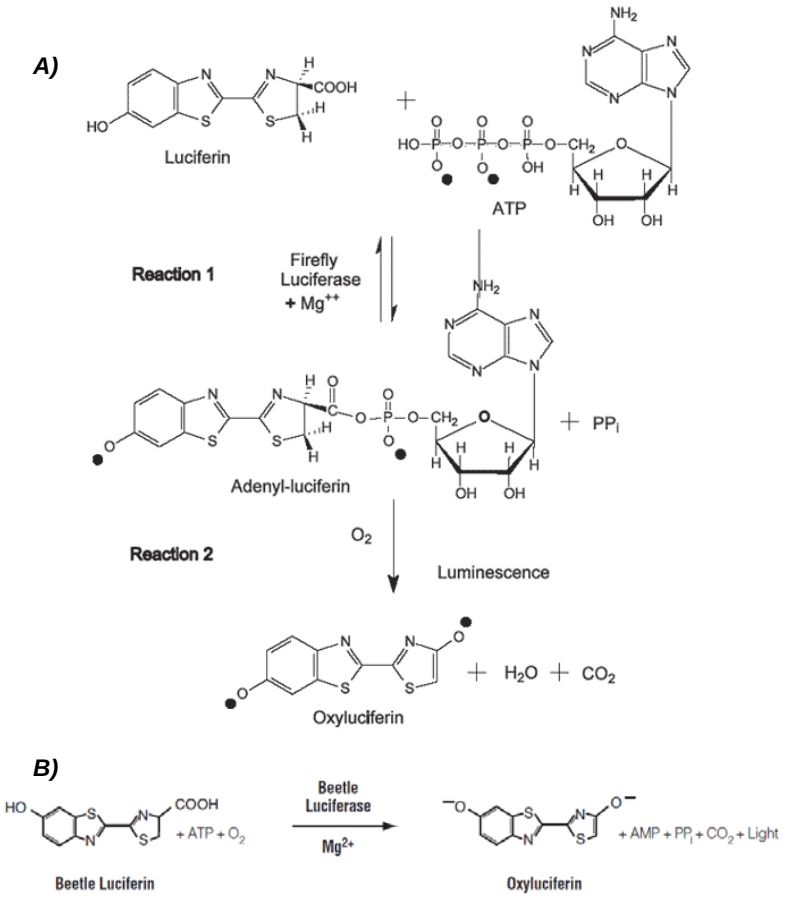


Figure 2.4: General bioluminescent reaction catalysed by luciferase enzyme

A) The picture shows the general biochemical mechanism of thebioluminescent reaction of luciferase. Specifically representing the reaction catalysed by firefly luciferase. Firefly luciferase protein catalyses the oxygenation of luciferin protein using adenosine tri-phosphate (ATP) and magnesium (Mg^{2+}) to yield an adenyl-luciferin intermediate with formation of inorganic pyrophosphate (PP_i). The intermediate is then converted to oxyluciferin, a highly unstable, singlet-excited compound with release of carbon dioxide (CO_2). When the energy levels of the oxyluciferin excited states fall to the ground states, oxyluciferin emits red or green light. Picture from Sigma-Aldrich ®

B) The picture shows the bioluminescence reaction catalyzed by the click beetle luciferase. The Chroma-Luc™ luciferase catalyses the oxygenation of click beetle luciferase in presence of ATP, Mg^{2+} and O_2. Picture from Chroma-Glo™ Luciferase Assay System- Technical Manual, Promega

2.13 Bioinformatics tools

All sequences were compared using the default settings of the Blast2seq web tool (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). The sequence alignment was performed using ClustalW program (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

2.13.1 ESE and ESS motifs *in silico* analysis

The *in silico* predictions of splicing enhancer elements were performed using web server tools such as RESCUE-ESE (http://genes.mit.edu/burgelab/rescue-ese/), ESE finder (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home) and FAS ESS (http://genes.mit.edu/fas-ess/). ESE and ESS motifs are specific short *cis*-regulatory elements that recruit protein factors for splicing regulation. Creation or loss of ESE or ESS motifs lead to an aberrant splicing compared to the conservative splicing mechanism affecting gene expression and causing human diseases (Faustino and Cooper, 2003). Human serine/arginine rich proteins (SR proteins) are a conserved family of nine proteins that regulate the RNA splicing machinery. SR proteins bind to ESE motifs through the N-terminal RNA-binding domain (RRM domain) and interact with other proteins through the C-terminal tail rich in Arg-Ser repeats (RS domain) (Shepard and Hertel, 2009).

A variety of bioinformatics programs have been developed to study or predict splice signals, so the potential effect of novel non-synonymous mutations was examined using RESCUE-ES, ESE-finder and FAS ESS web-server softwares. These tools have different approaches, for instance FAS-ESS web server predicts the effect of nucleotide polymorphism on the splicing silencer motif (ESS). This tool uses two ESS library called FAS-hex2 (176 sequences) and FAS-hex3 (103 sequences) (Wang et al., 2004). RESCUE-ESE is online server that compares an input sequence to a non-redundant list of 238 human ESE hexamers. A query sequence is blasted to this database of ESE motifs, as described by Fairbrother and colleagues (Fairbrother et al., 2002; Fairbrother et al., 2004). ESE-finder allows the identification of putative ESE motifs responsive to human serine/arginine rich proteins (SR proteins). SF2/ASF, SC35, SRp40 and SRp55 are members of this family and they are analysed by this web-server. The default threshold values were selected and the four SR proteins matrices were used simultaneously during ESE-finder analysis (Cartegni et al., 2003).

2.13.2 MicroRNA:3'UTR binding target site interaction *in silico* prediction

A growing number of different bioinformatic softwares are widely used to predict the interaction between microRNA and target mRNA. To investigate the possible effect of mutations in the *MECP2* 3'UTR, 5 different prediction software tools were used according to authors' instructions such as: microRNA.org (www.microrna.org/), MicroInspector (http://bioinfo.uni-plovdiv.bg/microinspector/), TargetScan (www.targetscan.org/), PicTar (http://pictar.mdc-berlin.de/) and Diana (http://diana.cslab.ece.ntua.gr/).

microRNA.org

The microRNA.org web-target prediction tool was based on the miRanda algorithm and more recently it is based on the mirSVR (Support Vector Regression) approach The miRanda algorithm analyses the sequence complementarily between a set of mature microRNAs and a given mRNA using a dynamic programming algorithm that weight match and mismatch of base pairs (Betel et al., 2008). Further, miRanda estimates also the free energy of formation of the microRNA:mRNA duplex using Vienna folding package. The mirSVR is a regression machine learning method trained to predict target site efficiency considering the secondary structure accessibility and conservation of the site. The MirSVR score is a probability of down regulation of canonical target genes where the more negative you set the mirSVR score cut-off the more you restrict the analysis.

PicTar

The PicTar is a "probabilistic identification of combinations of target sites" algorithm that identifies common targets for microRNAs in multiple species such as mice, vertebrates, flies and worms (Krek et al., 2005). The tool determines a final PicTar score which describes the likelihood of a gene being target to the given miRNA set and the algorithm also filters the results according to the free energy score of the microRNA:mRNA target match.

TargetScan

TargetScan was the first method to be applied for human miRNA target prediction, using mouse, rat and fish genomes for conservation analysis (Lewis et al., 2003). Initially, TargetScan searched for perfect complementarity match between the microRNA and the target sequence, and then the thermodynamic stability was evaluated using the Vienna folding websuite. The improved version of the algorithm is less accurate and imperfect seed matches and 3' compensatory pairing are now also predicted (Mazière and Enright, 2007).

DIANA

The DIANA web server search for a stringent nucleotide seed match between each microRNA and miRNA recognition element (MRE). The free energy of the potential binding site is calculated at each step and compared with the results obtained from shuffled sequences with the same dinucleotide composition (Krek et al., 2005; Maragkakis et al., 2009).

MicroInspector

The prediction tool uses a "dynamic algorithm" to align microRNAs to the target sequence that allow G:U wobble base pairs. The web-server also calculates the folding free energy for each predict duplex miRNA:mRNA target using Vienna folding web server. Subsequently, the structures with low energy value due to self-pairing are eliminated from the outputs though the use of a second filter. Interestingly, this web-server provides the opportunity to run the analysis on a specific 3'UTR sequence of interest (Rusinov et al., 2005).

2.14 Statistical analysis

To examine *MECP2* isoform expression levels in our cohort of patients and healthy controls and to examine the Lucifarase expression levels in different cells lines such as COS-7, PC12 and SH-SY5Y, a non-parametric Mann-Whitney U statistical test was performed which does not assume normal distribution. The Mann-Whitney U test, also called as the rank sum test, is a non-parametric test for two independent samples, which does not use calculation of variance as part of the testing hypothesis, but relies on ranking of the numerical values. Statistical analysis was performed with the GraphPad Prism® Software Inc.

The *null hypothesis* H_0 assumes that the two set of samples are stochastically equal and so there are no systematic differences between the samples. As such, there would be no difference in expression of *MECP2* isoforms between patients and normal controls; and there would be no difference in expression of the luciferase reporter gene between *MECP2* 3'UTR vectors carrying a mutation and the wildtype *MECP2* 3'UTR vector.

The *alternative hypothesis* H_1 assumes that the two set of samples do systematically differ; so differences in the expression of the *MECP2* isoforms between patients and normal controls would be observed; and there would be observed differences in the expression of the luciferase reporter gene between *MECP2* 3'UTR vectors carrying a mutation and the wildtype *MECP2* 3'UTR vector.

3. Results

3.1 The *MECP2* gene in patients with RTT and RTT-like phenotype

Quantitative Real-Time PCR: *PGK1* is the reference housekeeping gene

The expression levels of several reference genes were measured by TaqMan probe-based Real-Time PCR assay from control cDNA samples (n=3). We selected four housekeeping genes as candidate reference genes: β-actin (*ACTB,* OMIM **102630*), phosphoglycerate kinase 1 (*PGK1,* OMIM *311800), hypoxanthine-guanidine phosphoribosyl transferase (*HPRT,* OMIM *308000) and glyceraldehyde-3 -phosphate dehydrogenase (*GAPDH,* OMIM *138400). The expression profile of these genes were analysed with three Microsoft Excel applications called *geNorm*, *NormFinder* and *Bestkeeper*, to determine the most stable housekeeping gene from the set of candidates.

GeNorm graphically sorted our 4 housekeeping genes on the basis of their expression M-value from the least stable to the more stable: *HPRT* (M=0.820), *ACTB* (M=0.574), *GAPDH* and *PGK1* (M=0.409). The program reported that *GAPDH* and *PGK1* were the most stably expressed pair of reference genes with M-value equal to 0.409 (*Figure 3.1*).

Normfinder analysis on our data indicated that *GAPDH* and *PGK1* were the genes with the lowest stability value, and therefore deemed the most stable, a as outlined in *Table 3.1*.

The *Bestkeeper* analysis on our candidate reference genes showed a standard deviation value higher than 1 for *ACTB* and *GAPDH* genes (SD = 1.09) and then these genes were eliminated from further analysis. The analysis also highlighted *HPRT* and *PGK1* as the most suitable reference genes with r >0.97 and p-value = 0.001 (*Table 3.1*.

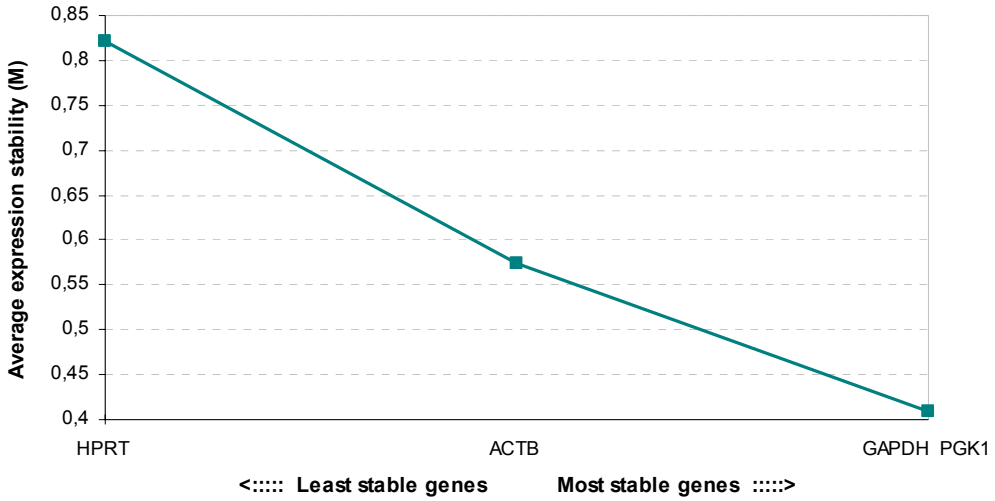


Figure 3.1: Expression stability of housekeeping genes calculated by GeNorm application

GeNorm algorithm calculates the gene expression stability (M-value) by step-wise elimination among 4 reference housekeeping genes. The M-value is the average pair-wise variation of a gene with all the other reference genes. The M-value is an indicator of the gene stability and the more stable gene has a low M-value.

Table 3.1: Normfinder and Bestkeeper housekeeping gene expression stability

Normfinder		
Gene name	**Stability value**	**Standard error**
HPRT	0.686	0.154
ACTB	0.460	0.122
GADPH	0.120	0.166
PGK1	0.147	0.144

Bestkeeper		
Gene name	**standard deviation SD**	**correlation coefficient r**
HPRT	0.85	0.972
ACTB	1.09	0.937
GADPH	1.09	0.975
PGK1	0.87	0.981

Real Time Quantification Assay: Comparative ΔΔCt analysis

Generally, the expression level of the gene of interest (GOI) can be measured by absolute or relative quantification during a Real-Time PCR assay. *Absolute quantification* is used to measure the input copy numbers using a standard curve, while *relative quantification* assumes that the amplification kinetics of the GOI and the housekeeping gene (reference gene) are approximately equal. This method is usually used when a large number of samples are analysed since it does not require a standard curve. Different mathematical methods are available to calculate the gene expression level. The Comparative Quantification Algorithm $2^{(-\Delta\Delta Ct)}$ was used in this study and it assumes that the housekeeping gene and the GOI are constantly and uniformly expressed in all samples. The amplification efficiency analysis was then carried out in the presence of a TaqMan probe specific to isoforms *MECP2_e1* and *MECP2_e2* and *PGK1* (*Figure 3.2*).

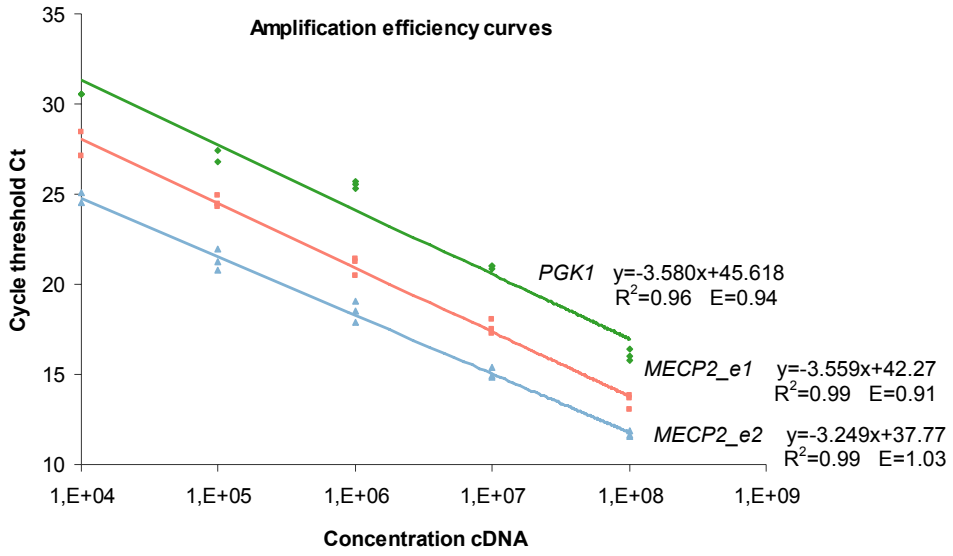


Figure 3.2: Amplification efficiency curves of *MECP2* isoforms and of the *PGK1* gene

Standard curves were determined by plotting the Ct values versus the log cDNA by serial dilutions of control samples using specific TaqMan Probe for *MECP2_e1* isoform (pink bar), *MECP2_e2* isoform (light blue bar) and the housekeeping gene *PGK1* (green bar). Amplification Efficiency = [$10^{\wedge(1/\text{slope})}$) – 1].

Real-Time PCR is characterized by a few important parameters such as the correlation coefficient (R^2), the efficiency of the reaction (E) and the slope value (m). The correlation coefficient represents the linearity of the curve and ideally its value is equal to 1. The data showed that the target and housekeeping gene have good R^2 values (*e1*: 0.99, *e2*: 0.99 and *PGK1*: 0.96, *Figure 3.2*). A second important parameter is the slope of the amplification curve used to calculate the efficiency of the reaction (E). Ideally the slope value (m) should be -3.32 which correspond to a 100% efficiency or two-fold amplification. The efficiency E [(E= $10^{\wedge\text{-}1/\text{slope}}$) -1] is hypothetically equal to 100% meaning a doubling of the amplicon at each cycle. The presence of inhibitors in the PCR, the length of the amplicon and the GC content can influence the E value; for these reasons, an amplification curve with a slope value between -3.58 and -3.10 corresponding to an E value of 90-110%, is considered a good reaction. Our data showed that *MECP2* isoforms and the housekeeping gene have good m and E values: *e1*: E=0.91, m= -3.559; *e2*: E=1.03 m=-3.249; *PGK1* E=0.94 m=-3.580.

Analysis of *MECP2* transcript levels in RTT patients

The expression pattern of *MECP2* isoforms were then determined in classical (n=9) and atypical (n=13) RTT patients with no known mutations in the *MECP2* or *CDKL5* genes.

The analyses showed that there was a significant difference in the expression levels of *MECP2* isoforms between normal controls and RTT patients with unknown mutations (*Figures 3.3* and *3.4, Table 3.2*). A statistical difference was seen in *MECP2_e1* expression levels in eight classical RTT patients (8/9) and in seven atypical RTT patients (7/13). Specifically, three classical patients (*Figure 3.3*) and two atypical patients (*Figure 3.4*) had elevated levels of *MECP2_e1,* while five classical patients and two atypical patients had low expression values in the *MECP2_e1* isoform, when compared to normal controls.

A statistical difference was seen in *MECP2_e2* isoform expression levels in eight classical (8/9) and in ten atypical patients (10/13) were statistically different from normal controls. Four classical patients and five atypical patients showed low expression values in *MECP2_e2,* while four classical and four atypical patients had elevated *MECP2*_e2 isoform expression levels. Three classical RTT patients (Samples 10, 37 and 260) and one atypical RTT patient (Sample 230) showed decreased expression levels in both *MECP2* isoforms. Two atypical RTT patients (samples 21 and 233) showed high expression levels in both *MECP2* isoforms. Statistically significant expression levels in each patients relative to normal controls, are indicated with asterisks, where * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 3.2: The *MECP2* isoform expression levels in classical and atypical RTT patients *MECP2*/*CDKL5* mutation negative

	Mean $2^{-\Delta\Delta Ct}$ *MECP2_e1*	Standard deviation	Fold change ↑↓	Mean $2^{-\Delta\Delta Ct}$ *MECP2_e2*	Standard deviation	Fold change ↑↓
Control 1	0,845	0,060		0,888	0,355	
Control 2	1,189	0,113		1,200	0,236	
Classical RTT patients						
Sample 4	0,520	0,0649	↓	2,440	0,1756	↑
Sample 10	0,075	0,03	↓	0,020	0,009	↓
Sample 15	0,504	0,0640	↓	3,194	0,7388	↑
Sample 37	0,190	0,0897	↓	0,551	0,0907	↓
Sample 50	3,168	0,78	↑	0,952	0,2842	=
Sample 126	2,799	0,0888	↑	0,237	0,0952	↓
Sample 193	0,681	0,1922	=	2,155	0,1829	↑
Sample 203	0,276	0,0472	↓	3,153	0,2662	↑
Sample 260	0,090	0,0412	↓	0,310	0,077	↓
Atypical RTT patients						
Sample 21	2,005	0,7186	↑	2,962	0,3294	↑
Sample 66	1,314	0,4964	=	0,129	0,1272	↓
Sample 91	0,742	0,2690	=	0,438	0,087	↓
Sample 164	0,567	0,1062	↓	2,485	0,1721	↑
Sample 184	0,339	0,1276	↓	1,005	0,1634	=
Sample 206	0,727	0,1587	=	0,166	0,0825	↓
Sample 215	0,810	0,1138	=	2,041	0,3179	↑
Sample 228	0,819	0,1477	=	2,454	0,9871	↑
Sample 230	0,310	0,0297	↓	0,223	0,0059	↓
Sample 263	1,209	0,2959	=	0,114	0,0177	↓
Sample 327	0,346	0,2386	↓	1,001	0,5687	=
Sample 386	0,417	0,1001	↓	1,039	0,2239	=
Sample 233	2,632	0,3576	↑	13,333	1,9066	↑

The fold change $2^{\wedge-\Delta\Delta Ct}$ is expressed as an average; the increased or decreased in the *MECP2* isoform expression levels are reported with up and downwards arrows respectively. The samples with high or low expression levels in both isoforms are represented in red.

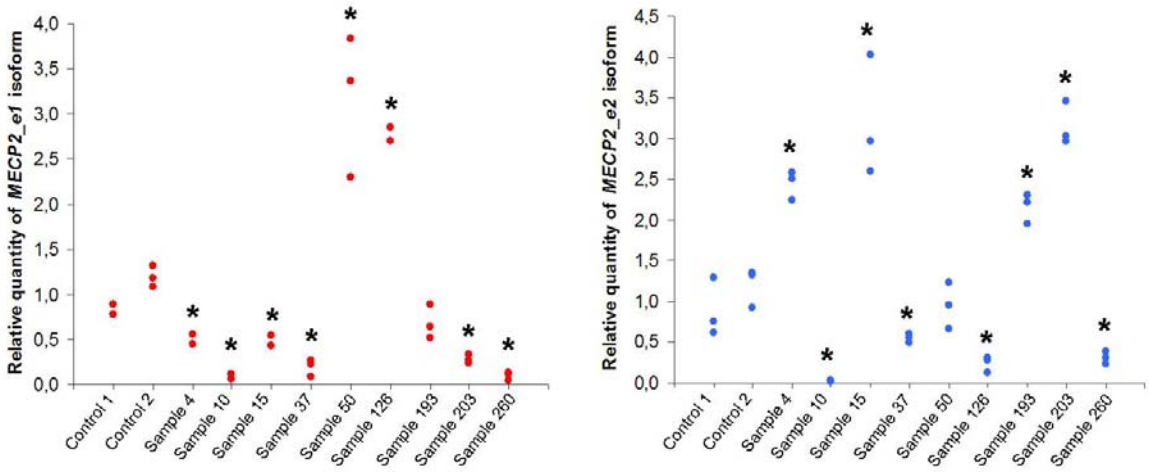


Figure 3.3: Expression levels of *MECP2* isoforms in classical RTT patients without *MECP2* mutations

The expression levels of both *MECP2_e1* (red dots) and *MECP2_e2* (blue dots) isoforms were calculated according to the Comparative $2^{(-\Delta\Delta Ct)}$ method. The expression level of the two isoforms was normalized to *PGK1* reference gene. The chart includes 9 female classical RTT patients mutation negative for the *MECP2* gene. The Y-axes represents the fold-change in *MECP2* mRNA expression. Statistical significant differences between normal controls and each of the patients are indicated with asterisk, where *p<0.05, **p<0.01, ***p<0.001.

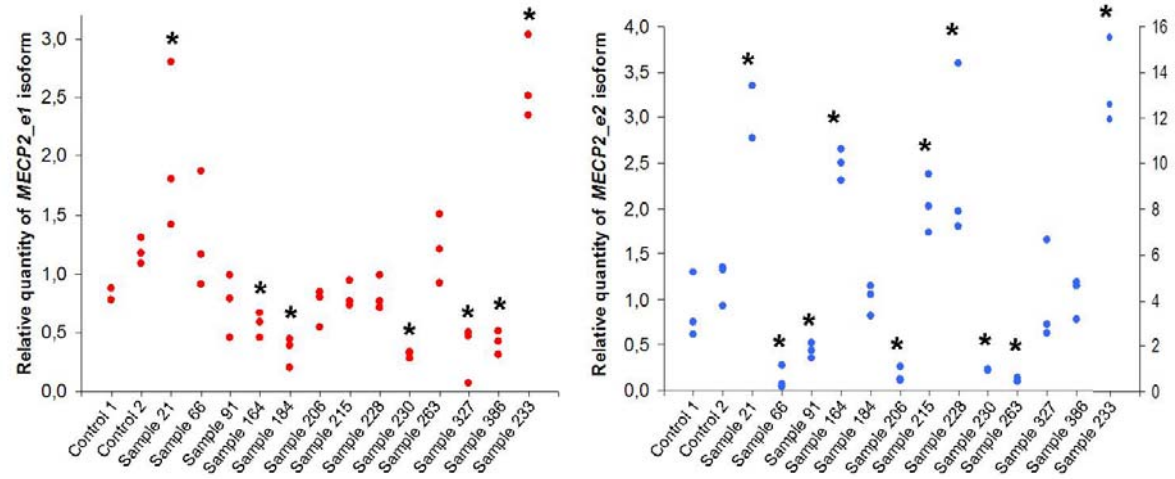


Figure 3.4: Expression levels of *MECP2* isoforms in atypical RTT patients without *MECP2* mutations

The expression levels of both *MECP2_e1* (red dots) and *MECP2_e2* (blue dots) isoforms were calculated according to the Comparative $2^{(-\Delta\Delta Ct)}$ method. The expression level of the 2 isoforms was normalized to *PGK1* reference gene. The chart includes 13 female atypical RTT patients, mutation negative for the *MECP2* gene. The Y-axes represents the fold-change in *MECP2* mRNA expression. Statistical significant differences between normal controls and each of the patients are indicated with asterisk, where *p<0.05, **p<0.01, ***p<0.001.

Screening of the *MECP2* 3'UTR in RTT patients

A fragment of 4512 bp of the 3'UTR of *MECP2* gene was screened for sequence variation by direct sequencing in a cohort of RTT patients (n = 17) which were *MECP2* or *CDKL5* mutation negative. The fragment was amplified and was then sequenced in a unidirectional forward walk for a total of 9 overlapping fragments per patients (*Figure 3.5*).

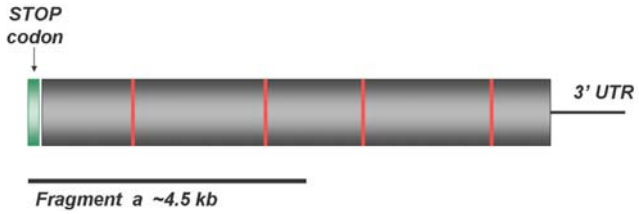


Figure 3.5: Schematic representation of the 3' UTR of the *MECP2*

The length of the 3' untranslated region of *MECP2* is represented as a dark grey bar, not in scale. The four polyadenylation signals are represented by red lines. The black line under the 3'UTR represents the region amplified. The stop codon is represented by a green box.

The screening results showed the presence of an already reported sequence variation called c.*371 G>C previously identified in an autistic male (Coutinho et al., 2007). A previously unreported alteration c.*1354 G>A has been identified in two unrelated RTT patients. Two Single Nucleotide Polymorphisms (SNPs) were identified: rs3027924 C/G and rs2734647 A/G. All 17 patients were carrier for the rs2734647 A/G SNP, while two patients had also rs3027924 C/G SNP (*Figure 3.6*).

The c.*371 G>C variation was detected in a RTT female (sample 126), which is characterized by significant high level of *MECP2_e1* and low level of *MECP2_e2* isoforms (*Figure 3.3*). The major problem with this analysis was the difficulty to test both parents to exclude the presence of this variant.

In order to study the conservation of the SNPs nucleotides, a comparative sequence analysis was performed using ClustalW alignment tool. The sequence alignments showed that the reference allele at position c.*3638 is Adenine and it is conserved in *H. sapiens* and *D. rerio*, while the allele Guanine is present in *G. gorilla*, *M. mulatta*, *M. musculus* and *R. Norvegicus*. The reference allele at position c.*878 is Cytosine and it is not conserved in the species analysed (*Figure 3.7*).

The two SNPs detected in our population are already listed in the NCBI SNPs database (www.ncbi.nlm.nih.gov/snp). Allele and genotype frequencies of the rs2734647 A/G are reported in the database, while information about the population diversity is incomplete for the rs3027924 C/G.

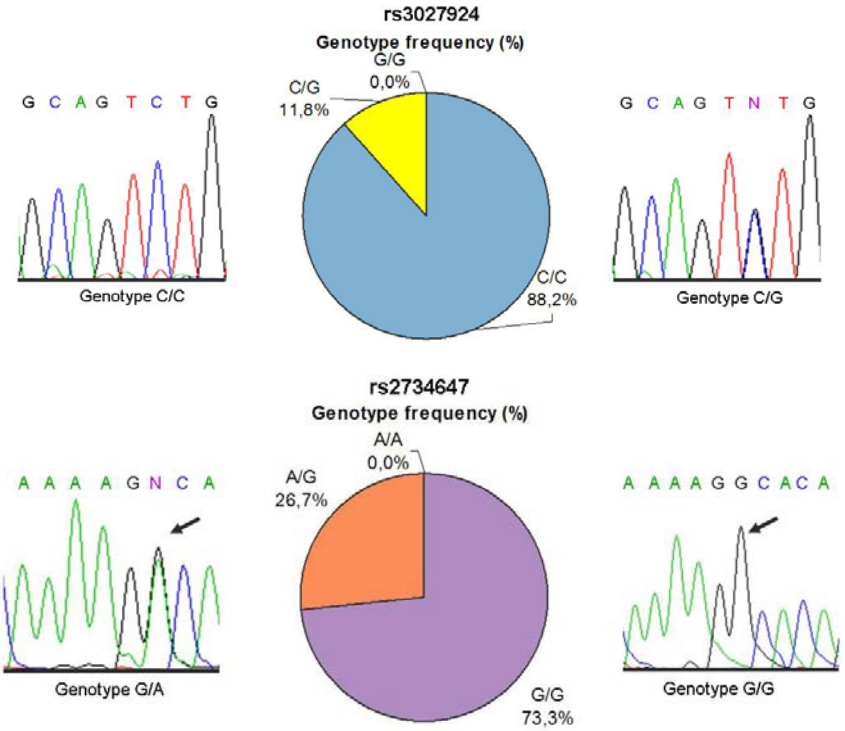


Figure 3.6: Sequence chromatograms and genotype frequency of the SNPs detected in our cohort of RTT patients

The DNA sequence chromatograms show the SNPs detected in RTT patients with no mutations in the *MECP2* or *CDKL5* genes. The sequence chromatograms of the rs3027924 C/G (on the top), the rs273467 A/G (on the bottom) and genotypes found in our cohort of patients are represented in the figure.

A)

```
H. sapiens          -CAGCTCCCTTCTGTTGATGCTGTATTCCCATATCAAAAGACACAGGGGACACCCAGAA 58
G. gorilla          -CAGCTTCCTTCTGTTGATGCTGTATTCCCATATCAAAAGGCACAGGGGACACCCAGAA 58
M. mulatta          -CAGCTCCCGTCTGCTGATGCTGTATTCCCATGTCAAAAGGCACAGGGGACACCCAGAA 58
M. musculus         ACAGCTCCTTTCTGCTGGTGCTGTATTTCCATATCAAAAGCACAGGGGAGATCTAGAAA 59
R. norvegicus       ACAGCTCCTTTCTGCTGGTGCTGTATTTCCATATCAAAAGGCACAGGGGAGATCTAGAA 59
                     ***** *  **** ** ********* **** ***************  * * ****
```

B)

```
H.sapiens           GAGCACAGCGGGGTGAGAGGGATTCCTAATCACTCAGAGCAGTCTGTGAC 66
G.gorilla           GAGCACAGCGGGGTGAGAGGGATTCCTAATCACTCAGAGCAGTGTGTGAC 66
M.mulatta           GAACACAGCGGGGTGAGAGGGATTCCTCGTC--TCAGAGCAGTCTGTGAC 64
M.musculus          TAGTGCCACAGAACAAGAG-----------------AGCAGTCT--GAC 46
R.norvegicus        TAGTGCCACAGAGCAAGAG-----------------AGCAGTCTCTGAC 48
                     *   *  * *    ****                 ******* *  ***
```

Figure 3.7: Sequence alignment of the *MECP2* 3'UTR

A cross-species alignment of the 3'UTR of the *MECP2* gene. A) The SNP rs2734647 A/G at position c.*3638 is highlighted in green. B) The SNP rs3027924 C/G at position c.*878 is highlighted in blue. The accession numbers and references for each species are as follows: *Homo sapiens* (NM_004992), *Gorilla gorilla* (ENSGGOT00000000224), *Macaca mulatta* (XM_001088566), *Mus musculus* (NM_010788) and *Rattus norvegicus* (NM_022673).

The International Haplotype Map (HapMap, http://hapmap.ncbi.nlm.nih.gov/) project provides genotype data observed in 11 different populations including Utah residents with Northern and Western European ancestry from the CEPH collection (CEU), Japanese in Tokyo (JPT), Yoruba in Ibadan in Nigeria (YRI), African ancestry in Southwest in USA (ASW), Han Chinese in Beijing in China (CHB), Chinese in Metropolitan Denver in

Colorado (CHD), Gujarati Indians in Houston in Texas (GIH), Luhya in Webuye in Kenya (LWK), Mexican ancestry in Los Angeles in California (MEX), Maasai in Kinyawa in Kenya (MKK) and Tuscans in Italy (TSI). (*Figure 3.8*).

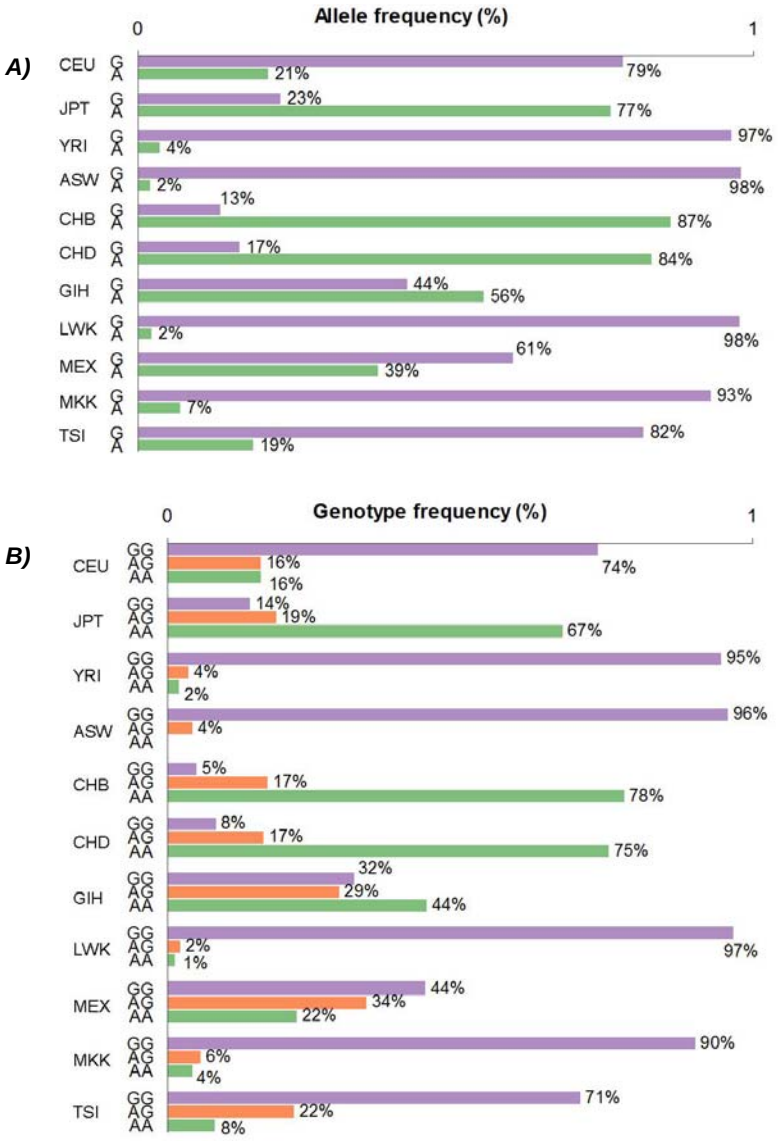


Figure 3.8: Allele and genotype frequencies of the SNP rs2734647 in the 3'UTR of the *MECP2*

Allele and genotype distributions among 11 populations are represented by two bar charts. Allele frequency is represented in the top graph (A), while genotype frequency is in the bottom graph (B). Both the distributions are reported as percentage. The purple bars represent the allele G/genotype GG; the green bars instead represent the allele A /genotype AA; the orange bars represent the genotype GA.

Regarding the SNP rs2734647, we observed 11 patients with genotype GG (63.7%) and 6 patients with AG genotype (54.5%). Our data revealed, although in a small cohort of patients, that G and A allele frequencies were respectively 82.4% and 17.6% and were very similar to Utah residents with Northern-Western European ancestry (CEU, G allele 79%, A allele 21%) and Tuscans (TSI, G allele 82% and A allele 19%) populations.

3.2 Sequence variations in the *MECP2* 3'UTR modulate cellular expression levels

Interactions between microRNAs and the *MECP2* 3'UTR target sites

It is known that mutations in the 3'UTR region of mRNA transcripts may create or destroy putative miRNA target sites. A large number of variations in the MECP2 3'UTR are listed in the RTT database, but the pathological effect is not known.

These tools are characterized by different algorithms and approaches. So it is difficult to certainly state which sequence variation leads to a loss/creation of microRNA binding. *In silico* software use different algorithms to predict: *a)* the complementarity (seed region complementarity) between microRNAs and 3'UTR target sites and *b)* the thermodynamic free energy of the duplex microRNA:3'UTR target. Therefore, these assumptions may lead out possible microRNAs with an imperfect seed match from the analysis and restrict the list of microRNA candidates.

Few known sequence variations in the *MECP2* 3'UTR were analysed with different *in silico* prediction tools to elucidate their potential effect in the microRNA:3'UTR target site interaction. In particular the only the sequence variations found previously in autistic or RTT-like patients are analysed. The sequence variation c.*1354G>A detected in our cohort of RTT patients was also analysed. The prediction results are listed in *Table 3.3*.

Our results demonstrate that few microRNAs can bind to the 3'UTR of the MECP2. In specific our results showed that has-miR-488 binds to the sequence with the nucleotide T at position c.*122; has-miR-218, 892 and 345 bind at position c.*204, hsa-miR-769-3p and hsa-miR-450b-3p bind to c.*544; hsa-miR-220b binds at c.*554. The microRNA hsa-miR-765 binds at position c.*1368; has-miR-7 and 939 target the 3'UTR at position c.*2556, two microRNAs bind at position c.2706 (has-miR-549, hsa-miR-490-3p). The base c.*2956 is targeted by has-miR-1226* and 491-5p. As a matter of fact, the *in silico* predictions should be linked with an *in vivo* functional analysis to better elucidate and comprehend the miRNA:3'UTR interaction effects.

Table 3.3. Comparison results of microRNA:mRNA interactions from different bioinformatic tools

	microRNA.org	PicTar	TargetScan	Diana	MicroInspector
c.*36 G					
hsa-miR-578	X		X		
c.*92 C					
hsa-miR-425*	X				X
hsa-miR-671-5p	X		X		
hsa-miR-30e*	X	X			
c.*122 T					
hsa-miR-186	X		X		
hsa-miR-488	***X***		***X***	***X***	
c.*204 G					
hsa-miR-363	X	X			
hsa-miR-218	***X***		***X***	***X***	
hsa-miR-4318	X				
hsa-miR-892a	***X***		***X***	***X***	
c.*204 G					
hsa-miR-943	X		X		
hsa-miR-345	***X***		***X***	***X***	
hsa-miR-490-3p	X	x	X		
hsa-miR-328		X		X	
hsa-miR-28		X		X	
hsa-miR-130a/b		X		X	
hsa-miR-372		X		X	
hsa-miR-106b				X	
hsa-miR-133					X
hsa-miR-4732-3p					X
c.*359 G					
c.*363 G					
hsa-miR-3183					X
c.*371 G					
hsa-miR-3183					X
c.*544 G					
hsa-miR-638	X		X		X
hsa-miR-92a-2*	X	X			
hsa-miR-541	X		X		
hsa-miR-769-3p	***X***		***X***		***X***
hsa-miR-450b-3p	***X***		***X***	***X***	
hsa-miR-654-5p	X		X		
hsa-miR-363*	X				
hsa-miR-1293	X		X		
hsa-miR-339-3p				X	X
c.*554 G					
hsa-miR-220b	***X***		***X***	***X***	
hsa-miR-449b*	X		X		
hsa-miR-138				X	
hsa-miR-637				X	X
c.*767 G					
hsa-miR-100	X				
hsa-miR-99a	X				
hsa-miR-99b	X				
hsa-miR-324-3p				X	X
c.861 T					
hsa-miR-651	X				
hsa-miR-1245	X				
hsa-miR-767-3p	X		X		

	microRNA.org	PicTar	TargetScan	Diana	MicroInspector
c.1134 G>A					
hsa-miR-628-5p	X				
hsa-miR-218	X				
hsa-miR-454*	X				
hsa-miR-4301	X				
hsa-miR-1225-3p	X		X		
hsa-miR-939					X
hsa-miR-1233			X		
c.*1354 G					
hsa-miR-4314	X				
hsa-miR-148b*	X				
hsa-miR-3168	X				
hsa-miR-637	X		X		
hsa-miR-let-7c*		X			
hsa-miR-759					
hsa-miR-187	X				X
hsa-miR-1225-3p	X				X
hsa-miR-339-5p					X
hsa-miR-4674					X
hsa-miR-1233					X
c.*1368 C					
hsa-miR-1197	X		X		
hsa-miR-765	***X***		***X***	***X***	
hsa-miR-761	X		X		
hsa-miR-214	X		X		
hsa-miR-885-3p	X		X		
hsa-miR-548a-3p				X	X
hsa-miR-548d-3p				X	X
c.*2556 T					
hsa-miR-7	***X***	***X***	***X***	***X***	
hsa-miR-623	X		X		
hsa-miR-3150	X				X
hsa-miR-637	X				
hsa-miR-939			***X***	***X***	***X***
c.*2657 G					
hsa-miR-665	***X***		***X***		***X***
c.*2706 G					
hsa-miR-619	X		X		
hsa-miR-490-3p	***X***		***X***	***X***	
hsa-miR-549	***X***		***X***		***X***
hsa-miR-542-3p	X		X		
c.*2956 G					
hsa-miR-22	X		X		
hsa-miR-125a-3p	X		X		
hsa-miR-1184	X		X		
hsa-miR-1205	X		X		
hsa-miR-1226*	***X***		***X***		***X***
hsa-miR-920	X			X	
hsa-miR-491-5p	***X***		***X***	***X***	
hsa-miR-637				X	
hsa-miR-490-3p				X	
hsa-miR-520a-5p				X	
c.*3658 C					
hsa-miR-229-3p	X				X

The black “X” symbol indicates the result of the microRNA:3’UTR binding target site interaction. The name and the “X” are in bold only if there is concordance of 3 predictions.

The 3'UTR of the *MECP2* gene in the Chroma-Glo reporter vector

A region of 1270 bp corresponding of the *MECP2* promoter was previously subcloned upstream the *Pyrophorus. plagiophthalamus* luciferase gene into the *pCBG99-Control vector.* Fragments corresponding to the first (called 1.8 kb polyA) and the second polyadenylation signals (called 5.4 kb polyA) were amplified from human genomic DNA and subcloned downstream of the luciferase gene into the *pCBG99-Control vector.* Expression of the reporter luciferase gene is then driven by the cloned *MECP2* promoter and different polyadenylation signals cause the termination of the transcription.

Five different mutations were generated with *in vitro* site-directed mutagenesis: c*92C>G, c*92C>T, c*122delT localized before the 1.8 polyA signal, while c*2556T>A and c*2956G>A localized before the 5.4 polyA signal as represented in *Figure 3.9.*

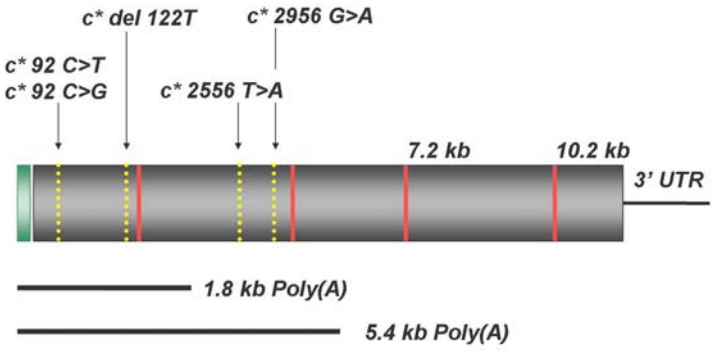


Figure 3.9: Schematic representation of the 3'UTR of the *MECP2* gene

The diagram show the *MECP2* 3'UTR and its 4 polyadenylation signals are represented in red. The two fragments cloned in the *pCBG99-Control vector* are represented by black lines, each fragments contained the *MECP2* stop codon (green). The yellow dotted lines represent the five mutations created by site-directed mutagenesis. Picture draw not in scale.

Mutations in the *MECP2* 3'UTR modulate luciferase expression levels in different cell lines

To test if sequence variations in the 3'UTR of the *MECP2* may affect luciferase reporter expression levels, we transfected three different cell lines such as monkey kidney fibroblast-like cell line (COS-7), rat adrenal gland pheochromocytoma cell line (PC12) and human neuroblastoma cell line (SH-SY5Y). We then performed an *in vitro* luciferase assay 24 hours post-transfection. The *pCBRed-Control vector* co-transfected into the cell lines was used as normalization control. Statistically significant levels in the *MECP2* 3'UTR vectors carrying mutations relative to wildtype *MECP2* 3'UTR vectors are indicated with asterisks, where *p<0.05, **p<0.01, ***p<0.001.

Statistical analysis was performed comparing the luciferase expression levels of the *MECP2* 3'UTR vector carrying a mutation and the wildtype *MECP2* 3'UTR vector (*Figure 3.10* and *3.11* and *Table 3.4*) in three different cell lines such as COS-7, PC12 and SH-SY5Y. The results did not show any significant difference in the luciferase expression levels of the *MECP2* 3'UTR vector containing the c*92C>G mutation in COS-7 and PC12 cell lines (*Figure 3.10*. Interestingly, the luciferase expression level showed a significant decreased of 10.10% in the human derived neuroblastoma cell line (SH-SY5Y) (*Table 3.0)*.

The *MECP2* 3'UTR vector containing the c*92C>T mutation analysis showed a significant increased in the expression level in all the three different cell lines (*Figure 3.10*) with respectively an increment of 36.37, 19.26 and 46.39% (*Table 3.4*).
Further, the *MECP2* 3'UTR vector containing the c*122detT mutation did not show a significant difference in COS-7 cell lines when compared to the wildtype *MECP2* 3'UTR vector, but it showed increased levels of 24.42 and 70.09% respectively in PC12 and SH-SY5Y cell lines (*Figure 3.10*.

A significant increased of 13.38, 35.28 and 21.58% in the expression level of the *MECP2* 3'UTR vector containing the c.*2556T>A was detected in COS-7, PC12 and Sh-SY5Y cell lines respectively (*Figure 3.11*). It was also observed an increased of 32.78% in COS-7, 9.53% in PC12 and of 36.07% in SH-SY5Y cell lines after transfection with the *MECP2* 3'UTR vector containing the c.*2956G>A (*Figure 3.11*).

These results showed that sequence variations in the 3'UTR can down-regulate or up-regulated the reporter expression level and this it may due to a change in the microRNA:3'UTR target site interaction.

Table 3.4 : Luciferase assay results in COS-7, PC12 and SH-SY5Y cell lines

Cell line	*MECP2* 3'UTR vectors	Normalized luciferase activity	Standard deviation (SD)	Statistical results	Difference in normalized luciferase activity (%)
COS-7	**1.8 UTRwt**	1	0		
	c*92C>G	0.90	0.15		-10.10
	c*92C>T	1.36	0.27	**	36.37
	c*122delT	1.18	0.13	**	18.26
PC12	**1.8 UTRwt**	1.00	0.00		
	c*92C>G	1.01	0.11		1.49
	c*92C>T	1.19	0.17	**	19.26
	c*122delT	1.24	0.21	**	24.42
SHSY5Y	**1.8 UTRwt**	1.00	0.00		
	c*92C>G	0.76	0.13	**	-23.75
	c*92C>T	1.46	0.12	**	46.39
	c*122delT	1.70	0.56	**	70.09

Cell line	*MECP2* 3'UTR vectors	Normalized luciferase activity	Standard deviation		Difference in normalized luciferase activity (%)
COS-7	**5.4UTRwt**	1.00	0.00		
	c*2556 T>A	1.13	0.12	**	13.38
	c*2956 G>A	1.33	0.21	**	32.78
PC12	**5.4UTRwt**	1.00	0.00		
	c*2556 T>A	1.35	0.42	**	35.28
	c*2956 G>A	1.10	0.16	**	9.53
SHSY5Y	**5.4UTRwt**	1.00	0.00		
	c*2556 T>A	1.22	0.41	**	21.58
	c*2956 G>A	1.36	0.14	**	36.07

The normalized luciferase assay data are expressed as an average of three biological replicates, the difference in the normalized expression levels is reported as percentage.

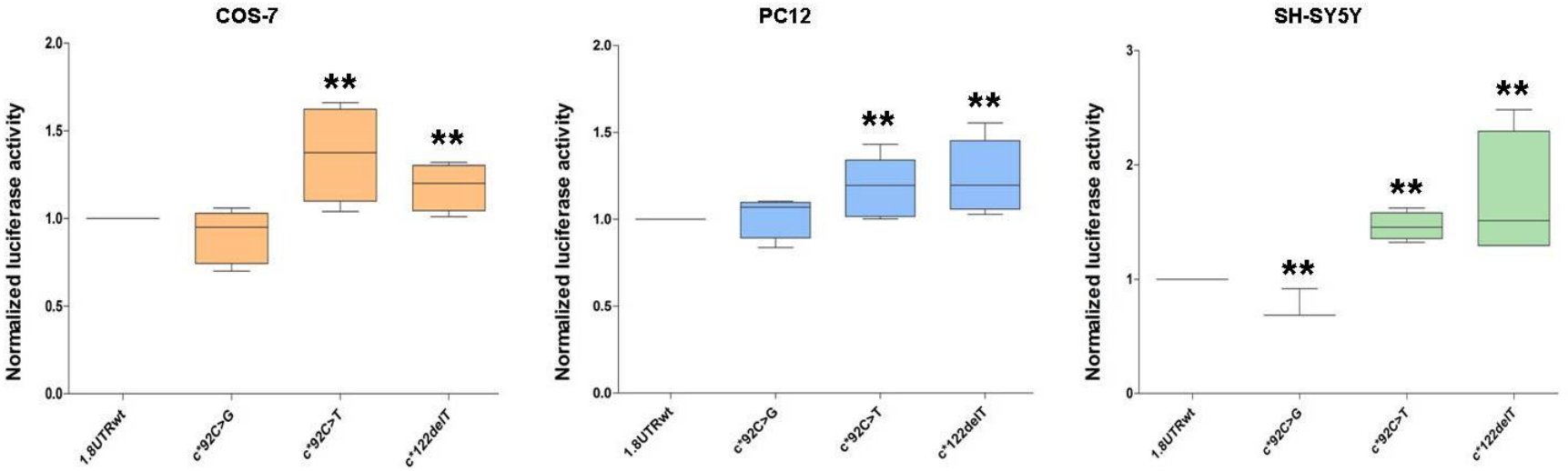


Figure 3.10: Luciferase assay in COS-7, PC12 and SH-SY5Y cell lines

The luciferase *CBG99luc* reporter gene activity was controlled by a region of the *MECP2* promoter (1270 bp). The wildtype *MECP2* 3'UTR fragment containing the first polyadenylation signal (1.8 kb wt) was cloned downstream the luciferase reporter gene. Three mutations were introduced by *in vitro* site directed mutagenesis: c*92 C>G, c*92 C>T and c*122delT. The constructs were transfected in COS-7 (light orange box plot), PC12 (light blue box plot) and SH-SY5Y (light green box plot) cell lines. Luciferase activity was normalized to luciferase pCBRed-control vector. Three independent experiments are represented on the box and whisker plot.

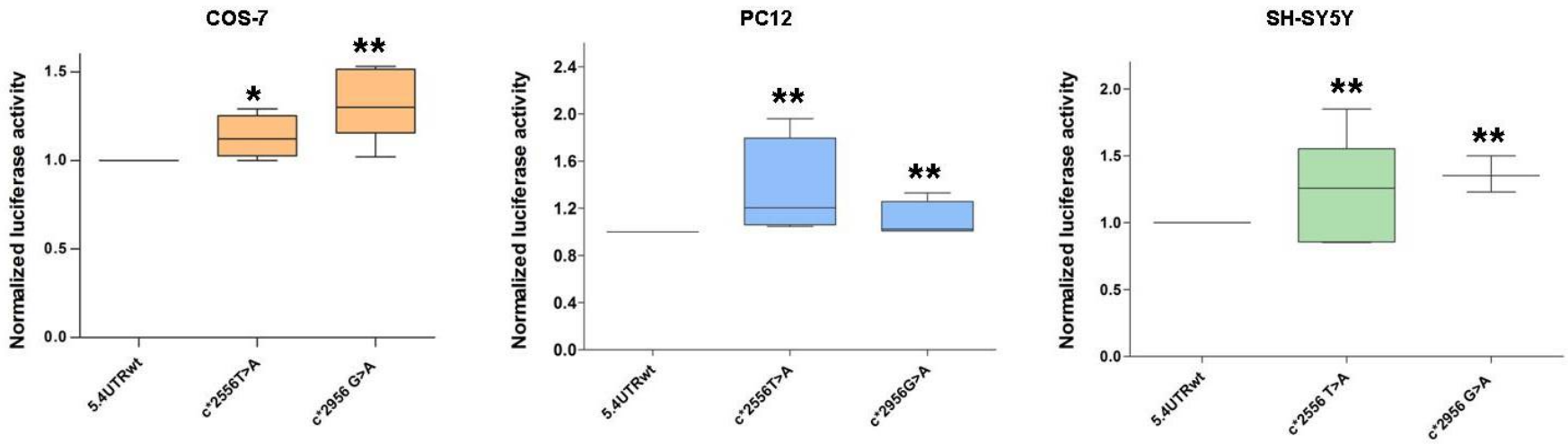


Figure 3.11: Luciferase assay in COS-7, PC12 and SH-SY5Y cell lines

The luciferase *CBG99luc* reporter gene is controlled by a region of the *MECP2* promoter (1270 bp). The wildtype *MECP2* 3'UTR fragment containing the second polyadenylation signal (5.4 kb wt) was cloned downstream the luciferase reporter gene. Two mutations were introduced by *in vitro* site directed mutagenesis: c*2556 T>A, c*2956 G>A. The constructs were transfected in COS-7 (light orange box plot), PC12 (light blue box plot) and SH-SY5Y (light green box plot) cell lines. Luciferase activity was normalized to luciferase pCBRed-control vector. Three independent experiments are represented on the box and whisker plot.

3.3 Mutation screening in patients with Early Infantile Epileptic Encephalopathy

Two new pathogenic sequence variations in the *ARX* gene

The entire open reading frame of the *ARX* gene, as well as the exon-intron boundaries, was screened for pathogenic sequence mutations by direct sequencing. Two missense mutations in the *ARX* gene were identified in three male subjects with epileptic encephalopathies with onset in the first year of life. Both mutations were localized on the exon 5 of this gene. Detailed information about the phenotypes of these patients is in *Appendix 4.*

A *de novo* non-synonymous mutation (c.1610T>C, p.Leu537Pro) was detected in a sporadic case of epileptic encephalopathy; while the second non-synonymous mutation (c.1604T>A, p.Leu535Gln) was detected in two related boys, born to monozygotic twin sisters. Carrier status was tested for the mutation identified in the proband using genomic DNA of parents and siblings (*Figure 3.12*). In addition, the identified variants were genotyped in controls from general population by direct sequencing (150 chromosomes). The missense mutations were not detected on the control X chromosomes.

In silico analysis of Exonic Splicing Enhancer and Exonic Splicing Silencer motifs

Point mutations within a coding sequence traditionally result in the corresponding amino acid change. However, some point mutations have been reported to affect the splicing mechanism, disrupting functional common motifs of 6-8 bp called exonic splicing enhancer (ESEs) or exonic splicing silencer (ESS).

The default setting was used in the web-server tools such as FAS-ESS, RESCUE-ESE and ESEfinder. Accordingly to FAS-ESS tool, no creation of any ESS motifs was detected using the FAS-hex3 list, while a new ESS motif was created blasting the query sequence with the c.1610T>C mutation to the FAS-hex2 motif list (*Table 3.5*).

Using RESCUE-ESE web-server, the non-synonymous mutation c.1604T>A was found to create a new ESE motif (CaGAGG) (*Table 3.6)* and, accordingly to ESEfinder, the

mutation increased the SF2/ASF protein score value for the same sequence from 2.40958 to 3.53210 (*Table 3.7*). In addition, this mutation seemed to create a new ESE motif (CGCaGAG) for the same protein. Moreover, this mutation created two new binding ESE sites (CCGCGCa and CGCaGAG) for the SF2/ASF (IgM-BRCA1) protein and a new binding site (CGCaGA) for SRp55 protein. The SF2/ASF (IgM-BRCA1) binding score for the sequence seemed to slightly increase (0.43 fold) (*Table 3.7*).

Regarding the c.1610T>C alteration, no new ESE motifs were detected with RESCUE-ESE tool (*Table 3.6*). Instead, this mutation slightly increased the ESE value for SF2/ASF (0.19 fold) and for SC35 (0.23 fold), while a small reduction was observed for the SRp55 (0.2) using ESE-finder (*Table 3.7*).

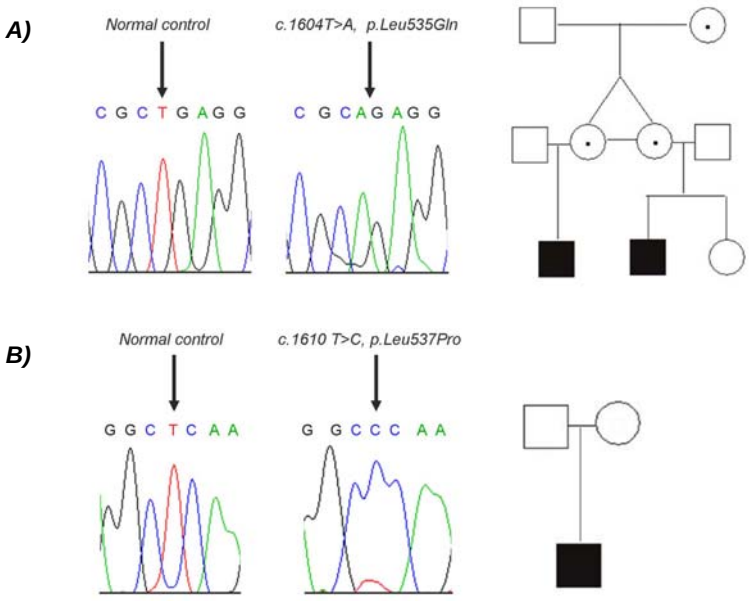


Figure 3.12: Family pedigrees and *ARX* non-synonymous mutations

The DNA sequence chromatograms show the non-synonymous mutations detected in three patients. The family pedigree tree of the two families studied is represented. Open symbols represent normal individuals, black solid squares represent affected males; the circles with a dot represent the female mutation carriers.
A) The sequence chromatogram of a normal unaffected individual (non-carrier) is shown on the left, while the sequence chromatogram of the non-synonymous mutation c.1604T>A, p.Leu535Gln is shown on the right. The family pedigree is represented on the right.
B) The sequence chromatogram of normal unaffected individual (non-carrier) is shown on the left and the chromatogram of the c.1610T>C, p.Leu537Pro is shown on the right.

Table 3.5: FAS-ESS *in silico* analysis of Exonic Splicing Silencer motifs in the *ARX* gene

wildtype sequence	CCAAGG AGCATAGCCGCGCtGAGGCtCAAGGCCAAGGAGCA
c.1604T>A alteration	CCAAGG AGCATAGCCGCGCaGAGGCTCAAGGCCAAGGAGCA
c.1611T>C alteration	CCAAGG CCAAGG AGCATAGCCGCGCTGAGGCcCAAGGCCAAGGAGCA

The wildtype and the two mutated sequences are shown. The results represented the FAS-ESS analysis with hex2 set up. The c. 1604T>A mutation is represented in red lower-case letter, while the c.1610T>C is represented in blue lower-case letter. The CCAAGG sequence is an ESS motif common between the three sequence analyzed.

Table 3.6: RESCUE-ESE *in silico* analysis of Exonic Splicing Enhancer motifs in the *ARX* gene

wildtype sequence	AAGGAG AGCATAGCCGCGCtGAGGCtCAAGGCCAAGGAGCA
c.1604T>A alteration	CAGAGG AAGGAG AGCATAGCCGCGCaGAGGCTCAAGGCCAAGGAGCA
c.1611T>C alteration	AAGGAG AGCATAGCCGCGCTGAGGCcCAAGGCCAAGGAGCA

The wildtype and the two mutated sequences are shown. The c. 1604T>A mutation is represented in red lower-case letter, while the c.1610T>C is represented in blue lower-case letter. The AAGGAG sequence is a ESE motif common between the three sequence analyzed.

Table 3.7: ESE-finder *in silico* analysis of Exonic Splicing Enhancer motifs in the *ARX* gene

wildtype sequence AGCATAGCCGCGCtGAGGCtCAAGGCCAAGGAGCA

SF2/ASF Threshold: 1.956	SF2/ASF (IgM-BRCA1) Threshold: 1.867	SC35 Threshold: 2.383	SRp55 Threshold: 2.676
Site/score	Site/score	Site/score	Site/score
CtGAGGC: 2.40958	CGCGCtG: 2.00071	GGCtCAAG: 3.92091	tGAGGC: 3.18234
CtCAAGG: 2.21605	CtGAGGC: 2.95606		
	CtCAAGG: 2.61997		

c.1604 T>A alteration AGCATAGCCGCGCaGAGGCTCAAGGCCAAGGAGCA

SF2/ASF Threshold: 1.956	SF2/ASF (IgM-BRCA1) Threshold: 1.867	SC35 Threshold: 2.383	SRp55 Threshold: 2.676
Site/score	Site/score	Site/score	Site/score
CGCaGAG: 2.3323	CCGCGCa. 2.07857	GGCtCAAG: 3.92091	CGCaGA: 3.51736
CaGAGGC: 3.53210	CGCaGAG: 3.27618		
CtCAAGG: 2.21605	CaGAGGC: 3.38732		
	CtCAAGG: 2.61997		

c.1610 T>C alteration AGCATAGCCGCGCTGAGGCcCAAGGCCAAGGAGC

SF2/ASF Threshold: 1.956	SF2/ASF (IgM-BRCA1) Threshold: 1.867	SC35 Threshold: 2.383	SRp55 Threshold: 2.676
Site/score	Site/score	Site/score	Site/score
CtGAGGC: 2.40958	CGCGCtG: 2.00071	GGCcCAAG: 4.15175	tGAGGC: 3.18234
	CtGAGGC: 2.95606		
	CcCAAGG: 2.37119		

The ESE-finder predictions score values are shown. The c. 1604T>A mutation is represented in red lower-case letter, while the c.1610T>C is represented in blue lower-case letter.

The *Aristaless* domain is highly conserved

The mutations detected were located in exon 5 of the *ARX* gene, which contains the *Aristaless* domain. The *Aristaless* domain consists of a 14 amino acid motif within the C-terminal sequence of the gene and its function is still poorly understood. This domain is also known as an OAR domain, using the initials, of the *otp*, *aristaless* and *rax* genes (Furukawa et al., 1997).

To analyse the conservation of this domain, a sequence alignment of 75 peptide sequences from 16 different species was performed using ClustalW multi protein sequence alignment tool (*Figure 3.13*). The *profile Hidden Markov Model* (HMM Logo) (*Figure 3.14)* graphically displays the protein conservation in a multiple alignment using stacks of letters for each position. Stacks of letters have different heights that are proportional to the frequency at the position. The letters are sorted by dimension and the largest letter is located at the top of the stack. Moreover, different colours are used to represent the biological properties of the letters/amino acid residues (Schuster-Bockler et al., 2004).

The amino acid Leucine at position 535 (Leu 535) appeared as the most conserved residue of the *Aristaless* domain among the 75 analyzed motif sequences (75/75). In fact, it was represented as a black conserved amino acid in the ClustalW analysis and as an orange largest letter in the HMM logo graphic. On the other hand, the amino acid Leucine at position 537 (Leu 537) did not show a high level of conservation among the amino acid motifs (41/75). Additionally, Leu 537 is replaced by other amino acids such as methionine (M, 19/73), arginine (R, 9/73), alanine (A, 5/73) and isoleucine (I, 1/73).

```
ALX1_HUMAN/306-319      SSIAVLRMKAKEHT
ALX1_MOUSE/306-319      SSIAVLRMKAKEHT
ALX1_RAT/306-319        SSIAVLRMKAKEHT
ALX1_XENLA/315-328      SSIAVLRMKAKEHT
ALX4_BOVIN/377-390      SSIAALRMKAKEHS
ALX4_HUMAN/391-404      SSIAALRMKAKEHS
ALX4_MOUSE/379-392      SSIAALRMKAKEHS
AL_DROME/378-391        SSIAALRLKAREHE
ARX_DANRE/421-434       SSIAALRLKAKEHS
ARX_HUMAN/530-543       SSIAALRLKAKEHA
ARX_MOUSE/532-545       SSIAALRLKAKEHA
ARX_RAT/534-547         SSIAALRLKAKEHA
DMBX1_HUMAN/359-372     TSIENLRLRAKQHA
DMBX1_MOUSE/358-371     TSIENLRLRAKQHA
DMX1A_DANRE/365-378     TSIENLRLRAKQHA
DMX1B_DANRE/346-359     TSIENLRLRAKQHA
DRGX_HUMAN/204-217      ASVATLRMKAREHS
DRGX_MOUSE/204-217      ASVAALRMKAREHS
DRGX_RAT/204-217        ASVAALRMKAREHS
OTBP_DANRE/296-309      TSIASLRRKALEHT
OTP_DROME/372-385       HSIAALRRRASELN
OTP_HELER/342-355       TSIASLRRKALEHA
OTP_HELTB/342-355       TSIASLRRKALEHA
OTP_HUMAN/306-319       TSIASLRRKALEHT
OTP_LYTVA/340-353       TSIASLRRKALEHA
OTP_MOUSE/306-319       TSIASLRRKALEHT
OTP_PARLI/344-357       TSIASLRRKLEHAA
OTP_SACKO/301-314       TSIAQLRRKALEHS
PITX1_CHICK/278-291     SSLASLALKSKQHS
PITX1_HUMAN/280-293     SSLASLRLKSKQHS
PITX1_MOUSE/281-294     SSLASLRLKSKQHS
PITX1_RAT/281-294       SSLASLRLKSKQHS
PITX1_XENLA/271-284     SSLASLRLKSKQHS
PITX2_CHICK/295-308     SSLASLRLKAKQHS
PITX2_DANRE/276-289     SSLASLRLKAKQHS
PITX2_HUMAN/279-292     SSLASLRLKAKQHS
PITX2_MOUSE/279-292     SSLASLRLKAKQHS
PITX2_RAT/286-299       SSLASLRLKAKQHS
PITX2_XENLA/288-301     SSLASLRLKAKQHS
PITX3_DANRE/255-268     SSLASLRLKAKQHA
PITX3_HUMAN/262-275     SSLASLRLKAKQHA
PITX3_MOUSE/262-275     SSLASLRLKAKQHA
PITX3_RAT/262-275       SSLASLRLKAKQHA
PITX3_XENLA/254-267     SSLASLRLKAKQHA
PITX_BRABE/289-302      SSIAALRLKAKQHS
PITX_DROME/460-473      SSIATLRLKAKQHA
PRRX1_CHICK/222-235     NSIANLRLKAKEYS
PRRX1_HUMAN/222-235     NSIANLRLKAKEYS
PRRX1_MOUSE/222-235     NSIANLRLKAKEYS
PRRX1_RAT/222-235       NSIANLRLKAKEYS
PRRX2_CHICK/142-155     NSIASLRLKAKEFS
PRRX2_HUMAN/230-243     NSIASLRLKAKEFS
PRRX2_MOUSE/224-237     NSIASLRLKAKEFS
RX1_ASTFA/311-324       SSIASLRMKAKEHI
RX1_CHICK/205-218       SSIASLRMKAKEHI
RX1_DANRE/306-319       SSIAALRMKAKEHI
RX2_CHICK/294-307       TSIASLRMKAKEHI
RX2_DANRE/304-317       SSIAALRMKAKEHI
RX2_ORYLA/304-317       SSIAALRMKAKEHI
RX3_DANRE/272-285       TSIASLRMKAKEHI
RXA_XENLA/302-315       NSIASLRMKAKEHI
RXB_XENLA/302-315       NSIASLRMKAKEHI
RX_DROME/849-862        NSIATLRIKAKEHL
RX_HUMAN/323-336        SSIAALRLKAKEHI
RX_MOUSE/319-332        SSIAALRLKAKEHI
RX_RAT/319-332          SSIAALRLKAKEHI
SHOX2_HUMAN/313-326     SSIADLRLKAKKHA
SHOX2_MOUSE/313-326     SSIADLRLKAKKHA
SHOX2_RAT/219-232       SSIADLRLKAKKHA
SHOX_HUMAN/274-287      SSIADLRLKARKHA
VSX2_CHICK/323-336      NSIAALRAKAQEHS
VSX2_DANRE/342-355      NSIAALRAKAQEHS
VSX2_HUMAN/304-317      NSIAVLRAKAQEHS
VSX2_MOUSE/304-317      NSIAALRAKAQEHS
VSX2_ORYLA/341-354      NSIAALRAKAQEHS
```

Figure 3.13: Sequence alignment of the *Aristaless* domain

A cross-species alignment of the *Aristaless* domain is represented in alphabetical order. Each entry contains gene name, organism name and residue numbers. The degree of conservation at each amino acid residue position is illustrated by intensity of shading: black shading with white letters indicates 100% identity, dark gray shading with white letters indicates 75-99% identity, light gray shading with black letters indicates 50-74% identity.

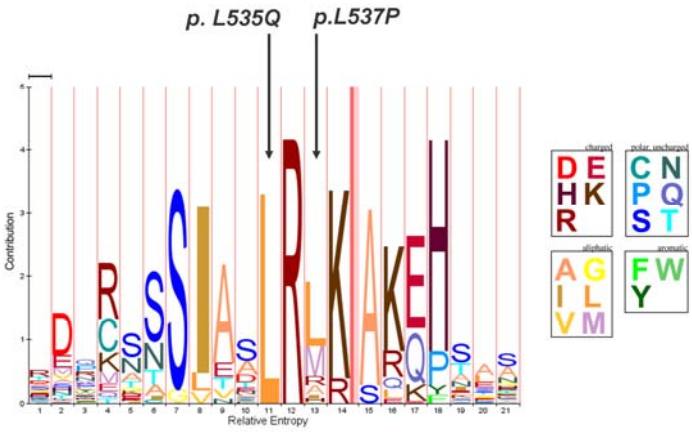


Figure 3.14: Visualization of the *Aristaless* domain with HMM Logos

The 14 amino acid *Aristaless* domain motif and flaking sequences are represented as HMM logo graphic image. The position number is displayed on the X-axis; the amino acid contribution (height) is displayed on the Y-axis. The arrows indicate the location of the amino acid mutations p. L535Q (Leu535Gln) and p.L537P (Leu537Pro). The pink bar does not contain letters, but represents an insertion with emission distribution very close or equal to the background.

4. Discussion

Neurodevelopmental disorders are a large group of common and complex disorders seen in paediatric, child-neurological and clinical genetic practice. These neurodevelopmental disorders show similar features such as brain dysfunction, abnormalities in motor and sensory systems, language and speech problems and cognitive impairments.

The clinical phenotypes of some of these neurodevelopmental disorders are well-established; however, knowledge about the genetic etiology is available only for some of them (for instance Down syndrome, Fragile X syndrome, Prader-Willi syndrome, Angelman syndrome), while for other disorders the underlying genetic mechanisms are by and large not completely elucidated (for instance many conditions associated with epilepsy, autistic spectrum disorder, developmental dyslexia, mental retardation and attention deficit hyperactivity disorder) (Bishop, 2010). Improvement of our understanding of the biology of neurodevelopmental disorders in recent years has highlight the existence of a network of genes that finely regulate the neuronal development and life-long maintenance of the brain function.

Historically, only sequence variations in the coding region of the *MECP2* have been linked to a pathological phenotype. Based on our study, we hypothesize that also sequence variations in the long *MECP2* 3'UTR can also modify the expression level of *MECP2* isoforms. Our results showed altered expression levels of both isoforms in a proportion of 22 RTT female patients, without sequence variations in the *MECP2* coding region and exon-intron boundaries. Our results were partly in agreement with previous data reported in literature, describing that both isoforms were overall down regulated in RTT patients; even thought, some of our atypical RTT patients showed high levels of both isoforms (2/22). On the other hand, in some of our cases the expression levels of both isoforms were variable. It should be noted that within our atypical RTT cohort we analysed patients with congenital variant (9/13), "forme fruste" (1/13) and no specific diagnosis other than atypical RTT (2/13), while the literature reported congenital variant (2/7), early seizures onset variant (2/7) and generally atypical RTT (3/7). Our results showed that two congenital RTT patients are characterized by increased expression levels of both *MECP2* isoforms, while only one patient had decreased expression of both isoforms. In literature, patients with the congenital variant have different expression patterns; one patient was

characterized by low expression of both isoforms, while the second patient was characterized by a *MECP2_e1* level equal to control and a *MECP2_e2* increased level. The results highlight the complicated pattern of expression of both *MECP2* isoforms. It is possible to hypothesize that the X-chromosome inactivation (XCI) might have an effect on the mRNA expression. Regarding our cohort, the XCI data were unfortunately not available, but for future direction it would be necessary to consider the analysis of this important epigenetic event since it modulates the expression of the X-linked genes in females.

Based on expression studies carried out in other 3'UTRs, we hypothesize that sequence variations in the *MECP2* 3'UTR binding target sites for microRNAs may alter the expression levels of both *MECP2* isoforms. As a matter of fact, variations in the microRNA binding target site lead to a functional dysregulation of microRNAs. As recent studies revealed, sequence variations in the *MECP2* 3'UTR correlate with reduced levels of expression in autistic male patients. Interestingly, our analysis carried out in females with RTT show that a known sequence variation (c.*371 G>C) determines a different expression level of *MECP2_e1* (increased) and *MECP2_e2* (decreased). Is it possible to hypothesize the presence of a compensatory dosage mechanism in the *MECP2* regulation?

Our screening results also showed the presence of a previously unreported sequence variation (c.*1354 G>A) in two girls (2/17). To determine the carrier frequency of these variations, the parents and the normal population should be screened. The samples analysed in this study are from a precious Australian collection of DNA and from an Italian cohort from which ha not been possible to obtained the parents DNA. The screening on the general population is currently on the way.

It should also be noted that quantification analyses are usually performed on RNA extracted from peripheral blood leukocytes (PBls) as an accessible practical and widely available clinical source of RNA. As *MECP2* transcripts are predominantly expressed in brain, changes in *MECP2* mRNA expression levels in the peripheral blood cell may not necessarily reflect the situation in brain neuronal cells of RTT patients.

The functional impact of sequence variations in the 3'UTR has been observed in our study and also in literature; in fact we identified variations in the proximal part (~4.5 kb) of the long and well conserved 3'UTR of the *MECP2* gene. Consequently, we are able to

hypothesize that among the pathogenic mechanisms that lead to *MECP2*-related disorders, there are sequence alterations in regulatory non-coding regions.

The complexity of the etiopathogenesis of the RTT syndrome is further highlighted by our gene reporter results. Some authors have previously analysed, from a transcriptional point of view, two variations (c.*2556 T>A and c.*2956 G>A) in the *MECP2* 3'UTR in autistic males. Our functional analysis on these variants showed that protein expression levels were increased in neuronal (PC12 and SH-SY5Y) and in non-neuronal cell lines (COS7). Our results also showed that other two variations reported in the RettBASE (c.*92C>T and c.*d122delT) lead to increased levels of the reporter gene in all three cell lines. Traditionally, sequence variations in the 3'UTR lead to an altered microRNAs interactions and consequently to a decreased level of protein. Recently instead it is emerging that variations in the 3'UTR can also lead to protein over expression (Simon et al., 2010). Our results agree with the more recent literature.

It should be noted that the cell lines used in this analysis may have different microRNAs and so far the specific microRNAs repertoire of these cell lines is not known. It may be possible to hypothesize that different microRNAs are expressed in non-neuronal and neuronal cell lines. The deep-sequencing technologies are rapidly increasing the number of mature microRNA sequences characterized. So nowadays, for instance, it is possible to discover the microRNAs expression profile of different tissues, such as whole blood, cultured cells, animal tissues, with a microRNA Chip Assay. Therefore, a panel of differently expressed micoRNAs can be obtained, and this way, a more specific list of microRNAs can be used to enforce and complete the *in silico* predictions.

As a matter of fact, the bioinformatic tools use different algorithms to predict the interaction between microRNAs and 3'UTR target sites, so it may be possible that some microRNAs, as the one with an imperfect seed match, are left out from the analyses. Our bioinformatic predictions led to discordant and unclear results about the effect of *MECP2* 3'UTR sequence variants. In fact, the results showed that few microRNAs are able to bind the *MECP2* 3'UTR wildtype sequence and more experimental data should be provided to certainly demonstrate the truthfulness of our predictions.

Although our results show that the sequence variations in the *MECP2* 3'UTR could modify the reporter gene expression levels, it is not clear how these variants modify the microRNAs:3'UTR target sites interactions leading to a protein dysregulation. For the future studies, first of all, it would be worth to differentiate the neuronal cell lines and

reproduce the gene reporter analysis. This way stronger results should be obtained about the protein expression levels.

Furthermore as continuation of this study, the microRNA Chip Assay on non-neuronal/neuronal cell lines and in undifferentiated/differentiate cell lines, can provide a list of cell line specific microRNAs. Some of these microRNA can subsequently be co-transfected either with wildtype 3'UTR vectors or mutated 3'UTR target sequences. Novel insights about microRNAs:*MECP2* 3'UTR target sites interactions can be obtained pulling together results from different experimental sources. By doing this, we hope that future strategies can contribute to elucidate the functional role of microRNAs on the etiopathogenesis of RTT.

The results we have obtained from the *ARX* gene analysis also provide further insights into the pathogenesis of early infantile epileptic encephalopathy (EIEE). We have identified two *de novo* single base substitutions in the *ARX* gene in a selected cohort of paediatric patients with EIEE. The position of these variants in the highly conserved domain called *Aristaless* domain, their molecular nature and the fact that they were *de novo*, allowed us to consider these alterations as being highly likely to be pathogenic sequence variants. It is well established that mutations in the *ARX* gene cause an heterogeneous spectrum of disorders characterized by epilepsy, infantile spasm with or without malformations. Our results agree with the more recent literature and show that this gene is responsible for EIEE and in specific for the Ohtahara syndrome. Initially, only alanine duplications were associated with Ohtahara syndrome and in particular, it seemed that mutations in this domain are responsible for a more severe epileptic encephalopathy phenotype. Our results, reported on two papers (Giordano et al., 2010; Sartori et al., 2011) followed by more recent data (Eksioglu et al., 2011; Kato et al., 2010) confirm and demonstrate that different type of mutations have a fundamental role in the determinism of Ohtahara syndrome.

This is the first study that analyse the role of the *MECP2* 3'UTR in association with RTT. Our results prove that a wide number of mechanisms may be responsible for the origin of these MECP2-related conditions and that mechanisms underlying these disorders are much more complex than previously identify. Our results improve our understanding of the etiology of these neurodevelopmental disorders, but more studies will be necessary to

expand our knowledge of their etiopathogenesis. In particular, it will be essential to clarify the role of microRNAs in the development of these devastating disorders.

Other analyses will also be necessary through new time and cost effective technologies. For example, the advent of the next-generation sequencing technologies will allows us to analyse the coding regions of many genes and the 3'UTR, the promoter region and the intronic sequences of each genes. Consequently, in a very near future, such large-scale sequencing studies will facilitate the identification of clinically significant sequence variations in the genome.

In particular, with new generation technologies improving our knowledge of, for instance, genome, transcriptome and epigenetic regulation, and the development of new targeted therapies based on a better understanding of the underlying disorders, we are at the cusp of a new era of personalized medicine, leading to effective personalized treatments to reduce symptoms in infancy, adolescence or adulthood. An increased understanding on the biological bases of the neurodevelopmental disorders will clarify the complexity of brain development. This understanding will also provides advanced tools for early and accurate diagnosis which will in turn lead to more effective clinical follow up protocols and a better genetic counselling for patients and families affected by these disorders.

5. References

Abelson, J.F., Kwan, K.Y., O'Roak, B.J., Baek, D.Y., Stillman, A.A., Morgan, T.M., Mathews, C.A., Pauls, D.L., Rasin, M.R., Gunel, M.*, et al.* (2005). Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science *310*, 317-320.

Absoud, M., Parr, J.R., Halliday, D., Pretorius, P., Zaiwalla, Z., and Jayawant, S. (2010). A novel ARX phenotype: rapid neurodegeneration with Ohtahara syndrome and a dyskinetic movement disorder. Dev Med Child Neurol *52*, 305-307.

Adams, B.D., Furneaux, H., and White, B.A. (2007). The micro-ribonucleic acid (miRNA) miR-206 targets the human estrogen receptor-alpha (ERalpha) and represses ERalpha messenger RNA and protein expression in breast cancer cell lines. Mol Endocrinol *21*, 1132-1147.

Adler, D.A., Quaderi, N.A., Brown, S.D., Chapman, V.M., Moore, J., Tate, P., and Disteche, C.M. (1995). The X-linked methylated DNA binding protein, Mecp2, is subject to X inactivation in the mouse. Mamm Genome *6*, 491-492.

Afonina, I., Zivarts, M., Kutyavin, I., Lukhtanov, E., Gamper, H., and Meyer, R.B. (1997). Efficient priming of PCR with short oligonucleotides conjugated to a minor groove binder. Nucleic Acids Res *25*, 2657-2660.

Amir, R.E., Fang, P., Yu, Z., Glaze, D.G., Percy, A.K., Zoghbi, H.Y., Roa, B.B., and Van den Veyver, I.B. (2005). Mutations in exon 1 of MECP2 are a rare cause of Rett syndrome. J Med Genet *42*, e15.

Amir, R.E., Van den Veyver, I.B., Schultz, R., Malicki, D.M., Tran, C.Q., Dahle, E.J., Philippi, A., Timar, L., Percy, A.K., Motil, K.J.*, et al.* (2000). Influence of mutation type and X chromosome inactivation on Rett syndrome phenotypes. Ann Neurol *47*, 670-679.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet *23*, 185-188.

Andersen, C.L., Jensen, J.L., and Orntoft, T.F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res *64*, 5245-5250.

Anvret, M., and Wahlstrom, J. (1994). Rett syndrome: random X chromosome inactivation. Clin Genet *45*, 274-275.

Archer, H.L., Evans, J.C., Millar, D.S., Thompson, P.W., Kerr, A.M., Leonard, H., Christodoulou, J., Ravine, D., Lazarou, L., Grove, L.*, et al.* (2006a). NTNG1 mutations are a rare cause of Rett syndrome. Am J Med Genet A *140*, 691-694.

Archer, H.L., Whatley, S.D., Evans, J.C., Ravine, D., Huppke, P., Kerr, A., Bunyan, D., Kerr, B., Sweeney, E., Davies, S.J.*, et al.* (2006b). Gross rearrangements of the MECP2 gene are found in both classical and atypical Rett syndrome patients. J Med Genet *43*, 451-456.

Ariani, F., Mari, F., Pescucci, C., Longo, I., Bruttini, M., Meloni, I., Hayek, G., Rocchi, R., Zappella, M., and Renieri, A. (2004). Real-time quantitative PCR as a routine method for screening large rearrangements in Rett syndrome: Report of one case of MECP2 deletion and one case of MECP2 duplication. Hum Mutat *24*, 172-177.

Armani, R., Archer, H., Clarke, A., Vasudevan, P., Zweier, C., Ho, G., Williamson, S., Cloosterman, D., Yang, N., and Christodoulou, J. (2011). Transcription Factor 4 and Myocyte Enhancer Factor 2C Mutations are not Common Causes of Rett Syndrome. American journal of medical genetics

Armstrong, D.D. (2001). Rett syndrome neuropathology review 2000. Brain Dev *23 Suppl 1*, S72-76.

Armstrong, D.D., Dunn, J.K., Schultz, R.J., Herbert, D.A., Glaze, D.G., and Motil, K.J. (1999). Organ growth in Rett syndrome: a postmortem examination analysis. Pediatr Neurol *20*, 125-129.

Artuso, R., Mencarelli, M.A., Polli, R., Sartori, S., Ariani, F., Pollazzon, M., Marozza, A., Cilio, M.R., Specchio, N., Vigevano, F., *et al.* (2010). Early-onset seizure variant of Rett syndrome: definition of the clinical diagnostic criteria. Brain Dev *32*, 17-24.

Ballestar, E., Ropero, S., Alaminos, M., Armstrong, J., Setien, F., Agrelo, R., Fraga, M.F., Herranz, M., Avila, S., Pineda, M., *et al.* (2005). The impact of MECP2 mutations in the expression patterns of Rett syndrome patients. Hum Genet *116*, 91-104.

Balmer, D., Goldstine, J., Rao, Y.M., and LaSalle, J.M. (2003). Elevated methyl-CpG-binding protein 2 expression is acquired during postnatal human brain development and is correlated with alternative polyadenylation. J Mol Med *81*, 61-68.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bauman, M.L., Kemper, T.L., and Arin, D.M. (1995). Pervasive neuroanatomic abnormalities of the brain in three cases of Rett's syndrome. Neurology *45*, 1581-1586.

Bebbington, A., Anderson, A., Ravine, D., Fyfe, S., Pineda, M., de Klerk, N., Ben-Zeev, B., Yatawara, N., Percy, A., Kaufmann, W.E., *et al.* (2008). Investigating genotype-phenotype relationships in Rett syndrome using an international data set. Neurology *70*, 868-875.

Bebbington, A., Percy, A., Christodoulou, J., Ravine, D., Ho, G., Jacoby, P., Anderson, A., Pineda, M., Ben Zeev, B., Bahi-Buisson, N., *et al.* (2010). Updating the profile of C-terminal MECP2 deletions in Rett syndrome. J Med Genet *47*, 242-248.

Beetz, C., Schule, R., Deconinck, T., Tran-Viet, K.N., Zhu, H., Kremer, B.P., Frints, S.G., van Zelst-Stams, W.A., Byrne, P., Otto, S., *et al.* (2008). REEP1 mutation spectrum and genotype/phenotype correlation in hereditary spastic paraplegia type 31. Brain *131*, 1078-1086.

Belichenko, P.V., Hagberg, B., and Dahlstrom, A. (1997). Morphological study of neocortical areas in Rett syndrome. Acta Neuropathol *93*, 50-61.

Belligni, E.F., Palmer, R.W., and Hennekam, R.C. (2010). MECP2 duplication in a patient with congenital central hypoventilation. Am J Med Genet A *152A*, 1591-1593.

Ben Zeev, B.B., Bebbington, A., Ho, G., Leonard, H., de Klerk, N., Gak, E., Vecsler, M., and Christodoulou, J. (2009). The common BDNF polymorphism may be a modifier of disease severity in Rett syndrome. Neurology *72*, 1242-1247.

Bennett, C.L., Brunkow, M.E., Ramsdell, F., O'Briant, K.C., Zhu, Q., Fuleihan, R.L., Shigeoka, A.O., Ochs, H.D., and Chance, P.F. (2001). A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA-->AAUGAA) leads to the IPEX syndrome. Immunogenetics *53*, 435-439.

Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., and Lai, E.C. (2007). Mammalian mirtron genes. Mol Cell *28*, 328-336.

Betel, D., Wilson, M., Gabow, A., Marks, D.S., and Sander, C. (2008). The microRNA.org resource: targets and expression. Nucleic Acids Res *36*, 149-150.

Beyer, K.S., Blasi, F., Bacchelli, E., Klauck, S.M., Maestrini, E., and Poustka, A. (2002). Mutation analysis of the coding sequence of the MECP2 gene in infantile autism. Hum Genet *111*, 305-309.

Bienvenu, T., and Chelly, J. (2006). Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. Nat Rev Genet *7*, 415-426.

Bienvenu, T., Philippe, C., De Roux, N., Raynaud, M., Bonnefond, J.P., Pasquier, L., Lesca, G., Mancini, J., Jonveaux, P., Moncla, A., *et al.* (2006). The incidence of Rett syndrome in France. Pediatr Neurol *34*, 372-375.

Bienvenu, T., Poirier, K., Friocourt, G., Bahi, N., Beaumont, D., Fauchereau, F., Ben Jeema, L., Zemni, R., Vinet, M.C., Francis, F., *et al.* (2002). ARX, a novel Prd-class-homeobox gene highly expressed in the telencephalon, is mutated in X-linked mental retardation. Hum Mol Genet *11*, 981-991.

Bird, A.P., and Wolffe, A.P. (1999). Methylation-induced repression--belts, braces, and chromatin. Cell *99*, 451-454.

Bishop, D.V. (2010). Which neurodevelopmental disorders get researched and why? PLoS One *5*, e15112.

Borg, I., Freude, K., Kubart, S., Hoffmann, K., Menzel, C., Laccone, F., Firth, H., Ferguson-Smith, M.A., Tommerup, N., Ropers, H.H., *et al.* (2005). Disruption of Netrin G1 by a balanced chromosome translocation in a girl with Rett syndrome. Eur J Hum Genet *13*, 921-927.

Bourdon, V., Philippe, C., Grandemenge, A., Reichwald, K., and Jonveaux, P. (2001a). Deletion screening by fluorescence in situ hybridization in Rett syndrome patients. Ann Genet *44*, 191-194.

Bourdon, V., Philippe, C., Labrune, O., Amsallem, D., Arnould, C., and Jonveaux, P. (2001b). A detailed analysis of the MECP2 gene: prevalence of recurrent mutations and gross DNA rearrangements in Rett syndrome patients. Hum Genet *108*, 43-50.

Brendle, A., Lei, H., Brandt, A., Johansson, R., Enquist, K., Henriksson, R., Hemminki, K., Lenner, P., and Forsti, A. (2008). Polymorphisms in predicted microRNA-binding sites in integrin genes and breast cancer: ITGB4 as prognostic marker. Carcinogenesis *29*, 1394-1399.

Brouwer, A., ten Berge, D., Wiegerinck, R., and Meijlink, F. (2003). The OAR/aristaless domain of the homeodomain protein Cart1 has an attenuating role in vivo. Mech Dev *120*, 241-252.

Buschdorf, J.P., and Stratling, W.H. (2004). A WW domain binding region in methyl-CpG-binding protein MeCP2: impact on Rett syndrome. J Mol Med *82*, 135-143.

Cai, X., Hagedorn, C.H., and Cullen, B.R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA *10*, 1957-1966.

Campos, M., Jr., Churchman, S.M., Santos-Reboucas, C.B., Ponchel, F., and Pimentel, M.M. (2010). High frequency of nonrecurrent MECP2 duplications among Brazilian males with mental retardation. J Mol Neurosci *41*, 105-109.

Carney, R.M., Wolpert, C.M., Ravan, S.A., Shahbazian, M., Ashley-Koch, A., Cuccaro, M.L., Vance, J.M., and Pericak-Vance, M.A. (2003). Identification of MeCP2 mutations in a series of females with autistic disorder. Pediatr Neurol *28*, 205-211.

Chabanon, H., Mickleburgh, I., and Hesketh, J. (2004). Zipcodes and postage stamps: mRNA localisation signals and their trans-acting binding proteins. Brief Funct Genomic Proteomic *3*, 240-256.

Chae, J.H., Hwang, Y.S., and Kim, K.J. (2002). Mutation analysis of MECP2 and clinical characterization in Korean patients with Rett syndrome. J Child Neurol *17*, 33-36.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. Science *320*, 1224-1229.

Chahrour, M., and Zoghbi, H.Y. (2007). The story of Rett syndrome: from clinic to neurobiology. Neuron *56*, 422-437.

Chang, Q., Khare, G., Dani, V., Nelson, S., and Jaenisch, R. (2006). The disease progression of Mecp2 mutant mice is affected by the level of BDNF expression. Neuron *49*, 341-348.

Chang, S., Wen, S., Chen, D., and Jin, P. (2009). Small regulatory RNAs in neurodevelopmental disorders. Hum Mol Genet *18*, R18-26.

Chapleau, C.A., Calfa, G.D., Lane, M.C., Albertson, A.J., Larimore, J.L., Kudo, S., Armstrong, D.L., Percy, A.K., and Pozzo-Miller, L. (2009). Dendritic spine pathologies in hippocampal pyramidal neurons from Rett syndrome brain and after expression of Rett-associated MECP2 mutations. Neurobiol Dis *35*, 219-233.

Cheadle, J.P., Gill, H., Fleming, N., Maynard, J., Kerr, A., Leonard, H., Krawczak, M., Cooper, D.N., Lynch, S., Thomas, N., *et al.* (2000). Long-read sequence analysis of the MECP2 gene in Rett syndrome patients: correlation of disease severity with mutation type and location. Hum Mol Genet *9*, 1119-1129.

Chen, C.Y., and Shyu, A.B. (1995). AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem Sci *20*, 465-470.

Chen, J.M., Ferec, C., and Cooper, D.N. (2006). A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. Hum Genet *120*, 1-21.

Christodoulou, J., Grimm, A., Maher, T., and Bennetts, B. (2003). RettBASE: The IRSA MECP2 variation database-a new mutation database in evolution. Hum Mutat *21*, 466-472.

Clayton-Smith, J., Watson, P., Ramsden, S., and Black, G.C. (2000). Somatic mutation in MECP2 as a non-fatal neurodevelopmental disorder in males. Lancet *356*, 830-832.

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibe, B., Bouix, J., Caiment, F., Elsen, J.M., Eychenne, F., *et al.* (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet *38*, 813-818.

Colvin, L., Leonard, H., de Klerk, N., Davis, M., Weaving, L., Williamson, S., and Christodoulou, J. (2004). Refining the phenotype of common mutations in Rett syndrome. J Med Genet *41*, 25-30.

Conne, B., Stutz, A., and Vassalli, J.D. (2000). The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology? Nat Med *6*, 637-641.

Coutinho, A.M., Oliveira, G., Katz, C., Feng, J., Yan, J., Yang, C., Marques, C., Ataide, A., Miguel, T.S., Borges, L., *et al.* (2007). MECP2 coding sequence and 3'UTR variation in 172 unrelated autistic patients. Am J Med Genet B Neuropsychiatr Genet *144B*, 475-483.

Couvert, P., Bienvenu, T., Aquaviva, C., Poirier, K., Moraine, C., Gendrot, C., Verloes, A., Andres, C., Le Fevre, A.C., Souville, I., *et al.* (2001). MECP2 is highly mutated in X-linked mental retardation. Hum Mol Genet *10*, 941-946.

Coy, J.F., Sedlacek, Z., Bachner, D., Delius, H., and Poustka, A. (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. Hum Mol Genet *8*, 1253-1262.

Cross, S.H., Meehan, R.R., Nan, X., and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet *16*, 256-259.

D'Esposito, M., Quaderi, N.A., Ciccodicola, A., Bruni, P., Esposito, T., D'Urso, M., and Brown, S.D. (1996). Isolation, physical mapping, and northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2. Mamm Genome *7*, 533-535.

De Smaele, E., Ferretti, E., and Gulino, A. (2010). MicroRNAs as biomarkers for CNS cancer and other disorders. Brain Res *1338*, 100-111.

Del Gaudio, D., Fang, P., Scaglia, F., Ward, P.A., Craigen, W.J., Glaze, D.G., Neul, J.L., Patel, A., Lee, J.A., Irons, M., *et al.* (2006). Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. Genet Med *8*, 784-792.

Demos, M.K., Fullston, T., Partington, M.W., Gecz, J., and Gibson, W.T. (2009). Clinical study of two brothers with a novel 33 bp duplication in the ARX gene. Am J Med Genet A *149A*, 1482-1486.

Deng, V., Matagne, V., Banine, F., Frerking, M., Ohliger, P., Budden, S., Pevsner, J., Dissen, G.A., Sherman, L.S., and Ojeda, S.R. (2007). FXYD1 is an MeCP2 target gene overexpressed in the brains of Rett syndrome patients and Mecp2-null mice. Hum Mol Genet *16*, 640-650.

Echenne, B., Roubertie, A., Lugtenberg, D., Kleefstra, T., Hamel, B.C., Van Bokhoven, H., Lacombe, D., Philippe, C., Jonveaux, P., and de Brouwer, A.P. (2009). Neurologic aspects of MECP2 gene duplication in male patients. Pediatr Neurol *41*, 187-191.

Einspieler, C., Kerr, A.M., and Prechtl, H.F. (2005a). Abnormal general movements in girls with Rett disorder: the first four months of life. Brain Dev *27 Suppl 1*, S8-S13.

Einspieler, C., Kerr, A.M., and Prechtl, H.F. (2005b). Is the early development of girls with Rett disorder really normal? Pediatr Res *57*, 696-700.

Erlandson, A., Samuelsson, L., Hagberg, B., Kyllerman, M., Vujic, M., and Wahlstrom, J. (2003). Multiplex ligation-dependent probe amplification (MLPA) detects large deletions in the MECP2 gene of Swedish Rett syndrome patients. Genet Test *7*, 329-332.

Erson, A.E., and Petty, E.M. (2008). MicroRNAs in development and disease. Clin Genet *74*, 296-306.
Fabbrini, G., Pasquini, M., Aurilia, C., Berardelli, I., Breedveld, G., Oostra, B.A., Bonifati, V., and Berardelli, A. (2007). A large Italian family with Gilles de la Tourette syndrome: clinical study and analysis of the SLITRK1 gene. Mov Disord *22*, 2229-2234.

Fendri-Kriaa, N., Mkaouar-Rebai, E., Moalla, D., Belguith, N., Louhichi, N., Zemni, R., Slama, F., Triki, C., and Fakhfakh, F. (2010). Mutational analysis of the MECP2 gene in Tunisian patients with Rett syndrome: a novel double mutation. J Child Neurol *25*, 1042-1046.

Filion, G.J., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P.A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol *26*, 169-181.

Freilinger, M., Bebbington, A., Lanator, I., De Klerk, N., Dunkler, D., Seidl, R., Leonard, H., and Ronen, G.M. (2010). Survival with Rett syndrome: comparing Rett's original sample with data from the Australian Rett Syndrome Database. Dev Med Child Neurol *52*, 962-965.

Friez, M.J., Jones, J.R., Clarkson, K., Lubs, H., Abuelo, D., Bier, J.A., Pai, S., Simensen, R., Williams, C., Giampietro, P.F., *et al.* (2006). Recurrent infections, hypotonia, and mental retardation caused by duplication of MECP2 and adjacent region in Xq28. Pediatrics *118*, e1687-1695.

Friocourt, G., Poirier, K., Rakic, S., Parnavelas, J.G., and Chelly, J. (2006). The role of ARX in cortical development. Eur J Neurosci *23*, 869-876.

Forster, J. (1948). [Not Available]. Wien Klin Wochenschr *60*, 692.

Fuks, F., Burgers, W.A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet *24*, 88-91.

Furukawa, T., Kozak, C.A., and Cepko, C.L. (1997). rax, a novel paired-type homeobox gene, shows expression in the anterior neural fold and developing retina. Proc Natl Acad Sci U S A *94*, 3088-3093.

Galliot, B., de Vargas, C., and Miller, D. (1999). Evolution of homeobox genes: Q50 Paired-like genes founded the Paired class. Dev Genes Evol *209*, 186-197.

Gecz, J., Cloosterman, D., and Partington, M. (2006). ARX: a gene for all seasons. Curr Opin Genet Dev *16*, 308-316.

Ghosh, R.P., Horowitz-Scherer, R.A., Nikitina, T., Shlyakhtenko, L.S., and Woodcock, C.L. (2010). MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. Mol Cell Biol *30*, 4656-4670.

Gibson, J.H., Slobedman, B., K, N.H., Williamson, S.L., Minchenko, D., El-Osta, A., Stern, J.L., and Christodoulou, J. (2010). Downstream targets of methyl CpG binding protein 2 and their abnormal expression in the frontal cortex of the human Rett syndrome brain. BMC Neurosci *11*, 53.

Gibson, J.H., Williamson, S.L., Arbuckle, S., and Christodoulou, J. (2005). X chromosome inactivation patterns in brain in Rett syndrome: implications for the disease phenotype. Brain Dev *27*, 266-270.

Gomot, M., Gendrot, C., Verloes, A., Raynaud, M., David, A., Yntema, H.G., Dessay, S., Kalscheuer, V., Frints, S., Couvert, P.*, et al.* (2003). MECP2 gene mutations in non-syndromic X-linked mental retardation: phenotype-genotype correlation. Am J Med Genet A *123A*, 129-139.

Goodlett, C.R., and Horn, K.H. (2001). Mechanisms of alcohol-induced damage to the developing nervous system. Alcohol Res Health *25*, 175-184.

Gronskov, K., Hjalgrim, H., Nielsen, I.M., and Brondum-Nielsen, K. (2004). Screening of the ARX gene in 682 retarded males. Eur J Hum Genet *12*, 701-705.

Gropman, A.L., and Batshaw, M.L. (2010). Epigenetics, copy number variation, and other molecular mechanisms underlying neurodevelopmental disabilities: new insights and diagnostic approaches. J Dev Behav Pediatr *31*, 582-591.

Group, T.R.S.D.C.W. (1988). Diagnostic criteria for Rett syndrome. The Rett Syndrome Diagnostic Criteria Work Group. Ann Neurol *23*, 425-428.

Guerrini, R., Moro, F., Kato, M., Barkovich, A.J., Shiihara, T., McShane, M.A., Hurst, J., Loi, M., Tohyama, J., Norci, V.*, et al.* (2007). Expansion of the first PolyA tract of ARX causes infantile spasms and status dystonicus. Neurology *69*, 427-433.

Hagberg, B. (1995). Rett syndrome: clinical peculiarities and biological mysteries. Acta Paediatr *84*, 971-976.

Hagberg, B., Aicardi, J., Dias, K., and Ramos, O. (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. Ann Neurol *14*, 471-479.
Hagberg, B., and Gillberg, C. (1993). Rett variants- rettoid phenotypes. from rett syndrome: clinical and biological aspect, 40-60.

Hagberg, B., Hanefeld, F., Percy, A., and Skjeldal, O. (2002). An update on clinically applicable diagnostic criteria in Rett syndrome. Comments to Rett Syndrome Clinical Criteria Consensus Panel Satellite to European Paediatric Neurology Society Meeting, Baden Baden, Germany, 11 September 2001. Eur J Paediatr Neurol *6*, 293-297.

Hagberg, B., and Witt-Engerstrom, I. (1987). Rett syndrome: epidemiology and nosology--progress in knowledge 1986--a conference communication. Brain Dev *9*, 451-457.

Hanefeld, F. (1985). The clinical pattern of the Rett syndrome. Brain Dev *7*, 320-325.

Harikrishnan, K.N., Chow, M.Z., Baker, E.K., Pal, S., Bassal, S., Brasacchio, D., Wang, L., Craig, J.M., Jones, P.L., Sif, S.*, et al.* (2005). Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nat Genet *37*, 254-264.

He, H., Jazdzewski, K., Li, W., Liyanarachchi, S., Nagy, R., Volinia, S., Calin, G.A., Liu, C.G., Franssila, K., Suster, S.*, et al.* (2005). The role of microRNA genes in papillary thyroid carcinoma. Proc Natl Acad Sci U S A *102*, 19075-19080.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol *18*, 6538-6547.

Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature *401*, 301-304.

Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P., and Walkinshaw, M.D. (2008). MeCP2 binding to DNA depends upon hydration at methyl-CpG. Mol Cell *29*, 525-531.

Hoffbuhr, K., Devaney, J.M., LaFleur, B., Sirianni, N., Scacheri, C., Giron, J., Schuette, J., Innis, J., Marino, M., Philippart, M., *et al.* (2001). MeCP2 mutations in children with and without the phenotype of Rett syndrome. Neurology *56*, 1486-1495.

Hoffbuhr, K.C., Moses, L.M., Jerdonek, M.A., Naidu, S., and Hoffman, E.P. (2002). Associations between MeCP2 mutations, X-chromosome inactivation, and phenotype. Ment Retard Dev Disabil Res Rev *8*, 99-105.

Horike, S., Cai, S., Miyano, M., Cheng, J.F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat Genet *37*, 31-40.

Huppke, P., Held, M., Hanefeld, F., Engel, W., and Laccone, F. (2002). Influence of mutation type and location on phenotype in 123 patients with Rett syndrome. Neuropediatrics *33*, 63-68.

Imessaoudene, B., Bonnefont, J.P., Royer, G., Cormier-Daire, V., Lyonnet, S., Lyon, G., Munnich, A., and Amiel, J. (2001). MECP2 mutation in non-fatal, non-progressive encephalopathy in a male. J Med Genet *38*, 171-174.

Ishii, T., Makita, Y., Ogawa, A., Amamiya, S., Yamamoto, M., Miyamoto, A., and Oki, J. (2001). The role of different X-inactivation pattern on the variable clinical phenotype with Rett syndrome. Brain Dev *23 Suppl 1*, S161-164.

Itoh, M., Ide, S., Takashima, S., Kudo, S., Nomura, Y., Segawa, M., Kubota, T., Mori, H., Tanaka, S., Horie, H., *et al.* (2007). Methyl CpG-binding protein 2 (a mutation of which causes Rett syndrome) directly regulates insulin-like growth factor binding protein 3 in mouse and human brains. J Neuropathol Exp Neurol *66*, 117-123.

Jellinger, K.A. (2003). Rett Syndrome -- an update. J Neural Transm *110*, 681-701.

Jensen, K.P., Covault, J., Conner, T.S., Tennen, H., Kranzler, H.R., and Furneaux, H.M. (2009). A common polymorphism in serotonin receptor 1B mRNA moderates regulation by miR-96 and associates with aggressive human behaviors. Mol Psychiatry *14*, 381-389.

Jian, L., Nagarajan, L., de Klerk, N., Ravine, D., Bower, C., Anderson, A., Williamson, S., Christodoulou, J., and Leonard, H. (2006). Predictors of seizure onset in Rett syndrome. J Pediatr *149*, 542-547.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet *19*, 187-191.

Jordan, C., and Francke, U. (2006). Ube3a expression is not altered in Mecp2 mutant mice. Hum Mol Genet *15*, 2210-2215.

Jordan, C., Li, H.H., Kwan, H.C., and Francke, U. (2007). Cerebellar gene expression profiles of mouse models for Rett syndrome reveal novel MeCP2 targets. BMC Med Genet *8*, 36.

Jorgensen, H.F., and Bird, A. (2002). MeCP2 and other methyl-CpG binding proteins. Ment Retard Dev Disabil Res Rev *8*, 87-93.

Kapeller, J., Houghton, L.A., Monnikes, H., Walstab, J., Moller, D., Bonisch, H., Burwinkel, B., Autschbach, F., Funke, B., Lasitschka, F., *et al.* (2008). First evidence for an association of a functional variant in the microRNA-510 target site of the serotonin receptor-type 3E gene with diarrhea predominant irritable bowel syndrome. Hum Mol Genet *17*, 2967-2977.

Kato, M., Das, S., Petras, K., Kitamura, K., Morohashi, K., Abuelo, D.N., Barr, M., Bonneau, D., Brady, A.F., Carpenter, N.J., *et al.* (2004). Mutations of ARX are associated with striking pleiotropy and consistent genotype-phenotype correlation. Hum Mutat *23*, 147-159.

Kato, M., Das, S., Petras, K., Sawaishi, Y., and Dobyns, W.B. (2003). Polyalanine expansion of ARX associated with cryptogenic West syndrome. Neurology *61*, 267-276.

Kato, M., Saitoh, S., Kamei, A., Shiraishi, H., Ueda, Y., Akasaka, M., Tohyama, J., Akasaka, N., and Hayasaka, K. (2007). A longer polyalanine expansion mutation in the ARX gene causes early infantile epileptic encephalopathy with suppression-burst pattern (Ohtahara syndrome). Am J Hum Genet *81*, 361-366.

Kerr, A.M., Nomura, Y., Armstrong, D., Anvret, M., Belichenko, P.V., Budden, S., Cass, H., Christodoulou, J., Clarke, A., Ellaway, C., *et al.* (2001). Guidelines for reporting clinical features in cases with MECP2 mutations. Brain Dev *23*, 208-211.

Kerr, A.M., and Prescott, R.J. (2005). Predictive value of the early clinical signs in Rett disorder. Brain Dev *27 Suppl 1*, S20-S24.

Khajuria, R., Gupta, N., Sapra, S., Gulati, S., Ghosh, M., Kalra, V., and Kabra, M. (2011). A novel MECP2 change in an indian boy with variant rett phenotype and congenital blindness: implications for genetic counseling and prenatal diagnosis. J Child Neurol *26*, 209-213.

Kim, V.N. (2005). MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol *6*, 376-385.

Kimura, H., and Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. J Biol Chem *278*, 4806-4812.

Kitamura, K., Yanazawa, M., Sugiyama, N., Miura, H., Iizuka-Kogo, A., Kusaka, M., Omichi, K., Suzuki, R., Kato-Fukui, Y., Kamiirisa, K., *et al.* (2002). Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. Nat Genet *32*, 359-369.

Klauck, S.M., Lindsay, S., Beyer, K.S., Splitt, M., Burn, J., and Poustka, A. (2002). A mutation hot spot for nonspecific X-linked mental retardation in the MECP2 gene causes the PPM-X syndrome. Am J Hum Genet *70*, 1034-1037.

Kleefstra, T., Yntema, H.G., Nillesen, W.M., Oudakker, A.R., Mullaart, R.A., Geerdink, N., van Bokhoven, H., de Vries, B.B., Sistermans, E.A., and Hamel, B.C. (2004). MECP2 analysis in mentally retarded patients: implications for routine DNA diagnostics. Eur J Hum Genet *12*, 24-28.

Klose, R.J., and Bird, A.P. (2004). MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. J Biol Chem *279*, 46490-46496.

Klose, R.J., Sarraf, S.A., Schmiedeberg, L., McDermott, S.M., Stancheva, I., and Bird, A.P. (2005). DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Mol Cell *19*, 667-678.

Knudsen, G.P., Neilson, T.C., Pedersen, J., Kerr, A., Schwartz, M., Hulten, M., Bailey, M.E., and Orstavik, K.H. (2006). Increased skewing of X chromosome inactivation in Rett syndrome patients and their mothers. Eur J Hum Genet *14*, 1189-1194.

Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., *et al.* (2005). Combinatorial microRNA target predictions. Nat Genet *37*, 495-500.

Kriaucionis, S., and Bird, A. (2004). The major form of MeCP2 has a novel N-terminus generated by alternative splicing. Nucleic Acids Res *32*, 1818-1823.

Kudo, S. (1998). Methyl-CpG-binding protein MeCP2 represses Sp1-activated transcription of the human leukosialin gene when the promoter is methylated. Mol Cell Biol *18*, 5492-5499.

Kutyavin, I.V., Afonina, I.A., Mills, A., Gorn, V.V., Lukhtanov, E.A., Belousov, E.S., Singer, M.J., Walburger, D.K., Lokhov, S.G., Gall, A.A., *et al.* (2000). 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. Nucleic Acids Res *28*, 655-661.

Laccone, F., Junemann, I., Whatley, S., Morgan, R., Butler, R., Huppke, P., and Ravine, D. (2004). Large deletions of the MECP2 gene detected by gene dosage analysis in patients with Rett syndrome. Hum Mutat *23*, 234-244.

Laget, S., Joulie, M., Le Masson, F., Sasai, N., Christians, E., Pradhan, S., Roberts, R.J., and Defossez, P.A. (2010). The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. PLoS One *5*, e11982.

Lam, C.W., Yeung, W.L., Ko, C.H., Poon, P.M., Tong, S.F., Chan, K.Y., Lo, I.F., Chan, L.Y., Hui, J., Wong, V., *et al.* (2000). Spectrum of mutations in the MECP2 gene in patients with infantile autism and Rett syndrome. J Med Genet *37*, E41.

Landi, D., Gemignani, F., Naccarati, A., Pardini, B., Vodicka, P., Vodickova, L., Novotny, J., Forsti, A., Hemminki, K., Canzian, F., *et al.* (2008). Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. Carcinogenesis *29*, 579-584.

Laperuta, C., Spizzichino, L., D'Adamo, P., Monfregola, J., Maiorino, A., D'Eustacchio, A., Ventruto, V., Neri, G., D'Urso, M., Chiurazzi, P., *et al.* (2007). MRX87 family with Aristaless X dup24bp mutation and implication for polyAlanine expansions. BMC Med Genet *8*, 25.

Larimore, J.L., Chapleau, C.A., Kudo, S., Theibert, A., Percy, A.K., and Pozzo-Miller, L. (2009). Bdnf overexpression in hippocampal neurons prevents dendritic atrophy caused by Rett-associated MECP2 mutations. Neurobiol Dis *34*, 199-211.

LaSalle, J.M., Goldstine, J., Balmer, D., and Greco, C.M. (2001). Quantitative localization of heterogeneous methyl-CpG-binding protein 2 (MeCP2) expression phenotypes in normal and Rett syndrome brain by laser scanning cytometry. Hum Mol Genet *10*, 1729-1740.

Lasalle, J.M., and Yasui, D.H. (2009). Evolving role of MeCP2 in Rett syndrome and autism. Epigenomics *1*, 119-130.

Laurvick, C.L., de Klerk, N., Bower, C., Christodoulou, J., Ravine, D., Ellaway, C., Williamson, S., and Leonard, H. (2006). Rett syndrome in Australia: a review of the epidemiology. J Pediatr *148*, 347-352.

Lavoie, H., Debeane, F., Trinh, Q.D., Turcotte, J.F., Corbeil-Girard, L.P., Dicaire, M.J., Saint-Denis, A., Page, M., Rouleau, G.A., and Brais, B. (2003). Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domains. Hum Mol Genet *12*, 2967-2979.

Le Meur, N., Holder-Espinasse, M., Jaillard, S., Goldenberg, A., Joriot, S., Amati-Bonneau, P., Guichet, A., Barth, M., Charollais, A., Journel, H., *et al.* (2010). MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. J Med Genet *47*, 22-29.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

Leonard, H., Silberstein, J., Falk, R., Houwink-Manville, I., Ellaway, C., Raffaele, L.S., Engerstrom, I.W., and Schanen, C. (2001). Occurrence of Rett syndrome in boys. J Child Neurol *16*, 333-338.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

Lewis, M.A., Quint, E., Glazier, A.M., Fuchs, H., De Angelis, M.H., Langford, C., van Dongen, S., Abreu-Goodger, C., Piipari, M., Redshaw, N., *et al.* (2009). An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. Nat Genet *41*, 614-618.

Li, Y., Lin, L., and Jin, P. (2008). The microRNA pathway and fragile X mental retardation protein. Biochim Biophys Acta *1779*, 702-705.

Lin, C., Franco, B., and Rosner, M.R. (2005). CDKL5/Stk9 kinase inactivation is associated with neuronal developmental disorders. Hum Mol Genet *14*, 3775-3786.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Lobo-Menendez, F., Sossey-Alaoui, K., Bell, J.M., Copeland-Yates, S.A., Plank, S.M., Sanford, S.O., Skinner, C., Simensen, R.J., Schroer, R.J., and Michaelis, R.C. (2003). Absence of MeCP2 mutations in patients from the South Carolina autism project. Am J Med Genet B Neuropsychiatr Genet *117B*, 97-101.
Lopez de Silanes, I., Quesada, M.P., and Esteller, M. (2007). Aberrant regulation of messenger RNA 3'-untranslated region in human cancer. Cell Oncol *29*, 1-17.

Lukhtanov, E.A., Mills, A.G., Kutyavin, I.V., Gorn, V.V., Reed, M.W., and Meyer, R.B. (1997). Minor groove DNA alkylation directed by major groove triplex forming oligodeoxyribonucleotides. Nucleic Acids Res *25*, 5077-5084.

Lynch, S.A., Whatley, S.D., Ramesh, V., Sinha, S., and Ravine, D. (2003). Sporadic case of fatal encephalopathy with neonatal onset associated with a T158M missense mutation in MECP2. Arch Dis Child Fetal Neonatal Ed *88*, F250-252.

Makedonski, K., Abuhatzira, L., Kaufman, Y., Razin, A., and Shemer, R. (2005). MeCP2 deficiency in Rett syndrome causes epigenetic aberrations at the PWS/AS imprinting center that affects UBE3A expression. Hum Mol Genet *14*, 1049-1058.

Maragkakis, M., Reczko, M., Simossis, V.A., Alexiou, P., Papadopoulos, G.L., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., *et al.* (2009). DIANA-microT web server: elucidating microRNA functions through target prediction. Nucleic Acids Res *37*, 273–276.

Mari, F., Azimonti, S., Bertani, I., Bolognese, F., Colombo, E., Caselli, R., Scala, E., Longo, I., Grosso, S., Pescucci, C., *et al.* (2005). CDKL5 belongs to the same molecular pathway of MeCP2 and it is responsible for the early-onset seizure variant of Rett syndrome. Hum Mol Genet *14*, 1935-1946.

Martin, M.M., Buckenberger, J.A., Jiang, J., Malana, G.E., Nuovo, G.J., Chotani, M., Feldman, D.S., Schmittgen, T.D., and Elton, T.S. (2007). The human angiotensin II type 1 receptor +1166 A/C polymorphism attenuates microrna-155 binding. J Biol Chem *282*, 24262-24269.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y.E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science *302*, 890-893.

Masuyama, T., Matsuo, M., Jing, J.J., Tabara, Y., Kitsuki, K., Yamagata, H., Kan, Y., Miki, T., Ishii, K., and Kondo, I. (2005). Classic Rett syndrome in a boy with R133C mutation of MECP2. Brain Dev *27*, 439-442.

Mazière, P., and Enright, A.J. (2007). Prediction of microRNA targets. Drug Discovery Today *Volume 12*, 452-458.

McGill, B.E., Bundle, S.F., Yaylaoglu, M.B., Carson, J.P., Thaller, C., and Zoghbi, H.Y. (2006). Enhanced anxiety and stress-induced corticosterone release are associated with increased Crh expression in a mouse model of Rett syndrome. Proc Natl Acad Sci U S A *103*, 18267-18272.

McKenzie, O., Ponte, I., Mangelsdorf, M., Finnis, M., Colasante, G., Shoubridge, C., Stifani, S., Gecz, J., and Broccoli, V. (2007). Aristaless-related homeobox gene, the gene responsible for West syndrome and related disorders, is a Groucho/transducin-like enhancer of split dependent transcriptional repressor. Neuroscience *146*, 236-247.

Meijlink, F., Beverdam, A., Brouwer, A., Oosterveen, T.C., and Berge, D.T. (1999). Vertebrate aristaless-related genes. Int J Dev Biol *43*, 651-663.

Meins, M., Lehmann, J., Gerresheim, F., Herchenbach, J., Hagedorn, M., Hameister, K., and Epplen, J.T. (2005). Submicroscopic duplication in Xq28 causes increased expression of the MECP2 gene in a boy with severe mental retardation and features of Rett syndrome. J Med Genet *42*, e12.

Mencia, A., Modamio-Hoybjor, S., Redshaw, N., Morin, M., Mayo-Merino, F., Olavarrieta, L., Aguirre, L.A., del Castillo, I., Steel, K.P., Dalmay, T.*, et al.* (2009). Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet *41*, 609-613.

Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002). Untranslated regions of mRNAs. Genome Biol *3*, REVIEWS0004.

Mishra, P.J., Humeniuk, R., Longo-Sorbello, G.S., Banerjee, D., and Bertino, J.R. (2007). A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. Proc Natl Acad Sci U S A *104*, 13513-13518.

Mnatzakanian, G.N., Lohi, H., Munteanu, I., Alfred, S.E., Yamada, T., MacLeod, P.J., Jones, J.R., Scherer, S.W., Schanen, N.C., Friez, M.J.*, et al.* (2004). A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. Nat Genet *36*, 339-341.

Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E.*, et al.* (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature *448*, 470-473.

Monros, E., Armstrong, J., Aibar, E., Poo, P., Canos, I., and Pineda, M. (2001). Rett syndrome in Spain: mutation analysis and clinical correlations. Brain Dev *23 Suppl 1*, S251-253.

Moretti, P., and Zoghbi, H.Y. (2006). MeCP2 dysfunction in Rett syndrome and related disorders. Curr Opin Genet Dev *16*, 276-281.

Moser, S.J., Weber, P., and Lutschg, J. (2007). Rett syndrome: clinical and electrophysiologic aspects. Pediatr Neurol *36*, 95-100.

Na, E.S., and Monteggia, L.M. (2011). The role of MeCP2 in CNS development and function. Horm Behav *59*, 364-368.

Nan, X., Campoy, F.J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471-481.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Nan, X., Tate, P., Li, E., and Bird, A. (1996). DNA methylation specifies chromosomal localization of MeCP2. Mol Cell Biol *16*, 414-421.

Nawara, M., Szczaluba, K., Poirier, K., Chrzanowska, K., Pilch, J., Bal, J., Chelly, J., and Mazurczak, T. (2006). The ARX mutations: a frequent cause of X-linked mental retardation. Am J Med Genet A *140*, 727-732.
Nectoux, J., Bahi-Buisson, N., Guellec, I., Coste, J., De Roux, N., Rosas, H., Tardieu, M., Chelly, J., and Bienvenu, T. (2008). The p.Val66Met polymorphism in the BDNF gene protects against early seizures in Rett syndrome. Neurology *70*, 2145-2151.

Nectoux, J., Girard, B., Bahi-Buisson, N., Prieur, F., Afenjar, A., Rosas-Vargas, H., Chelly, J., and Bienvenu, T. (2007). Netrin G1 mutations are an uncommon cause of atypical Rett syndrome with or without epilepsy. Pediatr Neurol *37*, 270-274.

Nemos, C., Lambert, L., Giuliano, F., Doray, B., Roubertie, A., Goldenberg, A., Delobel, B., Layet, V., N'Guyen M, A., Saunier, A., *et al.* (2009). Mutational spectrum of CDKL5 in early-onset encephalopathies: a study of a large collection of French patients and review of the literature. Clin Genet *76*, 357-371.

Neul, J.L., Fang, P., Barrish, J., Lane, J., Caeg, E.B., Smith, E.O., Zoghbi, H., Percy, A., and Glaze, D.G. (2008). Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. Neurology *70*, 1313-1321.

Neul, J.L., Kaufmann, W.E., Glaze, D.G., Christodoulou, J., Clarke, A.J., Bahi-Buisson, N., Leonard, H., Bailey, M.E., Schanen, N.C., Zappella, M., *et al.* (2010). Rett syndrome: Revised diagnostic criteria and nomenclature. Ann Neurol.

Nielsen, J.B., Henriksen, K.F., Hansen, C., Silahtaroglu, A., Schwartz, M., and Tommerup, N. (2001). MECP2 mutations in Danish patients with Rett syndrome: high frequency of mutations but no consistent correlations with clinical severity or with the X chromosome inactivation pattern. Eur J Hum Genet *9*, 178-184.

Nissenkorn, A., Gak, E., Vecsler, M., Reznik, H., Menascu, S., and Ben Zeev, B. (2010). Epilepsy in Rett syndrome---the experience of a National Rett Center. Epilepsia *51*, 1252-1258.

Nomura, Y., and Segawa, M. (2005). Natural history of Rett syndrome. J Child Neurol *20*, 764-768.

Nuber, U.A., Kriaucionis, S., Roloff, T.C., Guy, J., Selfridge, J., Steinhoff, C., Schulz, R., Lipkowitz, B., Ropers, H.H., Holmes, M.C., *et al.* (2005). Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. Hum Mol Genet *14*, 2247-2256.

Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell *105*, 487-497.

Orrico, A., Lam, C., Galli, L., Dotti, M.T., Hayek, G., Tong, S.F., Poon, P.M., Zappella, M., Federico, A., and Sorrentino, V. (2000). MECP2 mutation in male patients with non-specific X-linked mental retardation. FEBS Lett *481*, 285-288.

Ottens, A.K. (2009). The methodology of neuroproteomics. Methods Mol Biol *566*, 1-21.

Ozsolak, F., Poling, L.L., Wang, Z., Liu, H., Liu, X.S., Roeder, R.G., Zhang, X., Song, J.S., and Fisher, D.E. (2008). Chromatin structure analyses identify miRNA promoters. Genes Dev *22*, 3172-3183.

Pan, H., Li, M.R., Nelson, P., Bao, X.H., Wu, X.R., and Yu, S. (2006). Large deletions of the MECP2 gene in Chinese patients with classical Rett syndrome. Clin Genet *70*, 418-419.

Partington, M.W., Turner, G., Boyle, J., and Gecz, J. (2004). Three new families with X-linked mental retardation caused by the 428-451dup(24bp) mutation in ARX. Clin Genet *66*, 39-45.

Pasquini, M., Fabbrini, G., Berardelli, I., Bonifati, V., Biondi, M., and Berardelli, A. (2008). Psychopathological features of obsessive-compulsive disorder in an Italian family with Gilles de la Tourette syndrome not linked to the SLITRK1 gene. Psychiatry Res *161*, 109-111.

Peddada, S., Yasui, D.H., and LaSalle, J.M. (2006). Inhibitors of differentiation (ID1, ID2, ID3 and ID4) genes are neuronal targets of MeCP2 that are elevated in Rett syndrome. Hum Mol Genet *15*, 2003-2014.

Pelka, G.J., Watson, C.M., Christodoulou, J., and Tam, P.P. (2005). Distinct expression profiles of Mecp2 transcripts with different lengths of 3'UTR in the brain and visceral organs during mouse development. Genomics *85*, 441-452.

Petel-Galil, Y., Benteer, B., Galil, Y.P., Zeev, B.B., Greenbaum, I., Vecsler, M., Goldman, B., Lohi, H., Minassian, B.A., and Gak, E. (2006). Comprehensive diagnosis of Rett's syndrome relying on genetic, epigenetic and expression evidence of deficiency of the methyl-CpG-binding protein 2 gene: study of a cohort of Israeli patients. J Med Genet *43*, e56.

Pfaffl, M.W., Tichopad, A., Prgomet, C., and Neuvians, T.P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnol Lett *26*, 509-515.

Poirier, K., Eisermann, M., Caubel, I., Kaminska, A., Peudonnier, S., Boddaert, N., Saillour, Y., Dulac, O., Souville, I., Beldjord, C.*, et al.* (2008). Combination of infantile spasms, non-epileptic seizures and complex movement disorder: a new case of ARX-related epilepsy. Epilepsy Res *80*, 224-228.

Prescott, T.E., Rodningen, O.K., Bjornstad, A., and Stray-Pedersen, A. (2009). Two brothers with a microduplication including the MECP2 gene: rapid head growth in infancy and resolution of susceptibility to infection. Clin Dysmorphol *18*, 78-82.

Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev *15*, 1613-1618.

Purves, D., Augustine, G.J., Fitzpatrick, D., Katz, L.C., LaMantia, A.S., McNamara, J.O., and Williams, S.M. (2001). The Initial Differentiation of Neurons and Glia. Neuroscience 2nd edition Sinauer Associates, Inc.

Ravn, K., Nielsen, J.B., and Schwartz, M. (2005a). Mutations found within exon 1 of MECP2 in Danish patients with Rett syndrome. Clin Genet *67*, 532-533.

Ravn, K., Nielsen, J.B., Skjeldal, O.H., Kerr, A., Hulten, M., and Schwartz, M. (2005b). Large genomic rearrangements in MECP2. Hum Mutat *25*, 324.

Reichwald, K., Thiesen, J., Wiehe, T., Weitzel, J., Poustka, W.A., Rosenthal, A., Platzer, M., Stratling, W.H., and Kioschis, P. (2000). Comparative sequence analysis of the MECP2-locus in human and mouse reveals new transcribed regions. Mamm Genome *11*, 182-190.

Reish, O., Fullston, T., Regev, M., Heyman, E., and Gecz, J. (2009). A novel de novo 27 bp duplication of the ARX gene, resulting from postzygotic mosaicism and leading to three severely affected males in two generations. Am J Med Genet A *149A*, 1655-1660.

Rett, A. (1966). [On a unusual brain atrophy syndrome in hyperammonemia in childhood]. Wien Med Wochenschr *116*, 723-726.

Ro, S., Park, C., Young, D., Sanders, K.M., and Yan, W. (2007). Tissue-dependent paired expression of miRNAs. Nucleic Acids Res *35*, 5944-5953.

Rolando, S. (1985). Rett syndrome: report of eight cases. Brain Dev *7*, 290-296.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol *132*, 365-386.

Rusconi, L., Salvatoni, L., Giudici, L., Bertani, I., Kilstrup-Nielsen, C., Broccoli, V., and Landsberger, N. (2008). CDKL5 expression is modulated during neuronal development and its subcellular distribution is tightly regulated by the C-terminal tail. J Biol Chem *283*, 30101-30111.

Rusinov, V., Baev, V., Minkov, I.N., and Tabler, M. (2005). MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. Nucleic Acids Res *33*, W696-700.

Russo, S., Marchi, M., Cogliati, F., Bonati, M.T., Pintaudi, M., Veneselli, E., Saletti, V., Balestrini, M., Ben-Zeev, B., and Larizza, L. (2009). Novel mutations in the CDKL5 gene, predicted effects and associated phenotypes. Neurogenetics *10*, 241-250.

Saba, R., and Schratt, G.M. (2010). MicroRNAs in neuronal development, function and dysfunction. Brain Res *1338*, 3-13.

Samaco, R.C., Hogart, A., and LaSalle, J.M. (2005). Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. Hum Mol Genet *14*, 483-492.

Samaco, R.C., Nagarajan, R.P., Braunschweig, D., and LaSalle, J.M. (2004). Multiple pathways regulate MeCP2 expression in normal brain development and exhibit defects in autism-spectrum disorders. Hum Mol Genet *13*, 629-639.

Santos, M., Yan, J., Temudo, T., Oliveira, G., Vieira, J.P., Fen, J., Sommer, S., and Maciel, P. (2008). Analysis of highly conserved regions of the 3'UTR of MECP2 gene in patients with clinical diagnosis of Rett syndrome and other disorders associated with mental retardation. Dis Markers *24*, 319-324.

Sartori, S., Di Rosa, G., Polli, R., Bettella, E., Tricomi, G., Tortorella, G., and Murgia, A. (2009). A novel CDKL5 mutation in a 47,XXY boy with the early-onset seizure variant of Rett syndrome. Am J Med Genet A *149A*, 232-236.
Saudou, F., Amara, D.A., Dierich, A., LeMeur, M., Ramboz, S., Segu, L., Buhot, M.C., and Hen, R. (1994). Enhanced aggressive behavior in mice lacking 5-HT1B receptor. Science *265*, 1875-1878.

Saxena, A., de Lagarde, D., Leonard, H., Williamson, S.L., Vasudevan, V., Christodoulou, J., Thompson, E., MacLeod, P., and Ravine, D. (2006). Lost in translation: translational interference from a recurrent mutation in exon 1 of MECP2. J Med Genet *43*, 470-477.

Scala, E., Ariani, F., Mari, F., Caselli, R., Pescucci, C., Longo, I., Meloni, I., Giachino, D., Bruttini, M., Hayek, G., *et al.* (2005). CDKL5/STK9 is mutated in Rett syndrome variant with infantile spasms. J Med Genet *42*, 103-107.

Scala, E., Longo, I., Ottimo, F., Speciale, C., Sampieri, K., Katzaki, E., Artuso, R., Mencarelli, M.A., D'Ambrogio, T., Vonella, G., *et al.* (2007). MECP2 deletions and genotype-phenotype correlation in Rett syndrome. Am J Med Genet A *143A*, 2775-2784.

Schanen, C., Houwink, E.J., Dorrani, N., Lane, J., Everett, R., Feng, A., Cantor, R.M., and Percy, A. (2004). Phenotypic manifestations of MECP2 mutations in classical and atypical Rett syndrome. Am J Med Genet A *126A*, 129-140.

Schanen, N.C., Kurczynski, T.W., Brunelle, D., Woodcock, M.M., Dure, L.S.t., and Percy, A.K. (1998). Neonatal encephalopathy in two boys in families with recurrent Rett syndrome. J Child Neurol *13*, 229-231.

Scharf, J.M., Moorjani, P., Fagerness, J., Platko, J.V., Illmann, C., Galloway, B., Jenike, E., Stewart, S.E., and Pauls, D.L. (2008). Lack of association between SLITRK1var321 and Tourette syndrome in a large family-based sample. Neurology *70*, 1495-1496.

Scheffer, I., Brett, E.M., Wilson, J., and Baraitser, M. (1990). Angelman's syndrome. J Med Genet *27*, 275-277.

Schollen, E., Smeets, E., Deflem, E., Fryns, J.P., and Matthijs, G. (2003). Gross rearrangements in the MECP2 gene in three patients with Rett syndrome: implications for routine diagnosis of Rett syndrome. Hum Mutat *22*, 116-120.

Schuster-Bockler, B., Schultz, J., and Rahmann, S. (2004). HMM Logos for visualization of protein families. BMC Bioinformatics *5*, 7.

Schwartzman, J.S., Bernardino, A., Nishimura, A., Gomes, R.R., and Zatz, M. (2001). Rett syndrome in a boy with a 47,XXY karyotype confirmed by a rare mutation in the MECP2 gene. Neuropediatrics *32*, 162-164.

Sethupathy, P., Borel, C., Gagnebin, M., Grant, G.R., Deutsch, S., Elton, T.S., Hatzigeorgiou, A.G., and Antonarakis, S.E. (2007). Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the AGTR1 3' untranslated region: a mechanism for functional single-nucleotide polymorphisms related to phenotypes. Am J Hum Genet *81*, 405-413.

Seufert, D.W., Prescott, N.L., and El-Hodiri, H.M. (2005). Xenopus aristaless-related homeobox (xARX) gene product functions as both a transcriptional activator and repressor in forebrain development. Dev Dyn *232*, 313-324.

Shahbazian, M.D., Antalffy, B., Armstrong, D.L., and Zoghbi, H.Y. (2002a). Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. Hum Mol Genet *11*, 115-124.

Shahbazian, M.D., Sun, Y., and Zoghbi, H.Y. (2002b). Balanced X chromosome inactivation patterns in the Rett syndrome brain. Am J Med Genet *111*, 164-168.

Shoubridge, C., Fullston, T., and Gecz, J. (2010). ARX spectrum disorders: making inroads into the molecular pathology. Hum Mutat *31*, 889-900.

Silahtaroglu, A., and Stenvang, J. (2010). MicroRNAs, epigenetics and disease. Essays Biochem *48*, 165-185.

Simon, D., Laloo, B., Barillot, M., Barnetche, T., Blanchard, C., Rooryck, C., Marche, M., Burgelin, I., Coupry, I., Chassaing, N., *et al.* (2010). A mutation in the 3'-UTR of the HDAC6 gene abolishing the post-transcriptional regulation mediated by hsa-miR-433 is linked to a new form of dominant X-linked chondrodysplasia. Hum Mol Genet *19*, 2015-2027.

Singh, J., Saxena, A., Christodoulou, J., and Ravine, D. (2008). MECP2 genomic structure and function: insights from ENCODE. Nucleic Acids Res *36*, 6035-6047.

Smeets, E., Terhal, P., Casaer, P., Peters, A., Midro, A., Schollen, E., van Roozendaal, K., Moog, U., Matthijs, G., Herbergs, J., *et al.* (2005). Rett syndrome in females with CTS hot spot deletions: a disorder profile. Am J Med Genet A *132A*, 117-120.

Smyk, M., Obersztyn, E., Nowakowska, B., Nawara, M., Cheung, S.W., Mazurczak, T., Stankiewicz, P., and Bocian, E. (2008). Different-sized duplications of Xq28, including MECP2, in three males with mental retardation, absent or delayed speech, and recurrent infections. Am J Med Genet B Neuropsychiatr Genet *147B*, 799-806.

Stancheva, I., Collins, A.L., Van den Veyver, I.B., Zoghbi, H., and Meehan, R.R. (2003). A mutant form of MeCP2 protein associated with human Rett syndrome cannot be displaced from methylated DNA by notch in Xenopus embryos. Mol Cell *12*, 425-435.

Steffenburg, U., Hagberg, G., and Hagberg, B. (2001). Epilepsy in a representative series of Rett syndrome. Acta Paediatr *90*, 34-39.

Stromme, P., Mangelsdorf, M.E., Scheffer, I.E., and Gecz, J. (2002a). Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in Aristaless related homeobox gene, ARX. Brain Dev *24*, 266-268.

Stromme, P., Mangelsdorf, M.E., Shaw, M.A., Lower, K.M., Lewis, S.M., Bruyere, H., Lutcherath, V., Gedeon, A.K., Wallace, R.H., Scheffer, I.E., *et al.* (2002b). Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy. Nat Genet *30*, 441-445.

Taft, R.J., Pang, K.C., Mercer, T.R., Dinger, M., and Mattick, J.S. (2010). Non-coding RNAs: regulators of disease. J Pathol *220*, 126-139.

Takahashi, S., Ohinata, J., Makita, Y., Suzuki, N., Araki, A., Sasaki, A., Murono, K., Tanaka, H., and Fujieda, K. (2008). Skewed X chromosome inactivation failed to explain the normal phenotype of a carrier female with MECP2 mutation resulting in Rett syndrome. Clin Genet *73*, 257-261.

Tan, Z., Randall, G., Fan, J., Camoretti-Mercado, B., Brockman-Schneider, R., Pan, L., Solway, J., Gern, J.E., Lemanske, R.F., Nicolae, D., *et al.* (2007). Allele-specific targeting of microRNAs to HLA-G and risk of asthma. Am J Hum Genet *81*, 829-834.

Tao, J., Van Esch, H., Hagedorn-Greiwe, M., Hoffmann, K., Moser, B., Raynaud, M., Sperner, J., Fryns, J.P., Schwinger, E., Gecz, J., *et al.* (2004). Mutations in the X-linked cyclin-dependent kinase-like 5 (CDKL5/STK9) gene are associated with severe neurodevelopmental retardation. Am J Hum Genet *75*, 1149-1154.

Tejada, M.I., Penagarikano, O., Rodriguez-Revenga, L., Martinez-Bouzas, C., Garcia, B., Badenas, C., Guitart, M., Minguez, M., Garcia-Alegria, E., Sanz-Parra, A., *et al.* (2006). Screening for MECP2 mutations in Spanish patients with an unexplained mental retardation. Clin Genet *70*, 140-144.

Timchenko, L.T. (1999). Myotonic dystrophy: the role of RNA CUG triplet repeats. Am J Hum Genet *64*, 360-364.
Urdinguio, R.G., Sanchez-Mut, J.V., and Esteller, M. (2009). Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. Lancet Neurol *8*, 1056-1072.

Uyanik, G., Aigner, L., Martin, P., Gross, C., Neumann, D., Marschner-Schafer, H., Hehr, U., and Winkler, J. (2003). ARX mutations in X-linked lissencephaly with abnormal genitalia. Neurology *61*, 232-235.

Van den Veyver, I.B., and Zoghbi, H.Y. (2001). Mutations in the gene encoding methyl-CpG-binding protein 2 cause Rett syndrome. Brain Dev *23 Suppl 1*, S147-151.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol *3*, RESEARCH0034.

Van Esch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K., Lugtenberg, D., Bienvenu, T., Jensen, L.R., Gecz, J., *et al.* (2005). Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. Am J Hum Genet *77*, 442-453.
Villard, L. (2007). MECP2 mutations in males. J Med Genet *44*, 417-423.

Wade, P.A. (2005). SWItching off methylated DNA. Nat Genet *37*, 212-213.

Wallerstein, R., Sugalski, R., Cohn, L., Jawetz, R., and Friez, M. (2008). Expansion of the ARX spectrum. Clin Neurol Neurosurg *110*, 631-634.

Wan, M., Lee, S.S., Zhang, X., Houwink-Manville, I., Song, H.R., Amir, R.E., Budden, S., Naidu, S., Pereira, J.L., Lo, I.F., *et al.* (1999). Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. Am J Hum Genet *65*, 1520-1529.

Wang, G., van der Walt, J.M., Mayhew, G., Li, Y.J., Zuchner, S., Scott, W.K., Martin, E.R., and Vance, J.M. (2008). Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein. Am J Hum Genet *82*, 283-289.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat Genet *39*, 380-385.

Watson, P., Black, G., Ramsden, S., Barrow, M., Super, M., Kerr, B., and Clayton-Smith, J. (2001). Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein. J Med Genet *38*, 224-228.

Weaving, L.S., Christodoulou, J., Williamson, S.L., Friend, K.L., McKenzie, O.L., Archer, H., Evans, J., Clarke, A., Pelka, G.J., Tam, P.P., *et al.* (2004). Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. Am J Hum Genet *75*, 1079-1093.

Weaving, L.S., Ellaway, C.J., Gecz, J., and Christodoulou, J. (2005). Rett syndrome: clinical review and genetic update. J Med Genet *42*, 1-7.

Weaving, L.S., Williamson, S.L., Bennetts, B., Davis, M., Ellaway, C.J., Leonard, H., Thong, M.K., Delatycki, M., Thompson, E.M., Laing, N., *et al.* (2003). Effects of MECP2 mutation type, location and X-inactivation in modulating Rett syndrome phenotype. Am J Med Genet A *118A*, 103-114.

Wetmore, D.Z., and Garner, C.C. (2010). Emerging pharmacotherapies for neurodevelopmental disorders. J Dev Behav Pediatr *31*, 564-581.

White, R., Ho, G., Schmidt, S., Scheffer, I.E., Fischer, A., Yendle, S.C., Bienvenu, T., Nectoux, J., Ellaway, C.J., Darmanian, A., *et al.* (2010). Cyclin-dependent kinase-like 5 (CDKL5) mutation screening in Rett syndrome and related disorders. Twin Res Hum Genet *13*, 168-178.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.

Winter, J., Jung, S., Keller, S., Gregory, R.I., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol *11*, 228-234.

Xi, C.Y., Ma, H.W., Lu, Y., Zhao, Y.J., Hua, T.Y., Zhao, Y., and Ji, Y.H. (2007). MeCP2 gene mutation analysis in autistic boys with developmental regression. Psychiatr Genet *17*, 113-116.

Yaron, Y., Ben Zeev, B., Shomrat, R., Bercovich, D., Naiman, T., and Orr-Urtreger, A. (2002). MECP2 mutations in Israel: implications for molecular analysis, genetic counseling, and prenatal diagnosis in Rett syndrome. Hum Mutat *20*, 323-324.

Yasui, D.H., Peddada, S., Bieda, M.C., Vallero, R.O., Hogart, A., Nagarajan, R.P., Thatcher, K.N., Farnham, P.J., and Lasalle, J.M. (2007). Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc Natl Acad Sci U S A *104*, 19416-19421.

Ylisaukko-Oja, T., Rehnstrom, K., Vanhala, R., Kempas, E., von Koskull, H., Tengstrom, C., Mustonen, A., Ounap, K., Lahdetie, J., and Jarvela, I. (2005). MECP2 mutation analysis in patients with mental retardation. Am J Med Genet A *132A*, 121-124.

Yu, F., Thiesen, J., and Stratling, W.H. (2000). Histone deacetylase-independent transcriptional repression by methyl-CpG-binding protein 2. Nucleic Acids Res *28*, 2201-2206.

Zahorakova, D., Rosipal, R., Hadac, J., Zumrova, A., Bzduch, V., Misovicova, N., Baxova, A., Zeman, J., and Martasek, P. (2007). Mutation analysis of the MECP2 gene in patients of Slavic origin with Rett syndrome: novel mutations and polymorphisms. J Hum Genet *52*, 342-348.

Zappella, M., Gillberg, C., and Ehlers, S. (1998). The preserved speech variant: a subgroup of the Rett complex: a clinical report of 30 cases. J Autism Dev Disord *28*, 519-526.

Zappella, M., Meloni, I., Longo, I., Hayek, G., and Renieri, A. (2001). Preserved speech variants of the Rett syndrome: molecular and clinical analysis. Am J Med Genet *104*, 14-22.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev *13*, 1924-1935.

Zoghbi, H.Y. (2003). Postnatal neurodevelopmental disorders: meeting at the synapse? Science *302*, 826-830.

Zoghbi, H.Y., Percy, A.K., Schultz, R.J., and Fill, C. (1990). Patterns of X chromosome inactivation in the Rett syndrome. Brain Dev *12*, 131-135.

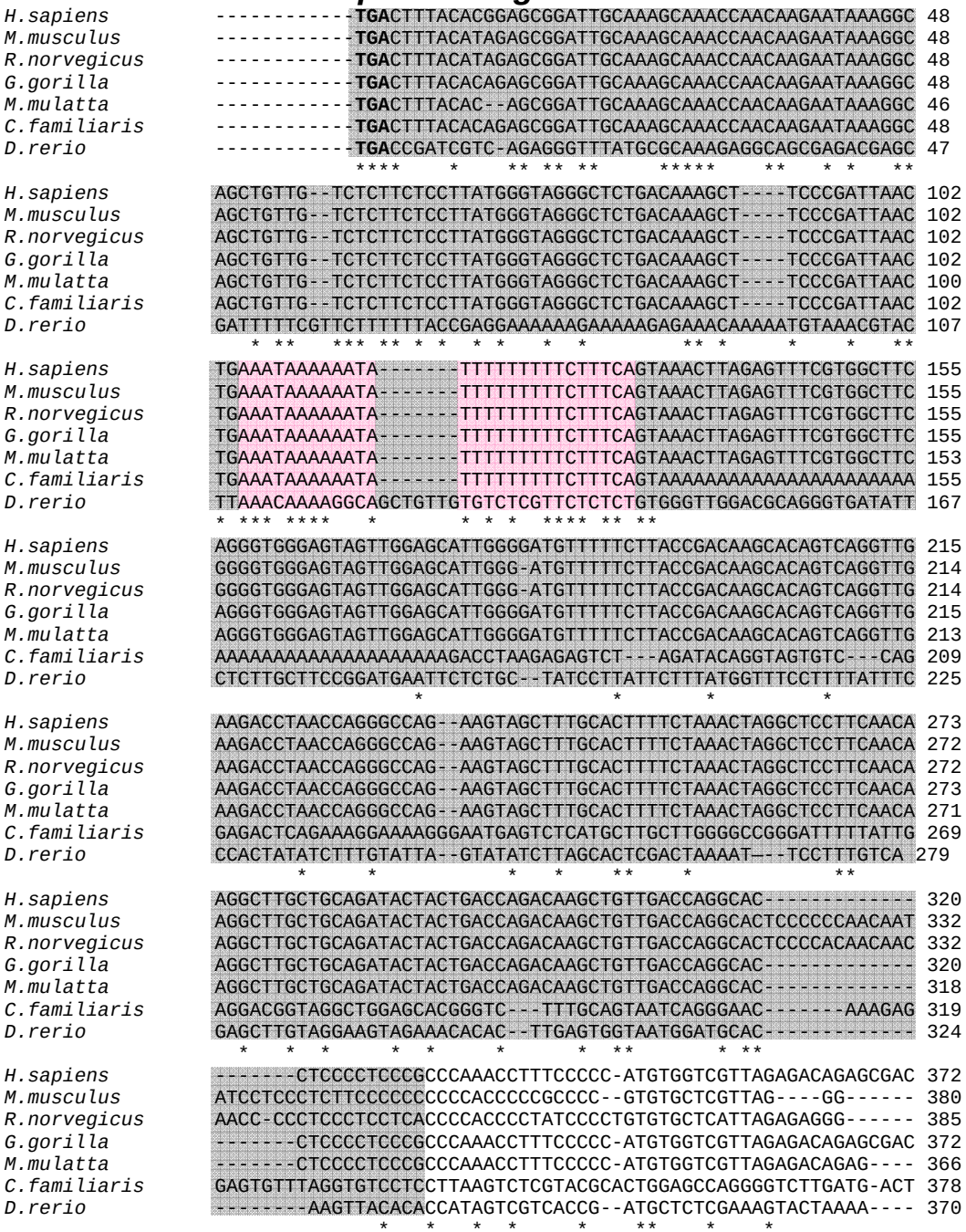
Zuchner, S., Wang, G., Tran-Viet, K.N., Nance, M.A., Gaskell, P.C., Vance, J.M., Ashley-Koch, A.E., and Pericak-Vance, M.A. (2006). Mutations in the novel mitochondrial protein REEP1 cause hereditary spastic paraplegia type 31. Am J Hum Genet *79*, 365-369.

Zweier, C., Peippo, M.M., Hoyer, J., Sousa, S., Bottani, A., Clayton-Smith, J., Reardon, W., Saraiva, J., Cabral, A., Gohring, I.*, et al.* (2007). Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). Am J Hum Genet *80*, 994-1001.

Zweier, M., Gregor, A., Zweier, C., Engels, H., Sticht, H., Wohlleber, E., Bijlsma, E.K., Holder, S.E., Zenker, M., Rossier, E.*, et al.* (2010). Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish MECP2 and CDKL5 expression. Hum Mutat *31*, 722-733.

Appendix 1

The MECP2 3'UTR sequence alignment



```
H.sapiens         ATGTTCTTCCAGTTACTTTCCAATTCT---CCTTTAGGGACAGCTTAGAATTATTTGCAC 984
M.musculus        ATATTCTTCCAATTACTTTCCAGTTCT---CCTTTAGGGACAGCTTAGAATTATTTGCAC 955
R.norvegicus      ATATTCTTCCAATTGCTTTCCAGTTCT---CCTTTAGGGACAGCTTAGAATTATTTGCAC 957
G.gorilla         ATGTTCTTCCAGTTACTTTCCAGTTCT---CCTTTAGGGACAGCTTAGAATTATTTGCAC 988
M.mulatta         ATGTTCTTCCAGTTACTTTCCAATTCTACTCCTTTAGGGACAGCTTAGAATTATTTGCAC 981
C.familiaris      CCAATCT----GTAGTGATGGAGGTCA----CCCTAGTGTTTACCGTCGGGGGTTGGTTT 961
D.rerio           GGAATTTAT-GTTTACCTGCTTGTTTT---CCATGAAGTACAATTCAAAA--AGTGGCCT 947
                      * *     *           *      *   *                 * *

H.sapiens         TATTGAGTCTTCA-TGTTCCCACTTCAAAACAAACAGATGCTCTGAGAGCAAACTGGCTT 1043
M.musculus        TATTGAGTCTTCA-TGTTCCCACTTCAAAACAAACAGATGCTCTGAAAGCAAACTGGCTT 1014
R.norvegicus      TATTGAGTCTTCA-TGTTCCCACTTCAAAACAAATAGATGCTCTGAAAGCAAACTGGCTT 1016
G.gorilla         TATTGAGTCTTCA-TGTTCCCACTTCAAAACAAACAGATGCTCTGAGAGCAAACTGGCTT 1047
M.mulatta         TATTGAGTCTTCA-TGTTCCCACTTCAAAACAAACAGATGCTCTGAGAGCAAACTGGCTT 1040
C.familiaris      GATGGGGCCTTCAGTGTGCATATGTGGAGAGTTCCGGTCGTCCCCCTAATATGCTCA-TA 1020
D.rerio           TGT---GTCTGCGCGACACAAATCCCTAACAAAAAAAA-------AAAGCATACCATTTC 0997
                    *   * ** *       *  *     *                   *  *  *    *

H.sapiens         GAATTGGTGACATTTAGTCCCTCAAGCCACCAGAT--G-TGACAGTGTTGAGAACTACCT 1100
M.musculus        GAAATGGTGACACT--GTCCCACAAGCCACCAGAC--A-TGGCAGTGTTCAGAACTACCT 1069
R.norvegicus      GAAATGGTGACACT--GTCCCACAAGCCACCAGAC--AATGGCAGTGTTCAGAACTACCT 1072
G.gorilla         GAATTGGTGACATTTAGTCCCTCAAGCCACCAGAT--G-TGACAGTGTTGAGAACTACCT 1104
M.mulatta         GAATTGGTGACGTTTAGTCCCTCAGGCCACCAGAT--G-TGATGGTGTTGAGAACTACCT 1097
C.familiaris      GAAAGAATTTCTGTCAGCTTCTAAGGCCGTAGGACTCACTTGACGCGCCATCAGCCTCCA 1080
D.rerio           AGAT--GTAGCTTTTGGTTTTATTGTGTGTGTGGTTATAGACTAATCCCATAGAGCTGCT 1055
                    *    *  *  *  *                *                    *

H.sapiens         GGATTTGTATATATACCTGCGCTTGTTTTAAAGTGGGCTCAGCA--CATAGGGTTCCCAC 1158
M.musculus        GTATCTGTATAT--ACCTGCGCTTGTTTTAAAGTGGGCTCAGCA--CATAGGATTCCCAA 1125
R.norvegicus      GTATATGTATAT--ACCTGGGCTTGTTTTAAAGTGGGCTCAGCA--CATAGGGTTCCCAC 1128
G.gorilla         GGATTTGTATATATACCTGCGCTTGTTTTAAAGTGGGCTCAGCA--CATAGGGTTCCCAC 1162
M.mulatta         GGATATGTATATATACCTGCGCTTGTTTTAAAGTGGGCTCAGCA—CATAGGGTTCCCAC 1155
C.familiaris      GTGTCATTATGT--ACTTGTTCCTATT--AAGGTGGTTCTAAGAATCTTAAAATTCTCAT 1136
D.rerio           GT-TCTAAAAATGAAACTG-----ACGGGAAAGTTAATAGATGG----CTATATTTATAT 1105
                  *  *    *  *  *  **          ** **      *            **   *

H.sapiens         GAAG-CTCCGAAACTCTAAGTGTTTGCTGCAATTTTATAAGGACTTCCTGATTGGTTTCT 1217
M.musculus        GAAG-CTCCGAAACTCTAAGTGTTTGCTGCAATTTTATAAGGACTTCCTGATTGCTTTCT 1184
R.norvegicus      GAAG-CTCCGAAACTCTAAGTGTTTGCTGCAATTTTATAAGGACTTCCTGATTGCTTTCT 1187
G.gorilla         GAAG-CTCCGAAACTCTAAGTGTTTGCTGCAATTTTATAAGGACTTCCTGATTGGTTTCT 1221
M.mulatta         GAAG-CTCCGAAACTCTAAGTGTTTGCTGCAATTTTATAAGGACTTCCTGATTGGTTTCT 1214
C.familiaris      CACGACTTTTAAAACTTTGCCATGTATGGCAATTGCATGTGCGCATGCTATGCG---CCT 1193
D.rerio           AATGTCTTGTCAGGCTGTAGC---TGTGTCAAATGCATGGGTTGCTGATGTTTGCGTGCT 1162
                   * *  **    *             *    *** *  **  *    *  *    *    **

H.sapiens         CTTCTCCCCTTCCATTTCTGCCTTTTGTTCATTTCATCCTTTCACTTCTTTCCCTTCCTC 1277
M.musculus        CTCTCGTCCTTCCATTTCTTCCTTCCTTCCATTTCATGCTTTCATTTCTTCCCCTAGCTT 1244
R.norvegicus      CTTTTGTCCTTCCATTTCTTCCTT----CCATTTCATGTTTTCATTTCTTCCCCTAGCTT 1243
G.gorilla         CTTCTCCCCTTCCATTTCTGCCTTTTGTTCATTTCATCCTTTCACTTCTTTCCCTTCCTC 1281
M.mulatta         CTTCTCTGCTTCCATTTCTGCCTTTTGTTCATTTCATTCTTTCAGTTCTTTCCCTTCCTC 1274
C.familiaris      CCGACCTAATTGCGCTGAGCTCCAAGAATGTGGTCTAACACCAAGTCTTTCTTTTATGTT 1253
D.rerio           TTGCGTGCTATTTATTACAGTACT----------CCGACATC--GTCTTTCCTGTACATA 1210
                        *    *                    *          *  **    *   *
```

```
H.sapiens         GTGTAGGGGCTTAGAGGCATGGGCTTGCTGTGGGTTTTTA-ATTGATCAGTTTTCATGTG 1976
M.musculus        GTGGTGAATTTTATAGACTTGACTTTGCTGTGGGGTTTTA-ATTGGTCAGTTTTAATTTG 1908
R.norvegicus      GTGGTGAATTTTATAAACTTGACTTTGCTGTGGGATTTTA-ATTGGTCAGTTTTCATTTG 1934
G.gorilla         GTGTAGGGGCTTAGAGGCATGGGCTTGCTGTGGGTTTTTA-ATTGATCAGTTTTCATTTG 1980
M.mulatta         GTGTAGGGGCTTCAAGGCATGGGCCTGCTGTGGGTTTTTA-ATTGATCAGTTTTCATTTG 1967
C.familiaris      GTGTCCCCGTCTGAAGTCCC--CTCAGAGGTGACATCT---GTTG-TAAGTGGTGGCTGT 1938
D.rerio           TTATCATGTACAAAAGTCAAATTTCACTTACAGTATTTGGTATCACACAATTGTGACCTT 1949
                   *             *  *                  * *    *     * *  *

H.sapiens         GGATCCCATC-TTTTTAACCTCTGTTCAGGAAGTCCTTATCTAGCTGCATATCTT-CATC 2034
M.musculus        GGATCCCAAA-GTTTTAACCTCCATTCAGGAAGTCCTTATCTAGCTGCATATCTT-CATC 1966
R.norvegicus      GGAT----------TTAACCTCTGTTCAGGAAGTCCTTATCTAGCTGCATATCTT-CATC 1983
G.gorilla         GGATCCCATC-TTTTTAACCTCTGTTCAGGAAGTCCTTATCTAGCTGCATATCTT-CATC 2038
M.mulatta         GGATCCCATC-TTTTTAACCTCTGTTCAGGAAGTCCTTATCTAGCTGCATATCTT-CATC 2025
C.familiaris      GGCT--CAGT-CCCTTGGGGGCGTCACAGGCTCGGCTCCTCTACCTGTTTGGGATGTGCC 1995
D.rerio           GGATCGCAAAACCTTTCTTAAGTGTGCAGTTTTCTCAAAGCTAAGTAAGTAATCT---TT 2006
                  ** *          **          ***      *    ***  *   *    *

H.sapiens         ATATTGGTATATCCTTTTCTGTGTTTACA-GAGATGTCTCTTAT--ATCTAAATCTGTCC 2091
M.musculus        ATATTGGTATATCCTTTTCTGTGTTTACA-GAGATGTCTCATATCTATCGAAATCTGTCT 2025
R.norvegicus      ATATTGGTATATCCTTTTCTGTGTTTACA-GAGATGTCTCATATCTATCGAAATCTGTCT 2042
G.gorilla         ATATTGGTATATCCTTTTCTGTGTTTACA-GAGATGTCTCTTAT--ATCTAAATCTGTCC 2095
M.mulatta         ATATTGGTATATCCTTTTCTGTGTTTACA-GAGATGTCTCTTAT--ATCTAAATCTGTCC 2082
C.familiaris      ATATAAGCA-ACAATCATCTGGGAAAACA-GCC-CATCCCA-AT--GCAGAGATGAGTTT 2049
D.rerio           TCATTAATTTATACTTTGTTGGGATTAGATGATATTTGGATAATATACAAATATATGAAT 2066
                    **      *   *    ** *   * * *     *     **      * **  *
```

```
H.sapiens         AACTGAGAAGTACCT-TATCAAAGTAGCAAATGAGACAGCAGTCTTATGCTTC-CAGAAA 2149
M.musculus        GA----GAAGTACCT-TATCAAAGTAGCAAATGAGACAGCAGTCTTATGCTTC-CAGAAA 2079
R.norvegicus      GACTGAGAAGTACCT-TATCAAAGTAGCAAATGAGACAGCAGTCTTACGCTTC-CAGAAA 2100
G.gorilla         AACTGAGAAGTACCT-TATCAAAGTAGCAAATGAGACAGCAGTCTTATGCTTC-CAGAAA 2153
M.mulatta         AACTGAGAAGTACCT-TATCAAAGTAGCAAATGAGACAGCAGTCTTATGCTTC-CAGAAA 2140
C.familiaris      CATGGGGAAACATCTCTGACTAACTCGCCCG-GCCAGGGCAGGGTTGTCTCCAGTGGAAG 2108
D.rerio           ATCTGAAATATGAAGGTTCAAAAATTGTAAATGGTGAGATAAGTGAAGCTCATCTGTCAT 2126
                         *        *    ** * *     *       *                 *

H.sapiens         CACCCACAGGCATGTCCCAT-GTGAGCTGCTG-CCATGAACTGTCAAGTGTGTGTTGTCT 2207
M.musculus        CACCCACAGGCACGTCCCAT-GTGAGCTGCTG-CCATGAACTGTCGAGTGTGTATTGTCT 2137
R.norvegicus      CACCCACAGGCACGTCCCAT-GTGAGCTGCTG-CCATGAACTGTCAAGTGTGTGTTGTCT 2158
G.gorilla         CACCCACAGGCATGTCCCAT-GTGAGCTGCTG-CCATGAACTGTCAAGTGTGTGTTGTCT 2211
M.mulatta         CACCCACAGGCATGTCCCAC-GTGAGCTGCTG-CCATGAACTGTCAAGTGCGTGTTGTCT 2198
C.familiaris      CGACGTTGGAGGACTCCCGT-GTCCACTCCTGGCCCCCAGCTCTTGCTAGGGCACCTACT 2167
D.rerio           CGTTAGTAAGCCAATCAGATCATTCCAAACTCACTATAAAGAGTCAGACTTGCAGAACTT 2186
                  *          **       *       **  *    *    *       *       *

H.sapiens         -TGTGTATTTCAGTTATTGT-CCCTG---GCTTCCTTACTATGGTGTAATCATGAAGGAG 2262
M.musculus        -TGTGTATTTTCGTTAACGTTCCCCA---GCTTCCTTCCTGCGGTGTAATCATGGAAGAG 2193
R.norvegicus      -TGTGTATTTTCACTAACGTTCCCCA---GCTTCCTTCCTGCGGTGTAATCATGGAAGAG 2214
G.gorilla         -TGTGTATTTCAGTTATTGT-CCCTG---GCTTCCTTACTATGGTGTAATCATGAAGGAG 2266
M.mulatta         -TGTGTATTTCAGTTATTGT-CCCTG---GCTTCCTTACTGCGGTGTAGTCATGAAGGAG 2253
C.familiaris      GTGGAGGTTCTGATTAGAGG-CTTTG---GAAATTTCAGAATGAGGGGCACCCGGGGGGC 2223
D.rerio           TAGTTTGAGCTTGCGAAAATCTGTCGTACGGCGCTGCGAAAAGGGACGGGATTAAACAGG 2246
                    *          *          *          *

H.sapiens         TGAAACATCATAGAAACTGTCTAGCACTT---CCTTG---CCAGTCTTTAGTGATCAGGA 2316
M.musculus        TGAAACATCATAGAAATCGTCTAGCACTT---CCTGG---CCAGTCCTTAGTGATCAGGA 2247
R.norvegicus      TGAAACATCATAGAATTCGTCTAGCACTT---CCTGG---CCAGTCTTTAGTGATCAGGA 2268
G.gorilla         TGAAACATCATAGAAACTGTCTAGCACTT---CCTTG---CCAGTCTTTAGTGATCAGGA 2320
M.mulatta         TGAAACATCATAGAAACTGTCTAGCACTCTTCCTTTG---CCAGTC--TAGTGATCAGGA 2308
C.familiaris      TCAG-CAGTTGAGCATCTGCCTTTGGCTCAGGTCGTGACCTCGGGGTCCCGGGATCGAGT 2282
D.rerio           ATGATTAGACATTAAAAAAGCGAGAGATTGGTCCATGTTTTAAATTTCCAGAGA---GGT 2303
                        *       *     *       *         *                * **    *

H.sapiens         ACCATAGTTGACAGTTCCAATCAG--TAGCTTAAGAAAAAACC---GTGTTTGTCTCTTC 2371
M.musculus        ACCGTAGTTGACAGTTCCAATTGA--TAGCTTAAGATAAAACC---ATGTTTGTCTCTTA 2302
R.norvegicus      ACCGTAGTTGACAGTTCGAATTGA--TAGCTTAAGATACAACC---ATGTTTGTCTCTTA 2323
G.gorilla         ACCATAGTTGACAGTTCCAATCAG--TAGCTTAAGAAAAAACC---GTGTTTGTCTCTTC 2375
M.mulatta         ACCATAGTTGACAGTTGCAGTCAG--TAGCTTAAGAAAAAAAC---GTGTTTGTCTCTTC 2363
C.familiaris      CCCACGTCGGGCTCCCTGCATGGAGCCTGCTTCTCCCTCTGCCTGAGTCTCTGCCTCTCT 2342
D.rerio           CCTGT-TTTGAT--CCTCGATTAGTCTCACGCAATCATATAAT-GCGATTTTGCAGGTCA 2359
                   *       *          *        *                    * **    *

H.sapiens         TGGAATG-GTT--AGAA----GTGAGGGAGTTTGCCCCGTTCTGTTTGTAGAGTCTCATA 2424
M.musculus        TGGAATG-GTT--AGAACTAAGTGAGAGATCTTGCCCCATTCTGTTTGCCGAAT--CATA 2357
R.norvegicus      TGGAATG-GTTTTAGAACTAAGTGAGAGATCTTGCCCCATTCTGTTTGCCGAAT--CATA 2380
G.gorilla         TGGAATG-GTT--AGAA----GTGAGGGAGTTTGCCCCATTCTGTTTGTAGAGTCTCATA 2428
M.mulatta         TGGAATG-GTT--AGAA----GTGAGGGAGTTTGCCCCATTCTGTTTGTAGAGTCTCATA 2416
C.familiaris      CTCTGTGTGTCTCATGA----ATAAATAAATAAAATCT-TAAAATTAAAAAAAAAACTCA 2397
D.rerio           GAGTTCACCAAACTTGAACTCTCCATTGCAACCAACTGCTAAACTTGCCGCATTACCTTG 2419
                            *        *                *    **      *    *

H.sapiens         GTTGGACTTTCTAGCATATATGTGTCCATTTCCTTATGCTGTAA-----AAGCAA--GTC 2477
M.musculus        GTTGGACTTT-TAGTGTATTTGTATCCATTTCCTTGTGCTATAA-----AAGCAA--ACC 2409
R.norvegicus      GCTGGACTTTCTAGTGTATTTGTATCCATTTCCTTGTGCTATAA-----AAGCAA--ACC 2433
G.gorilla         GTTGGACTTTCTAGCATATATGTGTCCATTTCCTTATGCTGTAA-----AAGCAA--GTC 2481
M.mulatta         GTTGGACTTTCTAGCATATATGTATCCATTTTCTTATGCTGTAA-----AAGCAA--GTC 2469
C.familiaris      GATAGTGTTTTCAAAGC-TTCACACCCACTAGAATG-GCCATAATG---AAGCAATAGTA 2452
D.rerio           CGTTTCCGGTCTGACGTATTCGCATGTATATGAATGGACGTCGAGGGGGAGGAAAGTGCA 2479
                    *      *         *         *      *   *    *      * *  **

H.sapiens         CTGCAACCAAACTCCCATCAGCCCAATCCCTGATCCCTG-ATCCCTTCCACCTGC----T 2532
M.musculus        CTGCAACCAGCTTTCTGTCAGGC-----------------AGTCCTTTTGCCTGC----T 2448
R.norvegicus      CTGCAACCAGCTTTCTGTCAGGC-----------------AATCCTTTTGCCTGC----T 2472
G.gorilla         CTGCAACCAAACTCCCATCAGCCCAATCCCTGATCCCTG-ATCCCTTCCACCTGC----T 2536
M.mulatta         CTGCAACCAAACTCCCATCAGCCCAATCCCTGATCCCTG-ATCCCTTCCACCTGC----T 2524
C.familiaris      ACAAAAGCCGGGACGGAATGGCCTTGGTGGCTGGTAGGA-AGCTAAATGGGCTGCG---C 2508
D.rerio           GTGTGACCGCAGCTTTACTCCCTCATCTTCGTTTTAGAAGAATCCCCTCATCCACCCCAT 2539
```

```
H.sapiens          GAATCTCTGAATTTTAAATC-ACTT--AGTAAGCGGCTC--AAGCCCAGGAGGGAGCAG- 4008
M.musculus         GCACCTCTGGATTTTAAATA-GGTTGTAATAAGTGGCTC--AAACCCATCCAGGAAAAA- 3810
R.norvegicus       GGACCTCTGGATTTTAAATA-AGTTGTAATAAGTGGATT--AAACCCATCCAGGAAAAA- 3827
G.gorilla          GAATCTCTGAATTTTAAATC-ACTT--AGTAAGCGGCTC--AAGCCCAGGAGGGAGCAG- 4021
M.mulatta          GAACCTCTGAATTTCAAATC-ACTT--AATAAGTGGCTC--AAGCCCAGGAGGGAGCAG- 3984
C.familiaris       ACGTCCCGTCCCTCCGAATCCGACTGCTCCGGGCACCTCGTAGACGCGGACTCCCGCGC- 3977
D.rerio            ACAAAAGCAAAAAAGGAAATGGTTTAGACTGATATTTATATGAAACCAGCCACAAAAGGT 4044
                                 **      *                     *

H.sapiens          AGTGAGAGCAGATGAGGTGAAA--AGGCCAAGAGGTTTGGCTCC--TGCCCACTGATAGC 5274
M.musculus         ACTGAAAGCAAATGCGCTGAAA--AGGCAAAGAGGTTTGGCTCC--TGCCCACTGATAGT 5277
R.norvegicus       ACTGAGAGCAAATGCGGTGAAA--AGGTAAAGAAGTTTGGCTCC—-TGCCCACTGATAGT 5270
G.gorilla          AGTGAGAGCAGATGAGGTGAAA--AGGCCAAGAGGTTTGGCTCC--TGCCCACTGATAGC 5290
M.mulatta          AGTGAGAGCAGATAAGGTGAAA--AGGCTAAGAGGTTTGGCTCC--TGCCCACTGATAGC 5244
C.familiaris       ACCGGGCACGGGGGTTCCGGCA--GGACCAGGGGCCCCGCCTCCACCGAGAGGCGGGGAG 5388
D.rerio            GAATAAAACAAGGGCTATGTAATCATTTCAAGTTGTCAAAAGCAGATGAAGGAAATGGAT 5477
                         *          *  *       * *          *    *

H.sapiens          CCCTCTC-CCCGCAGTGTTTGTGTGTCAAGTGGCA-AAGCTGTTCT-TCCTGGTGACCCT 5331
M.musculus         CCTT-TC-CCTGCAGTGTTTGTGTGTCAAGTGGCA-AAGCTGTTCT-TCCTGGTGACTCT 5333
R.norvegicus       CCTTCTC-CCTGCAGTGTTTGTGTGTCAAGTGGCA-AAGCTGTTCT-TCCTGGTGACTCT 5327
G.gorilla          CCCTCTC-CCTGCAGTGTTTGTGTGTCAAGTGGCA-AAGCTGTTCT-TCCTGGTGACCCT 5347
M.mulatta          CCCTCTC-CCTGCAGTGTTTGTGTGTCAAGTGGCA-AAGCTGTTCT-TCCTGGTGACCCT 5301
C.familiaris       GTGCGCC-CCAGCCACGTTGGTGTTTGGGGGCTCATAAACCGCCCGGTTCCGCTCATCCT 5447
D.rerio            TTTCACAGTCCTCAGTCCTCACTTTAAAGGCAACT-TGGTTAATTGCTTGAGGTCAAACA 5536
                         *  *     *    *     *   *              *   * * *  *

H.sapiens          GATTATATCCAGTAACACA-TAGACT-GTGCGCATAGGCCTGCTTTGTCTCCTCTATC-- 5387
M.musculus         GATTAGATCCAGTAACTTAAGAGATTTGTATGCATAGGTCTGCTTTGACTCTTCTATT-- 5391
R.norvegicus       GATTATATCTAGTAACTTA-GAGATT-GTATGCATTGGTCTGCTTTGACTCTTCTATT-- 5383
G.gorilla          GATTATATCCAGTAACGCA-TAGACT-GTGCGCATAGGCCTGCTTTGTCTCCTCTATC-- 5403
M.mulatta          GATTATATCCAGTAATGCA-TAGACT-GTGCGCATAGGCCTGCTTTGTCTCCTCTATC-- 5357
C.familiaris       AA--GATTTCACCTGCTGGAGATCTCCCGGGGCAGGGGGCTGCTCTACCTGTTGGAAGGG 5505
D.rerio            AAGAGACTTTAAGGGTTGCCAAAAACACCACCAACTCACTCTCTGTCCTCATCAGAGT-- 5594
                    *     *  *          *          *        ** *         *

H.sapiens          CTGGGCTTTT----GTTTTGCTTTTTAGTTTTGCTTTTAGTTTTTCTGTCCCTTTTATTT 5443
M.musculus         CTGGGCTTTT----GATTTGTTTTTCAGTTTTGCTTTTAGTTTTCCTAT---TTTTATTT 5444
R.norvegicus       CTGGGCTTTT----GATTTGTTTTTCAGTTTTGCTTTTAGTTTTCCTGT---TTTTATTT 5436
G.gorilla          CTGGGCTTTT----GTTTTGCTTTTTAGTTTTGCTTTTAGTTTTTCTGTCCCTTTTATTT 5459
M.mulatta          CTGGGCTTTT----GTTTTGCTTTTTAGTTTTGCTTTTAGTTTTTCTGTCCCTGTTATTT 5413
C.familiaris       CTGCGTCCCCCCAGGCTTCGCGGAAACCCCGGGCACAGAGCTTTGGGATCGCCTCGCCTC 5565
D.rerio            TTGAGCCCTCACTTAACCGGAGGCTGACGGTGAAACGTCTCCCCGGTCTGACTAAACTCG 5654
                    ** *              *                          *

H.sapiens          AACGCACCGACTAGACACACAAAGCAGTTGAATTTTTATATATATAT-----------C 5491
M.musculus         TATGCACCAACTAGACACACAAAGCAGTTGAATTTATATATATATATATATATATATATC 5504
R.norvegicus       TATGCACCAACTAGACACACAAAGCAGTTGAATTTATATATATATATATATATATATATC 5496
G.gorilla          AACGCACCGACTAGACACACAAAGCAGTTGAATTTTTATATATATAT-----------C 5507
M.mulatta          AACGCACTGACTAGACACACAAAGCAGTTGAATTTTTATATATATAT-----------C 5461
C.familiaris       GACGGGACG--TGACCCCAAGGAGC-GCCGGCTGACTACTACCGTCCGGCA------AAC 5616
D.rerio            AGCGGGAGAACGGAGGTTGAAGAACGGACAGGAAAAAAAAAACTCCAGA----------- 5702
                      *                  * * *          *

H.sapiens          TGTATATTGCACAATTATAAACTCATTTTGCTTGTGGCTCCACACACACAAAAAAA---- 5547
M.musculus         TGTATATTTCACAATTATAAACTCATTTTGCTTGTGACGCCACACACACACAAAAAGAAA 5564
R.norvegicus       TGTATATTGCACAATTATAAACTCATTTTGCTTGTGGCGCCACACACACAAAAAAA---- 5552
G.gorilla          TGTATATTGCACAATTATAAACTCATTTTGCTTGTGGCTCCACACACACAAAAA------ 5561
M.mulatta          TGTATATTGCACAATTATAAACTCATTTTGCTTGTGGCTCCACACACAAACAAA------ 5515
C.familiaris       TGTCACGTAGACAAGAACAGCTGAACGGTG-TTATTTCTAAATGTACTTTTTAAATT--- 5672
D.rerio            CATGAATTGAA--GTTACGACAACTCATAGTTTCGTTCTTAATCTGCGTCAGAAAG---- 5756
                     *    *  *     *             * **    *   *    *     **

H.sapiens          GACCTGTTAAAATTATACCTGTTGCTTAATTACAATATTTCTGATAACCATAGCATAGGA 5607
M.musculus         AACCTTTTAAAATTATACCTGTTGCTTAATTACAATATTTCTGATAACCATAGAGTAGGA 5624
R.norvegicus       ---CCTTTAAAATTATACCTGTTGCTTAATTACGATATTTCTGATAACCATAGAGTAGGA 5609
G.gorilla          GACCTGTTAAAATTATACCTGTTGCTTAATTACAGTATTTCTGATAACCATAGCATAGGA 5621
M.mulatta          GACCTGTTAAAATTATACCTGTTGCTTAATTACAATATTTCTGATAACCATAGCATAGGA 5575
C.familiaris       TTTTTTTTAAGATTTTTTAAATTTATTCAGAGAGAGAGG-CAGAGACCCA-GGCAGAGGG 5730
D.rerio            -----AGAAAGGGTAGTGCCTTTAATTTGGGAGATTACTAGAAGTGACCTTTGATTCTAA 5811
                         **   *        **  **         *          **   *

H.sapiens          CAAGGGAAAATAAAAAAAGAAAAAAAA-GAAAAAAAAACGACAAATCTGTCTGCTGG-TC 5665
M.musculus         CAAGGGAAAAAATTTAAAAAGAAAAAAAAAAAAAAGAAAAAACACATCTGTCTGCTGG-TC 5683
R.norvegicus       CAAGGGAAAAATTTTAAAAAGAAAAAAAAA----GAAAAAAAACATCTGTCTGCTGG-TC 5664
G.gorilla          CAAGGGAAAATAAAAAAAGAAAAAAAAAGAAAAAAAAACGACAAATCTGTCTGCTGG-TC 5680
M.mulatta          CAAGGGAAAATAAAAAAAGAAAAAAAA--GAAAAAAAACGACAAATCTGTCTGCTGG-TC 5632
C.familiaris       AGAAGCAGGCTCCATACAGGGAGCCTG--ATGCGGGACTCGATCCTGGGACTCCAGGATC 5788
D.rerio            ACAAACACCTTTTTCCCCTCAATGAAGCAAAGTGGGA---GGTGATTTTACTCGATGTTT 5868
                     *   *              *              *        *    **    * *
```

```
H.sapiens         ACTTCTTCTGTCCAAG-CAGATTCGTGGTCTTTT-----CCTCG----CTTCTTTCAAGG 5715
M.musculus        ACTTCTTCAATCCAAG-CAGATCTGTGATCTTT------CCTCG----CGTCTTTCAAAG 5732
R.norvegicus      ATTTCTTCAGTCCAAG-CAGATCTGTGATCTTT------CCTTG----CTTCTTTCAAAG 5713
G.gorilla         ACTTCTTCTGTCCAAG-CAGATTCGTGGTCTTTT-----CCTCG----CTTCTTTCAAGG 5730
M.mulatta         ACTTCTTCTGTCCAAG-CAGATTCGTGGTCTTTT-----CCTCG----CTTCTTTCAAGG 5682
C.familiaris      ATGCCCTGGGCCGAAGGCAGGTGCTAAAGCGCTGAGCCACCCAGGCGCCCCCGTTTTAAG 5848
D.rerio           GGGATTTTTATTCAGT-CGTTTTTCATAGGTAGGACGTGCACTGGTGTATCATTTTTTAT 5928
                        *      *   *   *                   *   *          **

H.sapiens         GCTTTCCTGTGCCAGGTGAAGGAGGCTCCAGGCAGCACCCAGGTTTTGCACTCTTGTTTC 5775
M.musculus        ACTTCCCTGTGCTAAGTGAAGGAAGCTCCAGGCTGCACCCAGGTTTTGTGCT-TTGTTTC 5791
R.norvegicus      GCTTCCCTGTGCTAAGTGAAGGAAGCTCCAAGCTGCACCCAGGTTTTGTGCT-TTGTTTC 5772
G.gorilla         GCTTTCCTGTGCCAGGTGAAGGAGGCTCCAGGCAGCACCCAGGTTTTGCAGTCTTGTCTC 5790
M.mulatta         GCTTTCCTGTGCCAGGTGAAGGAGGCTCCAGGCAGCACCCAGGTTTTGCACTCTTGTTTC 5742
C.familiaris      GATTTTTTATTT---ATGATTACCGCTTTAAATGATTTTTGAGC---GTACT-CTCTATG 5901
D.rerio           TTGTTTCTGTGTCCACCTAGAAATGATGTATACA--CTTCATTTTATTCATCCTTGTCCG 5986
                     *   * *       *     * *  *                        * *

H.sapiens         TCCCGTGCTTGTGAAAGAGGTCCCAAGGTTCTGGGTGCAGGA----------------- 5817
M.musculus        TCCTCTG-TTGTGAAAGGGGCCCCAAGATTCTGGGTACAGGACAGTTCATTTCAGCATGG 5850
R.norvegicus      TACTCTG-TTGTGAAAGGGGCCCCAAAATTCTGGGTATAGGACAGCTCATTTCAGCATGG 5831
G.gorilla         TCCCGTGCTTGTGAAAGAGGTCCCAAGGTTCTGGGTGCAGGA----------------- 5832
M.mulatta         TCCCATGCTTGTGAAAGAGGTCCCAAGGTTCTGGGTGCAGGA----------------- 5784
C.familiaris      CCCCACGTTGGGGAGATGACTTTCACGACCCTGAGATCAAGAGTCGCAGACTCCACC--- 5958
D.rerio           TTTTGTGTTCATCT-AGAAATGATGTATTGTTAAAACCCCTCTGCTTTAT---------- 6035
                        * *      *                  *
```

```
H.sapiens         GATGGA-GCGGGGCCCACCCGGTTC--AGTGTTTCTGGGGAGCTGGACAGTGGAGTGCAA 6755
M.musculus        GATGGA-GCAGG--CCACCTGGTTC--AATGTTTCTGGGCAGCTGGACAATGGAGTGCAA 6835
R.norvegicus      GATGGA-GCAGG--CTACCTGGTTC--AAGGTTTCTGGGCAGCTGGACAATGGAGTGCAA 6803
G.gorilla         GATGGA-GCGGGGCCCACCCGGTTC—-AGTGTTTCTGGGGAGCTGGACAGTGGAGTGCAA 6771
M.mulatta         GATGGA-GCAGGGCCCACCCGGTTC--AGTGTTTCTGGGGAGCTGGACAGTGGAGTGCAA 6726
C.familiaris      TGCAACTGTCGGCCAAGTCCAACTTCTAGAGTTTCCGTTTCTCTGAAACTCGGTGGGGGG 7008
D.rerio           GATGTACTCGGG-------CAATATTGACGATGTAAAAAAAGCTATAGGATTTTG-ATGT 7033
                           **                *   * *        **  *       *

H.sapiens         AAGGCTTGCAGAACTTGAAGCCTGCTCCTTCCCTTGCTACCACGGCCTCC-TTTCCGTTT 6814
M.musculus        AAGGCTTACAGAACTTGAAGCCTTTTCCTTACTTTGCTAGCACGGCCTCCTTTTCCATTT 6895
R.norvegicus      AAGGCTTGCAGAACTTGAAGCCTTTTCCTTACCTTGCTAGCACG-------TTTCCATTT 6856
G.gorilla         AAGGCTTGCAGAACTTGAAGCCTGCTCCTTCCCTTGCTACCACGGCCTCC-TTTCCGTTT 6830
M.mulatta         AAGGCTTGCAGAACTTGAAGCCTGCTCCTTCCCTTGCTACCACGGCCTCC-TTTCCGTTT 6785
C.familiaris      GGGGGCGGACAGCTCTGCAGATCCCTCCCAGCTTAGGGACCTCTGAACCGCTAGGGCCCA 7068
D.rerio           TTCGTTTTTCATCCCTGATA--TGTTGAATGTTTCACAGGTTATCGAGTGATGTGTGTTT 7091
                     *          **          *         *                *

H.sapiens         GATTTGTCACTG--CTTCAATCAATAACAGCCGCTCCA-GAGTCAGTAGTCAATGAATAT 6871
M.musculus        GATTTGTCACTG--CTTCAGTCAATAACAGCCGCTCCA-GAGTCAGTAGTTGATGAATAT 6952
R.norvegicus      GATTTGTCACTG--CTTCAATCAATAACAGACGCTCCA-GAGTCAGTAGTTGATGAATAT 6913
G.gorilla         GATTTGTCACTG--CTTCAATCAATAACAGCCGCTCCA-GAGTCAGTAGTCAATGAATAT 6887
M.mulatta         GATTTGTCACTG--CTTCAATCAATAACAGCCGCTCCA-GAGTCAGTAGTCGATGAATAT 6842
C.familiaris      AAGCAGGGAGGGGGCCCCGCTGGTGCCCAGGGGCTTCTCGGGCTTCTCGGGCAGGGGCGG 7128
D.rerio           G-TTGGGAAATATTCTCCATT--TTAACATGACCATAAACAAACAAAAGTTAAAACTTGG 7148
                     *  *     *  *  *      **    *                  *   *

H.sapiens         ATGACCAAATATCACCAGGACTGTTACTCAATGTGTGCCGAGCCCTTGC-CCATG-CTGG 6929
M.musculus        ATGACCAAATATCACCAGGACTGTTACTCAACGTGTGCCGAGCCCTTTCCTTGTG-CTGG 7011
R.norvegicus      ATGACCAAATATCACCAGGACTGTTACTCAATGTGTGCCGAGCCC-----TTGTG-CTGG 6967
G.gorilla         ATGACCAAATATCACCAGGACTGTTACTCAATGTGTGCCGAGCCCTTGC-CCATG-CTGG 6945
M.mulatta         ATGACCAAATATCACCAGGACTGTTACTCAATGTGTGCCGAGCCCTTGC-CCGTG-CTGG 6900
C.familiaris      GCGGTGGTAAGACGGAGAAGCGAGCACACGGGGCGGGGAGGACATGAAAGTGCCG-CAGG 7187
D.rerio           CCTCAAACATAATCTTCCAAC-GTTACATTTTATGACAAAAATACTTCAACCTTGACCGG 7207
                        *           *    **        *                  * * **

H.sapiens         GCTCCC-GTGTATCTGG-ACACTGTAACG--TGTGCTGTGTTTGCTCCCCTTCC--CCTT
6983
M.musculus        GCTCCCTGTGTACCTGG-ACACTGTAATG--TGTGCTGTGTTTGCTCTCCTTCC--TCTT 7066
R.norvegicus      GCTCCCTGTGTACCTGG-ACACTGTAATG--TGTGCTGTGTTTGCTCTCCTTCC--TCTT 7022
G.gorilla         GCTCCC-GTGTATCTGG-ACACTGTAACG—TGTGCTGTGTTTGCTCTCCTTCC--CCTT 6999
M.mulatta         GCTCCC-ATGTATCTGG-ACACTGTAACG--TGTGCTGTGTTTGCTCTCCTTCC--CCTT 6954
C.familiaris      GCCCAGAGGGGGTGGGGGACGGGGTGGGGA-CGGGCGGGGGTGGCACCGCCACCAGCCTT 7246
D.rerio           AAGGACAACGAGTCTGC-ACTTTATAAAGTATGTTTCCCGGCAGGACACCCATTC-ACTT 7265
                        *     *  **    *   *   *      *   *  *  *       ***

H.sapiens         CCTTCTTTGCCCTTTACTTGTCTTTCTGGGGTTTTTCTGTTT--GGGTTTGGTTTGGTTT 7041
M.musculus        CCTTCCTTGCCCTTTCCTTGTCTTTCTGGGGTTTTTCTGTT---GGGTTTGGTTTGGTTT 7123
R.norvegicus      CTTTCCTTGCCCTTTCCTTGTCTTTCTGGGGTTTTTCTGTTT--GGGTTTGGTTTGGTTT 7080
G.gorilla         CCTTCTTTGCCCTTTACTTGTCTTTCTGGGGTTTTTCTGTTT--GGGTTTGGTTTGGTTT 7057
M.mulatta         CCTTCTTTGCCCTTTACTTGTCTTTCTGGGGTTTTTCTGTTT--GGGTTTGGTTTGGTTT 7012
C.familiaris      CTTCCTCTAGCGCCT-CCAACGTTGCTCGCACCCTGGAGCCCAGGGGCTGGGTGGGGGCG 7305
D.rerio           TTATTTTTAATATGTATTTTTTATCTTTGTCCAAAGCTGTTT--TGTTTTCATATTATTT 7323
                        *      *        *  * *          *      *  *   *
```

```
H.sapiens         TTATTTCTCCTTTTGTGTTCCAAACATGAG-GTTCTCTCTACTGGTC-CTCTT-----AA 7094
M.musculus        T-ATTTTTCCTTTTGTGTTCCAAACATGAG-GTTTTCTCTACTGGTC-CTCTTT----AA 7176
R.norvegicus      T-ATTCTTCCTTTTGTGTTCCAAACATGAG-GTTCTCTCTACTGGTC-CTCTTT----AA 7133
G.gorilla         TTATTTCTCCTTTTGTGTTCCAAACATGAG-GTTCTCTCTACTGGTC-CTCTT-----AA 7110
M.mulatta         TTATTTCTCCTTTTGTGTTCCAAACATGAG-GTTCTCTCTACTGGTC-CTCTT-----AA 7065
C.familiaris      GTCAGCCCCGGAGGGGATGCAGGAAGCCGC-GTGGCCCATGCTGGTCACACTTTGAGGAA 7364
D.rerio           TTAAATTGAGTTTTAAATAAAAGGGATGGGTAGTATATATATCTAAACAACAACAACAAA 7383
                                  *                                *       *  **

H.sapiens         CTGTGGTGTTGAGGCTTATATTTGTGTAATTTTTGGTGGGTGAAAGGAAT-----TTTGC 7149
M.musculus        CTGTGGTGTTGAGGCTTCTATTTGTGTAATTTTTGGTGGGTGAAAGGAAC-----TTTGC 7231
R.norvegicus      CTGTGGTGTTGAGGCTTCTATTTGTGTAATTTTTGGTGGGTGAAAGGAAC-----TTTGC 7188
G.gorilla         CTGTGGTGTTGAGGCTTATATTTGTGTAATTTTTGGTGGGTGAAAGGAAT-----TTTGC 7165
M.mulatta         CTGTGGTGTTGAGGCTTATATTTGTGTAATTTTTGGTGGGTGAAAGGAAT-----TTTGC 7120
C.familiaris      ACAGGGGGCAGGGAGAGAAGGTTCTGGGCCTCCGGGTTCGGTTGGGTTAG-----CTGAT 7419
D.rerio           AGATGTTATTTTAACCAAACTATCCTACAGACTAGGTGTCAGTGGTTATTCGCCATCTGC 7443
                    *                 *          ***

H.sapiens         TAAGTAAATCTCT-TCTGTGTTTGAACTGAAGTCTGTA--TTGTAACTATGTTTAAAGTA 7206
M.musculus        TAAGTAAATCTCT-TCTGTGTTTGAAATGAAGTCTGTA--TTGTAACTATGTTTAAAGTA 7288
R.norvegicus      TAAGTAAATCTCT-TCTGTGTTTGAAATGAAGTCTGTA--TTGTAACTATGTTTAAAGTA 7245
G.gorilla         TAAGTAAATCTCT-TCTGTGTTTGAACTGAAGTCTGTA--TTGTAACTATGTTTAAAGTA 7222
M.mulatta         TAAGTAAATCTCT-TCTGTGTTTGAACTGAAGTCTGTA--TTGTAACTATGTTTAAAGTA 7177
C.familiaris      TGAGCAAATAAAAGTGTCTTTTTGAGTGCACCCAG----TTAAATCTGAATTTATT-TA 7474
D.rerio           CAAGCTAAGTCCTTGCCTTCCTGAAAGTGAAAAGTATAGTTTGTGTCTACTTTTATAGTA 7503
                    **  **          *  *   * ** *        **    **   ****   **

H.sapiens         ATTGTTCCAGAGACAAATATTTCTAGACACTTTTTCTTTACAAACAAAAGCATTCGGAGG 7266
M.musculus        ATTGTTCCAGAGACAAATGCTTCTAGGTACATTTTCATTACAAACAAA-GCATTTGAAGG 7347
R.norvegicus      ATTGTTCCAGAGACAAATGCTTCTAGGTACATTTTTATTACAAACGAA-GCATTGGAAGG 7304
G.gorilla         ATTGTTCCAGAGACAAATATTTCTAGACACTTTTTCTTTACAAAC----GCATTCGGAGG 7278
M.mulatta         ATTGTTCCAGAGACAAATATTTCTAGACACTTTTTCTTTACAAACAAAAGCATTCGGAGG 7237
C.familiaris      TTTATTTTAAAGA-------TTTTATTTATTTATCCAAGAGAGACAGAGACAGAGAGAGA 7527
D.rerio           -CTGTGTTACTG--------TATCATTTGTTTGCCGTATGTA-ACAGA--TACACAATAG 7551
                    * *   *  *        *   *     *         * **      *
```

```
H.sapiens         CTCCAGCCTGGAAACCTGTCTGAGGTTGGGAGAGGTGCACTTGGGGCACAGGGAGAGG-C 8222
M.musculus        CACCAGCTCAAAAACTCATCTAAGGTTGGGAGCAG-GCAGACAAGGCAGAGAGAAAGATC 8255
R.norvegicus      CACCAGCTCAAAAACCCATCTAAGGTTGGGA------CAGACAAGGCAGAGAGAAAGATT 8213
G.gorilla         CTCCAGCCTGGAAACCTGTCTGAGGTTGGGAGAGGTGCACTTGGGGCACGGGGAGAGG-C 8238
M.mulatta         CTCCAGCCTGGAAACCCGTCTGAGGTTGGGAGAGGTGCACTTGGGGCACGGGGAGAGG-A 8207
C.familiaris      AAACTTTTTTTAAATTTTTATTTATTTATGATAGTCACAGA-GAGAGAGAGAGAGAGAGG 8473
D.rerio           GACCAAACTTTGGGATCTCATAAATTTGCACCA-----GACTGTCATACAGACTTAAGGT 8439
                    *                 *    **                  *  *    *

H.sapiens         CGG-GACACACTTAGCTGG------AGATGT-CT-----CTAAAAGCCCTGTATCGTATT 8269
M.musculus        CAG-GACAGACCTAGCTGGGC-TGGAGGGGT-CT-----TGAAAAGCCCTCTGTCGTATT 8307
R.norvegicus      CAG-GACAGATATAGCTGG------AGGGGT-CT-----TGAAAAGCCCTCTGTGGTATT 8260
G.gorilla         CGG-GACACACTTAGCTGG------AGATGT-CT-----CTAAAAGCCCTGTATCGTATT 8285
M.mulatta         CAG-GACACACTTAGCTGG------AGATGT-CT-----CTAAAAGCCCTGCATCGTATT 8254
C.familiaris      CAGAGACACAGGCAGAGGGAG-AAGCAGGCT-CCATGCACCGGGAGCCTGACGTGGGATT 8531
D.rerio           TAC--ACTCATTTAACTGAGACTACCAATCTGCCAATCACTGATGACCTTTCAACTTACT 8497
                      **  *   *   *           * *           **          * *

H.sapiens         CACCTTCAGTTTTTGTGTTTTGGGACAATTACTTTAGAAAATAAGTAGGTCGTTTTAAAA 8329
M.musculus        CACCTTCAGTTTTTGTGCTTTGGGACAATTACTTTAGAAAATAAGTAGGTCGTTTTAAAA 8367
R.norvegicus      CACCTTCAGTTTTTGTGCTTTGGGACAATTACTTTAAAAAATAAGTAGGTCGTTTTAAAT 8320
G.gorilla         CACCTTCAGTTTTTGTGTTTTGGGACAATTACTTTAGAAAATAAGTAGGTCGTTTTAAAA 8345
M.mulatta         CACCTTCAGTTTTTGTGTTTTGGGACAATTACTTTAGAAAATAAGTAGGTCGTTTTAAAA 8314
C.familiaris      CGATCCCGGGTCTCCAGGATCGCGCCCTGGGCCAAAGGCAGGCGCTAAACCACTGCGCCA 8591
D.rerio           AGCCACAA---CCTAGCAACCACTTACAGTACCCTAGCAATTAACCCATAGACTTCCACT 8554
                                              *   *   *              *

H.sapiens         ACAAAAATTATTGAT--TGCTTTTTTGTAGTGTTCAGAA-AAAAGGTTCTTTGTGTATAG 8386
M.musculus        ACAAAA--TATTGAT--TGCTTTTTTGTAGTGTTCAAAACAAAAGGTTCTTTGTGTATAG 8423
R.norvegicus      ACAAAA--TATTGAT--TGCTTTTTTGTAGTGTTCAAAACAAAAGGTTCTTTGTGTATAG 8376
G.gorilla         ACAAAAATTATTGAT--TGCTTTTTTGTAGTGTTCAGAA-AAAAGGTTCTTTGTGTATAG 8402
M.mulatta         ACAAAAATTATCGAT--TGCTTTTTTGTAGTGTTCAGAA-AAAAGGTTCTTTGTGTATAG 8371
C.familiaris      CCCAGGGATCCCCCTCTTCTTTTTTTTTAAAGGTTTTATTTATTCATTCACGATAGACAG 8651
D.rerio           CTAAAAAGTGCCATT-----GACTTTACATAGGACATACTAATCCATGCTAACAATATAC 8609
                    *    *     *         ***  *  *     *   *    * *        * *

H.sapiens         CCAAATGACTGAAAGCACTGATATATTTAAAAACAAAA-GGCAATTTATTAAGGAAATTT 8445
M.musculus        CCAAATGACTGAAAGCACTGATATATTTAAAAACAAAA-GGCAATTTATTAAGGAAATTT 8482
R.norvegicus      CCAAATGACTGAAAGCACTGATATATTTAAAAACAAAA-GGCAATTTATTAAGGAAATTT 8435
G.gorilla         CCAAATGACTGAAAGCACTGATATATTTAAAAACAAAA-GGCAATTTATTAAGGAAATTT 8461
M.mulatta         CCAAATGACTGAAAGCACTGATATATTTAAAAACAAAA-GGCAATTTATTAAGGAAATTT 8430
C.familiaris      AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGCAGAG-ACCCAGGCAGAGGGAGAAGCA 8710
D.rerio           TAATCCATACAAACATACTAACAATATATTAATCATACTAGCAGCAAGCTAATTCATACT 8669
                    *          *   *   * *        *  ** *    *             *
```

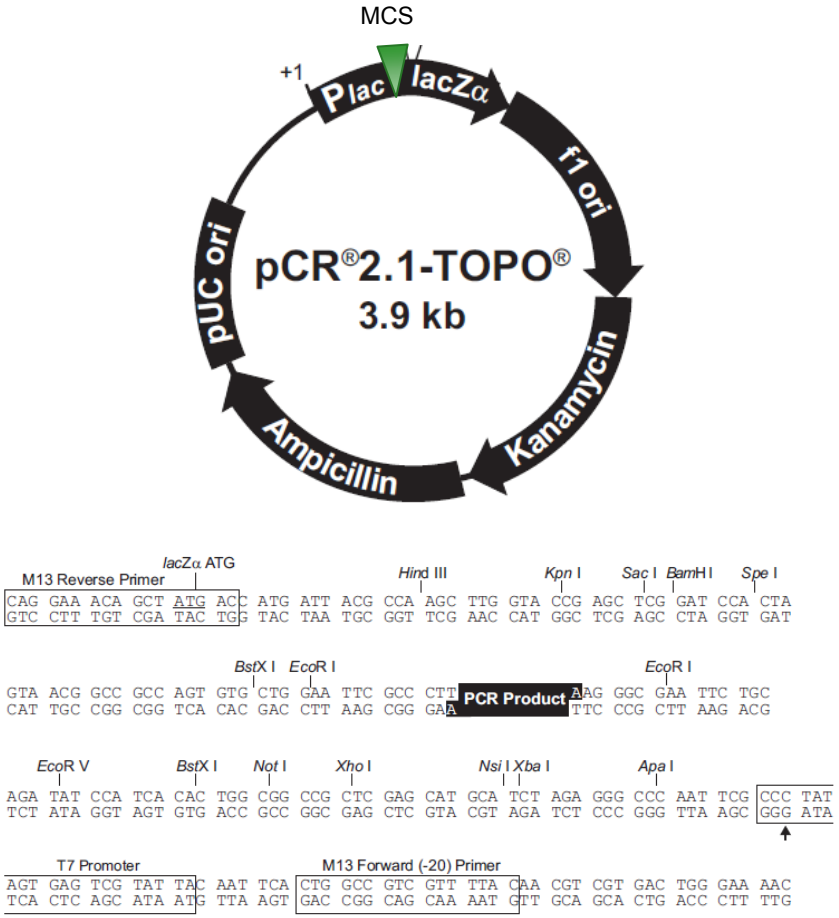
Figure 1: A cross-species alignment of the *MECP2* 3'UTR sequences showing the conserved blocks sequence and the localization of polyalanine tracts.

The four polyadenylation sequences are highlighted in pink. The conserved sequence blocks selected by Santos *et al*. are shown in grey (NM_004992: c.1667-2014, c.2620-2950, c.3610-3863; c.3827-4177; c.6911-7088; c.7175-7495; c.8441-8704; c.86664-8931; c.9903-10241).

The accession number and references for each species are as follows: *Homo sapiens* (NM_004992), *Mus musculus* (NM_010788), *Rattus norvegicus* (NM_022673), *Gorilla. gorilla* (ENSGGOT00000000224), *Macaca mulatta* (XM_001088566), and *Canis familiaris* (XM_848395) and *Danio rerio* (NM_21273).

Appendix 2

The pCR2.1-TOPO Vector Map



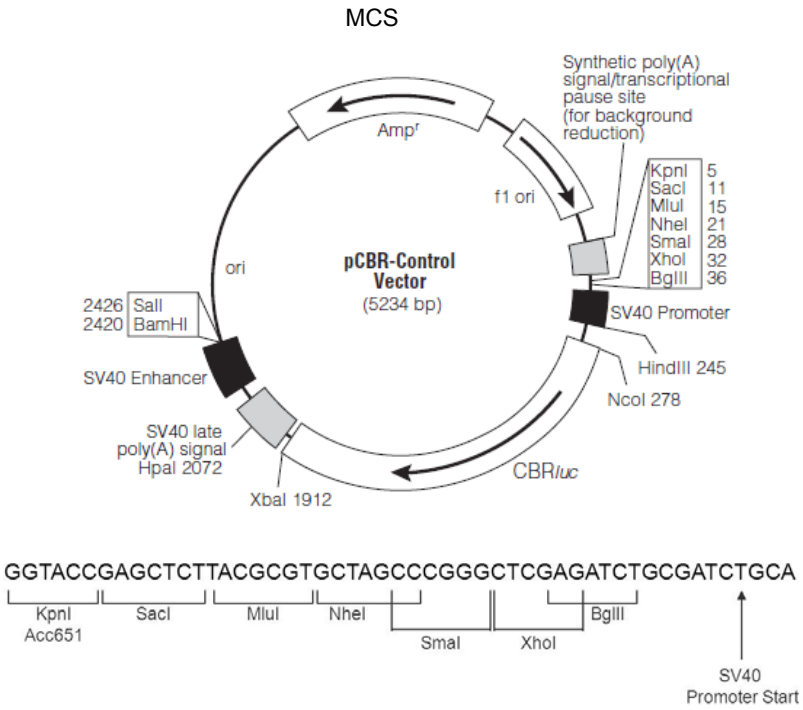
*LacZ*α fragments	1-547
M13 Reverse priming sites	205-221
Multiple Cloning Site	234-357
T7 promoter/priming site	364-383
M13 Forward (-20) priming site	391-406
F1 origin	548-985
Kanamiycin resistance ORF	1319-2113
Ampicillin resistance ORF	2131-2991
pUC origin	3136-3809

Figure 1: The pCR®2.1 TOPO Vector circle map.

The pCR®2.1-TOPO vector contains a TOPO® cloning sites for rapid and efficient cloning of PCR products. The vector is characterized by single, overhanging 3' deoxythymidine (T) residues for TA Cloning®. The vector has Ampicillin/Kanamycin resistance for easy selection in *E. coli* and a pUC origin for high copy replication of the plasmid in *E. coli*. Arrows within the genes indicate the direction of functionality. The sequence of the restriction enzyme positions along multiple cloning sites and the sequence reference points are shown below the vector map.

Appendix 3

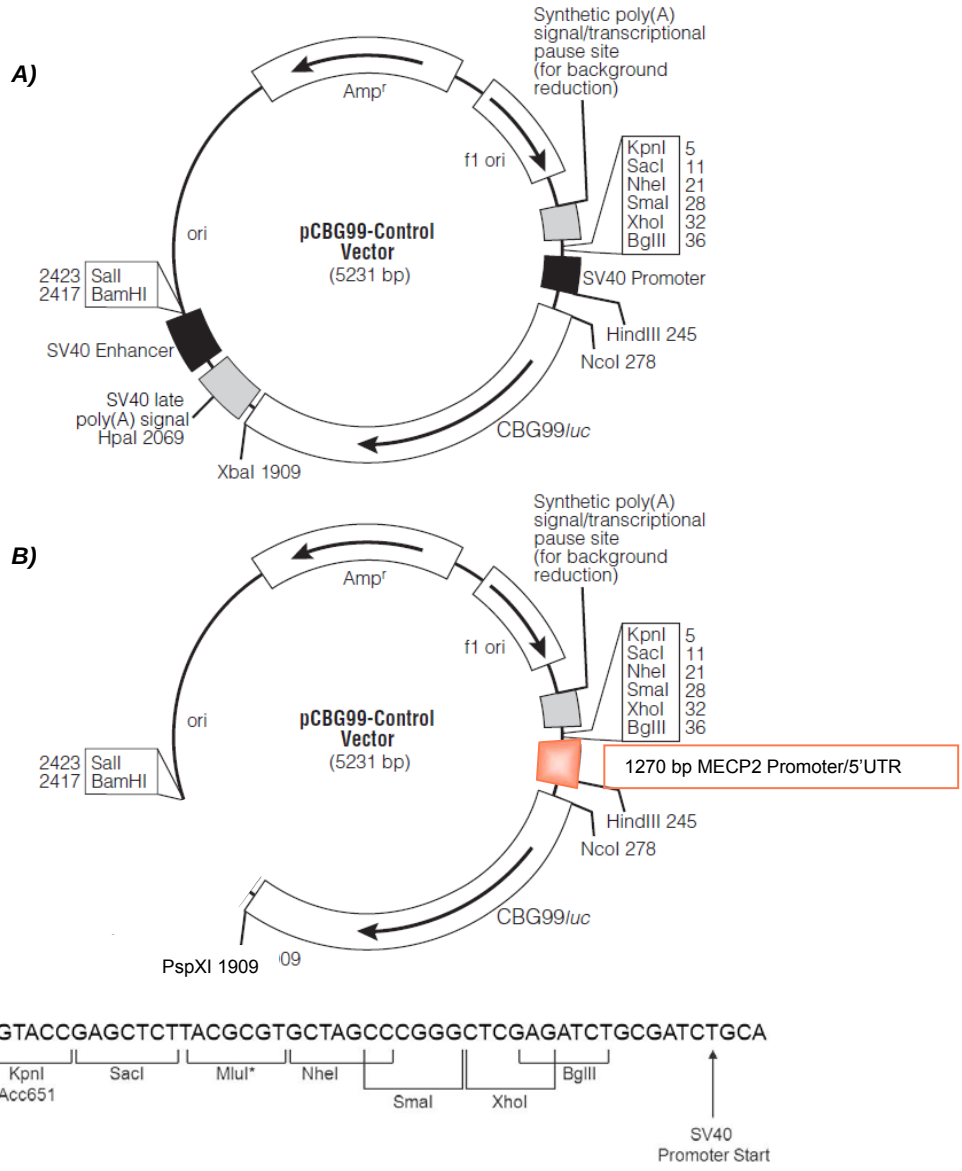
The Luciferase Expression Vector Maps



SV40 promoter	48-250
CBR*luc* reporter gene	280-1908
SV40 late poly(A) region	942-2163
SV40 Enhancer	2183-2419
Reporter Vector Primer 4 (RV Primer 4)	2477-2496
β-lactamase (Ampr) coding region	3496-4356
f1 origin of replication	4488-4943
Synthetic poly(A) signal	5074-5227
Reporter Vector Primer 3 (RV Primer 3)	5176-5195

Figure 1: The pCBR-Control Vector circle map.

The *pCBR*-Control Vector contains: *CBRluc*, a synthetic cDNA sequence encoding the red luciferase enzyme; *Amp*^{r}, gene conferring ampicillin resistance in *E. coli*; ori, origin of plasmid replication in *E. coli*. Arrows within the *CBRluc* and *Amp*^{r} genes indicate the direction of functionality. The sequences of the restriction enzymes positions along multiple cloning sites and the sequence reference points are shown below the vector map.



SV40 promoter	48-250
CBG99*luc* Reporter gene	280-1908
SV40 late poly(A) region	1939-2160
SV40 Enhancer	2180-2416
Reporter Vector Primer 3 (RV Primer 3)-	2474-2493
β-lactamase (Ampr) coding region	3493-4353
f1 origin of replication	4485-4940
Synthetic poly(A) signal	5071-5224
Reporter Vector Primer 4 (RV Primer 4)	5173-5192

Figure 2: The pCBR-Control Vector circle map.

A) The *pCBR*-Control Vector contains: CBG99luc, synthetic cDNA sequence encoding the green luciferase enzyme; *Amp*^{r}, gene conferring ampicillin resistance in *E. coli*; ori, origin of plasmid replication in *E. coli*. Arrows within the *CBG99luc* and *Amp*^{r} genes indicate the direction of functionality. The sequences of the restriction enzyme positions along multiple cloning sites and the sequence reference points are shown below the vector map.

B) The second vector is the backbone vector used in this thesis. An upstream fragment of 1270 bp of the *MECP2* promoter/5'UTR was previously cloned into the *pcBG99*-Control vector. A new enzyme restriction recognition site, PspXI, was created using site directed mutagenesis to allow the subsequent cloning of the *MECP2* 3'UTR fragments. The resulting vector *MECP2* promoter-*pCBG99* with and without SV40 late polyadenylation signal and with different *MECP2* 3'UTRs fragments were used for the experiments

Appendix 4

Clinical features of patients carrying ARX gene mutations

Table 1: Clinical and Electroencephalographic features in patients carrying *ARX* gene mutations

	Patient c.1611T>C p.Leu53Pro	**Patient c.1604T>A p.Leu535Gln**	**Patient c.1604T>A p.Leu535Gln**
Current age	16 years	2 years	3.5 years
Sex	Male	Male	Male
Neonatal auxologic parameters	BW 50 %ile L 50-75 %ile OCF 25 %ile	BW <3 %ile OCF <3%	 OCF 10%ile
Age at seizure onset	At birth	20 days	6 days
Seizures type at onset	Tonic seizures Spasms Massive myoclonus	Tonic spasms Erratic myoclonus	Hemiclonic seizures Generalized clonic seizures
EEG at onset	Suppression burst	Suppression burst	Suppression burst
Neurological evaluationat onset	Poor interaction Hypotonia	Poor interaction Marked hypotonia	Poor interaction Marked hypotonia
Psychomotor development	No acquired skills	No acquired skills	No acquired skills
Head growth	/	Congenital microcephaly	Acquired microcephaly
Epilepsy evolution	Atypical West syndrome Lennox-Gastaut like syndrome	Atypical West syndrome	Atypical West syndrome
Drug resistance	Yes	Yes	Yes
Current seizure type	Massive myoclonias Tonic seizures Focal seizures	Spasms Tonic Seizures	Spasms Focal seizures
EEG at last evaluation	Slow spike and wave Asynchronous polyspike and slow-wave complex often diffuse	Subcontinuous bilateral slow spike and wave	Subcontinuous bilateral slow spike and wave
Current neurological evaluation	Profound mental retardation Hypotonic tetra paresis	Profound mental retardation Hypotonic tetra paresis Pyramidal signs at lower limbs Dysphagia	Profound mental retardation Hypotonic tetra paresis axial hypotonia Dysphagia
Best performance	No head control	Head control	Head control
Speech	Absent	Absent	Absent
Visual contact	Absent	Absent	Absent
Genitalia abnormalities	No	No	No
Dysautonomia	No	No	Apnoea
Dystonia	Yes	No	No
Brain MRI	1 year: Normal 13 years: Brain atrophy (>F) Thinning of corpus callosum Arachnoid cyst	6 months: Normal 22 months: Diffuse brain atrophy	6& days: Normal 2 years: Diffuse brain atrophy

Abbreviations: EEG: electroencephalography; MRI: magnetic resonance imaging; OCF: occipito-frontal circumference; BW: birth weight; L: length; F: frontal region.

Acknowledgements

I would like to express my gratitude to Dr. Alessandra Murgia, who always believed in me and encouraged me and gave me a push to face a wonderful experience overseas.

I would like to show my gratitude to Prof. John Christodoulou, who advised me in the last two years. It was an honour for me to part of the "NSW Centre for Rett Syndrome Research" at the Children's Hospital, Westmead, Sydney, Australia.

Thank you to Dr Lina Artifoni for her thought-provoking discussions and precious advice. Many thanks to my brilliant colleagues Dr Cristina Busana, Dr Emanuela Leonardi, Dr Maddalena Martella and Dr Roberta Polli for their professional and personal encouragement and support through the last years.

Thank you to the members of the Rett Centre and Metabolic Research laboratory: Dr Roksana Armani, Dr Lisa Riley, Dr XingZhang Tong, Dr Sarah Williamson, Katrina Fisk, Simranpreet Kaur, Gladys Ho, Minal Menezes, Naz Al Hafid and Ahmad Alodaib. A special thank to Dr Wendy Gold for the help she have provided me over the past two years and for the homemade Italian food.

Many thanks to my family for their endless encouragement, trust and love especially while I was away.

I am enormously grateful to Julien for his truly love and constant patience.