UNIVERSITAT DE BARCELONA

Study of expression levels in *MECP2* related disorders using transcriptomics and proteomics: characterizing Rett syndrome and *MECP2* duplication syndrome

Ainhoa Pascual Alonso

Study of expression levels in *MECP2* related disorders using transcriptomics and proteomics: characterizing Rett syndrome and *MECP2* duplication syndrome

Ainhoa Pascual Alonso

2023

UNIVERSITAT DE BARCELONA

SJD Sant Joan de Déu Barcelona · Hospital

Study of expression levels in *MECP2* related disorders using transcriptomics and proteomics: characterizing Rett syndrome and *MECP2* duplication syndrome.

Ainhoa Pascual Alonso

2023

Universitat de Barcelona
Facultad de Biología; Departamento de Genética, Microbiología y Estadística

Study of expression levels in *MECP2* related disorders using transcriptomics and proteomics: characterizing Rett syndrome and *MECP2* duplication syndrome.

Memoria presentada por:
Ainhoa Pascual Alonso

Para optar al grado de
Doctora por la Universitat de Barcelona

Programa de Doctorado de Genética.
Tesis realizada en el Institut de Recerca Sant Joan de Déu.

Judith Armstrong Morón
Directora

Bru Cormand Rifà
Tutor

Ainhoa Pascual Alonso
Doctoranda

Barcelona, mayo de 2023

Gurasoei

Agradecimientos

Nunca me planteé hacer un doctorado y estoy a punto de acabarlo. Han sido unos años llenos de experiencias, de primeras veces, de momentos inolvidables, y alguno que otro para olvidar. Por suerte, en esta etapa no he estado sola; me he rodeado de personas maravillosas con las que he vivido momentos muy felices, pero que también me han cogido de la mano y me han ayudado a avanzar incluso en los momentos más oscuros. Por todo ello, muchas gracias de corazón.

Gracias a las madres y padres de la asociación Miradas que hablan: Duplicación de *MECP2,* por el esfuerzo titánico que hicisteis para reunir el dinero que hizo posible comenzar con este proyecto de doctorado. Ha sido un privilegio haberos conocido. Gracias también a las asociaciones de síndrome de Rett nacionales, y a todas las personas que se *mueven por los que no pueden*. Me gustaría agradecer en particular a todas las familias que han querido participar en el presente proyecto aportando esos milímetros de piel que han dado lugar a cientos de miles de datos genéticos que permitirán que la investigación de los trastornos relacionados con el gen *MECP2* siga avanzando a buen ritmo.

La oportunidad de realizar la tesis doctoral se la debo a mi directora de tesis, la Dra. Judith Armstrong. Muchas gracias por escogerme y confiar en mí para el proyecto que te propusieron desde la asociación Miradas que hablan. Han sido unos años muy intensos, llenos de retos y de contratiempos donde hemos tenido que ingeniárnoslas para salir adelante en más de una, y de dos, ocasiones; incluso nos topamos con una pandemia que descolocó nuestros planes. Gracias a ti he podido aprender de las y los mejores, me has proporcionado herramientas para crecer y consejos para seguir avanzando y mejorando. Echando la vista atrás, creo que hemos salido airosas y espero que estés tan orgullosa como yo del gran trabajo que hemos hecho.

Al personal del laboratorio de genética del Hospital Sant Joan de Déu, habéis logrado que ir a trabajar no se me hiciera cuesta arriba. Clara, Nuri, Carlota y Gonzalo, ha sido un verdadero placer poder compartir charletas profesionales y personales con vosotras entre tés y otros caldos. No se me ocurre un entorno más adecuado para haber descubierto que soy una mutante incapaz de disfrutar del cilantro. Espero que sigamos descubriendo rincones nuevos dentro y fuera de Barcelona. Silvia, Paola, Dèlia, Marta, Eloi, JM, Laia y el resto de compis del laboratorio de genética, arrays y citogenética, ha sido un lujo compartir departamento con vosotras y que hayamos podido ir celebrando todos los logros, cumpleaños y eventos como una gran familia. Preparaos, porque todavía nos

queda esto por celebrar. Ha sido una simbiosis en la que he aprendido muchas lecciones que sé que seguiré aplicando durante años, y me quedo con que no sólo he dejado un legado en forma de datos, técnicas y protocolos nuevos, sino que he despertado una faceta eco-warrior en algunos de vosotros.

I would also like to mention the people I collaborated with at the Prokisch lab in Munich. Those four months were a huge professional challenge, but at the same time I learnt so many other things that I will always remember those days with great affection. Dima, Riccardo, Christiane, Aiman, Fatemeh, Masaru, Nikita, Beryll, Robert, Catalina, Vicente, Dewi and Holger, thank you very much for your welcome, beyond the strictly professional. If those months were awesome, it was because of you. Danke schön!

Aquesta llista d'agraïments estaria incompleta si no esmentés aquelles persones que m'han acompanyat durant la meva estada a Catalunya. No me olvido de Mireia, increíble amiga y cómplice durante años en Catalunya y en tierras extranjeras. Recuerda que, amiga, estés donde estés, amb tu lluny és encara a prop.

Gràcies a Carla, Paola, Romina, Enric, Oriol, Ana, Adrián i Abel per construir amb mi el que nosaltres a Euskadi diem kuadrila. Junts hem descobert noves cultures, hem escapat de tota mena de llocs salvant al món en diverses ocasions, hem caçat raigs de sol *tumbados en la arena*, hem ballat amb grups desconeguts i fins i tot, hem saltat des d'un pont. Perquè encara que ens vagi a ploure, amb vosaltres sé que tot ens anirà bé i que riurem i gaudirem allà on estem.

Carlos, muchas gracias por estar a mi lado, por quererme, por hacerme reír y por transmitirme tu optimismo. Tu apoyo incondicional durante estos años, tanto en los momentos de éxito como en aquellos en los que las dificultades parecían insalvables, ha sido un regalo. Me siento muy afortunada de tenerte a mi lado.

Era berean, hitz samurrak eskaini nahi dizkiet gaztetatik nire bizitzan daudenei: Ibon, Ana, Irati, Janire, Maider, Julen eta besteei. Gasteizera itzultzen naizenetan zuekin egoteko irrikatan nagoelako, eta bitartean elkarbanatu ditugun mezu, postal, audio eta telefono dei amaigabeek, energiaz bete eta etxean bezala sentiarazten nautelako. Zuei ere eskerrik asko.

Eta azkenik, baina ez horregatik garrantzia gutxiagorekin, eskerrak eman nahi dizkiet gurasoei, nire erabakiak beti errespetatu dituzuelako eta bizitzan aurrera egiteko behar nituen lanabesak, indarrak, adorea eta besarkadak beti prest zenituztelako. Zalantzarik gabe, onenak zarete. Mila esker, bihotz-bihotzez.

"Somewhere, something incredible is waiting to be known."

— Carl Sagan.

Abstract

MECP2 is a multifunctional gene involved in multiple processes such as transcription regulation, chromatin remodelling, splicing and miRNA regulation. Malfunction of *MECP2* due to loss of function mutations leads to Rett syndrome (RTT) whereas its overexpression triggers *MECP2* duplication syndrome (MDS). Besides, variants in *MECP2* can cause a wide spectrum of phenotypes, from severe congenital encephalopathy with early death to mild intellectual disability (ID). RTT and MDS are two well characterized rare diseases with a partly overlapping phenotype consisting of neurodevelopmental delay, ID, impaired muscle tone, lack or unstable ambulation, little or absent speech, gastrointestinal problems, autism like behaviour and hand stereotypies. With next generation sequencing derived methodologies, gigantic breakthroughs have been done in diagnostics and research. These new omic strategies, such as transcriptomic or proteomic, can be applied to patient-derived samples to obtain answers to some of the still unknown aspects of the molecular effect of *MECP2* in RTT and MDS.

For the present thesis project, patients with alterations in *MECP2* were gathered and three cohorts were created and thoroughly studied and characterized: a classic RTT girls group with large deletions within *MECP2*, a group of patients with MDS together with their duplication carrier mothers, and a group of boys with ID and neurodevelopmental delay with variants in *MECP2*. Genotype-phenotype correlations were also attempted for these cohorts.

In order to further study patients with classic RTT and MDS we decided to use a multi-omic (transcriptomic and proteomic) approach. For that, 22 classic RTT, 17 MDS, 10 *MECP2* duplication carriers and 13 healthy controls were gathered and primary cultured cell lines were established from their skin biopsies. DNA, RNA and proteins were extracted from them all and RNA sequencing and tandem mass tag-mass spectrometry (TMT-MS) experiments were performed. The obtained data was analysed in a case-control approach.

The multi-omic analysis revealed shared and distinct altered biological processes for each cohort studied. The gene causing RTT and MDS is the same, but its downstream molecular effects might be opposite. Being able to obtain RNA and protein profiles from these patient cohorts seems to be a promising way to better understand *MECP2*'s role in the underlying pathomechanism triggering RTT and MDS. Differentially expressed genes and proteins involved in cytoskeleton, vesicular activity or immune system were found, and some of

them are highlighted as potential biomarker and therapeutic target candidates. Altogether, we aimed to fill the gap by exploring the patients' genetics, transcriptomics and proteomics in order to get closer to identifying therapeutic targets and biomarkers that could be used in future clinical trials.

Index

ABBREVIATIONS

AAV	Adeno-associated virus
ASD	Autism spectrum disorder
ASO	Antisense oligonucleotide
BH	Benjamini-Hochberg corrected p-value
CNS	Central nervous system
CPM	Counts per million mapped reads
CSS	Clinical severity score
CTD	C-terminal domain
Ddci	Dopa-decarboxylase inhibitor
DE	Differential expression
DEG	Differentially expressed gene
DEP	Differentially expressed protein
DPR	Deletion prone region
EEG	Electroencephalogram
HDAC	Histone deacetylase
hESC	Human embryonic stem cell
hiPSC	Human-induced pluripotent stem cell
ID	Intellectual disability
ID	Intervening domain
IL-1R	Interleukin-1 receptor
IRAK1	Human interleukin-1 receptor-associated kinase 1 gene
IRAK1	Human interleukin-1 receptor-associated kinase 1 protein
Irak1	Mouse interleukin-1 receptor-associated kinase 1 gene
LCR	Low copy repeat
LTD	Long-term depression
LTP	Long-term potentiation
MBD	Methyl-binding domain
mCG	Methylated CpG
MDS	MECP2 duplication syndrome
MECP2	Human methyl CpG binding protein 2 gene
MeCP2	Human methyl CpG binding protein 2 protein
Mecp2	Mouse methyl CpG binding protein 2 gene
Mecp2	Mouse methyl CpG binding protein 2 protein
MS	Mass spectrometry
NGS	Next generation sequencing
NLS	Nuclear localization signal domain
NTD	N-terminal domain
ORA	Overrepresentation analysis
PCA	Principal component analysis

PSM	Peptide spectrum match
PTM	Post-translational modification
RNAseq	RNA sequencing
ROS	Reactive oxygen species
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction
RTT	Rett syndrome
scRNA-seq	single-cell RNA sequencing
SNV	Single nucleotide variant
TF	Transcription factor
TLR	Toll-like receptor
TMT	Tandem mass tag
TRD	Transcriptional repression domain
VUS	Variant of unknown significance
WT	Wild type
XCI	X chromosome inactivation
XLMR	X-linked mental retardation

INTRODUCTION

1. *MECP2* gene

Methyl CpG binding protein 2 (*MECP2*; OMIM*300005) is a gene located on the long arm of the X chromosome (Xq28 region) and undergoes X chromosome inactivation (XCI) in females for gene dosage compensation [1]. The protein that generates, MeCP2, is expressed ubiquitously and its principal function is to regulate transcription, both activating and repressing it. Besides, MeCP2 is essential for proper neuronal maintenance and correct neuronal plasticity.

MECP2 has four exons and generates two well-known isoforms, *MECP2_e1* and *MECP2_e2*, which are translated to the proteins MeCP2_e1 and MeCP2_e2 [Figure 1]. *MECP2_e1* retains exons 1, 3 and 4 whereas *MECP2_e2* retains exons 2, 3 and 4 [2]. MeCP2_e1 is the ancestral form of the protein since its orthologues are present across vertebrate evolution, while orthologues of the exon 2 coding region have been found only in mammals [3]. MeCP2_e1 is the most abundant isoform in the brain but the ratio between the two isoforms varies across different tissues, being, for example, MeCP2_e2 more abundant in fibroblasts [2]. Although the two isoforms share most of their sequence, which includes the main functional domains, they are not completely redundant. That is, each isoform has its own properties, functions, spatial expression and interacting partners [3–5].

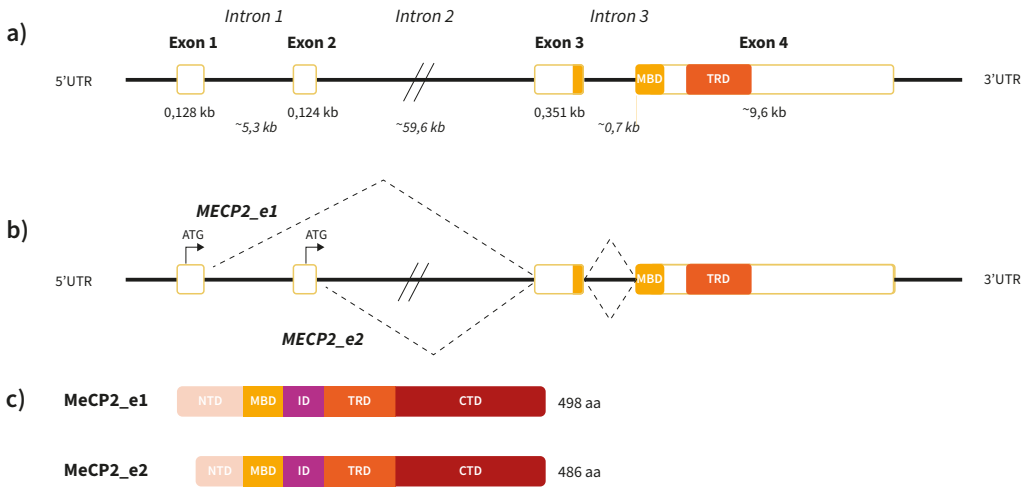

Figure 1: *MECP2* gene's scheme. a) *MECP2*'s four exons and the position of its two main functional domains: MBD and TRD. The lengths of the exons and introns are specified in kilobases (kb) b) Scheme of the alternative splicing that generates the isoforms *MECP2_e1* (containing exons 1, 3, 4) and *MECP2_e2* (containing exons 2, 3, 4). c) The two MeCP2 protein isoforms with their domains and lengths in amino acids (aa).

MeCP2 has several structural domains; an N-terminal domain (NTD), a methyl-binding domain (MBD), an intervening domain (ID), a transcriptional repression domain (TRD), a C-terminal domain (CTD) and a nuclear localization signal domain (NLS). Among them, MBD and TRD are the most crucial domains and where most of the disease-causing mutations are located [6]. At the beginning, it was thought that MBD only permitted the binding of MeCP2 to methylated CpG (mCG) dinucleotides, but later studies demonstrated that MBD binds to other methylated non-CG motifs such as mCAC, and even to unmethylated CAC/GTG containing DNA fragments [7,8]. On the other hand, TRD allows chromatin binding and recruits co-repressor complexes, such as NCoR-SMRT and histone deacetylase (HDAC)-Sin3A complex, to repress transcription [9,10]. These co-repressor complexes usually contain a HDAC protein that removes acetyl groups from the lysine residue of the histones, thereby modulating chromatin structure.

MeCP2 is a mainly unstructured protein that can adopt local secondary structures when interacting with other molecules, explaining its participation in multiple molecular interactions and functions [11–13]. Therefore, *MECP2* is a multifunctional gene implicated in transcription activation, repression and chromatin remodelling. The first function *MECP2* was associated with was transcription repression by interaction with Sin3A and HDACs via its TRD domain [14]. Binding to other co-repressor proteins such as NCoR, SMRT, TBL1, TBLR1 and HDAC3 through its NCoR-SMRT interaction domain within TRD, supported that initial function [10]. However, Zoghbi's group reported that MeCP2 was also responsible for transcription activation due to its association with CREB1, evidencing the complexity of having a dysregulated *MECP2* gene [15,16]. Thus, MeCP2 can alter the chromatin architecture and both activate and repress gene transcription depending on its molecular context. *MECP2* also interacts with microRNA processing machinery and with RNA splicing machinery and is thought to participate in both processes [17,18]. Recently, it was published that rather than being a global splicing regulator, MeCP2 may regulate splicing of only specific transcripts [19].

Post-translational modifications (PTMs) such as phosphorylation, SUMOylation or acetylation occur in MeCP2 and impact MeCP2's functions and its interactions with other proteins [20–22]. Besides, mouse models without specific phosphorylations in Mecp2's amino acids showed traits already seen in Rett syndrome murine models. For example, a mouse model with no phosphorylated Ser80 has an impaired locomotor activity and weight gain [23], and disruption of Ser421 phosphorylation leads to defects in synaptic and dendritic development *in vivo* [24]. These findings support the idea that PTMs can act as fine tune regulators of MeCP2 function [20]. These PTMs of Mecp2 have been further validated and studied with phosphorylation-defective

Mecp2 mouse models, and the importance of those phosphorylations for neuronal and synaptic development have been proved [25].

MeCP2 is ubiquitously expressed but it is more abundant in the brain, especially in neuronal cells. In the human brain, MeCP2's expression is first detected during mid embryogenesis and increases progressively during the postnatal period until 10 years of age. The increase in MeCP2 levels coincides with the stage of maturation of neurons, dendritic spine morphogenesis and development of synapse plasticity [26]. That implies that MeCP2 is important not only in early prenatal stages for neuronal development but also in later stages for neuronal maturation and maintenance [3,27,28] [Figure 2].

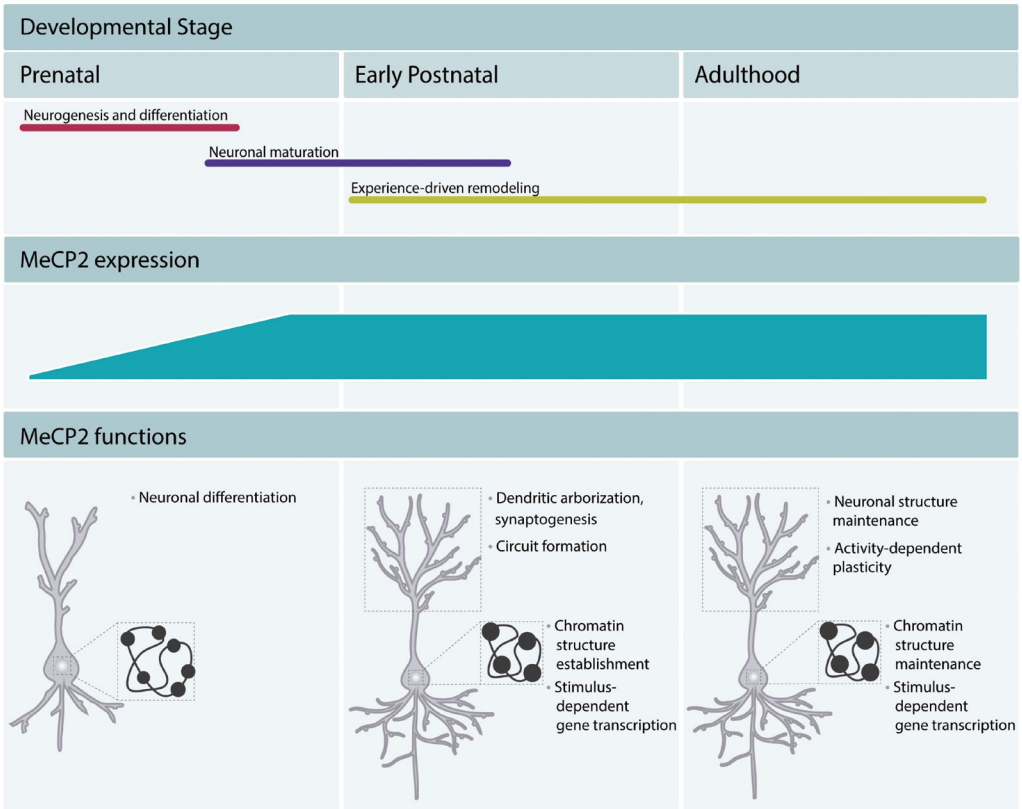

Figure 2: MeCP2 expression levels in different developmental stages, from embryogenesis until adulthood. MeCP2 regulates functions for brain development and maintains a correct functioning in mature neurons. In early embryonic development MeCP2 regulates neuronal differentiation while in postnatal stages, it regulates neuronal maturation and circuit formation. In adulthood, MeCP2 becomes critical for maintenance of neuronal structure and function. It maintains the chromatin structure, regulates the neuronal transcriptomic profile and the cognitive function. Modified from Gulmez Karaka et al. 2019 [26].

MeCP2 is also expressed outside the central nervous system (CNS) in the lung and spleen, but fewer studies have been carried in peripheral tissues. Ross et al. generated a Mecp2 knockout mouse in which Mecp2 was absent from peripheral tissues but normally expressed in CNS. As a result, the mice presented a pronounced exercise fatigue and defective bone properties [29].

2. Genetic alterations in *MECP2*

Genetic alterations in *MECP2* are associated with a variety of clinical phenotypes [Figure 3]. The most studied and the most common disorder among females is Rett syndrome (RTT; OMIM #312750) with its classic and atypical forms. Among males, *MECP2* duplication syndrome (MDS; OMIM #300260) is the most common syndrome, followed by severe neonatal encephalopathy (OMIM #300673). X-linked mental retardation (XLMR) syndrome (OMIM #300055) and autism susceptibility (OMIM #300496) have been reported in both females and males.

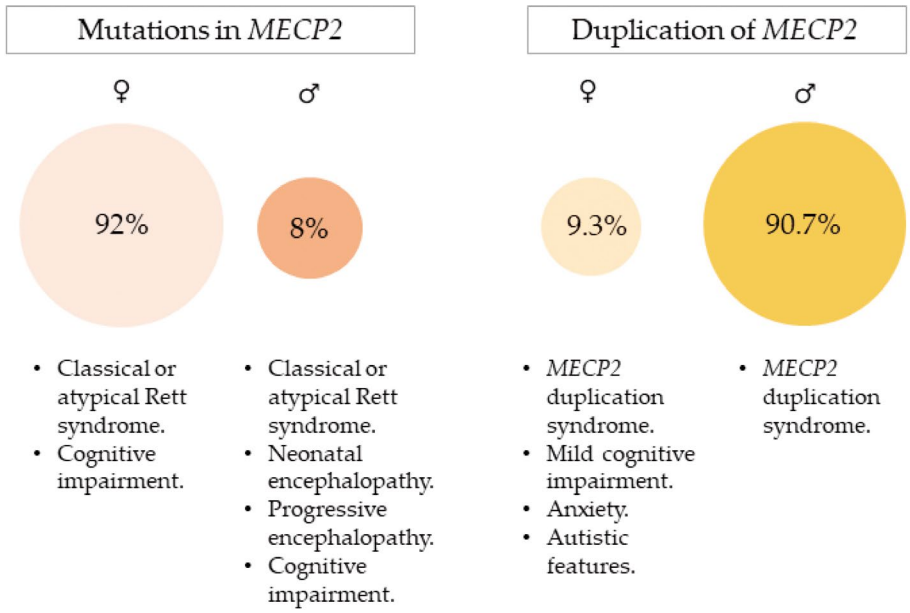

Figure 3: Different phenotypes that can be seen due to genetic alterations in *MECP2* in females and males. Sources of information are RettBASE for mutations in *MECP2* and 82 articles from PubMed containing the term "MECP2 duplication" for the patients with the duplication.

2.1. Rett syndrome

RTT is a neurodevelopmental disorder with early onset that affects mainly females. It is the second cause of intellectual disability (ID) in females after Down syndrome (OMIM #190685) and has an incidence of 1:10.000-15.000 born girls.

The vast majority of the RTT cases are caused by mutations in *MECP2*. RTT has a dominant X-linked inheritance, and while most of the pathogenic variants within *MECP2* in girls lead to RTT, boys can develop a broad spectrum of phenotypes.

In 1954 the paediatrician Andreas Rett spotted two girls with the same stereotyped hand movements in the waiting room of his clinic. Driven by curiosity, he realised that both girls had a similar development and clinical histories. In 1966, after further research, Andreas Rett published a report describing 22 girls with an uncomplicated birth history and a disruption in development around 1 year of age, an age at which girls exhibited motor dysfunction, speech loss, apraxic gait, epilepsy and hand stereotypies. That first publication remained unrecognised until in 1983 a European group of researchers published new cases, expanded the phenotype and renamed that disorder as Rett syndrome [30]. In 1999, *MECP2* was linked to RTT by Zoghbi's group [14] and since then, few thousands of cases have been published all over the world.

2.1.1. Clinical presentation

RTT is a well-characterized syndrome with approved and revised diagnostic criteria since the 1990s [31–34]. Alterations in *MECP2* are not always present in RTT patients and they are neither sufficient to make a firm diagnosis of RTT, that is what makes clinical criteria crucial for a proper RTT diagnosis [34]. These diagnostic criteria classify patients into the classic or typical RTT form, or in the atypical form [Table 1]:

- **Classic or typical RTT**. It needs the presence of a regression in the psychomotor development of the child, together with four main criteria. The main criteria appear in the first years of life, permitting a diagnosis when the child is from 2 to 5 years old. The supportive criteria can also be seen in classic RTT patients. No exclusion criteria should be present in the child since they could be related to other causes of neurological diseases. The majority of the classic RTT cases are females and have a mutation in *MECP2*, but as stated before, exceptions have been published.

- **Atypical RTT**. Patients must manifest at least 2 main criteria and at least 5 supportive criteria. Patients with atypical RTT can be classified in subgroups or variants according to the onset of the symptoms and the genes that are mutated. Boys diagnosed with some variants of the atypical RTT form are as frequent as females.

Table 1: Revised diagnostic criteria for typical and atypical RTT. Modified from Neul et al. 2010 [34].

RTT Diagnostic criteria 2010

Required for typical or classic RTT:

1. A period of regression followed by recovery or stabilisation.
2. All main criteria and no exclusion criteria.
3. Supportive criteria are not required, although they are often present in typical RTT.

Required for atypical or variant RTT:

1. At least 2 out of the 4 main criteria.
2. 5 out of the 11 supportive criteria.

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 (dyspraxic) or absence of ability.
4. Stereotypic hand movements such as hand wringing/squeezing, clapping/tapping, mouthing and washing/rubbing automatisms.

Exclusion criteria for typical RTT:

1. Brain injury secondary to trauma (peri-or postnatally), neurometabolic disease or severe infection that causes neurological problems.
2. Grossly abnormal psychomotor development in the first 6 months of life.

Supportive criteria for atypical RTT:

1. Breathing disturbances when awake.
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”.

Early RTT cohorts had a low rate of survival beyond the third decade of life [35]. Fortunately, improved symptom management has remarkably extended RTT patients' survival, and in present cohorts, 70% of patients reach 50 years of age [36]. Among the most recurrent causes of death are respiratory tract infection, respiratory failure, cardiac instability and seizure related illnesses [37,38].

Three severity scales had been created to measure RTT patients: Pineda's, Kerr's and Percy's severity scales [39–41]. Pineda's scale possesses the advantage that the scoring system is independent of the patient's age, and therefore, is the one that the Spanish Rett syndrome working group uses. There, nine clinical features are evaluated and scored according to the presence, the age at acquisition and the loss of them [Annex 1]. Pineda's score has been applied to classic RTT children as well as to atypical RTT children.

2.1.1.1. Classic RTT

A distinctive trait of classic RTT is that the disease progresses with a specific cadence. Disease progression in these children can be divided into four stages [42,43] [Figura 4]:

1. The first stage shows an apparent normal prenatal and perinatal development and covers the first months of life until 6 to 18 months of age. At the beginning, it was believed that during the first months of age RTT infants presented a normal psychomotor development, but later evidence showed otherwise. Parents' statements and subsequent objective evaluations of patients and mouse models during the first two years of life have demonstrated that behavioural and neurological alterations already occur during that time frame [42]. Gross motor performance starts to be impaired and the acquisition of the abilities to crawl, sit or walk alone are delayed. These infants also present few communicative gestures, linguistic impairments, stereotypic patterns of hands and arm tremors within the first couple of months of life. Altogether, RTT infants start to present mild alterations that might not be detected or raise suspicions of RTT until the second stage is reached.

2. Between 6 and 18 months of age, children with RTT enter the regression stage and psychomotor development is arrested leading to a partial or complete loss of previously acquired skills. Some of the impaired abilities include loss of spoken language, loss of hand use which is replaced with hand stereotypies, and loss of ambulation or abnormal

gait development. This stage is highly variable and can last from months to a few years.

3. After the regression phase, RTT children enter a plateau stage in which stabilisation of behavioural and cognitive functions happens. In some cases, a partial skill recovery has even been reported. However, other medical issues appear, such as breathing abnormalities, seizures and gastrointestinal problems. Epilepsy, for example, is present in around 60-80% of the patients and can be controlled in the majority of the cases by combining antiepileptic drugs [44,45]. Children with RTT can stay in the plateau stage for years.

4. Finally, around 10 years old, RTT patients enter the last stage. At that point, patients experience a progressive motor deterioration that affects mobility and motor performance and in which parkinsonian features and scoliosis appear. In some cases, ambulation can be completely lost. With adequate medical attention, women with RTT can live past their fifties. Unfortunately, around a quarter of the patients suffer a sudden death due to cardiac instability, respiratory infection or respiratory failure [46].

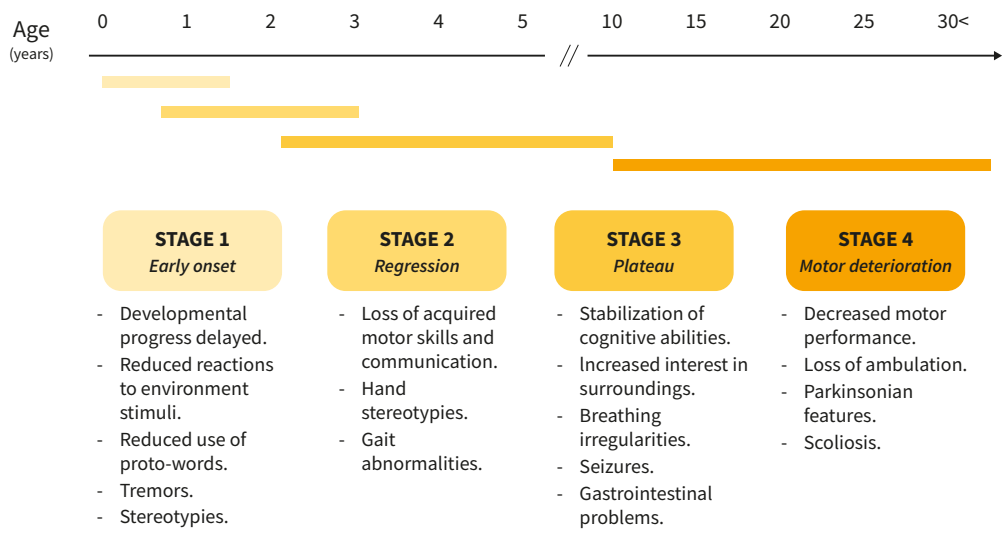

Figure 4: Stages of classic RTT disease evolution and the phenotypic features that are present during them. Note that variability exists in the duration of the stages between patients.

2.1.1.2. Atypical RTT and Rett-like

Patients that have a phenotypic resemblance to classic RTT but do not fulfil all the diagnostic criteria are generally grouped as atypical RTT. In depth characterizations of these atypical RTT children have shown that some of them can also be re-classified into two more specific groups or RTT variants. The remaining children that possess some clinical characteristics of RTT but cannot be further grouped into any variant are referred to as Rett-like patients:

- **Early onset seizure variant.** Seizures begin within the first 3 months of life. Severe gross motor delay, hand stereotypies, gastrointestinal problems and sleep problems are among frequent features [47]. The mutated gene is usually *CDKL5* (OMIM *300203), a gene also located in the X chromosome.

- **Congenital variant.** The neurodevelopmental impairment is evident soon after birth. These children present impaired or absent ambulation, lack of speech, hypotonia, spasticity, sleep problems and hand stereotypies [48]. The mutated gene is usually *FOXG1* (OMIM *164874) which is located in chromosome 14 and therefore, it affects both females and males.

- **Rett-like.** This category comprises children presenting a different combination of distinct features related to RTT. Several genes have been found mutated, some of which have already been described in other neurodevelopmental syndromes with overlapping phenotypes [49,50].

2.1.2. Genetic characterization

The vast majority of the RTT cases are *de novo* and only less than 1% are inherited from an asymptomatic carrier mother with a completely skewed XCI. Approximately 98% of typical RTT cases and 86% of atypical RTT cases are caused by mutations in *MECP2* [51]. There are around 750 RTT causing pathogenic variants located in *MECP2* reported in RettBASE [6]. In fact, 46% of the RTT cases are caused by eight recurrent mutations (R106W, R133C, T158M, R168X, R255X, R270X, R294X, R306C) that are located in the MBD and TRD domains [Figure 5]. Mutations in *MECP2* were first reported in 1999 in female and male patients with RTT [14,52]. Since then, genetic alterations that range from single nucleotide variants (SNVs) to large deletions have been reported in *MECP2* in RTT patients. In fact, large deletions in *MECP2* account for 8-15% of the alterations found in RTT patients [53]. Exon 4 has a well characterized deletion prone region (DPR) [54]. The repetitive structure

of DPR causes genomic instability and leads to both smaller deletions inside exon 4 or to larger deletions that can encompass other exons or even adjacent genes such as *interleukin-1 receptor-associated kinase 1* (*IRAK1*). Besides, half of the sequence of intron 2 consists of interspersed repeats such as Alu repetitive elements. Interaction between Alu elements and the DPR generates some of the large rearrangements found in *MECP2* [54]. Small duplications spanning from one to seven nucleotides have also been reported in RettBASE. The majority of the pathogenic variants causing RTT are located within *MECP2*'s exons 3 and 4. A few variants causing RTT have been reported in exon 1 but no one in exon 2. From all the published cases with classic RTT, only few are males and they are mosaics for the mutation or have Klinefelter syndrome (47, XXY karyotype) [55–58]. Males not fulfilling these two criteria, usually develop other phenotypes.

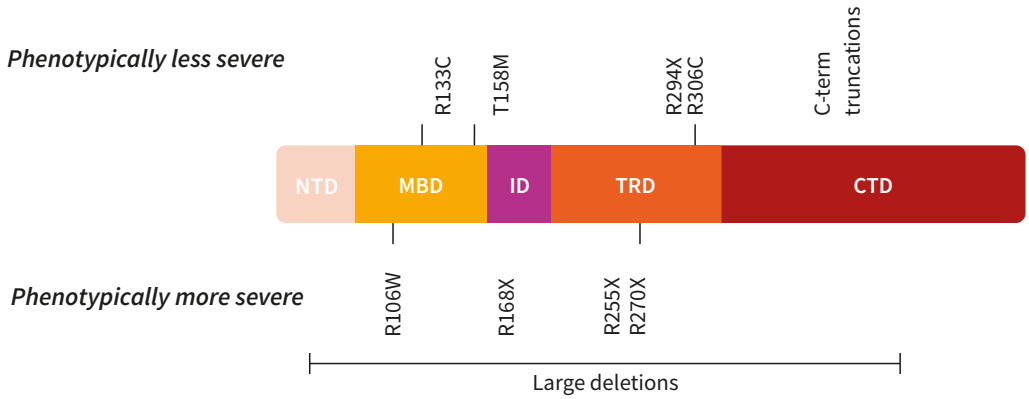

Figure 5: Schematic representation of the position of the 8 recurrent mutations that lead to RTT. They have been divided according to the existing evidence that associates them with more or less severe phenotypes. Mutations are annotated according to NP_004983.1.

2.1.3. Genotype-phenotype correlation

Exhaustive clinical and molecular studies have been performed to properly characterize RTT children (both with classic and atypical forms). Establishing a genotype-phenotype correlation could permit paediatricians to foresee the evolution of the disease and to give parents a more personalised prognosis. Until date, no definitive genotype-phenotype correlation has been found for RTT. An important reason for it is that there is a high variability in the clinical

phenotype among RTT children, even between individuals sharing the same mutation. Several contributors have been suggested to explain it:

- XCI is a phenomenon that occurs in every cell of the female body in order to compensate for gene dosage. During early female embryogenesis, each cell is supposed to inactivate one of the X chromosomes stochastically, but sometimes, the inactivation is aimed at a specific chromosome and a skewed XCI occurs. A completely skewed XCI of the chromosome carrying the mutation causing RTT could lead to the expression of the wild type (wt) allele and thus, to a healthier phenotype. Therefore, it is hypothesised that depending on the inactivation percentage of the mutated allele and the tissues in which the XCI is skewed, the phenotype will be more or less severe. Unfortunately, the mainly affected tissue in RTT is the brain, and getting XCI ratios from there is complicated. XCI of more accessible tissues is conducted, but it has been demonstrated that the XCI ratio is unique for each tissue, so no clear correlation has been obtained between XCI and RTT phenotype severity [59].

- The impact of different *MECP2* mutations on RTT phenotype has been reported and a tendency to severity has been obtained. Mutations R133C, R294X, R306C, exon 1 and C-terminal truncations have been related to milder phenotypes, whereas R106W, R168X, R255X, R270X, splice sites and large deletions have been related to more severe phenotypes [60] [Figure 5]. Other groups have also reported that mutations towards the C-terminal of *MECP2* lead to milder phenotypes. What's more, RTT patients with mutations in R306C and C-terminal truncations are more likely to retain the use of language and ambulation [61].

- Genetic modifiers, such as epigenetic factors, could explain the variability found between patients with the same mutations or even between familiar cases by influencing *MECP2* directly or indirectly [62]. Genetic screens looking for secondary mutations that ameliorate the RTT phenotype have been conducted in RTT mouse models and some positive results have been obtained [63].

- Environmental factors can also modulate RTT phenotype. Environmental enrichment applied since early stages of RTT has shown improvements in gross motor abilities in RTT mouse models and patients [64,65]. What's more, improvements were observed in children with different capacities at baseline, and abilities such as sitting or walking were achieved. Besides,

BDNF levels, that are diminished in the RTT population, were threefold increased in the serum of those RTT patients. *BDNF* (OMIM *113505) is needed for proper neuronal development and synapse function, and increased levels of *BDNF* have shown improvements in behavioural features in RTT patients and mouse models.

2.1.4. RTT models

The acquisition of target tissue samples of children affected by neurodevelopmental disorders as RTT is complicated. Animal and cell models have been created to expand the current knowledge about RTT and *MECP2* and they have proved to be very valuable tools. In particular, RTT mouse models are the major source for findings related to *MECP2*'s function in RTT, although in recent years, human derived models are starting to be widely studied.

The first two murine models, a female and a male *Mecp2*-null mice, were generated by Bird's lab and by Jaenisch's lab in the early 2000s [66,67]. The female mouse model is heterozygous for the null allele. They are born normal and as they get older they gain stiffness, uncoordinated gait, breathing difficulties and hindlimb clasping. Males are hemizygous and exhibit a similar disease development but with a faster regression [66,67]. Being mainly neurological features altered in RTT, a mouse model with the null allele only in the CNS was created, and these conditional mutant mice turned to be indistinguishable from the global knockout mice [67]. Jaenisch's lab also generated other conditional mice with the *Mecp2* deficiency restricted to neurons. That mouse model showed stiff gait and decreased activity, similar to what other null mouse models present.

To test whether *MECP2* deficiency is only needed for initial brain development or is required continuously in the developed brain, *Mecp2* was eliminated in adult mice using a tamoxifen-inducible *Mecp2* knockout strategy. Those adult mice started to exhibit hypoactivity, abnormal gait, hindlimb clasping, impaired memory and learning ability and premature death, and it was confirmed that *MECP2* is needed for brain development and maintenance [68]. Then, the reverse question was asked, if reactivation of *Mecp2* after symptom onset could reverse the observed phenotype. Both Bird's hemizygous male and heterozygous female *Mecp2* mutant mice showed an improved gait, hindlimb clasping, motor behaviour, breathing pattern and lifespan, together with neuronal morphology [69,70]. These experiments have provided the bases for a potential therapy.

A mouse model with *Mecp2* deficiency restricted to astrocytes was also created and showed breathing irregularities, but locomotion, anxiety behaviour and lifespan were normal [71]. Subsequent studies found that *Mecp2*-null astrocytes are toxic to neurons and can deteriorate neuronal functions due to abnormal glutamate metabolism, indicating that glial cells also play a role in RTT [72]. Besides, astrocyte specific *Mecp2* reactivation in global *Mecp2*-null mice rescued aberrant respiration, hypoactivity and decreased dendritic complexity together with increasing lifespan [71].

In addition, numerous mouse models have been engineered to answer different questions involving *MECP2*'s role in the pathophysiology of RTT. As a result, selective reduction and reactivation of *Mecp2* has been achieved in different brain regions, such as forebrain, hypothalamus and cerebellum, and cell types, for example GABAergic, glutamatergic or serotonergic neurons [43]. Mouse models harbouring *MECP2* mutations seen in patients, like the 8 common *MECP2* point mutations, have been created and studied as well.

Alternative models as primary cell cultures of patients' peripheral tissues, human embryonic stem cells (hESCs), reprogrammed human-induced pluripotent stem cells from somatic cells (hiPSCs) or even region-specific brain organoids have also been created [73–76]. 3D brain organoid models have been engineered to recapitulate neurogenesis, neuronal differentiation and maturation in RTT patient derived cells. Among others, reduced cell migration and proliferation, impairment in dendritic complexity, decreased synaptogenesis, decrease of VGLUT1, a specific protein expressed in glutamatergic neurons, and impairments in calcium signalling have been observed in RTT brain organoids [76]. These results prove 3D models to be a perfectly valid technology not only to gain more insight into the pathology itself, but also to discover and test drugs before entering clinical trials [74]. Trujillo et al. tested several pharmacological compounds in a 2D hiPSC derived neuronal culture first, and the most promising compounds were later tested in 3D RTT cortical organoids. The improvements detected in the organoid upon treatment with the two compounds were confirmed by RNA sequencing (RNAseq) analysis, which showed an improvement in the expression of genes related to synaptic function and neurotransmitter markers [77].

At brain level, models with MeCP2 deficiency have revealed alterations in neuronal differentiation and maturation, axonal and dendritic growth and morphology, synaptic formation and function and neurotransmission itself [78,79]. Besides, RTT mice and human neurons have smaller soma size, impaired neuronal maturation, impaired dendritic spine density and dendritic branching complexity,

fewer synapses, and an increased neuronal density [76]. Altered calcium signalling and an excitatory/inhibitory imbalance have also been seen [43,80]. Altogether, dysregulated cellular excitability together with deficits in neuronal functions and synaptic plasticity can affect higher-order cognitive processes such as different types of learning and memory, explaining part of the RTT phenotype [81].

2.1.5. Treatments and therapies

There is no cure for RTT so medical management is mainly symptomatic. Seizures are controlled with antiepileptic drugs and anxiety with selective serotonin reuptake inhibitors. Nutritional management is also common and a ketogenic diet is recommended to children with refractory epilepsy. Ketogenic diet has demonstrated to prevent seizures, and it even slightly improves behaviour, social abilities and motor skills in RTT children [82]. Diet recommendations to avoid constipation and other gastrointestinal problems are also considered. Rehabilitation through different therapies (e.g. physiotherapy for movement and strength, or speech and language therapy for communication) is also crucial in order to delay the last stage of the syndrome and improve these children's quality of life.

The obtained information about the altered pathomechanisms in RTT has led to the beginning of some clinical trials and the first treatment for RTT has just been approved:

- IGF-1 (OMIM *147440) is downregulated in RTT and is needed for brain development and synaptic regulation. IGF-1 administration to RTT mouse models and hiPSCs and hESC-derived neurons with different *MECP2* mutations increases the number of glutamatergic synapses, rescuing soma size and dendritic complexity deficits through inducing AKT/mTOR pathway [83]. A clinical trial with Mecasermin or rhIGF-1 showed safety and tolerability in RTT girls and improved breathing and behavioural abnormalities in phase 1 (NCT01253317). However, in phase 2 (NCT01777542), no significant improvements were registered among the patients [84]. Another analog of IGF-1, Trofinetide (NNZ-2566) also underwent clinical studies. Phases 2 and 3 for Trofinetide (NCT02715115, NCT04181723) revealed safety, tolerability and an improvement in breathing problems, abnormal behaviours and seizure frequency [85,86]. Recently, the FDA has approved Trofinetide for RTT

treatment under the name of DAYBUE and it will be available in the US by the end of April 2023.

- Another important synaptic regulator is *BDNF* whose expression is diminished in RTT patients the moment the phenotypic abnormalities are evident [87]. Two clinical trials with *BDNF*-boosters reached phase 2. The clinical trial with Fingolimod (NCT02061137) did not show a significant improvement at phase 2 [88], whereas glatiramer acetate (NCT02153723) improved gait velocity, memory and breath holding, and overall, a trend towards an ameliorated quality of life was seen, although it was not significant [89]. In a posterior trial, multiple adverse reactions were observed and its safety has been reconsidered [90].

- The administration of cannabidiol analogue cannabidivarin demonstrated to reduce seizure frequency, improved memory and upregulated BDNF, IGF-1 and cannabinoid receptor expression in RTT models. First phases of the clinical trial have shown safety, tolerability and improved seizure control [91] and currently, phase 3 results are being evaluated (NCT03848832).

- Sarizotan, an agonist of the serotonergic 5HT1A receptor that binds dopamine and serotonin neurotransmitters, also entered clinical trials, but phases 2 and 3 (NCT02790034) did not show significant efficacy [76].

- Other treatments that focus on normalising secondary changes caused by a deficient MeCP2 were tested too. Omega-3 fatty acids were used to address damaging effects of the oxidative stress. Mitochondrial dysfunction was targeted (NCT02696044) as well as altered cholesterol metabolism (NCT02563860), but none of those trials entered phase 3.

Ideally, replacement of the mutant *MECP2* allele would be the best treatment for RTT, and thus, gene therapy approaches have been developed for RTT as well [Figure 6]. Adeno-associated viral (AAV) vectors have proven to be good vehicles for the CNS due to its capacity to cross the blood-brain barrier. Systemic delivery of AAV9 bearing *Mecp2* rescued several behavioural and neuronal aspects of RTT male and female mice. The same study noted that intracranial injection of the vector improved some symptoms but also caused parkinsonism in the mice, possibly, due to local overexpression of MeCP2 [92].

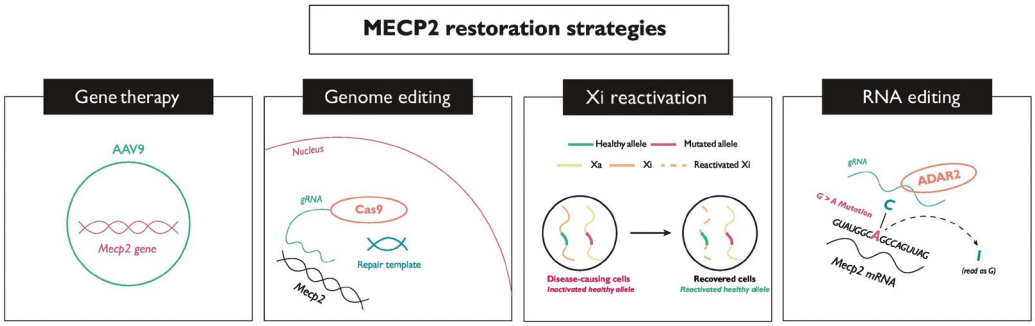

Figure 6: Strategies to restore *MECP2*. Multiple strategies to increase the amount of functional MeCP2 are being developed. Among them, MeCP2 delivery platforms, editing approaches or XCI reactivation can be highlighted due to a wider applicability. Modified from Grimm et al. 2022 [93].

Better vectors that evenly distribute the cargo throughout the brain need to be developed to avoid side effects and toxicity. Moreover, neuronal heterogeneity caused by XCI complicates successfully targeting only the cells containing the dysfunctional MeCP2 and not altering the cells with the wt allele, since overexpression of MeCP2 is detrimental too.

Site-directed gene repair using CRISPR-Cas9 could be an alternative [Figure 6]. Huong Le et al. were able to introduce the *MECP2* mutation R270X in iPSCs with CRISPR-Cas9 and then, to correct the mutation with the same technology with around 25% of efficacy [94]. Proven that CRISPR-Cas9 can repair the mutated *MECP2* allele, the next step would be to prove its validity reversing RTT traits. Future improvements in the CRISPR-Cas9 technology efficiency and safety will indeed enable fruitful therapeutic opportunities for RTT.

As XCI status seems to be the reason why some females are asymptomatic, X chromosome reactivation has been proposed as a therapeutic tool [Figure 6]. That approach aims to reactivate the wt *MECP2* allele in female neurons to restore the functional MeCP2 protein levels. It has been seen that reactivation of the inactive *MECP2* in iPSC neurons corrects defects of RTT neurons such as soma size [95]. Notwithstanding, the effect of reactivating other X-linked genes when reactivating *MECP2* should be studied. Besides, the amounts of *MECP2* that need to be reactivated are still unknown [93].

Finally, a strategy to edit RNA has been applied to a RTT mouse model with a 317G>A mutation [Figure 6]. G>A transitions are present in 30% of the RTT

patients [93]. The G>A was corrected in 50% of the transcripts by injecting an AAV with an adenosine deaminase enzyme in the mice hippocampus [96]. Unfortunately, the used enzyme, ADAR2, can generate adverse immune responses, so some improvements must be done.

2.2. Males with mutations in *MECP2*

Male individuals with mutations in *MECP2* show a varied scenario. Having only a few cases of males with mutations in *MECP2* reported, together with the difficulty of creating a male mouse model suggested that mutations in *MECP2* could be lethal in boys. Nowadays, we know that alterations in *MECP2* can cause from mild intellectual and psychomotor impairment (as happens with NM_004992 c.608C>T) [97] to severe neonatal encephalopathies and premature death (as with NM_004992 c.806delG) [98]) [Figure 3]. Having such a wide spectrum of phenotypes in male patients complicates giving an accurate diagnosis, which is not always reached.

2.2.1. Clinical presentation

From the studies carried by Ravn et al., Huppke and Gärtner, Villard and Neul et al. four groups have been specified to classify males with point mutations, small duplications or deletions in *MECP2* [57,58,98,99]:

- **Boys with classical or atypical RTT:** When consensus diagnostic criteria for each form are fulfilled. These boys are mosaics or have Klinefelter syndrome (47, XXY), and thus, a percentage of the wt allele is expressed as happens in females. Usually, the same mutations cause RTT in girls as well. Since the clinical evolution of males with classic RTT is more severe than observed in females, the term "male RTT encephalopathy" has been recommended [99].

- **Boys with neonatal encephalopathy:** When impairment of the clinical traits happens from birth. These boys present a very severe encephalopathy, seizures, frequent respiratory and feeding problems, do not develop ambulation and usually die young. The boys have normal chromosomal complements (46, XY). Those mutations cause RTT in girls.

- **Boys with progressive encephalopathy:** When impairment of the clinical traits appears with the years. In these children, hand use is limited, communication is impaired and an abnormal gait and muscle tone has been reported. Some cases also develop seizures.

- **Boys with cognitive impairment and/or psychiatric manifestations:** When no progressive worsening is detected. These boys present a milder phenotype with cognitive impairment, behavioural problems, poor hand use and limited motor skills, although some boys gain ambulation. Same mutations in girls could not affect them, cause mild ID or have not been reported. Some of these mutations are inherited from asymptomatic carrier mothers.

2.2.2. Genetic characterization

Males with mutations in *MECP2* are thought to be underdiagnosed. The broad phenotype described for these boys does not necessarily make paediatricians think of *MECP2,* and thus, the gene was not often screened. Fortunately, the implementation of next generation sequencing (NGS) as a routine test has brought more cases to light. However, variant classification for *MECP2* in males without the RTT clinical traits is not always straightforward and *in silico* studies are not always sufficient to classify a variant as pathogenic or benign. For the variants of unknown significance (VUS) performing functional studies is recommended. The results could tilt the balance towards a pathogenic or benign variant and enable a proper diagnosis.

2.2.3. Genotype-phenotype correlation

A genotype-phenotype correlation is not easy to achieve with these boys either since few detailed cases have been published. Neul et al. gathered a group of 30 males with mutations in *MECP2* and reported that males with early mutations (before codon 271, based on NP_004983.1) had a higher score in the RTT Clinical Severity Scale (CSS) compared to the ones with later mutations [99]. Detailed clinical and molecular characterization of more patients is needed to enlarge the current databases and facilitate future variant interpretations and phenotype-genotype correlations.

2.2.4. Models

Although no animal model has been designed for boys with mutations in *MECP2*, they can benefit from the mouse models engineered for RTT, as most of them are male. XCI makes the phenotype of female mice more variable and researchers prefer to work with simplified models [100]. Most of the mouse models created are null for *MECP2* or have mutations described in RTT patients, but the results obtained from these animal models and the subsequent therapies may benefit both female and male patients.

2.3. *MECP2* duplication syndrome

Contrary to what happens in RTT, MDS affects mainly males, while females with MDS can be asymptomatic, present mild neuropsychiatric traits, learning difficulties or the same severe phenotype as MDS males develop [Figura 3]. The duplication can happen *de novo* but it is frequently inherited from a carrier mother that is apparently healthy and whose XCI is skewed. The frequency of MDS varies depending on the studied cohort, ranging from 1% in a screening of males with XLMR to 2.6% in a cohort with multiple congenital anomalies and ID [101,102]. To date, more than 600 MDS cases have been reported worldwide.

2.3.1. Clinical presentation

The core phenotype of MDS includes developmental delay, moderate to severe ID, hypotonia, epilepsy, recurrent infections, gastrointestinal problems, progressive lower extremity spasticity, poor speech development and autistic features.

- Hypotonia, present in almost all the children, is thought to contribute to the delay in acquiring developmental milestones such as crawling and sitting. Most of the MDS females and a few males achieve ambulation without support, but the rest of the males usually develop an ataxic gait that causes lumbar hyperlordosis as a compensation mechanism [103]. Motor regression can happen as they get older, which causes progressive loss of ambulation [103–105].

- Seizures appear in these children with age and are one of the causes that triggers regression. The percentage of MDS children with seizures is believed to be underestimated, since the onset varies from childhood to the second decade of life [103]. Developing drug-resistant epilepsy is, unfortunately, common [104]. 32 different anti-seizure medications have been given to MDS children with variable effectiveness [106].

- Recurrent infections are, together with epilepsy and the difficulties in the gross and fine motor skills, the main cause of concern for parents [107]. Infections affect mainly the respiratory tract, tend to require hospitalizations and are responsible for premature death in MDS patients. Vaccination against pneumococci and periodical evaluation of post-vaccination antibody levels combined with extra boosters have been recommended to these patients [103,108,109]. Nevertheless, there are some patients in which a reduction in the frequency of infections has been noted with age [110].

- Feeding difficulties due to hypotonia can be detected shortly after birth. Among the gastrointestinal problems severe constipation is the most frequent one, but gastro-esophageal reflux, swallowing dysfunction and excessive drooling have also been reported. Besides, excessive drooling contributes to having more respiratory infections [111,112].

- Hand stereotypies have been seen in MDS patients. In RTT hand stereotypies replace purposeful hand use. However, in MDS, hand stereotypies still allow hand usage whenever it has been learnt [13,104].

- Some MDS patients have been described to possess reduced pain sensitivity as well [104].

- MDS patients present poor expressive language skills, impaired social affect, repetitive behaviours and anxiety, characteristics that can mislead to an autism spectrum disorder diagnosis if considered alone. However, a considerable proportion of MDS patients are able to interact and are described as social kids, even though speech is not always achieved and the ones who speak a few words tend to lose that ability [104].

- Dysmorphic features have been listed for MDS. Some of them are brachycephaly, deep-set eyes, strabismus, midface hypoplasia, a small open mouth, thick lower lip, large prominent ears, prominent nasal bridge, pointed nose, prominent chin, thick and dense hair, teeth anomalies and

tapered fingers [104,112]. Dysmorphic features are slightly different in males and females and are more subtle in females.

Even if some of the mentioned clinical aspects have been reported to improve in some patients, overall, the syndrome worsens with age, especially when epilepsy appears. As a result, Peters et al. and Takeguchi et al.'s cohorts showed that their older MDS patients had more severe clinical symptoms [112,113]. The survival of these children has not been closely monitored. Van Esch revised clinical information from 88 published MDS male patients and found that around 40% died before the age of 25 [103]. On the contrary, a French cohort of 59 patients studied around 10 years later than Van Esch's cohort, reported that 15% died before the age of 25 years old [104]. Earlier diagnosis and a more personalised medical monitoring may be improving these children's life expectancy and quality of life.

2.3.2. Genetic characterization

The duplications can be *de novo* or inherited from asymptomatic carrier mothers, being the second option more frequent. The carrier mothers are asymptomatic although some authors have reported that few mothers present neuropsychiatric features such as anxiety or depression [114]. It is thought that a skewed XCI in brain tissues is the reason why these carriers are asymptomatic and can live a normal life.

The minimum duplicated region contains the genes *MECP2* and *IRAK1* (OMIM *300283), but the reported MDS duplications range from 0.079 Mb to 15.8 Mb, involving different combinations of genes [115,116]. Thus, *MECP2* duplications are non-recurrent and vary between families. Usually, the duplications are stable and the same gene content is duplicated inside each family, but Yi et al. saw that sometimes, duplications can increase or decrease in size when transmitted from mothers to children, making each duplication unique [117]. Xq28 is a region with a high GC content and an elevated number of repetitive elements such as Alu or low copy repeats (LCR). Those elements increase the genomic instability and make the region prone to suffer rearrangements such as duplications [118]. What's more, triplications of *MECP2* have also been reported in males [119–123].

MDS causing duplications have been localised in different genomic positions. Duplicated segments can be in the same Xq28 region in tandem, on the short arm of the X chromosome, translocated to the Y chromosome or translocated to any other autosome [115,124] [Figure 7].

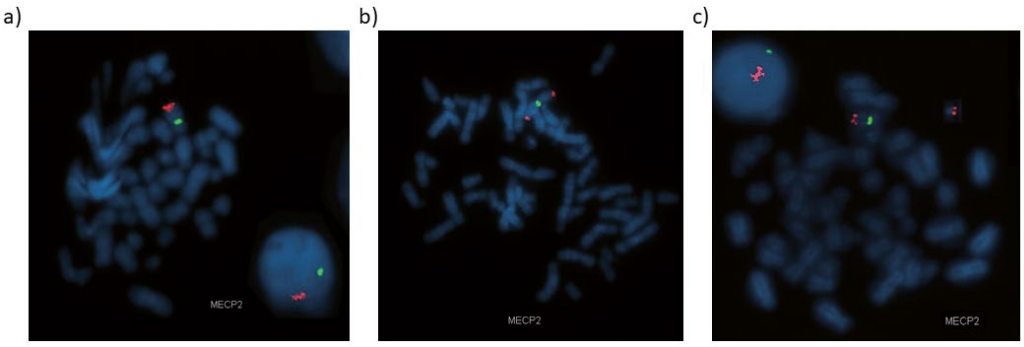

Figure 7: The different chromosomal positions observed for the *MECP2* duplication in patients with MDS. a) The *MECP2* duplication is located in tandem in the Xq28 region. b) The *MECP2* duplication is located in the Xp region. c) The *MECP2* duplication is translocated outside the X chromosome; in this case, it is located in the Y chromosome. All three images correspond to male MDS patients. The green signal shows the centromere of the X chromosome, whereas the red signal shows the *MECP2* gene.

2.3.3. Genotype-phenotype correlation

Several attempts have been done to discover a genotype-phenotype correlation, although the lack of large cohorts and long-term studies have made it a challenge.

Gene dosage has been reported to contribute to the severity in MDS, since children bearing a triplication have a more severe phenotype [119–121,125].

The location and size of the duplications have also been studied, but no significant correlations have been found [104,117,126]. It was thought that in females, having the duplication translocated outside the X chromosome would generate a more severe phenotype since that duplicated segment would escape XCI and be constitutionally active. However, the reported girls with translocations into autosomes present completely different clinical courses [127].

It has been hypothesised that the gene content of the duplication can explain the differences in the phenotype among children with MDS [Figure 8]:

- ***IRAK1***, which is part of the minimal duplicated region, is located in the Xq28 region right next to *MECP2*. *IRAK1* is composed of 14 exons, is ubiquitously expressed and forms homodimers or heterodimers with other IRAK members. *IRAK1* plays a key role in two signalling cascades: Toll-like receptors (TLRs) and interleukin-1 receptors (IL-1Rs). TLRs are

pathogen recognition receptors that initiate the response to infections. IL-1R and its family members are cytokine receptors that initiate and modulate the inflammatory and immune responses. In addition, the protein IRAK1 participates in interferon induction and gene transcription of inflammatory target genes [128]. Thus, a dysregulated *IRAK1* could be contributing to the recurrent infections MDS patients present.

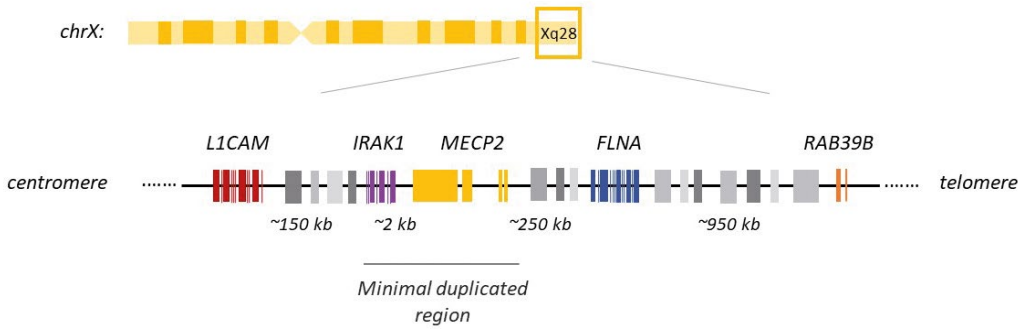

Figure 8: Relative position of genes that are often part of the duplication and have been studied for genotype-phenotype correlation in MDS. The approximate distance between the genes has been added. The minimal duplicated region comprises the genes *MECP2* and *IRAK1*.

- ***FLNA*** (OMIM *300017) has been reported to be associated with intestinal pseudo-obstruction problems and could be explaining part of the gastrointestinal dysfunctions of MDS patients [129]. However, patients with gastrointestinal problems and without *FLNA* duplication have also been published.

- ***L1CAM*** (OMIM *308840) has been associated with hypoplasia of the corpus callosum, which has been seen in MDS brain imaging [130,131]. Unfortunately, brain imaging results in MDS are scarce and larger cohorts should be studied to confirm this genotype-phenotype correlation.

- Duplications of ***RAB39B*** (OMIM *300774) are related to ID. In a study conducted by Peters et al. they saw that MDS patients with duplicated *RAB39B* had higher CSS and a more severe phenotype [132]. The CSS they used was designed for RTT and, even though MDS and RTT share some clinical traits, various core traits of MDS are missed by that scale. A MDS severity score would be needed in order to confirm these findings.

Sometimes, relatives with MDS that share the duplication, such as brothers or cousins, present a different phenotype. Those cases complicate the finding of a correlation. Interestingly, in monozygotic twins, somatic mutations have aided explaining the phenotypic differences between them [133]. Disruptions of specific genes or regulatory regions due to the molecular rearrangements causing the duplications cannot be discarded either. Anyway, further studies are still needed to decipher the current challenge of establishing a genotype–phenotype correlation in MDS, especially when cohorts tend to be small.

2.3.4. MDS model

Mouse models for MDS have been created to overcome the same challenges as for RTT. Transgenic mice overexpressing mouse and human *MECP2* genes at variable levels (*MECP2*-TG1 two-fold and *MECP2*-TG3 from 3-5 fold) or only human *MECP2* (hDup) have been engineered [134,135]. Those mice exhibit a progressive neurological disorder with hypoactivity, autistic features, anxiety, seizures and enhanced motor learning. A mouse model with *MECP2* overexpression restricted to neurons (*Tau-Mecp2* mice) was also generated [136]. *Tau-Mecp2* mice show an impaired motor coordination, anxiety-like behaviour, impaired learning ability and memory and altered synaptic plasticity. Recently, a new mouse model that contains the duplication in tandem of not only *Mecp2* but also *Irak1* has been generated (*Mecp2 Dup*) [137]. This animal model considers the minimal duplicated region of all MDS patients and exhibits a neurological phenotype like the other models but also abnormal immune responses to infections, a critical trait among MDS patients that has not been reported in other MDS models. In addition, transgenic rats and monkeys that overexpress MeCP2 have also been created to better study autism behaviour in MDS [138,139].

Primary cultures of patients' fibroblasts, whole blood and hESCs or hiPSCs reprogrammed from patients' somatic cells have also been successfully established and used for further studies [140,141].

While neurons without MeCP2 present a decreased number of excitatory synapses and impaired long-term potentiation (LTP), neurons overexpressing MeCP2 have an increased density of excitatory synapses (such as the glutamatergic synapses), increased synaptogenesis and enhanced LTP [136,142]. LTP and long-term depression (LTD) models can induce axonal bouton formation and elimination, events that occur continuously in the adult sensory cortex as part of the learning

and memory formation processes. In MDS mice, the bouton elimination rate does not change during motor training while it significantly decreases in wt littermates, explaining the enhanced motor learning abilities observed in MDS mouse models. Strikingly, Jiang et al. found that in L5 pyramidal neurons of the cortex, the spine density is significantly higher during the first 12 weeks but then, falls below control levels [142]. Altogether, the synaptic plasticity process has been demonstrated to be disrupted in the MDS mouse models too.

2.3.5. Therapies

No cure has been found for MDS and there are no registered ongoing clinical trials with MDS patients either. However, several treatment strategies have been tested in MDS models. Pharmacologic inhibitors for ERK pathway reverse molecular and behavioural changes in MDS animal models, suggesting that Ras-ERK signalling hyperactivity contributes to the MDS phenotype [143]. A histone deacetylase inhibitor, NCH-51, could also be a potential therapeutic candidate since it reverted neuronal morphology of MDS patient derived iPSCs [140].

Theoretically, treating MDS should be easier than RTT because the treatment should focus on reducing the levels of *MECP2* with no need of correcting it. Anyways, getting the right quantity of MeCP2 is still the main challenge. Antisense oligonucleotides (ASOs) have arisen as a potential tool to normalise MeCP2's expression. ASOs bind target RNA molecules and by silencing them, regulate subsequent protein expression [Figure 9]. Administration of ASOs has been performed in MDS mouse models and MDS patient derived cells, and reversing of molecular, behavioural and electrophysiological defects have been seen [144]. Notably, a phenotypic improvement in *MECP2*-TG1 mice after ASO administration was achieved after 4 weeks, and the ASO effects disappeared after several weeks. The same group engineered the hDup humanised MDS mouse model and their results showed that CNS administration of *MECP2*-ASO is tolerated and improves some behavioural traits 9 weeks after ASO administration [135]. Shao et al. performed RNAseq to analyse the response of the *MECP2* regulated genes to the ASO treatment. The altered transcriptomic profile of hDup mice was rescued 2 weeks after the administration of the ASOs and it could last from 5 to 16 weeks after injection. The observed time window requires this kind of molecular tracking to register molecular changes and allow fine tuning of the MeCP2 levels, as well as to maintain and prolong the optimal ASO levels and benefit with an improved phenotype.

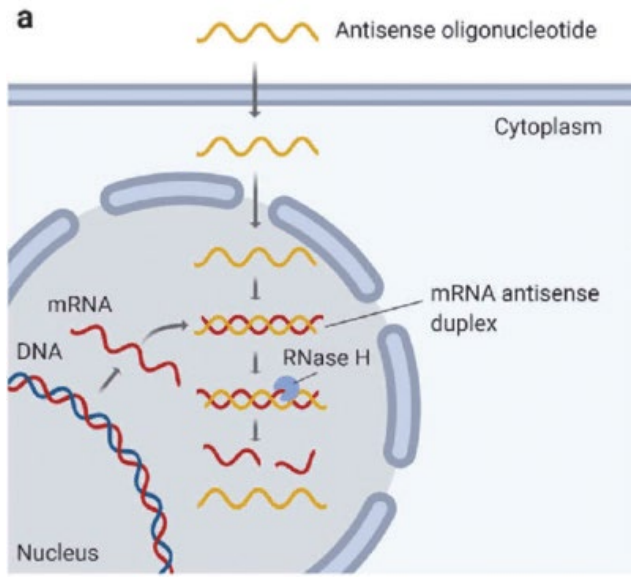

Figure 9: One of the mechanisms by which antisense oligonucleotides can alter gene expression. ASO intervention can occur in the nucleus or in the cytoplasm. However, the ASO designed for MDS hybridises into intron 2 of *MECP2* and therefore acts in the nucleus. The typical mechanism of action of an ASO is that after entering the nucleus, it binds to the complementary mRNA. The resulting mRNA-ASO duplex is recognised by RNase H, which cleaves the mRNA and prevents protein translation. The image is modified from Tromp et al. 2020 [145].

MeCP2 overexpression has been shown to affect neurotransmission in the brain, resulting in a reduced probability of neurotransmitter release [136]. Na et al. attempted to rebalance neurotransmission in *Tau-Mecp2* mice using a GABAA receptor antagonist. Treatment with low doses of the antagonist improved motor coordination and synaptic plasticity in mice [146].

CRISPR-Cas has also been applied to MDS mouse models and human cells. Wotjal et al. were able to correct a duplication spanning *MECP2* and *IRAK1* in a MDS patient derived human fibroblasts [141]. Qiu's group successfully used CRISPR-Cas9 in *MECP2*-TG1 mice's medial prefrontal cortex and hippocampal CA1 region, and showed improvements in social behaviour [147,148].

3. Omic technology: transcriptomics and proteomics

As *MECP2* is a transcriptional regulator, its downstream target genes have been studied extensively. Until the first decade of this century, the limitations of experimental technology meant that selected genes could only be studied one at a time to confirm whether they were MeCP2 targets. Important progress was made, but it was a time-consuming approach. The introduction of NGS, with its ability to multiplex DNA samples and genes, made it possible to interrogate up to thousands of genes simultaneously. The same technology was soon adapted to sequence RNA molecules.

With the rapid technological improvements of the last 15 years, omic technologies have advanced tremendously. Apart from the main omics, which are genomics, transcriptomics and proteomics, other omics are also developing rapidly. For the purposes of this work, however, we will focus only on transcriptomics and proteomics.

3.1. Transcriptomics

The resulting RNAseq technology has been applied not only to sequence the molecule to search for alterations, but also to quantify the relative amounts of transcripts. This allows differential expression, alternative splicing and monoallelic expression analysis to be performed. In general, the main steps for a short-read RNA sequencing performed on the Illumina sequencing platform (the most popular platform for short-read RNAseq) include RNA purification, fragmentation, cDNA synthesis, ligation of the adaptors, PCR amplification, sequencing and data analysis [Figure 10]. As a result, mRNA fragments of maximum 200 bp are obtained and sequenced to get an average of 30 million reads per sample. Afterwards, the data is computationally processed to determine the fraction of reads associated with each gene or transcript and finally, it is statistically analysed.

To overcome the limitation of correctly mapping highly variable transcript isoforms, long-read RNAseq technology has been created by sequencing platforms such as Oxford Nanopore or Pacific Biosciences. A long-read RNAseq protocol starts with full-length cDNA synthesis by using template switching reverse transcription, followed by an optional step of PCR amplification and a subsequent adaptor binding in order to create the library, which is the input for

the sequencing. These long-reads can sequence up to 15 kb. Oxford Nanopore adapted their sequencing technology to sequence directly RNA without the need of synthesising cDNA and PCR amplification, and thus, avoiding biases caused by these steps and retaining epigenetic information [149]. Long-read RNAseq technologies present two major limitations compared to short-read RNAseq: a lower throughput, which puts genes expressed at medium or low levels in disadvantage, and a higher rate of sequencing errors [149]. The previous three RNAseq technologies (short-read, long-read and direct long-read RNAseq) are performed with RNA obtained from several cells and is referred to as "bulk RNAseq". However, bulk RNAseq cannot easily study specific cell types and spatial information gets lost. Luckily, single-cell RNA sequencing (scRNAseq) and spatially resolved RNAseq methods have been developed.

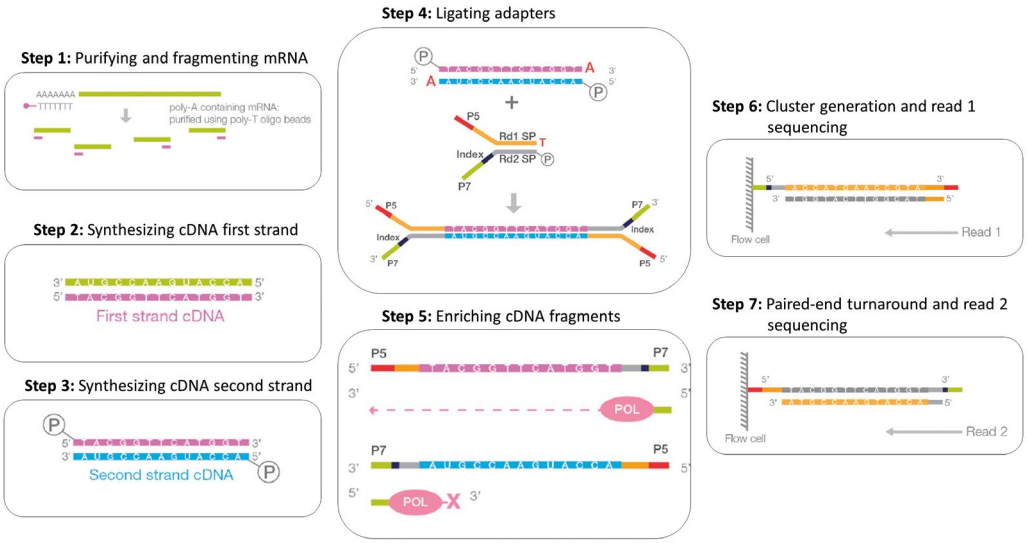

Figure 10: Workflow for a short-read mRNA library preparation for RNA sequencing. Depending on the used technology some discrepancies can be found with the present protocol. Modified from the TruSeq Stranded mRNA reference guide from Illumina company (San Diego, CA, USA).

scRNAseq methodology parts from solid tissues that are mechanically and enzymatically dissociated. Then, the cells are separated, normally with a flow cytometer, and lysed so as the RNA is released and reverse transcription can be done. The cDNA is generally amplified by PCR as part of the library preparation steps. scRNAseq can be performed to generate short or long reads,

with the same advantages and disadvantages mentioned above. Apart from the transcript length, two other properties should be considered when performing scRNAseq: the breadth (the amount of profiled cells) and the depth (the quantity of transcripts per cell) of the experiment [149]. Usually, scRNAseq experiments are viewed in terms of two dimensions, since, as the number of profiled cells rises, the depth of transcriptome profiling decreases. Given the circumstance, the choice of which methodology should be used is left to the researcher, and will mainly depend on the research question.

3.2. Proteomics

Studying proteomics is becoming a powerful technique to further understand the functioning of the cells. For example, proteomic can be used to track changes in the levels of a particular protein or its PTMs. Quantification in proteomic is based in mass spectrometry (MS) analysis, which detects ions and measures its mass-to-charge ratio (m/z). Therefore, the resolution and sensitivity of a mass spectrometer depends on the mass analyser's ability to separate ions. Coupling tandem mass analysers in a single mass spectrometer allows two or more sequential ion separations, improving the technique's accuracy. The term proteomic encompasses different approaches that need to be combined to obtain the most adequate data for the biological question [150]:

- Proteomic can be global or targeted. The first approach is ideal to discover new biomarkers but the second, is more sensitive, reproducible and faster.

- Proteomic can be addressed using a peptide-centric (or bottom-up) approach, or a protein-centric (or top-down) approach. The peptide-centric strategy permits identifying and quantifying thousands of proteins, although it eliminates the connections between peptides and proteins and leads to less accurate protein inference whenever two or more proteins share the same peptide. Protein-centric approach is more used for quantification of modified and unmodified proteins.

- Proteomic quantification methods can be label-free or label-based. Labelling-based methods can introduce chemical, metabolic or enzymatic labels. Tandem mass tags (TMTs) are a type of chemical-labels that consist of an amine-reactive, a balance and a reporter group that are released after fragmentation emitting an intensity that is captured and used to calculate

relative peptide amounts between samples. TMT labelling enables multiplexing up to 16 samples in the same run [Figure 11].

- Quantification in proteomic can be either relative or absolute. Relative quantification measures relative changes in a protein by comparing the amount of protein between samples. Absolute quantification determines the amounts of a peptide or protein within a sample.

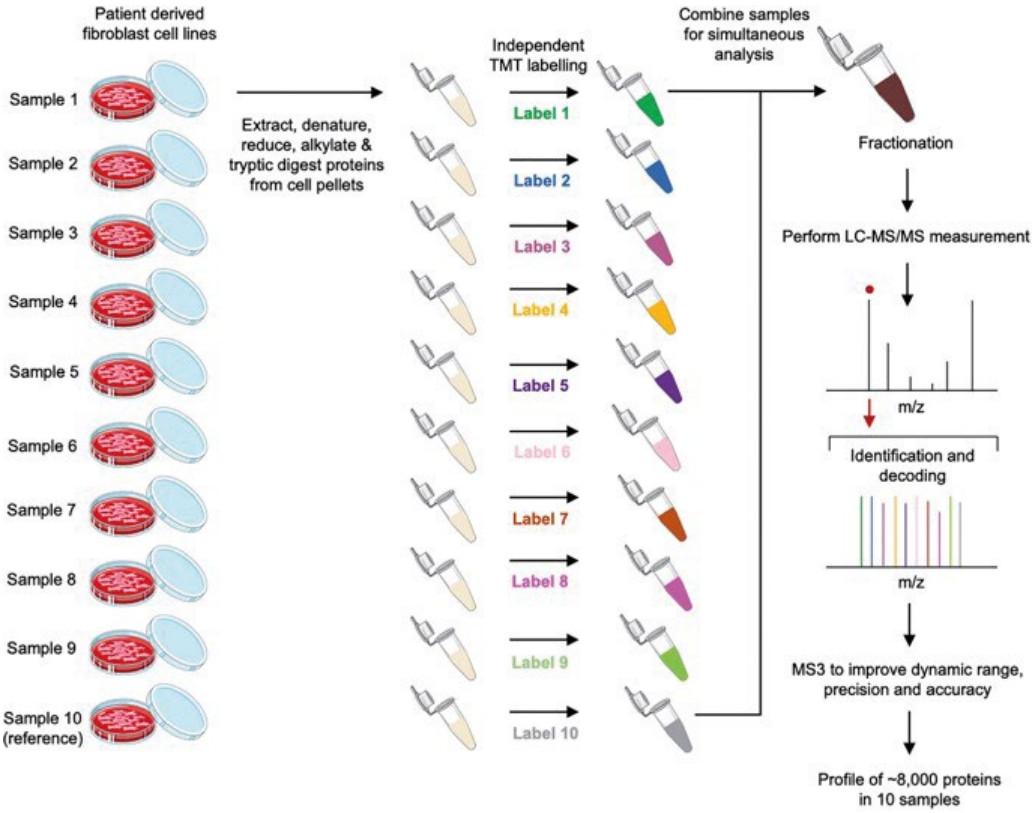

Figure 11: Schematic representation of a multiplexed proteomic experiment labelled with TMT. The multiplexing of samples reduces missing values and improves quantitative precision together with statistical analysis. Modified from Stenton et al. 2019 [151].

Similar to what happens with transcriptome studies, the use of bulk proteomics can lead to misleading results due to the heterogeneity of the cells that comprise the analysed sample. Single-cell proteomic can be the solution to

this issue. As with bulk proteome profiling, label-free and label-based methods have their advantages and limitations. On the other hand, the additional steps during sample preparation and the need to combine samples in larger wells for the analysis generate additional surface exposure that can result in cell loss [152]. Although single-cell proteomic technology is still in its dawn, it is advancing very rapidly and will become a widely used tool such as scRNAseq.

3.3. Analysis pipelines

The utility of omics is a fact, and several data analysis tools have been developed in recent years. However, using different combinations of tools for analysis can lead to significant differences in the biological conclusions drawn from the data. The optimal set of tools to use depends on the specific biological question being addressed. For RNAseq analysis, at least the following four steps should be included in a pipeline [149]:

- *Alignment and assembly*. After sequencing, the data files containing the sequenced reads, usually FASTQ files, are collected. These reads are mapped to known transcriptomes or annotated genomes to obtain the genomic coordinates of the reads. Among the alignment tools, TopHat, STAR or HISAT are commonly used. They use a reference genome as a template, and since cDNA-derived reads lack introns, these tools perform a spliced alignment, allowing for gaps in the reads. Different mapping tools differ in how they designate the reads. These effects are particularly noticeable for multi-mapped reads, which could come from more than one distinct gene, pseudo-gene or transcript.

- *Quantification of transcript abundance*. Once mapped, the reads are assigned to genes or transcripts to determine their abundance. Some of the more common quantification tools are CuffLinks, HTSeq, MMSeq and RSEM. In addition, there is a group of tools termed "alignment-free" tools that directly assign reads to transcripts and quantify them. Some of these alignment-free tools are Kallisto, Salmon and Sailfish. They have shown good performance in characterizing highly abundant and longer transcripts, but are less accurate for less abundant or short transcripts [153]. These alignment-free tools tend to include the multi-mapped reads in their transcript count estimates, whereas other quantification tools discard them.

- *Filtering and normalisation.* Genes with low read abundance can be removed to improve the detection of true differential expression [149]. Normalisation of the expression matrix can be done using several approaches, such as RPKM, quartile or median normalisation methods.

- *Differential expression modelling*. Once the expression matrix is ready, the experimental data can be modelled to determine which genes or transcripts are likely to have changed expression levels. Tools that model read counts for gene-level expression use generalised linear models. This is the case for edgeR, DESeq2 and limma+voom, which provide comparable results and are computationally efficient. Tools that model differential isoform expression, such as CuffDiff, MMSEQ and Ballgown, produce more variable results and require more computational power [149].

On the other hand, the analysis of proteomic experiments requires the following steps [154]:

- *Peptide and protein identification.* There are two strategies to identify the peptides: searching in a fragmentation spectra database and *de novo* peptide sequencing. Fragmentation spectra databases contain all the resulting protein sequences generated from an *in silico* digestion of the expressed proteins. For each fragmentation spectra obtained from the MS experiment, a peptide spectrum match (PSM) score is calculated. The highest PSM score indicates the best candidate for the peptide. Some of the frequently used databases are Andromeda, Mascot or SEQUEST. In the *de novo* peptide sequencing, the peptide sequence is determined from fragmentation spectra information and fragmentation method properties. The analysis strategy has been run with PepNovo and NovoHMM, but new algorithms are emerging to overcome the current limitations. Hybrid approaches are also being developed. Once the peptides are identified, the original protein sequence is reconstructed. For this protein inference step, degenerate peptides, the ones that can be part of several proteins, should be addressed with caution and again, several tools based on different probabilistic models exist.

- *Protein abundance quantification.* Several software packages have been developed to quantify all kinds of proteomics experiments. Softwares such as IsobariQ, MaxQuant or PVIEW can quantify label-based experiments, whereas MaxLFQ or VIPER are designed for label-free proteomic experiments.

- *Data processing and normalisation.* It is common practice to log transform intensities prior to normalisation [154]. Normalisation removes non-biological variation and makes the results reliable and consistent across samples, which is critical for downstream analysis. There are several types of normalisation methods based on different statistical assumptions.

- *Statistical analysis of quantitative protein data.* The most straightforward statistical analysis to answer the question of which proteins have significantly altered levels between two groups is to perform a t-test. However, due to the limited multiplexity in proteomic, sample sizes are relatively small and the statistical power of the test is compromised. The LIMMA model has demonstrated more robust and accurate results than t-tests on small datasets, but other methods have also been developed [155]. Finally, an FDR threshold has to be applied to effectively reduce the number of false positives in the statistical analysis.

3.4. Current transcriptomic and proteomic results for RTT and MDS

More than two decades of research have led to tremendous advances in the understanding of MeCP2 functions. Transcriptomic profiling, both using microarrays or RNAseq, and proteomic profiling aim to quantify and compare gene or protein expression between affected and healthy samples to detect differentially expressed genes (DEGs) and proteins (DEPs). In addition to detecting individual genes and proteins, the identification of dysregulated functional pathways may provide biological insights that contribute to the disease process and could potentially be reversed.

To date, about 70 transcriptomic studies have been performed to study RTT, whereas only 6 of them also studied MDS. More than two-thirds of the RTT transcriptomic studies and 5 MDS studies were conducted in mouse models, mainly in male mice. The remaining studies used human derived tissues such as post-mortem brain, blood tissue or fibroblasts that were sometimes reprogrammed to generate neurons. Researchers have applied different strategies and technologies to decipher the transcriptional profile of RTT and MDS, and because of this lack of uniformity, each experiment has obtained an almost exclusive set of DEGs. However, the following biological processes have been consistently found to be dysregulated in typical RTT:

- Dysregulation of neuronal maturation and function, synaptic transmission and neurotransmitter imbalances have been widely reported, particularly in brain tissue [61,156]. These processes are mainly downregulated, which is consistent with the impaired neurological scenario of RTT.

- Mitochondrial dysfunction is present in the brains of RTT patients and mouse models, and the DEGs are usually upregulated. Reactive oxygen species (ROS) are also upregulated as a result of mitochondrial malfunction and are thought to be the major cause of oxidative stress reported in RTT patients and mouse models [156].

- Lipid metabolism is altered, particularly cholesterol metabolism. Cholesterol is an abundant lipid in the brain and is essential for synaptic plasticity, neurite outgrowth, synaptogenesis and myelination. Enzymes involved in cholesterol biosynthesis are downregulated [156].

- The cytoskeleton and cell adhesion are altered in several tissues and are usually downregulated. Dysregulation of genes involved in actin dynamics and cell adhesion have been reported to alter synaptic plasticity and dendritic spine morphology required for proper synaptic function [157] [Figure 12].

- The NFkB pathway has been found to be upregulated. NFkB regulates, among other things, neurite growth and dendritic complexity [156].

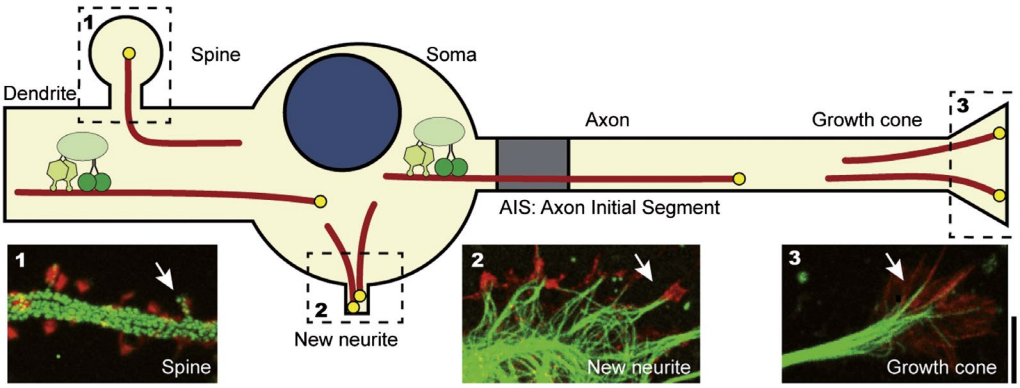

Figure 12: The role of the cytoskeleton in neurons. The cytoskeleton is involved in dendritic spine morphodynamics, neurite initiation and outgrowth, axon elongation and neuronal migration. Modified from Kapitein and Hoogenraad 2015 [158].

The only transcriptomic experiment aimed at revealing altered biological processes in an MDS mouse model was performed in 2009. They showed that neuronal development, synaptic transmission and neurotransmitter transport were altered and, in contrast to RTT mouse models, these processes were upregulated in MDS transgenic mice [16].

Proteomics experiments are scarcer. To date, about 30 publications have analysed the proteome of RTT in mice and human-derived samples, but there is no work on MDS. The data obtained supports the transcriptomic findings, and the following processes have been found to be altered in RTT [159]:

- Synaptic function, neuronal morphology and neurotransmission have been found to be altered [160].
- Mitochondrial organisation, biogenesis and stress defence have been reported to be altered in human fibroblasts. Proteins participating in these processes are upregulated [161].
- The protein network involved in cholesterol metabolism is altered in RTT patients and mouse models [162].
- The actin cytoskeleton and cell adhesion have been found to be upregulated [163] but specific proteins such as beta-actin are downregulated [164].
- Inflammation and immunity have also been reported to be altered in RTT, and the proteins involved are upregulated [160].

OBJECTIVES

The main objective of the present doctoral thesis is to study the expression levels of the RTT and MDS patient cohorts by using transcriptomic and proteomic approaches and to obtain a thorough picture of all the dysregulated genes and molecular processes of these two syndromes in order to look for biomarker and therapeutic target candidates.

The specific objectives are the following:

1. Characterization of new patients with alterations in the gene *MECP2*.

    1.1. RTT patients with large deletions.

    1.2. MDS patients and carrier mothers.

    1.3. Male patients with point mutations in *MECP2*.

2. Characterization of dysregulated molecular processes in RTT patients using transcriptomic and proteomic technologies.

3. Characterization of dysregulated molecular processes in MDS patients and carrier mothers using transcriptomic and proteomic technologies.

RESULTS

PhD thesis director's report

The doctoral thesis entitled "Study of expression levels in *MECP2* related disorders using transcriptomics and proteomics: characterizing Rett syndrome and *MECP2* duplication syndrome", presented by Ainhoa Pascual-Alonso, is a compilation of 6 publications, 3 of which have already been published, one is under revision, one was submitted for publication and one is in preparation. The PhD candidate, Ainhoa Pascual-Alonso, actively participated in the experimental part, data interpretation, bioinformatic analysis and writing of all the publications derived from this thesis. All the publications listed here are part of the core of this thesis and will not be used in other doctoral theses, with the exception of article 5, which will also be used in Clara Xiol's thesis, since both are listed as co-authors and have contributed equally.

Publication 1: *Characterization of large deletions of the MECP2 gene in Rett syndrome patients by gene dosage analysis.* Silvia Vidal, Ainhoa Pascual-Alonso, Marc Rabaza-Gairí, Edgar Gerotina, Núria Brandi, Paola Pacheco, Clara Xiol, Mercè Pineda, Rett working group and Judith Armstrong. **Molecular Genetics and Genomic Medicine**. 2019;e793. Impact Factor (2019): 1.92 (Q3).

The PhD student has been responsible for much of the experimental work (MLPA, Sanger sequencing, quantitative PCR, long PCR) and the cases spanning *IRAK1*. She has been involved in the interpretation of the results, in the phenotype-genotype correlation and in the writing of the manuscript.

Publication 2: *Molecular characterization of Spanish patients with MECP2 duplication syndrome.* Ainhoa Pascual-Alonso, Laura Blasco, Silvia Vidal, Esther Gean, Patricia Rubio, Mar O'Callaghan, Antonio F. Martínez-Monseny, Alba Aina Castells, Clara Xiol, Vicenç Català, Núria Brandi, Paola Pacheco, Carlota Ros, Miguel del Campo, Encarna Guillén, Salva Ibañez, María J. Sánchez, Pablo Lapunzina, Julián Nevado, Fernando Santos, Elisabet Lloveras, Juan D. Ortigoza-Escobar, María I. Tejada, Hiart Maortua, Francisco Martínez, Carmen Orellana, Mónica Roselló, María A. Mesas, María Obón, Alberto Plaja, Joaquín A. Fernández-Ramos, Eduardo Tizzano, Rosario Marín, José L. Peña-Segura, Soledad Alcántara and Judith Armstrong. **Clinical Genetics**. 2020;97:610–620. Impact Factor (2020): 4.438 (Q1).

The PhD student has been involved in the conduct of this study by collecting samples and patient data from all participating centres throughout Spain. The custom MLPA was designed by her. The genomic studies (custom MLPA, XCI)

and the establishment of primary fibroblast cell lines were also performed by the PhD student. Thus, the PhD student has participated in the collection, analysis, interpretation, genotype-phenotype correlation and preparation of the manuscript.

Publication 3: *MECP2-Related Disorders in Males.* Ainhoa Pascual-Alonso, Antonio F. Martinez-Monseny, Clara Xiol, and Judith Armstrong**. International Journal of Molecular Science**. 2021, 22, 9610. Impact Factor (2021): 6.208 (Q1).

The PhD student performed the clinical and bibliographic data collection for the writing of this review of *MECP2*-related disorders in males, in which we were invited to participate. She was also in charge of the drafting, actively participated in the elaboration of Table 1 and the design of Figure 1, and the critical discussion of the manuscript.

Publication 4: *Further study of males carrying variants in MECP2: beyond Rett syndrome.* Ainhoa Pascual-Alonso, Noelia Rodríguez, Mar O'Callaghan, Juan Darío Ortigoza and Judith Armstrong. **Manuscript in preparation.**

The PhD student has been responsible for the collection of clinical data and samples from patients and their carrier mothers, validation of the findings, analysis and interpretation of the results. She has actively participated in the elaboration of tables and figures and in the writing of the manuscript. The manuscript is pending review by all authors and submission for publication.

Publication 5: *Identification of molecular signatures and pathways involved in Rett syndrome using a multi-omics approach*. Ainhoa Pascual-Alonso*, Clara Xiol*, Dmitrii Smirnov, Robert Kopajtich, Holger Prokisch and Judith Armstrong. **Human Genomics. Under revision.**

The present publication has been a co-authored work: Ainhoa Pascual-Alonso studied *MECP2* duplication syndrome patients in relation to Rett syndrome patients, and Clara Xiol studied the Rett-like patients in relation to Rett syndrome patients. PhD candidate Ainhoa Pascual-Alonso has participated in the establishment of fibroblast primary cell lines derived from skin biopsies, in the DNA, RNA and protein extraction, XCI, RNAseq library preparation, bioinformatics analysis, data interpretation, figure and table creation and writing of the manuscript.

Publication 6: Molecular profiles of *MECP2* duplication syndrome patients and carriers. Ainhoa Pascual-Alonso, Clara Xiol, Dmitrii Smirnov, Robert Kopajtich, Holger Prokisch and Judith Armstrong. **Experimental and Molecular Medicine. Submitted.**

The PhD student has been responsible for all clinical data collection, patient selection, analytical and bioinformatics procedures, result interpretation, figure creation and has been in charge of writing the manuscript.

Barcelona, April 25th 2023

Judith Armstrong Morón

Bru Cormand Rifà

Chapter 1

Chapter 1: Characterization of new patients with alterations in the gene *MECP2*.

The first chapter of the thesis corresponds to the first objective which is the characterization of new patients with alterations in the gene *MECP2*. Four publications will be shown through the chapter, each of them focused in a specific cohort. The first publication targets RTT patients with large deletions; the second publication characterizes the Spanish MDS cohort; the third publication presents the current knowledge of males with *MECP2*-related disorders at the molecular and clinical level; and the fourth article studies seven cases of males with a different degree of developmental delay and ID harbouring variants of unknown significance in *MECP2*.

Obtaining a thorough characterization of a group of patients is essential, and especially in rare diseases, since the number of reported patients is scarce. The better a cohort is known, the easier it will be to merge different groups of patients and enlarge the sample size to gain statistical power. That could be incredibly useful to, for example, find a genotype-phenotype correlation that will lead to a more personalised genetic counselling. Providing an accurate diagnosis is crucial for the patient's wellbeing and their families, since a clearer idea of the disease's evolution and prognosis could be achieved. Therefore, the current step of patient characterization lays the groundwork for future research and will end benefiting the patients in multiple ways.

Publication 1: Characterization of large deletions of the *MECP2* gene in Rett syndrome patients by gene dosage analysis.

Authors: Silvia Vidal, Ainhoa Pascual-Alonso, Marc Rabaza-Gairí, Edgar Gerotina, Núria Brandi, Paola Pacheco, Clara Xiol, Mercè Pineda, Rett working group and Judith Armstrong.

Reference: Mol Genet Genomic Med. 2019; e793. https ://doi.org/10.1002/mgg3.793

Abstract:

Around 98% of typical RTT cases are caused by alterations in the gene *MECP2* and approximately 5% of them are large rearrangements that range from the deletion of part of an exon to the entire gene. For the present study 21 girls diagnosed with typical RTT and with a deletion in *MECP2* detected by MLPA were selected.

Breakpoints were narrowed down by qPCR until the amplification of the deleted allele was possible by long-PCR. With our approach we were able to confirm the deletions at nucleotide level in 14 patients. The remaining 7 patients' deletions were enclosed in two restricted regions of the gene that are prone to suffer rearrangements. In fact, all our cases support the current evidence claiming that most of the breakpoints within *MECP2* occur in two restricted regions: an area with repetitive elements in intron 2 and in the DPR of the exon 4. The gathered molecular information together with the clinical information of the patients enabled us to suggest a genotype-phenotype correlation that could help to improve the genetic counselling given to these patients.

Received: 24 August 2018 | Revised: 24 January 2019 | Accepted: 16 May 2019

DOI: 10.1002/mgg3.793

ORIGINAL ARTICLE

Molecular Genetics & Genomic Medicine Open Access WILEY

Characterization of large deletions of the *MECP2* gene in Rett syndrome patients by gene dosage analysis

Silvia Vidal[1,2] | Ainhoa Pascual-Alonso[1,2] | Marc Rabaza-Gairí[1,2] | Edgar Gerotina[1,2] | Nuria Brandi[3] | Paola Pacheco[4] | Clara Xiol[1,2] | Mercè Pineda[1] | Rett Working Group | Judith Armstrong[2,4,5]

[1]Sant Joan de Déu Research Foundation, Barcelona, Spain

[2]Sant Joan de Déu Research Institute (IRSJD), Hospital Sant Joan de Déu, Esplugues de Lobregat (Barcelona), Spain

[3]Facultat de Medicina, Universitat de Barcelona, Barcelona, Spain

[4]Molecular and Genetics Medicine Section, Hospital Sant Joan de Déu, Barcelona, Spain

[5]CIBER-ER (Biomedical Network Research Center for Rare Diseases), Instituto de Salud Carlos III, Madrid, Spain

Correspondence
Judith Armstrong, Molecular and Genetics Medicine Section, Hospital Sant Joan de Déu, Pg. Sant Joan de Déu 2, planta 0, 08950 Esplugues de Llobregat, Barcelona.
Email: jarmstrong@sjdhospitalbarcelona.org

Funding information
This work was supported by grants from the Spanish Ministry of Health (Instituto de Salud Carlos III/FEDER, PI15/01159), Catalan Association for Rett Syndrome, Fondobiorett and Mi Princesa Rett.

Abstract

Background: Rett syndrome (RTT) is a developmental disorder with an early onset and X-linked dominant inheritance pattern. It is first recognized in infancy and is seen almost always in girls, but it may be seen in boys on rare occasions. Typical RTT is caused by de novo mutations of the gene *MECP2* (OMIM*300005), and atypical forms of RTT can be caused by mutations of the *CDKL5* (OMIM*300203) and *FOXG1* (OMIM*164874) genes.

Methods: Approximately 5% of the mutations detected in *MECP2* are large rearrangements that range from exons to the entire gene. Here, we have characterized the deletions detected by multiplex ligation-dependent probe amplification (MLPA) in the gene *MECP2* of 21 RTT patients. Breakpoints were delineated by DNA-qPCR until the amplification of the deleted allele by long-PCR was possible.

Results: This methodology enabled us to characterize deletions ranging from 1,235 bp to 85 kb, confirming the partial or total deletion of the *MECP2* gene in all these patients. Additionally, our cases support the evidence claiming that most of these breakpoints occur in some restricted regions of the *MECP2* gene.

Conclusion: These molecular data together with the clinical information enable us to propose a genotype–phenotype correlation, which is essential for providing genetic counseling.

KEYWORDS
large deletions, *MECP2*, Phenotype-genotype correlations, Rett syndrome

1 | INTRODUCTION

Rett syndrome (RTT; OMIM#312750) is a neurodevelopmental disorder with early onset that is most often found in girls. It is first recognized in infancy; a period of apparently normal development (up to the age of 6–18 months) is followed by a stagnation-regression characterized by a loss of purposeful hand use and speech, motor apraxia that may be associated with epilepsy and dysautonomic features, including disturbed breathing, sleep, and gastrointestinal motility(Hagberg, Aicardi, Dias, & Ramos, 1983). RTT has a worldwide incidence of 1:10,000 live female births and

is the second leading cause of severe mental retardation in females.

Since 1999, numerous reports have supported the evidence that mutations in the *Methyl CpG binding protein 2* gene (*MECP2*; OMIM*300005) are the primary cause of classic RTT (Amir et al., 1999). MeCP2 is a transcriptional regulatory protein, and in its absence, a large number of genes exhibit abnormal expression with implications in the balance between synaptic excitation and inhibition (Kron et al., 2012).

The *MECP2* gene is localized in Xq28, contains four exons, and encodes two major functional domains namely: the methyl binding domain (MBD) (Nan, Meehan, & Bird, 1993) and the transcription repression domain (TRD), which contains a nuclear localization signal (NLS) (Singh, Saxena, Christodoulou, & Ravine, 2008). The *MECP2* translational initiation site was originally identified in exon 2, but a second translation initiation site was described in exon 1, which led to a new MeCP2 isoform (Mnatzakanian et al., 2004). MeCP2E1 is comprised of the exons 1, 3, and 4 while MeCP2E2 contains exons 2, 3, and 4; both forms comprise the MBD and TRD domains. MeCP2E1 is much more abundant in the brain while MeCP2E2 has a higher transcriptional expression level in the skeletal muscles, placenta, liver, and prostate gland (Liyanage & Rastegar, 2014).

No clear phenotype–genotype correlation has been identified in RTT patients (Bebbington et al., 2012; Neul et al., 2008; Scala et al., 2007). It has been reported that 95% of individuals affected by classic RTT have a loss of function in *MECP2*, but is less frequently seen in atypical RTT (Neul et al., 2010). There are eight common mutations of this gene that constitute approximately two-thirds of all mutations. Another small number of the patients carry a deletion ranging between 1 and 338 bp, in the C-terminal region (see RettBASE; http://mecp2.chw.edu.au/).

Soon after *MECP2* was identified as causative of RTT, several groups started to study the gene dosage of the cases that were negative for point mutations or small *indels* in the coding sequence of the gene. Thus, Southern Blot or MLPA followed by qPCR, long-PCR, and Sanger sequencing to narrow down the rearrangement, proved to be helpful in explaining approximately 10% of the mutations in those cases (Archer et al., 2006; Erlandson et al., 2003; Laccone et al., 2004; Ravn et al., 2005; Yaron et al., 2002). At our hospital, when taking into account all the cases diagnosed as RTT that have a mutation in *MECP2*, 4.5% of them have large rearrangements (Vidal et al., 2017), which is consistent with what has been reported in the literature (Hardwick et al., 2007).

Here, we present the molecular characterization of the breakpoints of the deletions detected in *MECP2* by MLPA in 21 RTT patients. The patients' clinical information was gathered as well, when available, in order to assess their severity with Pineda's score to determine a genotype–phenotype correlation and attempt to improve the genetic counseling for these and similar families.

2 | MATERIALS AND METHODS

2.1 | Patients and DNA samples

2.1.1 | Ethical compliance

Written informed consent was obtained from individuals legally responsible for the patients in accordance with appropriate ethics protocols for the analysis of genes related to RTT.

This study involved 21 patients clinically diagnosed with classic RTT who were negative for *MECP2* point mutations and small *indels* in the coding sequence. To evaluate the severity of the clinical presentation of each patient, a set of symptoms were measured using the clinical severity scores designed by Dr. Pineda (Monrós et al., 2001).

DNA was extracted from peripheral blood leukocytes using the Puregene DNA Isolation kit (Gentra System, Minneapolis, USA).

2.2 | MLPA analysis

All patients were analyzed by MLPA. MECP2-MLPA was performed with SALSA P015-D1, P015-E1 or P015-F1 kits (MRC-Holland, Amsterdam, The Netherlands) in accordance with the manufacturer's instructions. This assay covers all four *MECP2* exons and the flanking genes *IRAK1* (OMIM*300283), *L1CAM* (OMIM*308840), and *VAMP7* (OMIM*300053).

2.3 | Quantitative-PCR analysis (qPCR)

To narrow down the deletion breakpoints in each patient, we used real-time qPCR to test the relative copy number of various strategically designed amplicons located along the *MECP2* gene. Primers were designed from the genomic clone NM_004992.3 using Primer3 program (primer sequences and annealing sites in Supplementary Data S1). Briefly, our qPCR strategy was based on generating standard curves for each *MECP2* amplicon and for the autosomal reference gene *MTHFR* (OMIM*607093). These standard curves defined the relationship between the input DNA concentration and the C_t value.

The real-time qPCR was performed with the GoTaq Master Mix kit (Promega Corp., USA) for ABI 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA) and the PowerUp SYBR Green Master Mix kit for the QuantStudio 6 Flex Real-Time PCR System (both from Applied Biosystems, USA). All reactions were conducted in triplicate with the average of each triplicate group used for quantitative analysis. Product specificity was assessed by melting curve analysis.

TABLE 1 All the collected data from the patients; genomic information is based on the GRCh38/hg38 and the accession number NM_004992.3 (NG_007107.2) for *MECP2* gene

Patient ID	Clinical phenotype	Clinical severity score	Age of scoring, yr.	*MECP2* exons deleted	Breakpoint	Deletion Size	ICX
P1	Classic	8	11	4	c.1153_2387del	1,235 bp	72:28
P2	Classic	NA	NA	4	c.1157_3664del	2,508 bp	88:22
P3	Classic	11	10	4	c.1041_4447del	3,407 bp	64:36
P4	Classic	10[a]	17	4	c.1164_4665del	3,502 bp	87:13
P5	Classic	15	2	4	c.1164_1461+3282del	3,580 bp	58:42
P6	Classic	NA	NA	3 and 4	c.27-1125_1146del	3,001 bp	55:45
P7	Classic	NA	NA	3 and 4	c.27-5834_1166del	7,720 bp	58:42
P8	Classic	NA	NA	3 and 4	c.27-10677_1192del	12,599 bp	88:12
P9	Classic	NA	NA	3 and 4	c.(26+1_27-1)_(378_1385)del	≈ 11.5 kb	82:18
P10	Classic	17	10	3 and 4	c.27-7985_1209del	9,924 bp	51:49
P11	Classic	10	6	3 and 4	c.27-2950_1170del	4,850 bp	62:38
P12	Classic	13[a]	6	3 and 4	c.27-6312_1301delinsTG	8,343 bp	68:32
P13	Classic	14	4	3 and 4	c.(26+1_27-1)_(989_1241)del	≈19.15 kb	74:26
P14	Classic	NA	NA	3 and 4	c.27−15131_1461 + 7939del;insGGATCAGGT	25,261 bp	76:24
P15	Classic	13	8	3 and 4	c.(26+1_27-1)_(1461+4313_1461+5337)del	≈22.5 kb	83:17
P16	Classic	14	3	3, 4 and *IRAK1*	g.(154092184_154032557)_(154000172_153997133) del	≈ 50.7 kb	94:6
P17	Classic	10	10	3, 4 and *IRAK1*	g.(154092184_154032557)_(153986984_153969656) del	≈ 77.9 kb	88:12
P18	Classic	15	21	3, 4 and *IRAK1*	g.(154092184_154032557)_(154021813_154018990) del	≈ 28.3 kb	64:36
P19	Classic	15	16	3, 4 and *IRAK1*	c.[27-16409_1201delinsGGGGGCC; 1202_1460+2170inv; 1460+2171_1460+12766delinsTCTGCACGGGG]	18339bp and 10596bp	97:03
P20	Classic	14[a]	NA	4 and *IRAK1*	g.154030942_154003453del	27,589 bp	74:26
P21	Classic	17	NA	1 and 2	g.(154128954_154097731)_(154092184_154032557) del	≈ 85 kb	73:27

Abbreviation: NA, not available.

[a]Indicates the lack of information for some of the clinical features (See Supplementary Data S4); so the given score is the result of the data we were given by the clinicians.

The *MECP2* amplicon of interest and the *MTHFR* reference amplicon were amplified separately for each patient and for three normal female controls, yielding a copy number variant for each.

2.4 | Long-range PCR amplification and Sanger sequencing of deletion junctions

Once the deletions' breakpoints had been narrowed down to a sufficiently small region by qPCR, primer sites in the regions immediately flanking the breakpoints were selected for long-range PCR amplification. As the precise size of the junction fragment in each patient was unknown, several different PCR conditions were tested and optimized. Long-range PCR was performed with the Expand High Fidelity PCR System kit (Roche, Mannheim, Germany). This protocol was carried out in accordance with the manufacturer's instructions on a SimpliAmp Thermal cycler (Applied BioSystems, Waltham, MA). The PCR products were sequenced using a Big-Dye® Terminator version 3.1 Cycle Sequencing Kit in an Applied Biosystems 3,730/DNA Analyzer (Applied BioSystems, Waltham, MA). The raw data were analyzed with Chromas trace viewer (http://technelysium.com.au/wp/chromas/). The sequences of the junction fragments were aligned to the reference sequence of *MECP2* (NM_004992.3) using Genomatix diAlign® program (local multiple alignment; http://www.genomatix.de/cgi-bin/dialign/dialign.pl).

2.5 | X chromosome inactivation assay (XCI)

The XCI status of all 21 female patients was determined by the analysis of the methylation status of the highly polymorphic trinucleotide X-linked androgen receptor (*AR*; OMIM*313700) locus. For each subject, 50 ng of genomic DNA was digested separately with *Hpa*II restriction enzyme (New England Biolabs, Beverly, MA) in accordance with the manufacturer's instructions. A region between 252 and 327 bp of the locus was PCR amplified from digested and undigested DNA using fluorochrome-labeled primers. Samples were electrophoresed on an ABI Prism Genetic Analyzer 3130, and the peak areas were quantified using Gene Mapper v4.0 software (Applied Biosystems, Foster City, CA).

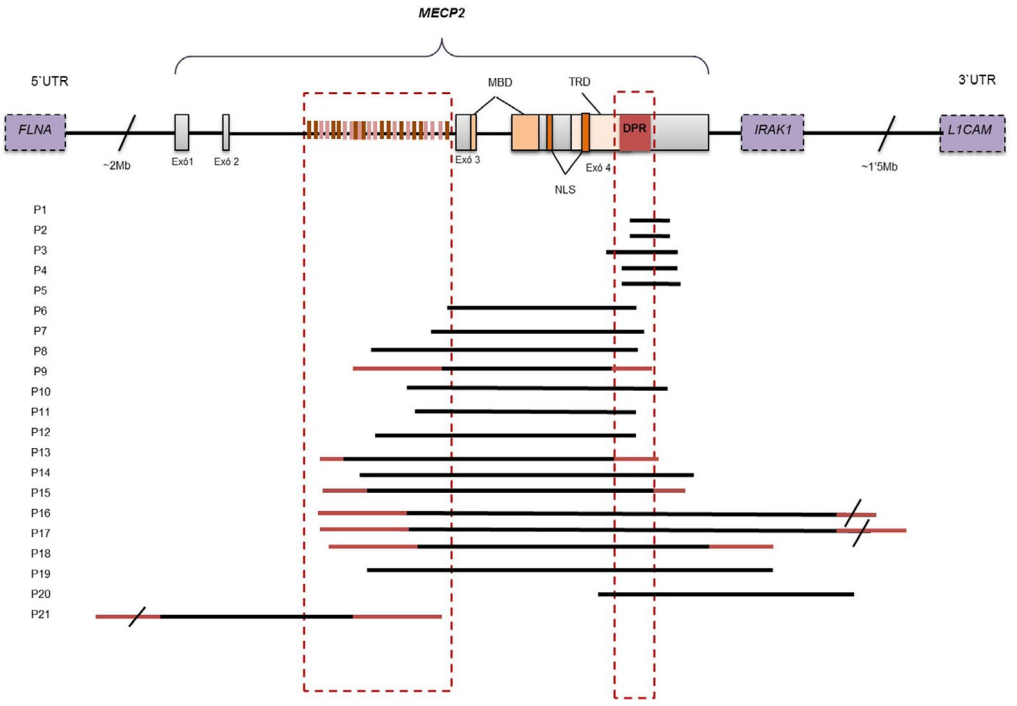

FIGURE 1 Position of the deletions in *MECP2*. Note that there are two regions prone to harbor a breakpoint. The black line indicates the region known with certainty to be deleted; the red line designates the region where only qPCR information is available

3 | RESULTS

For a total of 21 patients without a point mutation or small *indel* detected in the *MECP2* coding region, MLPA was carried out, and at least one exonic probe was missing in each patient (see Supplementary Data S2). These patients were classified depending on the affected exons: only exon 4 of *MECP2* was affected in five patients (P1 to P5), exons 3 and 4 in nine patients (P6 to P15), exons 3, 4, and *IRAK1* gene in four patients (P16 to P19), exon 4 and *IRAK1* in one patient (P20), and exons 1 and 2 in one patient (P21) (see Table 1). All deletions have been confirmed to be de novo.

To validate the MLPA technique, quantitative-PCR (qPCR) was performed in all patients with suspected deletions. qPCR analysis of the respective regions showed results compatible with deletion. Relative ratios of 0.5 ± 0.2 were suggestive of a deletion, whereas ratios of 1 ± 0.2 were indicative of a normal copy number for that region (for more information about the narrowing down of the deletion in each case, see Supplementary Data S3). Several PCR primer sets were evaluated to identify the ones that flank the deletion junction and could amplify, such that Sanger sequencing could be performed. For patients P1, P2, P3, P4, P5, P6, P7, P8, P10, P11, P12, P14, P19, and P20, different pairs of primers (Supplementary Data S1) successfully amplified the junction fragment that was subsequently sequenced (Figure 1). The deletions we have characterized range from 1,235 bp to 85 kb and involve different exons of *MECP2*, sometimes even ending in nearby genes. In case P19, we found a large inversion alongside the deletion.

With the patients' clinical information, we assigned each patient a severity score based on Pineda score. This scoring system gathers information about clinical features for classic RTT such as the patient's age of onset of the first sign, the presence of microcephaly, the ability to sit alone, ambulation, epilepsy, hand use, onset of stereotypies, respiratory function, and language. Complete information of approximately 12 patients was available; three more patients' reports lack one of those aspects, and no information was available about the other six girls. Even if each patient has a unique phenotype, there are some characteristics that are present more frequently among them: 60% of the girls present the first signs before the age of 12 months, 92.3% of the girls have microcephaly, 64.3% have respiratory problems, 93.3% suffers from epilepsy, 93.3% lost their hand use, 80% began with the stereotypies before 36 months of age and 60% before 24 month, and 86.6% lost their language (see Supplementary Data S4). No clear correlation was identified between the size of the deletion in base pairs and the severity of the phenotype, although there seems to be a trend when taking into account the deleted exons: patients with deletions comprising only one exon of *MECP2* have milder symptoms than those with deletions that involve both exons 3 and 4 or contain *IRAK1* as well.

All 21 patients were heterozygous at the *AR* locus and were thus informative for the assay. Patients P2, P4, P8, P9, P15, P16, P17, and P19 have skewed XCI (defined here as ≥80% activity of one X chromosome). However, 13 of 21 girls have an XCI of >70%; even if we do not consider these results completely skewed, they may indicate a cellular trend to inactivate the mutant allele. This phenomenon could account for why some of our patients do not have such a high score as we may have expected for the size of their deletions (for example P15 or P17). Unfortunately, we lack the clinical information about some of the girls with skewed X. The results for each subject are listed in Table 1 along with a summary of all other results obtained.

4 | DISCUSSION

In this study, we have screened a cohort of 21 classic RTT patients with large *MECP2* deletions detected by MLPA. Subsequent qPCR analysis has confirmed the presence of large deletions in all of them. The deletion breakpoints were further characterized by qPCR and long-range PCR with the aim of defining the precise endpoints at the nucleotide level. That last step was achieved in 14 out of 21 patients. The large number of GCs and all the repetitive sequences found in the intronic region of the gene and in the intergenic zones may have increased the difficulty for the polymerase to amplify our targeted products in those cases. Additionally, after characterizing the case of P19 in which an inversion occurred between two different deletions, we cannot dismiss the possibility that more complex rearrangements are present in the genome of those patients interfering with the correct hybridization of the primers. With the introduction of next generation sequencing and specifically, the whole genome sequencing (WGS), the delineation of those unresolved cases could be achieved; although WGS is still not affordable in order to use it as a routine technique for Rett Syndrome patient testing.

Our results showed a wide range of genotypes, from deletions affecting only a single exon to others involving almost the entire *MECP2* gene and the gene located downstream, *IRAK1*. We found only one patient with a deletion in exons 1 and 2 and part of the promoter region of the *MECP2* gene. This is in accordance with previous findings, although a small number of deletions have been reported affecting exons 1 and 2(Archer et al., 2006; Erlandson et al., 2003; Hardwick et al., 2007; Ravn et al., 2005).

Nine patients (P1, P2, P4, P5, P6, P7, P8, P11, and P12) whom we successfully characterized had a breakpoint in the “deletion-prone region” (DPR, GRCh38/hg38 chrX:154,030,619-154,030,770), as defined by Laccone et

al. (2004)(Laccone et al., 2004). Another two patients (P3 and P10) had their breakpoint close to this region (less than 80 bp away). There are two patients (P9 and P13) who could have one of their breakpoints in the DPR as well, but since they are not fully characterized, we cannot confirm this conclusion (see Figure 1). Our finding together with previous studies (Hardwick et al., 2007; Laccone et al., 2004; Ravn et al., 2005; Schollen, Smeets, Deflem, Fryns, & Matthijs, 2003) could better define the junction sequence of the large MECP2 deletions, since 22 of 42 (52.3%) rearrenged alleles have the breakpoint in the DPR. This region is also the hotspot for the smaller deletions (<500 bp) confined within exon 4. The repetitive nature of the DPR has been considered the major cause of genomic instability there; these include the presence of direct and inverted small repeats, the abundance of polypurine residues in the antisense strand and the presence of the χ-sequence GCTGGTGG, which has been found to be highly recombinogenic in the *Escherichia coli* genome (Stahl, Kobayashi, Stahl, & Huntington, 1983). It has been suggested that this sequence stimulates the recombinase BC-dependent system and is responsible of certain deletions that cause human diseases (Amor, Parkert, Globerman, New, & White, 1988; Marshall, Isidro, & Boavida, 1996).

In addition to the DPR, eight patients whom we successfully characterized and seven in whom long PCR failed had a breakpoint in the same intron 2 region (GRCh38/hg38 version chrX:154033244-154052415). The RepeatMasker program (http://www.repeatmasker.org) revealed that 48.9% of this intronic region consists on interspersed repeats, and 17.9% of them are Alu elements. It has been previously hypothesized that those abundant Alu elements interact with the χ-sequence near the DPR making these types of large rearrangements in *MECP2* possible and recurrent (Laccone et al., 2004; Rüdiger, Gregersen, & Kielland-brandt, 1995). Additionally, Alu has proven to be involved in other genomic rearrangements in different genes (Gu et al., 2015; López et al., 2015; Peixoto et al., 2013). The data we provide contribute to strengthen the theory that all these rearrangements do not occur randomly across the gene and its surroundings but in focal areas. Once the deletion is precisely delimited, studies to correct this mutation by CRISPR-Cas9 technology could be considered to regain a complete and functional *MECP2,* among other strategies the cell possess such as homology repair.

We have attempted to establish genotype–phenotype correlations with our patient cohort. Although no clear correlation between the deleted exons and the clinical severity has emerged from this study, we can appreciate some trends (see Figure 2, left). The patient with the deletion involving exons 1 and 2 has a severe phenotype, which seems reasonable because those exons contain the starting sites for both isoforms of *MECP2* and that without that signal, no product could be generated a priori. Patients with a deletion in exon 4 show the mildest phenotype compared to the remaining combinations. This finding can be explained because none of our five deletions in exon 4 occurs in any of the main functional domains of the protein. An exception could be P5 who has the highest score of the group but, in this particular case, the patient was only 2 years old when the score was set, so it can still improve in the following years. Some authors have claimed that deletions involving *IRAK1* generate a more severe phenotype (Hardwick et al., 2007). *IRAK1* is the interleukin 1 receptor-associated kinase and plays a critical role in initiating an innate immune response against foreign pathogens. In our cohort, little difference can be seen when *IRAK1* is added to the deletion but we must admit that the used checklist and scoring system does not take into account the severity or recurrence of the infections of the patients, features that may allow differentiating the effect of having or not *IRAK1* deleted. If we use to determine the severity of the phenotype the Pineda's clinical score, patients reported by Harwick et al. (2007) and patients from our cohort trends to be similar (see Figure 2 right). However, we are aware that exceptions exist and that, sometimes, patients with the same or very similar deletion present a very different phenotype. For example, P4 and P5 have a similar deletion but their score differs by five points. Other examples are the cases described

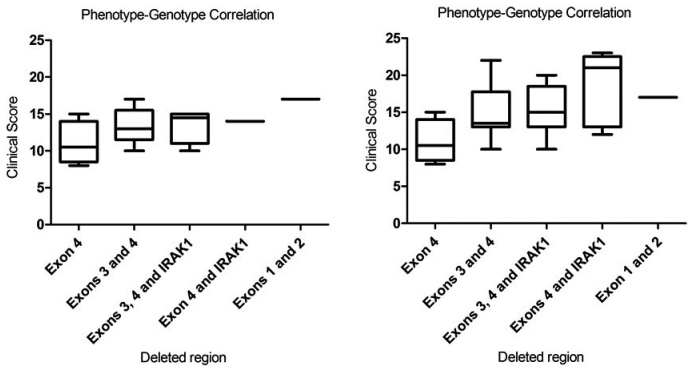

FIGURE 2 Phenotype–genotype correlation according to the deleted region. The left side corresponds to the correlation based on our patients. The right side shows the same correlation based on our patients and the ones reported by Hardwick et al. (2007)

by Bebbington et al. (2012), Mittal, Kabra, Juyal, and BK (2011) or Erlandson et al. (2003). All these cases may suggest another mechanism that alters the direct effect caused by the deletions, possibly a specific methylation pattern causing another molecular alteration in another gene or regulatory domain. Such is the case of the brain-derived neurotrophic factor, *BDNF* (OMIM*113505), gene which is known to protect the carriers of the polymorphism p.Val66Met against early onset epilepsy (Li & Pozzo-Miller, 2014).

Considering the molecular and clinical effect that the lack of a noteworthy part of the coding region of the gene can cause, we were expecting very severe phenotypes. However, the scores in our cohort were not always correlated. The XCI issue has frequently been considered in research on RTT as a potential explanation for the diverse phenotypes generated from the same genotype (Shahbazian, Sun, & Zoghbi, 2002). It has been shown that different cell types can have a different XCI pattern and that the one observed in blood lymphocytes may not be the same as in the brain, the organ in which the majority of symptoms of RTT occur (De Hoon, Monkhorst, Riegman, Laven, & Gribnau, 2015). This phenomenon has accounted for how some females carry a mutation in *MECP2* and are asymptomatic, because of the extreme inactivation of the X chromosome that harbors the aberrant allele (Shahbazian et al., 2002). A similar scenario occurs in mothers carrying a duplication of *MECP2,* who have a skewed XCI and are phenotypically normal unlike their affected offspring who develop *MECP2* duplication syndrome (Van Esch, 2011; Lim, Downs, Wong, Ellaway, & Leonard, 2017). Eight of our patients have a skewed XCI pattern, nine if we lower the threshold to <75% like other authors have done (Hardwick et al., 2007). However, if we lower it to <70% of XCI, four more girls can be included, making a total of 13 patients without a complete random pattern, which could suggest positive selection of cells with an inactivated copy of the defunct *MECP2* allele as a protective mechanism against such large deletions. This hypothesis could explain the relatively mild phenotypes of our cohort. Additionally, we could perform allele specific XCI in two of our patients, P8 and P20 (Personal Data). P8 presents an allele-specific inactivation of 6:94, so most of the mutated allele was inactive, as we expected. Unfortunately, no clinical data were available for this patient. In the case of P20, this technique showed a random inactivation of the gene, so the score might not be so high because no functional domain is present in the deletion and, therefore, the molecular implications for the loss might not be as critical as if they were.

In conclusion, molecular characterization of large rearrangements in *MECP2* is possible in the majority of the cases using the methodology we have exposed. Analysis of that information supports the theory that the 3' end of exon 4 and intron 2 are prone to suffer breaks that can lead to these deleterious big deletions. In addition, gathering clinical data enabled us to define a new set of features that are present in patients with large deletions, such as microcephaly, epilepsy, loss of hand use, loss of language, or onset of stereotypies before 36 months. These data will be very helpful for genetic counseling. A correlation between the severity of the patient and the position of the deletion shows that it is milder when only one exon is deleted and more severe when exons 3 and 4 and *IRAK1* are also involved. In addition, it seems that there is a cellular trend that inactivates the chromosome with the aberrant allele alleviating the final phenotype.

ACKNOWLEDGMENTS

We thank all patients and their families, who contributed to this study. This work was supported by grants from the Spanish Ministry of Health (Instituto de Salud Carlos III/FEDER, PI15/01159), Catalan Association for Rett Syndrome, Fondobiorett and Mi Princesa Rett.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

J.A. and M.P. conceived and supervised the study. S.V., A.P.A., M.R., E G., N.B., and P.P. performed the experiments and collected the data. J.A., S.V., A.P.A., M.R., E.G., N.B., and P.P. analyzed the results. J.A, M.P., and the Rett Working Group provided the patients' samples and clinical and genetic information. A.P.A., S.V., and J.A. wrote the manuscript. All the authors reviewed the article critically for intellectual content.

RETT WORKING GROUP

Hospital Sant Joan de Déu (Barcelona): Maria del Mar O'Callaghan (mocallaghan@sjdhospitalbarcelona.org); Àngels Garcia-Cazorla (agarcia@sjdhospitalbarcelona.org)

Hospital San Borja Arriaran, Santiago de Chile (Chile), Facultad de Medicina Universidad de Chile: Dra Monica Troncoso (monicatroncososch@gmail.com); Dr Guillermo Fariña (guillermofarina@gmail.com)

Hospital Clínico Universitario Virgen de la Arrixaca (Murcia): Dra María Rosario Domingo (mrosario.domingo@gmail.com); Dr. Salvador Ibañez (salibmi@hotmail.com)

Hospital Infantil Universitario Niño Jesús (Madrid): Dr. García Peñas (jgarciadelarape.1961@gmail.com); Dra López Marín (laural.marin@hotmail.com); Dr. González Gutierrez-Solana (luisggsolana@hotmail.com)

Hospital Materno Infantil- Hospital Infanta Cristina (Badajoz): Dr. Enrique Galán Gomez (egalan@unex.es)

Hospital Universitario Fundación Alcorcón (Madrid): Miguel Ángel Martínez Granero (MAMartinezg@fhalcorcon.es)

Hospital Universitario de Girona Dr. Josep Trueta (Girona): Dra María Obon (mobon.girona.ics@gencat.cat)

Parc Taulí Hospital Universitari (Barcelona): Dra Neus Baena Díez (NBaena@tauli.cat)

ORCID

Judith Armstrong https://orcid.org/0000-0003-0588-9307

REFERENCES

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, *23*, 185–188. https://doi.org/10.1038/13810

Amor, M., Parkert, K. L., Globerman, H., New, M., &White, P. C. (1988). Mutation in the CYP21B gene (Ile-172-%3eAsn) causes steroid 21-hydroxylase deficiency. *Genetics*, *85*, 1600–1604.

Archer, H. L., Whatley, S. D., Evans, J. C., Ravine, D., Huppke, P., Kerr, A., … Clarke, A. J. (2006). Gross rearrangements of the MECP2 gene are found in both classical and atypical Rett syndrome patients. *Journal of Medical Genetics*, *43*, 451–456. https://doi.org/10.1136/jmg.2005.033464

Bebbington, A., Downs, J., Percy, A., Pineda, M., Zeev, B. B., Bahi-Buisson, N., & Leonard, H. (2012). The phenotype associated with a large deletion on MECP2. *European Journal of Human Genetics*, *20*, 921–927. https://doi.org/10.1038/ejhg.2012.34

De Hoon, B., Monkhorst, K., Riegman, P., Laven, J. S. E., & Gribnau, J. (2015). Buccal swab as a reliable predictor for X inactivation ratio in inaccessible tissues. *Journal of Medical Genetics*, *52*(11), 784–790. https://doi.org/10.1136/jmedgenet-2015-103194

Erlandson, A., Samuelsson, L., Hagberg, B., Kyllerman, M., Vujic, M., & Wahlström, J. (2003). Multiplex ligation-dependent probe amplification (MLPA) detects large deletions in the MECP2 gene of Swedish Rett Syndrome patients. *Genetic Testing*, *7*(4), 329–332.

Gu, S., Yuan, B. O., Campbell, I. M., Beck, C. R., Carvalho, C. M. B., Nagamani, S. C. S., … Lupski, J. R. (2015). Alu-mediated diverse and complex pathogenic copy-number variants within human chromosome. *Human Molecular Genetics*, *24*(14), 4061–4077. https://doi.org/10.1093/hmg/ddv146

Hagberg, B., Aicardi, J., Dias, K., & 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. *Annals of Neurology*, *14*(4), 471–479. https://doi.org/10.1002/ana.410140412

Hardwick, S. A., Reuter, K., Williamson, S. L., Vasudevan, V., Donald, J., Slater, K., … Christodoulou, J. (2007). Delineation of large deletions of the MECP2 gene in Rett syndrome patients, including a familial case with a male proband. *European Journal of Human Genetics*, *15*, 1218–1229. https://doi.org/10.1038/sj.ejhg.5201911

Kron, M., Howell, C. J., Adams, I. T., Ransbottom, M., Christian, D., Ogier, M., & Katz, D. M. (2012). Brain activity mapping in Mecp2 mutant mice reveals functional deficits in forebrain circuits, including key nodes in the defauit mode network, that are reversed with Ketamine treatment. *Journal of Neuroscience*, *32*(40), 13860–13872. https://doi.org/10.1523/JNEUROSCI.2159-12.2012

Laccone, F., Jünemann, I., Whatley, S., Morgan, R., Butler, R., Huppke, P., & Ravine, D. (2004). Large deletions of the MECP2 gene detected by gene dosage analysis in patients with Rett Syndrome. *Human Mutation*, *23*, 234–244. https://doi.org/10.1002/humu.20004

Li, W., & Pozzo-Miller, L. (2014). BDNF deregulation in Rett syndrome. *Neuropharmacology*, *76*, 737–746. https://doi.org/10.1016/j.neuropharm.2013.03.024

Lim, Z., Downs, J., Wong, K., Ellaway, C., & Leonard, H. (2017). Expanding the clinical picture of the MECP2 duplication syndrome. *Clinical Genetics*, *91*, 557–563. https://doi.org/10.1111/cge.12814

Liyanage, V. R. B., & Rastegar, M. (2014). Rett syndrome and MeCP2. *NeuroMolecular Medicine*, *16*(2), 231–264. https://doi.org/10.1007/s12017-014-8295-9

López, E., Casasnovas, C., Giménez, J., Matilla-Dueñas, A., Sánchez, I., & Volpini, V. (2015). Characterization of Alu and recombination-associated motifs mediating a large homozygous SPG7 gene rearrangement causing hereditary spastic paraplegia. *Neurogenetics*, *16*, 97–105. https://doi.org/10.1007/s10048-014-0429-6

Marshall, B., Isidro, G., & Boavida, M. G. (1996). Insertion of a short Alu sequence into the hMSH2 gene following a double cross over next to sequences with chi homology. *Gene*, *174*, 175–179. https://doi.org/10.1016/0378-1119(96)00515-X

Mittal, K., Kabra, M., Juyal, R., & BK, T. (2011). De novo deletion in MECP2 in a monozygotic twin pair: A case report. *BMC Medical Genetics*, *12*, 113. https://doi.org/10.1186/1471-2350-12-113

Mnatzakanian, G. N., Lohi, H., Munteanu, I., Alfred, S. E., Yamada, T., MacLeod, P. J. M., … Minassian, B. A. (2004). A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nature Genetics*, *36*(4), 339–341. https://doi.org/10.1038/ng1327

Monrós, E., Armstrong, J., Aibar, E., Poo, P., Canós, I., & Pineda, M. (2001). Rett syndrome in Spain: mutation analysis and clinical correlations. *Brain and Development*, *23*, S251–S253. https://doi.org/10.1016/S0387-7604(01)00374-6

Nan, X., Meehan, R. R., & Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Research*, *21*(21), 4886–4892. https://doi.org/10.1093/nar/21.21.4886

Neul, J. L., Fang, P., Barrish, J., Lane, J., Caeg, E. B., Smith, E. O., … Glaze, D. G. (2008). Specific mutations in Methyl-CpG-Binding Protein 2 confer different severity in Rett syndrome. *Neurology*, *70*(16), 1313–1321. https://doi.org/10.1212/01.wnl.0000291011.54508.aa

Neul, J. L., Kaufmann, W. E., Glaze, D. G., Clarke, A. J., Leonard, H., Bailey, M. E. S., … Renieri, A. (2010). Rett syndrome: Revised diagnostic criteria and nomenlcature. *Annual Neurology*, *68*(6), 944–950. https://doi.org/10.1002/ana.22124.Rett

Peixoto, A., Pinheiro, M., Massena, L., Santos, C., Pinto, P., Rocha, P., … Teixeira, M. R. (2013). Genomic characterization of two large Alu-mediated rearrangements of the BRCA1 gene. *Journal of Human Genetics*, *58*, 78–83. https://doi.org/10.1038/jhg.2012.137

Ravn, K., Nielsen, J. B., Skjeldal, O. H., Kerr, A., Hulten, M., & Schwartz, M. (2005). Large genomic rearrangements in MECP2.

Human Mutation, *25*(3), 324–329. https://doi.org/10.1002/humu.9320

Rüdiger, N. S., Gregersen, N., & Kielland-brandt, M. C. (1995). One short well conserved region of Alu-sequences is involved in human gene rearrangements and has homology with prokaryotic chi. *Nucleic Acids Research*, *23*(2), 256–260. https://doi.org/10.1093/nar/23.2.256

Scala, E., Longo, I., Ottimo, F., Speciale, C., Sampieri, K., Katzaki, E., … Ariani, F. (2007). MECP2 deletions and genotype-phenotype correlation in Rett syndrome. *American Journal of Medical Genetics, Part A*, *43A*, 2775–2784. https://doi.org/10.1002/ajmg.a.32002

Schollen, E., Smeets, E., Deflem, E., Fryns, J. P., & Matthijs, G. (2003). Gross rearrangements in the MECP2 gene in three patients with Rett syndrome: Implications for routine diagnosis of Rett syndrome. *Human Mutation*, *22*, 116–120. https://doi.org/10.1002/humu.10242

Shahbazian, M. D., Sun, Y., & Zoghbi, H. Y. (2002). Balanced X chromosome inactivation patterns in the Rett syndrome brain. *American Journal of Medical Genetics*, *111*(2), 164–168. https://doi.org/10.1002/ajmg.10557

Singh, J., Saxena, A., Christodoulou, J., & Ravine, D. (2008). MECP2 genomic structure and function: Insights from ENCODE. *Nucleic Acids Research*, *36*(19), 6035–6047. https://doi.org/10.1093/nar/gkn591

Stahl, M. M., Kobayashi, I., Stahl, F. W., & Huntington, S. K. (1983). Activation of Chi, a recombinator, by the action of an endonuclease at a distant site. *Genetics*, *80*, 2310–2313.

Van Esch, H. (2011). MECP2 duplication syndrome. *Molecular Syndromology*, *2*, 128–136. https://doi.org/10.1159/000329580

Vidal, S., Brandi, N., Pacheco, P., Gerotina, E., Blasco, L., Trotta, J.-R., … Armstrong, J. (2017). The utility of Next Generation Sequencing for molecular diagnostics in Rett syndrome. *Scientific Reports*, https://doi.org/10.1038/s41598-017-11620-3

Yaron, Y., Ben Zeev, B., Shomrat, R., Bercovich, D., Naiman, T., & Orr-Urtreger, A. (2002). MECP2 mutations in Israel: Implications for molecular analysis, genetic counseling, and prenatal diagnosis in Rett syndrome. *Human Mutation*, *20*(4), 323–324. https://doi.org/10.1002/humu.9069

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Vidal S, Pascual-Alonso A, Rabaza-Gairí M, et al; Rett Working Group. Characterization of large deletions of the *MECP2* gene in Rett syndrome patients by gene dosage analysis. *Mol Genet Genomic Med*. 2019;e793. https://doi.org/10.1002/mgg3.793

Publication 2: Molecular characterization of Spanish patients with *MECP2* duplication syndrome.

Authors: Ainhoa Pascual-Alonso, Laura Blasco, Silvia Vidal, Esther Gean, Patricia Rubio, Mar O'Callaghan, Antonio F. Martínez-Monseny, Alba Aina Castells, Clara Xiol, Vicenç Català, Núria Brandi, Paola Pacheco, Carlota Ros, Miguel del Campo, Encarna Guillén, Salva Ibañez, María J. Sánchez, Pablo Lapunzina, Julián Nevado, Fernando Santos, Elisabet Lloveras, Juan D. Ortigoza-Escobar, María I. Tejada, Hiart Maortua, Francisco Martínez, Carmen Orellana, Mónica Roselló, María A. Mesas, María Obón, Alberto Plaja, Joaquín A. Fernández-Ramos, Eduardo Tizzano, Rosario Marín, José L. Peña-Segura, Soledad Alcántara and Judith Armstrong.

Reference: Clin Genet. 2020; 97:610–620. https://doi.org/10.1111/cge.13718

Abstract:

MDS is a rare disease with a few hundred cases reported worldwide. Due to its rarity, most of the publications describe single cases or small familiar cases rather than large cohorts. We contacted neuropaediatricians from other Spanish medical centres and were able to gather 21 MDS patients and 16 carrier mothers and to perform a thorough molecular characterization of them all.

We designed a custom MLPA covering *MECP2* and the flanking genes that lay most recurrently inside the duplications to confirm the MDS diagnosis and to determine each duplication's gene content. It was verified that *MECP2* and *IRAK1* need to be duplicated for a correct MDS diagnosis. In line with the literature, in the majority of our cases the duplication was inherited from an asymptomatic mother. The location of the duplication, its size and gene content were determined for the patients and the carrier mothers. In addition, XCI was studied in all the females. The clinical features of the children were also collected and genotype-phenotype correlation attempts were done. Correlation was analysed based on duplication size and position and a trend was found, although in our small cohort it was not statistically significant. Finally, we also measured the mRNA expression levels of *MECP2* and *IRAK1* and found that lymphocyte derived RNA was highly variable whereas results from fibroblast derived RNA were more reproducible. Although less fibroblast samples were available for expression analysis, the results were consistent with higher doses of both genes. The present characterization has enabled a profound understanding of the Spanish MDS patients, which turns out to be one of the largest reported MDS cohorts.

Received: 24 October 2019 | Revised: 4 February 2020 | Accepted: 5 February 2020

DOI: 10.1111/cge.13718

ORIGINAL ARTICLE

CLINICAL GENETICS WILEY

Molecular characterization of Spanish patients with *MECP2* duplication syndrome

Ainhoa Pascual-Alonso[1] | Laura Blasco[1] | Silvia Vidal[1] | Esther Gean[2] | Patricia Rubio[2] | Mar O'Callaghan[3] | Antonio F. Martínez-Monseny[2] | Alba Aina Castells[1,4] | Clara Xiol[1] | Vicenç Català[5] | Nuria Brandi[6] | Paola Pacheco[6] | Carlota Ros[6] | Miguel del Campo[7] | Encarna Guillén[8] | Salva Ibañez[8] | María J. Sánchez[8] | Pablo Lapunzina[9,10] | Julián Nevado[9,10] | Fernando Santos[9] | Elisabet Lloveras[11] | Juan D. Ortigoza-Escobar[3] | María I. Tejada[10,12] | Hiart Maortua[10,12] | Francisco Martínez[13] | Carmen Orellana[13] | Mónica Roselló[13] | María A. Mesas[14] | María Obón[15] | Alberto Plaja[16] | Joaquín A. Fernández-Ramos[17] | Eduardo Tizzano[18] | Rosario Marín[19] | José L. Peña-Segura[20] | Soledad Alcántara[4] | Judith Armstrong[6,10,21]

[1]Fundación San Juan de Dios, Servicio de Medicina Genética y Molecular, Barcelona, Spain

[2]Departamento de Medicina Genética y Molecular, Hospital Universitario San Juan de Dios, Barcelona, Spain

[3]Departamento de Neurología Pediátrica, Hospital Universitario San Juan de Dios, Barcelona, Spain

[4]Neural Development Lab, Departament de Patologia i Terapèutica Experimental, Institut de Neurociències, Universitat de Barcelona, IDIBELL, l'Hospitalet de Llobregat, Barcelona, Spain

[5]Unidad de Biología Celular y Genética Médica, Departament of BCFyI, Universidad Autónoma de Barcelona, Barcelona, Spain

[6]Servicio de Medicina Genètica i Molecular, Hospital Universitario San Juan de Dios, Barcelona, Spain

[7]Pediatrics, Genetic Epidemiology, Hospital Valle Hebrón, Barcelona, Spain

[8]Unidad de Genética, Hospital Virgen de la Arrixaca, Murcia, Spain

[9]Instituto de Genética Médica y Molecular, Hospital Universitario La Paz, Madrid, Spain

[10]CIBERER (Biomedical Network Research Center for Rare Diseases), Instituto de Salud Carlos III, Madrid, Spain

[11]Departamento de Genètica, LABCO-Iberia, Barcelona, Spain

[12]Laboratorio de Genética Molecular, Servicio de Genética, Instituto de Investigación Sanitaria Biocruces, Hospital Universitario de Cruces, Barakaldo, Spain

[13]Unidad de Genética, Hospital Universitario y Politécnico La Fe, Valencia, Spain

[14]Gastroenterologia, Hospital Xanit, Málaga, Spain

[15]Area de Genètica Clínica i Consell Genètic, Laboratoris ICS, Girona, Spain

[16]Institut de Recerca (VHIR), Universitat Autònoma de Barcelona, Hospital Universitari Vall d'Hebron, Barcelona, Spain

[17]Unidad de Neuropediatría, Hospital Universitario Reina Sofía, Córdoba, Spain

[18]Area Genética Clínica y Molecular, Hospital Universitari Vall d'Hebron, Barcelona, Spain

[19]Hospital Universitario Puerta del Mar Unidad de Genética, Cádiz, Spain

[20]Unidad de Neuropediatría, Hospital Universitario Miguel Servet, Zaragoza, Spain

[21]Institut de Recerca Pediàtrica, Hospital Sant Joan de Déu, Barcelona, Spain

Laura Blasco and Silvia Vidal equally contributed to this study.

Correspondence
Judith Armstrong, Molecular and Genetics Medicine Section, Hospital Sant Joan de Déu, Pg. Sant Joan de Déu 2, Planta 0, 08950 Esplugues de Llobregat, Barcelona, Spain.
Email: jarmstrong@sjdhospitalbarcelona.org

Funding information
Miradas que hablan-Sindrome de Duplicación MECP2, Grant/Award Number: PFNR0085

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1111/cge.13718.

Abstract

MECP2 duplication syndrome (MDS) is an X-linked neurodevelopmental disorder characterized by a severe to profound intellectual disability, early onset hypotonia and diverse psycho-motor and behavioural features. To date, fewer than 200 cases have been published. We report the clinical and molecular characterization of a Spanish MDS cohort that included 19 boys and 2 girls. Clinical suspicions were confirmed by array comparative genomic hybridization and multiplex ligation-dependent probe amplification (MLPA). Using, a custom in-house MLPA assay, we performed a thorough study of the minimal duplicated region, from which we concluded a complete duplication of both *MECP2* and *IRAK1* was necessary for a correct MDS diagnosis, as patients with partial *MECP2* duplications lacked some typical clinical traits present in other MDS patients. In addition, the duplication location may be related to phenotypic severity. This observation may provide a new approach for genotype-phenotype correlations, and thus more personalized genetic counselling.

KEYWORDS
genotype-phenotype correlation, hypotonia, intellectual disability, *IRAK1*, *MECP2* duplication, *Methyl-CpG-binding protein 2 (MECP2)*, recurrent infections, Xq28-duplication

1 | INTRODUCTION

Methyl-CpG-binding protein 2 (MECP2) (OMIM*300005) is located at Xq28 and is involved in multiple functions, including chromatin architecture, RNA splicing and transcriptional regulation.[1-3] Alterations in *MECP2* are related to several syndromes, including *MECP2* duplication syndrome (OMIM#300260), X-linked mental retardation syndrome (OMIM#300055), severe neonatal encephalopathy (OMIM#300673), autism susceptibility (OMIM#300496) and Rett Syndrome (OMIM#312750). Rett Syndrome (RTT) is a neurodevelopmental disorder with an incidence of 1:10 000 that occurs mostly in women, although some male cases have been reported.[4] RTT is characterized by normal development between 6 and 18 months, followed by a regression of neurological traits, including loss of speech, stereotypic movements, microcephaly, seizures and intellectual disability.[5,6] *MECP2* duplication syndrome (MDS) is an X-linked neurodevelopmental disorder that predominantly affects males. It is characterized by severe to profound intellectual disability, early onset hypotonia, poor or absent speech, autistic features, epilepsy, progressive spasticity, constipation, decreased pain sensitivity, scoliosis, recurrent infections and mild dysmorphic features.[7-9] Worldwide, there are less than 200 MDS cases described; however, the syndrome may be underdiagnosed because it is estimated to be the cause of 1% of cases with unexplained X-linked intellectual disability.[10] Because each patient harbours a specific duplication that contains an exclusive set of genes inserted at different genomic loci, identifying both a phenotype-genotype correlation and a molecular pathological basis is not straightforward. To date, there are no treatments for MDS, but improvements have been observed in animal models using antisense oligonucleotides[11,12] or epigenetic compounds,[13] both of which require further study.

Here, we present the clinical and molecular characterization of 21 Spanish patients with MDS, comprising 19 boys and 2 girls. Moreover, our preliminary data show that the duplication position appears to be closely related to the size of the duplication, which opens a new approach for genotype-phenotype correlation analysis.

2 | MATERIAL AND METHODS

2.1 | Patients

Our cohort consisted of 21 patients diagnosed with MDS, identified from several tertiary hospitals across Spain. All patients were recruited after genetic tests confirmed *MECP2* gene duplication. A specific checklist was designed to study MDS to assess the degree of severity for each case (Table S1). The features in the checklist were annotated according to the Human Phenotype Ontology (HPO) database.[14]

The study was approved by the ethical committees of Hospital Sant Joan de Déu, CEIC: Comitè d'Ètica d'Investigació Clínica-Fundació Sant Joan de Déu (internal code: PIC-56-16). Written informed consent was obtained from the legal guardians of patients in accordance with ethical protocols on performing genetic studies. Tissue samples from patients and controls were obtained according to the Helsinki Declaration of 1964, as revised in 2001.[15]

2.2 | Fibroblast primary cell culture

Skin biopsies were obtained from 12 patients and 7 carrier mothers, and primary fibroblast cell lines were established. DNA and RNA were

extracted from fibroblasts using the DNeasy Blood & Tissue Kit and RNeasy Fibrous Tissue Mini Kit (both from Qiagen, Hilden, Germany) according to manufacturer's instructions. The cDNA was generated using the SuperScript III First-Strand Synthesis SuperMix for qRT-PCR from Invitrogen (Thermo Fisher, Waltham, Massachusetts) following manufacturer's instructions.

2.3 | Blood samples: DNA and RNA extraction

DNA was extracted from peripheral blood leukocytes using the Puregene DNA Isolation kit (Gentra System, Minneapolis, Minnesota). RNA extraction was performed using TRI-Reagent (Sigma-Aldrich Co., St. Louis, Missouri) or the PAXgene Blood RNA Kit (PreAnalytiX, Hombrechtikon, Switzerland), depending on whether the blood sample came from Hospital Sant Joan de Déu or another hospital.

2.4 | Array CGH

Different array comparative genomic hybridization (aCGH) platforms were used to detect duplications. Array CGH platforms were assigned by clinicians in hospitals were the children were diagnosed; copies of results were forwarded to us for this study (Table S2).

3 | MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION

To confirm aCGH results and to identify the origin of the duplication (by testing parents), multiplex ligation-dependent probe amplification (MLPA) was performed using SALSA MLPA, P015-D2, E1, F1 or P245-B1 kits (MRC-Holland, Amsterdam, The Netherlands), all of which hybridize in the Xq28 region. The protocol was performed according to manufacturer's instructions. Furthermore, we designed a custom set of probes to specifically analyse the region that was usually duplicated in our cohort. We followed the "Designing synthetic MLPA probes and probe mixes" manual by MRC-Holland and designed custom probes for the following genes: *Zinc finger protein 275 (ZNF275)*, *L1 cell adhesion molecule* (*L1CAM*; OMIM*308840), *N(alpha)-acetyltransferase 10 (NAA10*; OMIM*300013), *Transmembrane protein 187 (TMEM187*; OMIM*300059), *Interleukin 1 receptor-associated kinase 1* (*IRAK1*; OMIM*300283), *Methyl-CpG-binding protein 2 (MECP2*; OMIM*300005), *Filamin A* (*FLNA*; OMIM*300017) and *Vesicle-associated membrane protein 7* (*VAMP7*; OMIM*300053) (Table S3). As reference probes, we used the SALSA MLPA Probemix P300-B1 Reference-2, and the reagents were SALSA MLPA EK (MRC-Holland). We followed the protocol provided by MRC-Holland.

3.1 | Real-time quantitative PCR (qPCR)

To confirm the increased dosage of *MECP2* and *IRAK1*, we performed qPCR analyses of several strategically designed amplicons. Primers were designed using the Primer3 programme (Table S4). We generated standard curves for the *MECP2* and *IRAK1* amplicons and for the autosomal reference gene *MTHFR* (OMIM*607093). Real-time qPCR was performed using the GoTaq Master Mix kit (Promega Corp., Madison, Wisconsin) on the ABI 7500 Real-Time PCR System (Applied Biosystems, Foster City, California) and the PowerUp SYBR Green Master Mix kit for the QuantStudio 6 Flex Real-Time PCR System (both from Applied Biosystems). The analysis was performed using the ΔΔCt relative quantification method. Relative ratios of 1.5 ± 0.2 or more were suggestive of duplication, whereas ratios of 1 ± 0.2 were indicative of a normal copy number for the given region. Three control samples from subjects of the same gender as the patient were used to normalize the dosage copy. All reactions were conducted in triplicate with the average used for quantitative analysis. Product specificity was assessed by melting curve analysis.

3.2 | Reverse transcription-quantitative PCR

Double-stranded cDNA was generated from 500 ng total RNA in the presence of random hexamers using the SuperScript III First-Strand Synthesis SuperMix for qRT-PCR kit (Invitrogen, Carlsbad, California) according to manufacturer's instructions. The GoTaq Master Mix kit (Promega Corp.) for the ABI 7500 Real-Time PCR System (Applied Biosystems) and the PowerUp SYBR Green Master Mix kit for the QuantStudio 6 Flex Real-Time PCR System (both from Applied Biosystems) were used to quantitatively analyse the expression levels of the two isoforms of *MECP2* and *IRAK1*. *Phosphoglycerate kinase 1* (*PGK1;* OMIM *311800) and *ribosomal protein lateral stalk subunit P0 (RPLP0;* OMIM*180510) were used as reference housekeeping genes. Primers for all genes were designed at exon junctions to ensure that only cDNA was amplified (the primer sequences are provided; Table S4). Analyses were performed using the ΔΔCt relative quantification method as described for qPCR. All reactions were performed in triplicate, with the average of each triplicate group used for analysis. Product specificity was assessed using melt curve analysis.

3.3 | X chromosome inactivation (XCI) analysis

The XCI status of female patients and all carrier mothers was determined by studying the methylation status of the highly polymorphic trinucleotide X-linked *androgen receptor* (*AR*) locus, as described by Allen et al (1992). [16] Inactivation was considered skewed (non-random) if the ratio was >80:20.

3.4 | Fluorescence in situ hybridization

Fluorescence in situ hybridization (FISH) was performed in 19 out of 21 patients and 13 out of 14 carrier mothers. FISH determined the chromosomal position of the duplication following

TABLE 1 Clinical characterization of our cohort

Patient ID		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
Age at examination		15y	11m	8y	1y4m	31y	8y	11y	2y9m	6y	12y	9y	4y	24y	20y	14y	6y	2y6m	5y	4y	3y	2y
Gender		F	M	M	M	M	M	M	M	M	M	M	M	M	M	M	F	M	M	M	M	M
HPO	**Characteristic**																					
HP:0008935	Generalized neonatal hypotonia	–	+	+	+	+	+	+	+	+	–	+	+	+	+	+	–	+	+	+	+	+
HP:0001249	Intellectual disability	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
HP:0000750	Delayed speech and language development	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HP:0002191	Progressive spasticity	–	–	–	–	–	+	–	+	+	–	–	–	–	–	+	–	–	–	–	–	–
HP:0001250	Seizures	+	+	+	–	+	+	+	–	+	–	+	+	–	+	+	–	–	–	+	–	–
HP:0012719	Functional abnormality of the gastrointestinal tract	+	+	–	+	+	–	+	+	+	–	–	–	+	+	+	–	+	n.a.	–	–	–
HP:0002376	Developmental regression	+	+	–	–	+	+	+	+	+	n.a.	–	–	+	+	+	–	+	–	–	–	–
HP:0007328	Impaired pain sensation	+	+	+	–	–	–	+	+	–	+	+	–	+	+	+	+	–	+	–	–	–
HP:0000733	Stereotypic behavior	+	+	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HP:0000735	Impaired social interactions	–	+	+	–	+	+	–	–	+	–	–	+	+	+	+	+	+	+	+	+	+
HP:0000817	Poor eye contact	–	+	+	–	+	+	+	–	+	+	+	+	+	–	+	–	+	+	+	+	–
HP:0002205	Recurrent respiratory infections	–	+	+	–	+	+	+	+	+	–	n.a.	+	–	+	+	–	+	+	+	+	+
HP:0002020	Gastroesophageal reflux	–	+	+	+	+	+	–	+	+	–	+	+	–	+	+	–	+	+	–	–	–
HP:0002019	Constipation	–	+	+	+	+	–	+	–	+	+	+	–	+	+	+	–	+	+	+	+	–

Note: The table specifies the age of each patient when the last clinical characterization was performed. The symptoms are coded using the Human Phenotype Ontology database.
Abbreviations: +, present characteristic; – absent characteristic; n.a., not avaible data; y, years; m, months; F, female; M, male.

manufacturer's instructions. *MECP2* was tagged with BAC RP11-119A22 (Illumina, San Diego, CA) labelled in the red spectrum, and with a *MECP2* FISH probe labelled in the red spectrum (Empire Genomics, Williamsville, New York). *IRAK1* was tagged with an *IRAK1* FISH probe labelled in the red spectrum (Empire Genomics). The centromere probe was Vysis CEP X (DXZ1), which was labelled in the green spectrum (Abbott Laboratories, Chicago, Illinois).

TABLE 2 Frequencies of the observed clinical features ordered based on the duplication locations

Clinical features	Chr Xq28	Chr Xp	Outside chr X
Number of patients	**15**	**3**	**3**
Gender	**13M, 2F**	**3M**	**3M**
Generalized neonatal hypotonia	12/15	3/3	3/3
Intellectual disability	14/15	3/3	3/3[a]
Delayed speech and language development	15/15	3/3	3/3[a]
Progressive spasticity	2/15	1/2	1/3
Seizures	8/15	3/3	2/3[a]
Developmental regression	6/14	2/3	2/3
Stereotypical behaviour	14/15	3/3	2/3[a]
Poor eye contact	9/15	3/3	2/3
Impaired social interactions	9/15	3/3	2/3
Pain/Temperature hyposensitivity	10/15	0/3	2/3
Constipation	8/15	1/3	3/3
Gastroesophageal reflux	7/15	3/3	3/3
Functional abnormality of the gastrointestinal tract	9/14	1/3	3/3
Recurrent respiratory infections	10/14	3/3	2/3

Abbreviations: F, female; M, male.
[a]Indicates features that appeared with age.

4 | RESULTS

4.1 | Clinical data

Twenty-one patients were included in this study, including 2 girls and 19 boys. The mean age at evaluation was 9.02 years (SD = 7.8 years). We had two pairs of brothers (P5-P6 and P10-P11) and a pair of cousins (P20-P21). Four patients died due to respiratory complications: P5 and P6 (brothers), died at 31 and 8 years old, respectively; P2 died at 11 months old; and P9 died at 6 years old.

We summarized the main clinical participant features (Table 1) using the Human Phenotype Ontology database. P1 was already described by Mayo et al (2011).[17] P7 was also described elsewhere.[18,19] The most prominent clinical features of patients were delayed speech and language development, intellectual disability and generalized neonatal hypotonia (observed in approximately 95% of cases). Recurrent respiratory infections and constipation were present in 15 of the 20 patients, together with gastroesophageal reflux and functional abnormalities of the gastrointestinal tract (13/21 and 11/20, respectively). Seizures and impaired pain sensations were reported in 12 patients and developmental regression was observed in 11 patients. The least manifested clinical feature was progressive spasticity, which was observed in four patients. Finally, stereotypic behaviour was observed in nearly all patients (20/21), and impaired social interactions and poor eye contact were observed in 15 patients. The frequency of clinical features depending on duplication locations are shown in Table 2.

4.2 | Detection, confirmation and origin of duplications

All patients had duplications detected via aCGH. Duplications were confirmed in our laboratory by qPCR and our custom MLPA assay. Duplication sizes ranged from 221 kB to 14.3 MB, all including *MECP2* and *IRAK1* (Figure 1 and Table 3). To identify the origins of the

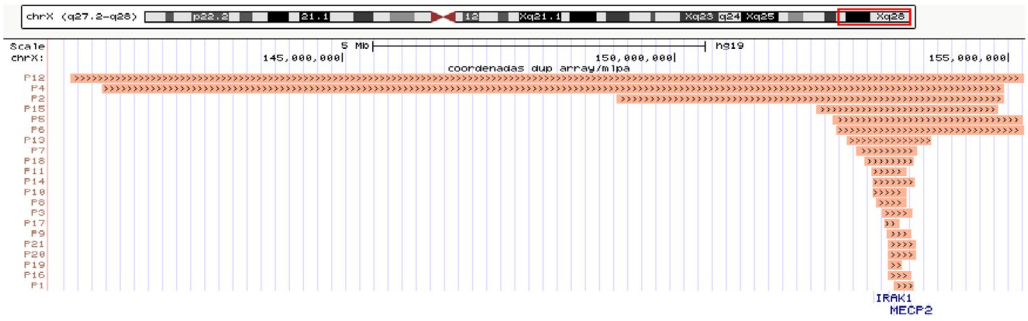

FIGURE 1 Genomic locations of the 21 duplications based on the results obtained from array comparative genomic hybridization and multiplex ligation-dependent probe amplification assays. The annotation is based on GRCh37/hg19

TABLE 3 Detailed information on our cohort, including the origins of the duplications, the positions of the new copies, as well as their sizes and genomic coordinates based on the array comparative genomic hybridization (aCGH) results

Patient ID	Gender	Origin	Position	Duplication size	Genomic coordinates aCGH results (GRCh37/hg19)	XCI	Mother's ID	Carrier mother's XCI
P1	F	*de novo*	tandem	0,299Mb	Xq28(153277239 -153576940)x3; Xq28(153740375 -153877929)x2; Xq28(154535119 -154560375)x3	63:37	M1	-
P2	M	*de novo*	chr Y	5,8Mb	Xq28(149116213-154929279)x2; Yq11.222q11.23(20826207-28629893)x0	-	M2	-
P3	M	maternal	tandem	0,465Mb	Xq28(chrX:153101077-153565901)x2	-	M3	**99:1†**
P4	M	*de novo*	chr Y	13,52Mb	Xq27.2q28(141408852-154929420)x2; Yq11.221q11.23(19562686-28581619)x0	-	M4	-
P5	M	maternal	chr Xp	2,857Mb	Xq28(152369516-155226073)x2; Xp22.33(287562-1728512)x0	-	M5/6	n.i
P6	M	maternal	chr Xp	2,806Mb	Xq28(152422669-155228958)x2; Xp22.33(60701-1719057)x0	-		
P7	M	maternal	tandem	0,913Mb	Xq28(152723597-153636703)x2; 7p14.3(34058830-34098040)x1		M7	**89:11**
P8	M	maternal	tandem	0,45Mb	Xq28(153023556-153473892)x2	-	M8	**100:0**
P9	M	maternal	tandem	0,373Mb	Xq28(153176362-153549367)x2	-	M9	**95:5**
P10	M	maternal	tandem	0,511Mb	Xq28(152962751-153473892)x2	-	M10/11	n.i†
P11	M	maternal	tandem	0,524Mb	Xq28(152949788-153473892)x2	-		
P12	M	maternal	chr Xp	14,3Mb	Xq27.2q28(140928466-155232885)x2	-	M12	**95:5†**
P13	M	maternal	tandem	1,268Mb	Xq28(152567819-153836222)x2	-	M13	**100:0**
P14	M	maternal	tandem	0,637Mb	Xq28(152957295-153594098)x2 1q25.3(183996990-184265525)x3	-	M14	95:5†
P15	M	*de novo*	chr 18	2,7Mb	Xq28(152111224-154841455)x2; 15q13.2(30783615-31089985)x3; 18p11.32(159550-496915)x1	-	M15	-
P16	F	*de novo*	tandem	0,346Mb	Xq28(153194797-153541289)x3	51:49†	M16	-
P17	M	maternal	tandem	0,88Mb	Xq28(153130116-153363136)x2	-	M17	**99:1**
P18	M	maternal	tandem	0,74Mb	Xq28(152832700-153576940)x2	-	M18	n.i†
P19	M	maternal	tandem	0,221Mb	Xq28(153194797-153406233)x2	-	M19	**99:1**
P20	M	maternal	tandem	0,416Mb	Xq28(153194797-153611490)x2	-	M20	76:24†
P21	M	maternal	tandem	0,416Mb	Xq28(153194797-153611490)x2	-	M21	70:30†

Note: X chromosome inactivation (XCI) results of the patients and carrier mothers as measured in peripheral blood lymphocytes or fibroblast cells (†) depending on availability. The results showing skewed patterns are in bold. N.i. stands for non-informative results.

duplications, the parents were studied using MLPA, with data confirmed by qPCR and our custom MLPA assay. Only five of 21 cases were de novo, while the remaining 16 cases were inherited from 14 asymptomatic carrier mother (Table 3).

4.3 | Duplication locations

We performed FISH to identify the chromosomal positions of duplications. We tagged *MECP2* and *IRAK1* and observed they were always duplicated and inserted together into the genome (Figure S1). Four different locations were found: 15 duplications were located in tandem in the Xq28 region, three were located on the short arm of the X chromosome, at Xp, two were located on the Y chromosome, and one was located on chromosome 18 (Figure 2). When applying FISH analysis to the mothers, we distinguished the 14 carrier mothers, validating once more the previous results. Furthermore, the positions of the duplications were the same in the carrier mothers and their offspring Figure S1). We observed that duplications located in the Xq28 regions were smaller than those located in Xp, or outside the X chromosome (Figure 1). Several attempts were made to identify breakpoints in the tandem duplications, but we only succeeded in some cases (Figure S2).

4.4 | XCI results

Because *MECP2* is located on the X chromosome which undergoes X chromosome inactivation, we studied the inactivation status in the

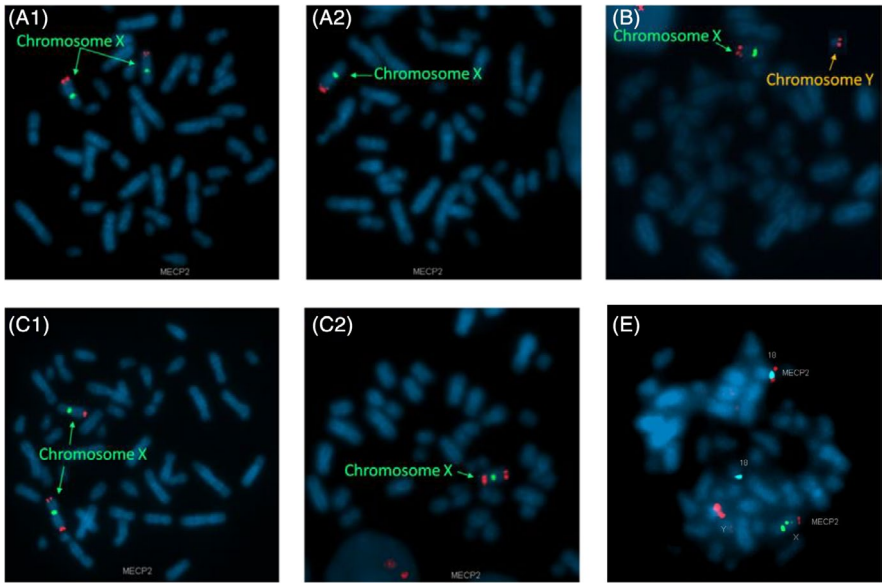

FIGURE 2 The four different chromosomal positions observed in our cohort. In the images in A to C, the red signal shows the *MECP2* gene and the green signal shows the chrX centromere. A, A female (1) and male (2) with *MECP2* tandem duplications. B, A male with the *MECP2* duplication in chrY. C, A female (1) and male (2) with *MECP2* duplications in chrXp. In image D, the green signal shows chrX. The red signal shows chrY and *MECP2*, and the blue signal shows chr18

14 carrier mothers and the two female patients. Among the mothers, nine had a completely skewed pattern, two had results close to the skewed threshold (80:20), and three had non-informative results (Table 3). The mothers M20 and M21 did not reach the skewed threshold, and upon rechecking clinical history and direct evaluation, both exhibited some learning difficulties suggesting a possible mild intellectual disability (observed by the clinician, not measured). This experiment was performed with peripheral blood lymphocyte DNA and when possible, with DNA from cultured fibroblast cells. In cases where both analyses were possible, the data were identical for all patients except for P16, who showed a skewed pattern in blood tissue and a random pattern in fibroblast cells. Patient P1 showed a random XCI pattern.

4.5 | RNA expression

To evaluate whether gene duplications were associated with increases in mRNA expression, we measured the expression levels of *MECP2* and *IRAK1* mRNA in our patients and their carrier mothers. Initially, peripheral blood lymphocytes were used as the RNA source, which was retro-transcribed to generate cDNA for the qPCR analyses. We detected normal or increased expression levels in all patients, except for P4 and P12 (Table 4). Among the carrier mothers, we observed cases with decreased, normal or increased mRNA expression levels. Due to blood sample data variability, we measured expression levels in cultured patient fibroblast cells. In these samples, *MECP2* and *IRAK1* mRNA expression levels were increased, with levels consistent with duplications or triplications seen by aCGH. The mRNA levels in carrier mothers were variable, ranging from normal to increased expression (Table 4).

5 | DISCUSSION

5.1 | Phenotype-genotype correlations

This study expands the literature on the molecular and clinical characterization, and genotype-phenotype correlations for MDS. Some of the common MDS clinical traits appear with age, for example, motor regression, therefore maintaining an updated checklist is important for correct patient supervision, and developing a better understanding of MDS aetiology.

We observed an apparent correlation between the size of duplications and their chromosomal positions, with larger duplications corresponding to translocations outside the X chromosome (Xq28 region < Xp region < non-X chromosome). We believe what we observe in our small cohort could be statistically significant in a larger patient cohort; our findings indicate that the chromosomal position of duplications could be used as an indicator of the size of the duplicated

TABLE 4 RNA expression levels of *MECP2e1*, MECP2*e2*, *and IRAK1* in our patients and their carrier mothers

				Peripheral blood lymphocytes						Fibroblasts					
				Patients			Mothers			Patients			Mothers		
Patient ID	Gender	Origin	Position	*MECP2e1*	*MECP2e2*	*IRAK1*	*MECP2e1*	*MECP2e2*	*IRAK1*	*MECP2e1*	*MECP2e2*	*IRAK1*	*MECP2e1*	*MECP2e2*	*IRAK1*
P1	F	*de novo*	tandem	1.56	1,44	1.28	-	-	-	-	-	-	-	-	-
P3	M	maternal	tandem	1.36	1,44	1.27	0.96	0,70	1,00	1,96	2	2,5	1,71	1,31	1,61
P7	M	maternal	tandem	1,27	1,21	1,877	0,85	0,93	1,15	-	-	-	-	-	-
P8	M	maternal	tandem	1,8	1,46	1,87	0.67	0,48	0.39	-	-	-	-	-	-
P9	M	maternal	tandem	1,02	0,84	1,32	0,65	0,73	1,18	-	-	-	-	-	-
P10	M	maternal	tandem	1.46	1,35	1.30	0.79	0,70	0.87	1,43	2,05	1,43	1,25	1,2	1,22
P11	M	maternal	tandem	1.45	1,28	1.80				1,29	1,98	1,43	1,25	1,2	1,22
P13	M	maternal	tandem	1,27	1,32	1,04	1,38	1,08	1,64	-	-	-	-	-	-
P14	M	maternal	tandem	1.50	1,47	1.18	0.47	0,43	1.39	1,4	2,3	2	1,58	1,48	1,57
P16	F	*de novo*	tandem	-	-	-	-	-	-	1,62	1,22	1,61	-	-	-
P17	M	maternal	tandem	-	-	-	-	-	-	-	-	-	-	-	-
P18	M	maternal	tandem	-	-	-	-	-	-	-	-	-	-	-	-
P19	M	maternal	tandem	-	-	-	-	-	-	-	-	-	-	-	-
P20	M	maternal	tandem	-	-	-	-	-	-	-	-	-	-	-	-
P21	M	maternal	tandem	-	-	-	-	-	-	-	-	-	-	-	-
P5	M	maternal	chrXp	-	-	-	0.96	0,9	0.96	-	-	-	-	-	-
P6	M	maternal	chr Xp	-	-	-				-	-	-	-	-	-
P12	M	maternal	chr Xp	0.83	0,64	0.56	0.24	0,23	0.26	1,8	2,29	1,78	1,48	1,14	0,78
P2	M	*de novo*	chr Y	1.07	0,80	1.33	-	-	-	2	1,97	1,37	-	-	-
P4	M	*de novo*	chr Y	0,666	0,752	1,61	-	-	-	-	-	-	-	-	-
P15	M	*de novo*	chr 18	-	-	-	-	-	-	2,56	2,66	3,08	-	-	-

Note: The measurements were made using cDNA generated from RNA isolated from peripheral blood lymphocytes or fibroblast cells, depending on availability. The patients are classified according to the chromosomal positions of the duplications.

region. Still, we are aware of the modest size of our cohort so more studies revealing the locations of duplications in other patients are required to validate our observations.

Several studies have sought to identify phenotype-genotype correlations based on duplication size, however, no correlations have been found until now.[10,20,21] Recently, Peters et al (2019) declared a correlation existed between duplication size and phenotype severity when applying the RTT severity scale.[22] We strongly believe, and we agree with the authors, that even if MDS and RTT shared some clinical features, a specific MDS severity scale would be more appropriate than the RTT severity scale, as the RTT phenotype does not exhibit, for example, recurrent infections. When looking at chromosomal positions of duplications, we observed some trends. We considered the following clinical traits as more severe: presence of seizures (HP:0001250), generalized neonatal hypotonia (HP:0008935), delayed speech and language development (HP:0000750) and recurrent respiratory infections (HP:0002205). Despite lacking one clinical feature from 3 of our patients, we could observe that patients with duplications translocated outside the chromosome X were more severely affected than the patients with tandem duplications (Table 2). From this data (Table 2), we see that the percentage of patients presenting each feature increases as we move from individuals with duplications in the same Xq28 region to patients having duplications outside the X chromosome. Most likely, insertion of duplications elsewhere fatally disrupts other genes or interactions between genes and their promoters, which may play important roles in the severity of the phenotypes.

Furthermore, it has been suggested that genetic modifiers may explain the variable phenotypes seen between brothers or relatives with the same duplication.[10] Here, we assembled three familial cases: brothers P5-P6, brothers P10-P11, and cousins P20-P21. The brothers P5 and P6 died from infection and respiratory problems. For P10 and P11, some clinical features were different even though both brothers shared the same duplication. We also observed a similar situation for the cousins, and while both are still very young (2 and 3 years old), they may develop new characteristics over the coming years. Our data supports the hypothesis that genetic modifiers may explain the phenotypic variability within some families.

5.2 | Origin of the duplications

We characterized 21 Spanish individuals with MDS. The duplications were inherited from carrier mothers in 16 cases. Among the de novo cases, two out of five patients showed an extra copy of *MECP2* and *IRAK1* on the Y chromosome. We hypothesize that non-homologous recombination events may have occurred in the healthy father's germ line cells, leading to an aberrant Y chromosome recombining with the duplication (Figure S3). The two de novo females could have generated the alterations at some point during zygote formation, or its first divisions. In these two de novo cases, it appears that cells may not have detect the altered copy and did not develop the skewed inactivation of the harbouring chromosome, as occurred in the carrier mothers of the other cases.

5.3 | Implication of duplicated genes

Because duplications in our patients ranged from 221 kB to 14.3 MB, several gene combinations are duplicated. However, we noted that only two genes were present in all duplications in our cohort; *MECP2* and *IRAK1*. In addition, when performing FISH on both genes, we confirmed they were duplicated and translocated together and furthermore, both genes were overexpressed at the RNA level.

IRAK1 encodes an interleukin receptor-associated kinase that acts on the Toll-like receptor transduction pathway, involved in immunity.[8] Dose alterations of this gene may be implicated in recurrent infections, experienced by MDS patients. In our cohort, 15 patients suffered with recurrent infections. It has been observed that MDS patients are more susceptible to infections, because of IgA or IgG2 deficiencies, low antibody titres against encapsulated bacteria, and strong acute-phase responses.[23] According to Bauer et al (2015), *IRAK1* may be implicated in elevated acute responses.

Apart from *IRAK1*, other genes may be involved in this phenotype. Some studies have suggested that *FLNA* and *L1CAM* may be responsible for the intestinal pseudo-obstruction observed in some patients.[20,24] In our cohort, 10 patients had *FLNA* and *L1CAM* duplications, five patients had only *L1CAM* duplication, and two had only *FLNA* duplications. Among these 17 patients, 16 showed various gastrointestinal abnormalities. The other patient was 2 years old, therefore these gastrointestinal issues may not become manifest for several more years. Thus, *FLNA* and *L1CAM* may explain part of the MDS phenotype; however, other factors must be implicated because three of our patients did not have these gene duplications, but showed the same gastrointestinal issues.

6 | X CHROMOSOME INACTIVATION

We also studied the XCI status of carrier mothers and found that nine of the 14 had completely skewed inactivation of one X chromosome. Two mothers showed inactivation patterns close to the threshold, whereas data from three mothers were non-informative. We conclude that XCI events may explain why the carrier mothers were asymptomatic, as the chromosome containing the duplication may have been silenced. M20 and M21 did not reach the skewed threshold, but both exhibited some learning difficulties, which may have been due to aberrant *MECP2* activity. One study has reported neuropsychiatric features, such as depression or anxiety in carrier females.[7] Bijlsma et al (2012) reported that M7 also suffered from depression and compulsions. Although her XCI pattern results were variable at the time of the first characterization, these authors detected a random pattern, in contrast to our skewed data. The female patient P1 had a random XCI, which was consistent with her phenotype.[17] For P16, we obtained different data from the two different tissues, that is, blood lymphocytes and fibroblasts. Previous evidence has suggested that different XCI patterns depend on the tissue type.[25] Furthermore, P16 was suspected to be mosaic, thanks to her FISH data. This fact, along with skewed XCI in some of her tissues may explain her mild phenotype. However, a recent study measured XCI in Rett syndrome patients with point mutations in

MECP2, and showed that inactivation levels in certain tissue types did not explain the observed phenotypes.[26] Some authors have stated that some female patients present MDS because their duplications lie in autosomal chromosomes that escape XCI, leading to severe phenotypes.[8,21,27] Our two cases showed there was no requirement to have translocations outside the X chromosome to trigger this syndrome in females. To provide genetic counselling, we noted that the two symptomatic females with MDS had de novo origins, while the asymptomatic carrier mothers had transmitted the duplications through generations without inherited clinical presentation.

6.1 | MECP2 expression

We also assessed expression levels of *MECP2* and *IRAK1* mRNA. Our first attempt was in peripheral blood lymphocytes; however, mRNA expression was highly variable in both patients and carrier mothers. Blood is a dynamic tissue that continuously renews and adapts itself in response to external stimuli, for example, infections, diet or drugs. Because *MECP2* and *IRAK1* are expressed in almost all cell types, we searched for another tissue type,[28] for example, fibroblasts, a more stable cell type. Results were more reproducible in these cells, although fewer samples were available. As expected for both genes, we found increased mRNA levels in patients, correlating with double or triple genomic doses, and consistent with previous studies.[29,30] In carrier mothers, we detected increased expression levels, but there was variation (Table 4). Because increased *MECP2* expression was detected, we do not expect another protein of the same family to be compensating for its dysregulation. As previously stated, our fibroblast sampling cohort was small, therefore it is premature to speculate on these mRNA expression levels in these patients and their affected mothers. A more exhaustive transcriptome and proteome study in our cohort would undoubtedly shed more light on *MECP2* and *IRAK1* functions in this syndrome. Kishi et al (2016) detected *IRAK1* upregulation in neurons as part of the NF-κβ pathway, which can be regulated by inhibitors to ameliorate the RTT phenotype in mice.[31] If *IRAK1* overexpression is confirmed in other tissue types, we could propose a similar approach as used in Rett mice for MDS.

In conclusion, we analysed and verified *MECP2* duplications, potentially providing accurate MDS diagnoses for affected patients, and avoiding confusion in comparing this phenotype with other X-linked intellectual disability syndromes. In addition, because *IRAK1* was present in all duplications, we hypothesize it must be implicated in this disorder. Our preliminary results show that duplication location and size could be used as "severity parameters" to describe this phenotype. Because our number of patients is small, a wider and larger patient cohort is required to definitively determine the role of these duplications in MDS.

ACKNOWLEDGEMENTS

We thank all patients and their families for their collaboration and support in this study. We would like to thank the Cytogenetics Department of the Hospital Sant Joan de Déu, especially, Teresa, Lluisa, Mar, Carlos and Isabel for their role in preparing samples for FISH studies, and starting some of the fibroblast primary cell culture. We are indebted to the "Biobanc de l'Hospital Infantil Sant Joan de Déu per a la Investigació" integrated in the Spanish Biobank Network of Instituto de Salud Carlos III (ISCIII) for sample and data procurement.

This study was funded by the parents association "Miradas que hablan-Sindrome de Duplicación MECP2," through Fundació Sant Joan de Déu.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

All data from this article is available in the Supporting Information.

ORCID

Antonio F. Martínez-Monseny https://orcid.org/0000-0002-0960-3153
María I. Tejada https://orcid.org/0000-0002-7334-1864
Eduardo Tizzano https://orcid.org/0000-0002-7116-6310
Soledad Alcántara https://orcid.org/0000-0003-3313-2715
Judith Armstrong https://orcid.org/0000-0003-0588-9307

REFERENCES

1. Ehrhart F, Coort SLM, Cirillo E, Smeets E, Evelo CT, Curfs LMG. Rett syndrome—biological pathways leading from MECP2 to disorder phenotypes. *Orphanet J Rare Dis*. 2016;11(1):158. https://doi.org/10.1186/s13023-016-0545-5.
2. Sharma K, Singh J, Frost EE, Pillai PP. MeCP2 in central nervous system glial cells: current updates. *Acta Neurobiol Exp (Wars)*. 2018;78:30-40. https://doi.org/10.21307/ane-2018-007.
3. Horvath PM, Monteggia LM. MeCP2 as an activator of gene expression. *Trends Neurosci*. 2018;41(2):72-74. https://doi.org/10.1016/j.tins.2017.11.005.
4. Villard L. MECP2 mutations in males. *J Med Genet*. 2007;44(7):417-423. https://doi.org/10.1136/jmg.2007.049452.
5. Neul JL, Kaufmann WE, Glaze DG, et al. Rett syndrome: Revised diagnostic criteria and nomenclature. *Ann Neurol*. 2010;68(6):944-950. https://doi.org/10.1002/ana.22124.Rett.
6. Liyanage VRB, Rastegar M. Rett syndrome and MeCP2. *Neuromolecular Med*. 2014;16(2):231-264. https://doi.org/10.1007/s12017-014-8295-9.
7. Ramocki MB, Tavyev YJ, Peters SU. The MECP2 duplication syndrome. *Am J Med Genet Part A*. 2010;152A(5):1079-1088. https://doi.org/10.1002/ajmg.a.33184.
8. Van Esch H. MECP2 duplication syndrome. *Mol Syndromol*. 2011;2:128-136. https://doi.org/10.1159/000329580.
9. Lim Z, Downs J, Wong K, Ellaway C, Leonard H. Expanding the clinical picture of the MECP2 duplication syndrome. *Clin Genet*. 2017;91:557-563. https://doi.org/10.1111/cge.12814.
10. Lugtenberg D, Kleefstra T, Oudakker AR, et al. Structural variation in Xq28: MECP2 duplications in 1% of patients with unexplained XLMR and in 2% of male patients with severe encephalopathy. *Eur J Hum Genet*. 2009;17:444-453. https://doi.org/10.1038/ejhg.2008.208.
11. Sztainberg Y, Chen H, Swann JW, et al. Reversal of phenotypes in MECP2 duplication mice using genetic rescue or antisense oligonucleotides. *Nature*. 2015;528(7580):123-126. https://doi.org/10.1038/nature16159.
12. Lombardi LM, Zaghlula M, Sztainberg Y, et al. An RNA interference screen identifies druggable regulators of MeCP2 stability. *Sci Transl Med*. 2017;9(404):1-22. https://doi.org/10.1126/scitranslmed.aaf7588.

13. Nageshappa S, Carromeu C, Trujillo CA, et al. Altered neuronal network and rescue in a human MECP2 duplication model. *Mol Psychiatry*. 2016;21(2):178-188. https://doi.org/10.1038/mp.2015.128.
14. Köhler S, Carmody L, Vasilevsky N, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res*. 2019;47:D1018-D1027. https://doi.org/10.1093/nar/gky1105.
15. Carlson RV, Boyd KM, Webb DJ. The revision of the Declaration of Helsinki: past, present and future. *Br J Clin Pharmacol*. 2004;57(6):695-713. https://doi.org/10.1111/j.1365-2125.2004.02103.x.
16. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet*. 1992;51:1229-1239. https://doi.org/10.1158/1538-7445.AM2014-CT404.
17. Mayo S, Monfort S, Roselló M, et al. De novo interstitial triplication of MECP2 in a girl with neurodevelopmental disorder and random X chromosome inactivation. *Cytogenet Genome Res*. 2011;135:93-101. https://doi.org/10.1159/000330917.
18. Madrigal I, Rodríguez-Revenga L, Armengol L, et al. X-chromosome tiling path array detection of copy number variants in patients with chromosome X-linked mental retardation. *BMC Genomics*. 2007;8:443. https://doi.org/10.1186/1471-2164-8-443.
19. Bijlsma EK, Collins A, Papa FT, et al. Xq28 duplications including MECP2 in five females: expanding the phenotype to severe mental retardation. *Eur J Med Genet*. 2012;55(6–7):404-413. https://doi.org/10.1016/j.ejmg.2012.02.009.
20. Miguet M, Faivre L, Amiel J, et al. Further delineation of the MECP2 duplication syndrome phenotype in 59 French male patients, with a particular focus on morphological and neurological features. *J Med Genet*. 2018;55:359-371. https://doi.org/10.1136/jmedgenet-2017-104956.
21. El Chehadeh S, Touraine R, Prieur F, et al. Xq28 duplication including MECP2 in six unreported affected females: what can we learn for diagnosis and genetic counselling? *Clin Genet*. 2017;91:576-588. https://doi.org/10.1111/cge.12898.
22. Peters SU, Fu C, Suter B, et al. Characterizing the phenotypic effect of Xq28 duplication size in MECP2 duplication syndrome. *Clin Genet*. 2019;95(5):575-581. https://doi.org/10.1111/cge.13521.
23. Bauer M, Kölsch U, Krüger R, et al. Infectious and immunologic phenotype of MECP2 duplication syndrome. *J Clin Immunol*. 2015;35:168-181. https://doi.org/10.1007/s10875-015-0129-5.
24. Clayton-Smith J, Walters S, Hobson E, et al. Xq28 duplication presenting with intestinal and bladder dysfunction and a distinctive facial appearance. *Eur J Hum Genet*. 2009;17:434-443. https://doi.org/10.1038/ejhg.2008.192.
25. De Hoon B, Monkhorst K, Riegman P, Laven JSE, Gribnau J. Buccal swab as a reliable predictor for X inactivation ratio in inaccessible tissues. *J Med Genet*. 2015;52:784-790. https://doi.org/10.1136/jmedgenet-2015-103194.
26. Xiol C, Vidal S, Pascual-Alonso A, et al. X chromosome inactivation does not necessarily determine the severity of the phenotype in Rett syndrome patients. *Sci Rep*. 2019;9(1):11983. https://doi.org/10.1038/s41598-019-48385-w.
27. Makrythanasis P, Moix I, Gimelli S, et al. De novo duplication of MECP2 in a girl with mental retardation and no obvious dysmorphic features. *Clin Genet*. 2010;78:175-180. https://doi.org/10.1111/j.1399-0004.2010.01371.x.
28. Song C, Feodorova Y, Guy J, et al. DNA methylation reader MECP2: cell type- and differentiation stage-specific protein distribution. *Epigenetics Chromatin*. 2014;7(17):1-16. https://doi.org/10.1186/1756-8935-7-17.
29. Meins M, Lehmann J, Gerresheim F, et al. 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*. 2005;42:e12. https://doi.org/10.1136/jmg.2004.023804.
30. Van Esch H, Bauters M, Ignatius J, et al. Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am J Hum Genet*. 2005;77:442-453. https://doi.org/10.1086/444549.
31. Kishi N, MacDonald JL, Ye J, Molyneaux BJ, Azim E, Macklis JD. Reduction of aberrant NF-κB signalling ameliorates Rett syndrome phenotypes in Mecp2-null mice. *Nat Commun*. 2016;7(10520):1-13. https://doi.org/10.1038/ncomms10520.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Pascual-Alonso A, Blasco L, Vidal S, et al. Molecular characterization of Spanish patients with *MECP2* duplication syndrome. *Clin Genet*. 2020;97:610–620. https://doi.org/10.1111/cge.13718

Publication 3: *MECP2*-related disorders in males.

Authors: Ainhoa Pascual-Alonso, Antonio F. Martínez-Monseny, Clara Xiol and Judith Armstrong.

Reference: Int. J. Mol. Sci. 2021; 22:9610. https://doi.org/10.3390/ijms22179610

Abstract:

MECP2 is a multifunctional gene located on the X chromosome. Loss-of-function mutations in *MECP2* are associated with RTT, a well-known syndrome that affects mainly females. Since the majority of the cases reporting mutations in *MECP2* showed affected females, it was thought that they were lethal in males because no wt allele could be partially expressed as in females. Sporadic reports of boys harbouring mutations in this gene have proved that they can present a wide spectrum of clinical presentations from mild ID to severe neonatal encephalopathy and premature death. As a result, the diagnosis of males harbouring variants in *MECP2* is not always straightforward. Conversely, the entire duplication of the *MECP2* gene is associated with MDS, which predominantly affects males. The following review addresses the different clinical presentations caused by alterations in *MECP2* in males, providing a thorough description and aiding in the diagnosis and classification of these patients.

International Journal of
Molecular Sciences

MDPI

Review

MECP2-Related Disorders in Males

Ainhoa Pascual-Alonso [1,2], Antonio F. Martínez-Monseny [2,3], Clara Xiol [1,2] and Judith Armstrong [2,3,4,*]

1 Fundació Per la Recerca Sant Joan de Déu, Santa Rosa 39-57, 08950 Esplugues de Llobregat, Spain; ainhoa.pascual@sjd.es (A.P.-A.); clara.xiol@sjd.es (C.X.)
2 Institut de Recerca Sant Joan de Déu, Santa Rosa 39-57, 08950 Esplugues de Llobregat, Spain; antoniofederico.martinez@sjd.es
3 Clinical Genetics, Molecular and Genetic Medicine Section, Hospital Sant Joan de Déu, 08950 Esplugues de Llobregat, Spain
4 CIBER-ER (Biomedical Network Research Center for Rare Diseases), Instituto de Salud Carlos III (ISCIII), 28029 Madrid, Spain
* Correspondence: judith.armstrong@sjd.es; Tel.: +34-93-600-9451; Fax: +34-93-600-9760

Abstract: Methyl CpG binding protein 2 (*MECP2*) is located at Xq28 and is a multifunctional gene with ubiquitous expression. Loss-of-function mutations in *MECP2* are associated with Rett syndrome (RTT), which is a well-characterized disorder that affects mainly females. In boys, however, mutations in *MECP2* can generate a wide spectrum of clinical presentations that range from mild intellectual impairment to severe neonatal encephalopathy and premature death. Thus, males can be more difficult to classify and diagnose than classical RTT females. In addition, there are some variants of unknown significance in *MECP2*, which further complicate the diagnosis of these children. Conversely, the entire duplication of the *MECP2* gene is related to *MECP2* duplication syndrome (MDS). Unlike in RTT, in MDS, males are predominantly affected. Usually, the duplication is inherited from an apparently asymptomatic carrier mother. Both syndromes share some characteristics, but also differ in some aspects regarding the clinical picture and evolution. In the following review, we present a thorough description of the different types of *MECP2* variants and alterations that can be found in males, and explore several genotype–phenotype correlations, although there is still a lot to understand.

Keywords: *MECP2*; Rett syndrome; *MECP2* duplication syndrome; encephalopathy; point mutation; loss-of-function; males

Citation: Pascual-Alonso, A.; Martínez-Monseny, A.F.; Xiol, C.; Armstrong, J. *MECP2*-Related Disorders in Males. *Int. J. Mol. Sci.* **2021**, *22*, 9610. https://doi.org/10.3390/ijms22179610

Academic Editors: Nicoletta Landsberger and Angelisa Frasca

Received: 30 July 2021
Accepted: 1 September 2021
Published: 4 September 2021

1. *MECP2* Gene

1.1. General Characteristics of MECP2

Methyl CpG binding protein 2 (*MECP2*) (OMIM *300005) encodes the protein MeCP2 and is located in the Xq28 region which can be inactivated for gene dosage compensation of the X chromosome in females [1]. *MECP2* has four exons and undergoes alternative splicing from which two well-characterized isoforms are generated—isoform e1 and isoform e2. Isoform e1 retains exons 1, 3, and 4 whereas isoform e2 retains exons 2, 3, and 4. *MECP2_e1* is conserved across vertebrates while e2 appeared later in the class Mammalia [2]. *MECP2_e1* is the most abundant isoform in the brain although the ratio between the two isoforms varies across different tissues; for example, *MECP2_e2* is more abundant in fibroblasts [2]. Even though both isoforms share the majority of their sequence and the main functional domains, they are not completely redundant. Each of them has their own properties, spatial expression, function, and interacting partners [3–5].

MECP2 has several structural domains—N-terminal domain (NTD), methyl-binding domain (MBD), intervening domain (ID), transcriptional repression domain (TRD), and C-terminal domain (CTD). MBD and TRD are considered crucial functional domains. MBD enables the binding to methyl CpG dinucleotides and is where most of disease-causing mutations are located. TRD is needed for the binding and posterior recruitment of co-repressor

proteins, such as NCoR, SMRT, and HDAC3, in order to repress transcription [6]. The protein has a nuclear localization signal (NLS) domain as well. MeCP2 is an unstructured protein that can adopt local secondary structures when binding to other molecules, which explains its involvement in multiple molecular interactions and thereby, functions [7,8]. Thus, *MECP2* is a multifunctional gene that acts as a transcriptional regulator (both activating and repressing) and a chromatin remodeler; it also interacts with the RNA splicing machinery and with microRNA processing machinery, among others [9]. Post-translational modifications are also implicated in regulating its activity and interactions with other proteins [10,11]. The resultant protein MeCP2 is ubiquitously expressed even though it is more abundant in the brain, especially in neuronal cells. It is noteworthy that the level of expression correlates with the maturation of neurons, indicating the importance of MeCP2 not only in neuronal development but also in neuronal maturation and maintenance [12,13].

1.2. One Gene, Multiple Phenotypes

Since mutations in the *MECP2* gene were first reported in 1999 in female and male patients with Rett syndrome (RTT) (OMIM #312750) [14,15], genetic alterations ranging from single nucleotide mutations to large deletions have been described and associated with RTT. As the majority of the reported cases described affected females, it was suggested that mutations in *MECP2* lead to embryonic lethality or early postnatal death in males, since no wildtype allele can be partially expressed as in females. However, sporadic reports of boys with mutations in this gene have shown otherwise [15–19].

Anyway, *MECP2* is not only associated with RTT, but is also related to severe neonatal encephalopathy (OMIM #300673), autism susceptibility (OMIM #300496), X-linked mental retardation syndrome (OMIM #300055), and *MECP2* duplication syndrome (MDS) (OMIM #300260). All these phenotypes, and others not so well defined, have been reported in males carrying variants in *MECP2.*

As can be inferred, *MECP2* is a dosage-sensitive gene because loss-of-function mutations lead to RTT, but whole gene duplication leads to MDS. This must be taken into consideration when looking for a treatment. Here, we would like to take a deeper look into the male cases harboring mutations and alterations in *MECP2* since much remains to be understood.

2. Mutations in *MECP2*

Whenever a mutation in *MECP2* is found in a patient, RTT becomes a possible diagnosis. RTT was first clinically described in 1966 by Andreas Rett in girls. In 1999, Zoghbi's group linked *MECP2* to RTT [14,15]. Since then, groups all over the world have reported patients, reaching a few thousands of cases. In fact, *MECP2* mutations cause 97% of classic RTT cases [20,21].

Several specific *MECP2* mutation screenings have been performed in males affected by neurological disorders. In all of them, a low frequency of variants in *MECP2* was found [22–25]. RettBASE is an international curated database, which gathers the genetic variation found in individuals with RTT and related clinical disorders. To date, there have been 3924 female cases with mutations in *MECP2* and 345 male cases. As in females, in males, mutations range from single nucleotide changes to larger deletions involving up to 240 nt [26,27]. Small duplications from one to seven nucleotides have also been reported in RettBASE. The disparity of cases for each sex and the difficulties in creating the first male mouse model suggested that mutations in *MECP2* in males were lethal. Fortunately, different groups have reported new patients and, nowadays, we know that the effects of these mutations range from severe neonatal encephalopathies and premature death (as in the case of c.806delG [26]) to mild intellectual and psychomotor impairment (as in c.608C > T [19]) and that they are not always related to RTT.

Int. J. Mol. Sci. 2021, 22, 9610

2.1. Clinical Presentation

RTT is a very well-characterized syndrome in females with approved and revised diagnostic criteria since the early 1990s [28–31] and established severity scores [32–34]. However, in males, the diagnosis sometimes remains unclear. Several groups have studied these male patients and classified them according to clinical and genetic criteria (Table 1).

Table 1. Different methods of classifying male cases with mutations in *MECP2* depending on clinical and/or genetic criteria.

Classification Groups	Reference
1. **Boys with severe neonatal encephalopathy.** When having normal chromosomal complement (46, XY) they die at an early age. When being mosaic or having Klinefelter syndrome (47, XXY) they develop classic RTT. The same mutations cause RTT in girls. 2. **Boys with non-specific mental retardation.** They have normal chromosomal complement. The same mutations do not affect girls or cause mild ID.	Ravn et al., 2003
1. **Boys with RTT.** RTT consensus criteria are fulfilled. They are mosaic or have Klinefelter syndrome (47, XXY). The same mutations cause RTT in girls. 2. **Boys with severe neonatal encephalopathy and early death.** They have a normal chromosomal complement (46, XY). The same mutations cause RTT in girls. 3. **Boys with less severe neurological and/or psychiatric manifestations.** They have normal chromosomal complement. The same mutations do not affect girls, cause mild ID, or have not been reported in female patients.	Huppke and Gärtner 2005, Moretti et al., 2006, Villard 2007, Neul et al., 2012
1. **Boys with Classical or atypical RTT.** Consensus criteria for each category are fulfilled. When classical RTT is diagnosed the term "male RTT encephalopathy" is suggested. 2. **Neonatal encephalopathy.** Impairment of the clinical traits is noted from birth. 3. **Progressive encephalopathy.** Impairment of the clinical traits appears over the years. 4. **Cognitive impairment.** No progressive worsening is detected.	Neul et al., 2019

In 2003, Ravn et al. compared the first group of 18 male patients with pathogenic variants in *MECP2*. They classified the variants into two groups—mutations causing RTT in girls and mutations that do not affect or cause mild intellectual disability (ID) in females. This genetic classification corresponds with the phenotypes of the boys, which are divided into two groups as well—cases with severe neonatal encephalopathy and cases with non-specific mental retardation. They pointed out that, among the patients harboring RTT mutations, two kinds of patients could be found. Whenever the patient has Klinefelter syndrome (47, XXY) or is a mosaic, the boy develops an RTT phenotype. However, if the chromosomal complement is normal and no mosaicisms are found, the boy usually dies at a very early age [26].

Huppke and Gärtner classified the male patients with mutations in *MECP2* into three groups: (1) boys presenting severe encephalopathy and infantile death with *MECP2* mutations seen in RTT girls; (2) boys with RTT, who are mosaic or have a 47, XXY karyotype and harbor mutations already seen in girls with RTT; and (3) boys with less severe neurological and/or psychiatric manifestations. These mutations have not been found in RTT girls [35]. This classification was used in subsequent articles to group the cases of males with variants in *MECP2* [36–38]. Although the previous classifications stated that an abnormal chromosomal complement is needed to have RTT as a male, there are reported cases with a normal 46, XY karyotype and diagnosed RTT such as the one reported by Masuyama et al. [39] In those cases, due to limitations of the technique, mosaicisms cannot be excluded.

Recently, Neul et al., performed a thorough examination of males with *MECP2* mutations enrolled in the RTT Natural History Studies RTT5201 and RTT5211 [40]. Cases were divided into four groups according to their phenotype: (1) classical or atypical RTT when consensus criteria were met, (2) neonatal encephalopathy when the impairment was seen from birth, (3) progressive encephalopathy when the worsening of the clinical traits was delayed, and (4) cognitive impairment when no progressive worsening was detected during the study. They also emphasized that the clinical evolution of males meeting the diagnostic criteria for typical RTT is more severe than that observed in females. This includes more impaired initial development, ventilatory requirement, and early death. They suggested a new diagnostic category named "male RTT encephalopathy". The new term must include all three following criteria: (1) meeting all the criteria for RTT (having a regression pattern, two of the four main criteria, and five of eleven supportive criteria), (2) having a mutation in *MECP2*, and (3) being male. This new term includes sufficient clinical features for an RTT diagnosis but acknowledges that the progression and pattern of the disease is different from that seen in females.

We encourage clinicians and researchers to start applying the classifications mentioned above depending on the kind of information possessed. Whenever the available information is clinical, Neul's group's classification could be used. Otherwise, if genetic information is as well known, Huppke and Gärtner's classification could be used (see Table 1). Knowing the chromosomal complement or the mosaicism status of the patient, together with a thorough exploration considering the consensus criteria for RTT diagnosis and information about whether the mutation has already been identified in RTT girls, would improve diagnosis of these individuals and would lead to a more personalized visit. Males carrying *MECP2* variants have a very wide spectrum of clinical presentation and evolution and are difficult to classify, unlike classical RTT females.

2.2. Genotype–Phenotype Correlation

Improvements in the technology for molecular diagnosis have proven to be helpful. Since the discovery of *MECP2* as a causative gene of RTT, the denaturing high-performance liquid chromatography (DHPLC) technique and direct sequencing of the coding exons has been performed to confirm the clinical suspicion [18,35,41]. At the beginning, the search for alterations in the sequence was focused on exons 3 and 4 since most of the coding region is located in them. When the second isoform was found and the first mutations were reported in exon 1, the search expanded to cover all four exons [2].

The incorporation of next generation sequencing (NGS), especially of gene panels, has helped reduce the time needed for a molecular diagnosis in patients with rare diseases because of its ability to multiplex genes and patients. NGS has enabled the finding of the molecular cause in patients with either a more recognizable RTT phenotype and for whom traditional techniques were unable to detect a variation, or a more ambiguous phenotype such as X-linked intellectual disability [42–44]. The implementation of NGS as a diagnostic tool has found new patients with *MECP2* variations, especially males, who otherwise might never have been redirected for a *MECP2* direct sequencing test [own data]. In addition, NGS-based methods possess a high read coverage for the amplified genes which makes them a technique to take seriously into consideration for mosaicism detection rather than Sanger sequencing [45].

More and more variants of different types are being reported in *MECP2*. Some of them coincide with the mutations found in girls diagnosed with RTT. Others might seem pathological because of the effect they cause on MeCP2. But there are cases in which the relationship between the variant and the phenotype is not clear and in silico studies together with functional studies must be performed in order to assess the variant's pathogenicity. We encourage geneticists not to dismiss a variant only because it is inherited from an apparently asymptomatic mother. XCI could explain the mild phenotype of the carrier mothers. In general, the study of different molecular levels should be taken into consideration before making conclusions about the pathogenicity of a variant.

Int. J. Mol. Sci. 2021, 22, 9610

Genotype–phenotype correlations have been difficult to determine because of the low number of reported cases. Since new cases are reported individually or in small series, usually, a comparison between the new case (or cases) and genetically similar ones is conducted in order to stablish a correlation. That approximation has led to diverse outcomes [26,46]. The lack of always having a straight correlation highlights the molecular complexity within these boys. Neul et al., 2019 found a correlation in their cohort and observed that males with early RTT mutations (before codon 271) had a higher RTT Clinical Severity Scale (CSS) score compared to the ones with later RTT mutations [40]. That could be accomplished because a large group of patients was analyzed.

3. Duplication of *MECP2*

Duplication of the entire *MECP2* and *IRAK1* (OMIM *300283) genes causes *MECP2* duplication syndrome (MDS). Classical MDS phenotype shows mainly in males while in females the severity ranges from anxiety, autistic features, and mild intellectual impairment to a severe phenotype similar to that reported in males (Figure 1). Although the duplication can occur *de novo*, it is usually inherited from a carrier mother with a skewed XCI who is apparently healthy.

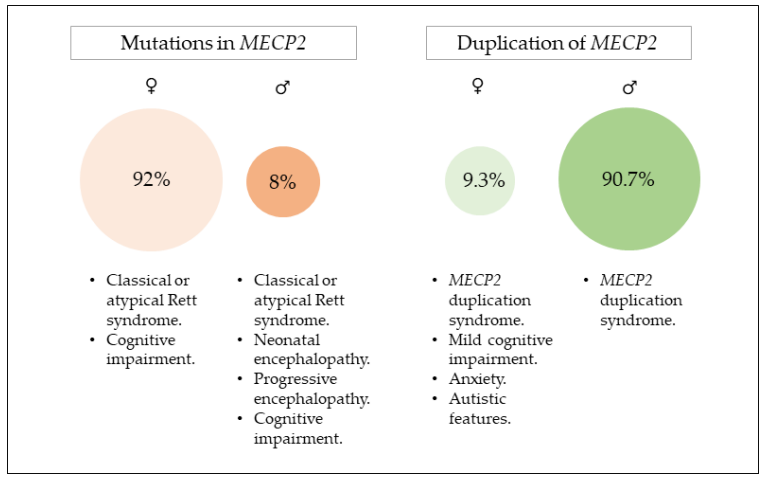

Figure 1. Frequency of each sex among patients harbouring mutations in *MECP2* and those with a duplication of the entire *MECP2*. Sources of information are RettBASE for mutations in *MECP2* and 82 articles from PubMed containing the term "*MECP2* duplication" for the patients with the duplication.

In the late 1990s, several groups were trying to link patients with X-linked mental retardation (XLMR) to specific genetic alterations or genes. During this search, several cases with X chromosome distal duplications were reported [47]. In particular, Lubs et al., described a family of five affected boys with an Xq28 duplication inherited from carrier mothers which later were confirmed to be proper cases of MDS [48,49]. Because of that first article, MDS was originally named as Lubs X-linked mental retardation syndrome, a name that still can be found in OMIM (MIM #300260). In 2005, publications of well-characterized new cases with duplications encompassing *MECP2* led to the establishment of the MDS as we know it nowadays [47,50,51].

More publications with further MDS cases followed those first articles and scientists and clinicians started to wonder about the frequency of the syndrome. Shao et al., Ramocki et al., and de Brouwer et al., analyzed the results of 5380 and 4683 male cases and 600 family members, respectively; the frequency of MDS was 0.30%, 0.41%, and 0.5% [52–54].

Shao et al., pointed out that duplication of Xq28, including *MECP2*, was the most common duplication they found in their cohort. Whereas the screening of males with an X-linked inheritance and idiopathic XLMR without expansion of *FMR1* and normal karyotypes set the frequency of MDS at 1%, the screening of males with ID and severe, mostly progressive, neurological symptoms raised the frequency to 2.2% [55]. In a similar cohort of multiple congenital anomalies with ID, Honda et al., set the frequency of MDS at 2.6%, and among families with ID, 2.3% [56]. The frequency of MDS in small male cohorts with specific inclusion criteria ranges from 1.6% to 17.6% [53,57,58].

Even though MDS is a rare syndrome and most of the articles describe sporadic or small familiar cases rather than large cohorts, to date, there are more than 600 cases reported worldwide [47–51,55–131].

3.1. Clinical Presentations

MDS is characterized by a wide variety of symptoms. The core phenotype of MDS patients includes hypotonia, developmental delay (DD), mostly moderate-severe intellectual disability, autistic features, epilepsy, progressive lower extremity spasticity, poor speech development, recurrent infections, and gastrointestinal problems. Descriptions of individual cases and, specially, of large cohorts have helped us better understand the main and constant symptoms of MDS and expand its phenotype [53,102,109,113,116,119,121,123,129,130,132]. Even so, there are no publications focused on knowing the natural history of this disorder.

3.1.1. Neurological Aspects

Hypotonia, which is only absent in few children, is associated with delay in attaining developmental milestones including sitting and crawling. Some males achieve ambulation without support, but it is usually an ataxic gait that generates a lumbar hyperlordosis as a compensation [132]. Unfortunately, with age, many MDS patients manifest a motor regression that will lead to progressive loss of ambulation in some of them [116,121,132].

Many individuals with MDS meet the formal criteria for a diagnosis of autism spectrum disorder due to poor expressive language skills, abnormal social affect, and restricted/repetitive behaviors. Mood disorders such as anxiety sometimes occur. Two studies comparing a group of 9 and 10 boys with MDS and a matched group of boys with idiopathic autism spectrum disorder (ASD) revealed similarities between them [70,133]. Another study, however, highlighted that a considerable proportion of their MDS cohort lacked social impairment since they interacted during the consultations and that they frequently smiled meaningfully [116]. Moreover, most affected children do not develop speech and the ones who at some point speak a few words usually lose that ability.

Hand stereotypies and bruxism are traits shared between RTT and MDS [53,109,116,130]. However, while hand stereotypies in RTT replace purposeful hand use, they do not in MDS and, when present, they appear later than in RTT [8,116].

It is believed that refractory epilepsy and recurrent infections trigger the regression in these patients. Seizure types that have been reported in individuals with MDS include head/neck and trunk-drop attacks, absence seizures, myoclonic seizures, and generalized or secondarily generalized tonic-clonic or simply tonic seizures (once known as grand mal seizures). Up to half of individuals develop recurrent seizures in childhood or adolescence. Seizures may start at a very variable age and the prevalence might be underestimated, since, in some patients, the onset occurs during the second decade of life [132]. Unfortunately, more than half will possess drug-resistant epilepsy [116]. A decreased pain sensitivity has been found in these children as well [116,123].

3.1.2. Immunological Aspects

Together with epilepsy and neurological deterioration, recurrent infections, mainly in the respiratory tract, are the main cause of concern and are a major contributing factor to reduced life-expectancy. Affected individuals may develop recurrent pneumonia that

is sometimes severe, requiring mechanical ventilation. This is responsible for the many hospitalizations and the premature death of many MDS patients. Middle ear infections (otitis media) and sinusitis are also common. Additional infections have been reported including meningitis or urinary tract infections. Bauer et al., recommend vaccination against pneumococci and evaluation of post-vaccination titers [98]. MDS patients show irregularities in their antibody levels, sometimes even after shorts periods of time post-vaccination; in those cases, extra boosters are required [98,115,132]. Nevertheless, there are patients who have shown a reduction in the frequency of infections with age [67].

3.1.3. Gastrointestinal Aspects

Feeding difficulties due to hypotonia may be evident shortly after the child is born. Frequent gastrointestinal problems are gastro-esophageal reflux, swallowing dysfunction, and excessive drooling, which may contribute to their recurrent respiratory infections. Affected children will often fail to gain weight or will grow at the expected rate for age and gender but may be at a risk of aspiration; however, some infants experience no recognized problems in the neonatal period and concern is not raised until other developmental milestones are missed. The majority of patients suffer from severe constipation [109,130]. Bladder dysfunction has also been seen. Fortunately, Miguet et al., observed in their cohort that, with age, these difficulties often disappear [116].

3.1.4. Dysmorphological Aspects

MDS patients have some dysmorphic features as well. Among them are brachycephaly, deep-set eyes, strabismus, midface hypoplasia, a small open mouth, thick lower lip, large prominent ears, prominent nasal bridge, pointed nose, prominent chin, thick and dense hair, teeth anomalies, and tapered fingers [116,130]. Early in life, head size appears to be normal, but as the boys get older, some might develop microcephaly. However, macrocephaly has also been observed, and overall, these are not consistent features of the syndrome. Some boys have hypoplastic genitalia, cryptorchidism, and hypospadias. Some degree of growth retardation has been described as well. Lim et al., found that 21.7% of their studied males suffer from scoliosis, a trait that they estimated half of the patients will develop by the age of 22 [109].

Due to the potential facial gestalt of MDS male patients, the Face2Gene platform could be trained to identify MDS patients. Through automated image analysis, Face2Gene (FDNA., Boston, MA, USA) has been widely used as the tool based on pattern recognition of frontal photographs in the field of syndromic rare diseases (https://face2gene.com (accessed on 14 June 2021)) [134]. This work consisted of training of the algorithm within the tool from frontal facial photographs of MDS males and females and observing differences between them.

3.1.5. Evolution of the Syndrome with Age

As mentioned, some clinical features, such as epilepsy, respiratory tract infections, or constipation, appear or worsen with age. As a result, Peters et al., reported that their older participants have more severe clinical symptoms, specifically regarding motor dysfunction (e.g., dystonia, scoliosis, and/or rigidity) and functional skills (e.g., motor skills, communication skills, chewing, and swallowing) [129]. A longitudinal Japanese study found a similar outcome by comparing the clinical traits of MDS boys at their first medical visit and some years later [130].

3.2. Characteristics of the Duplication

Duplications can be located at different genomic positions. They have been reported in tandem in the same Xq28 region, translocated to the Xp arm, or even outside the X chromosome [47,87,92,123]. Cases of *MECP2* triplications have also been described in males [57,68,82,84,93,104].

Duplications can be inherited from mothers to their children and the location and gene content is mainly maintained. However, detailed studies of the duplications between different families have revealed that MDS duplications are non-recurrent, as each rearrangement is of a different size and has different breakpoints [62,68,72,82,97,123]. Moreover, Yi et al., found that the transmission of the duplication is not always stable and can increase or decrease in size when passed from mothers to children [105]. The Xq28 region is an unstable region because of the high GC content and the elevated number of repetitive elements as Alu or low copy repeats (LCR). These factors are involved in the breakpoints that lead not only to DNA duplications around *MECP2* but also to deletions and inversions within it [62,135]. Nevertheless, what all the duplications of MDS have in common is that they always contain the genes *MECP2* and *IRAK1*, which comprise the minimal duplicated region. Reported duplications range from 0.079 Mb to 15.8 Mb [87,92].

In order to detect duplications (or triplications) in the Xq28 region and to establish their length, most laboratories use the microarray-based comparative genomic hybridization (array-CGH) as it is broadly used to study patients with developmental delay [132]. Real-time quantitative PCR and MLPA techniques can also be used to diagnose or to confirm the result of an array-CGH. The location of the extra copy can be determined by performing a FISH study.

3.3. Genotype–Phenotype Correlation

Establishing a genotype–phenotype correlation has been a challenge in MDS because of the lack of large cohorts and long-term studies on them. Several traits have been taken into consideration.

Males with *MECP2* triplication seem to be more severely affected, which points to gene dosage as a contributor to the severity in MDS [57,82,84]. A similar result has been seen in girls [81].

The location of the duplications has been studied as well. It was thought that girls with translocations of the duplications to autosomal chromosomes were more severely affected than girls with interstitial duplications, since the former escape XCI [74,106]. However, females with duplications in tandem and a severe phenotype have also been reported [86]. In our personal experience with a Spanish cohort, a boy with a duplication in the Y chromosome was the most severely affected individual and died before one year of age. Nevertheless, we know another boy with an even larger duplication in the Y chromosome who is still alive [123].

The size of the duplication has not succeeded in grouping patients according to their clinical severity [55,57,68,70,87,92,94,105,116,123]. Recently, Peters et al., studied a cohort of 48 MDS individuals and applied the CSS used in RTT [122]. They saw a correlation between CSS and severity. It should be noted, though, that a specific scale would have been more adequate since the phenotype of MDS and RTT, especially in males, has little overlap. A bespoke scale that incorporates the actual symptoms and complications of MDS patients is required to help describe the natural history of the disorder, establishing genotype–phenotype correlations and monitoring the evolution and the response to future potential treatments within clinical trials.

The role of *MECP2* as the main causative gene of the syndrome is clear. Furthermore, it has been hypothesized that the gene content of each duplication could contribute in different degrees to the clinical phenotype. *IRAK1*, which is part of the minimal duplicated region, encodes for the interleukin-1 receptor-associated kinase 1, which participates in the TLR/IL1R signaling pathway and its overexpression could be related to recurrent infections [98]. The *FLNA* (OMIM *300017) gene could contribute to the intestinal pseudo-obstruction problems of these patients [66]. Loss of function of *L1CAM* (OMIM *308840) is associated with hypoplasia of the corpus callosum, which is a common finding in MDS brain imaging [102,136]. However, larger cohorts should be studied in order to find a genotype–phenotype correlation for this trait. *RAB39B* (OMIM *300774) has been highlighted as well, since duplications of it are related with ID. In the study conducted

by Peters et al., they found that MDS patients with duplications harboring *RAB39B* have higher CSS [122].

There are cases of relatives sharing the same duplication and presenting a variable phenotype [97,105,113,117,123]. Interestingly, somatic mutations have recently helped explain the phenotypic differences between monozygotic twins [117]. This event adds more complexity to the current challenge of establishing a genotype–phenotype correlation in small cohorts of rare diseases such as those gathered in this review. The possibility that the clinical phenotype of these patients varies due to the disruption of specific genes or their regulatory regions cannot be ruled out. In any case, further studies are still needed in order to obtain robust results about other genes that contribute to the phenotype.

4. Modeling RTT and MDS for Future Therapies

We would not like to end this review without briefly mentioning the advantages of in vitro and preclinical models. The difficulty in accessing target tissue samples from children affected by neurodevelopmental disorders has encourage researchers to create specific animal and cellular models to gain knowledge of rare diseases as RTT and MDS. In RTT, the most frequently used animal model has been the male *Mecp2*-null mouse (*Mecp2*$^{-/y}$) which manifests the early severe phenotype seen in humans [137]. Despite being the major source of findings related to mechanisms and pathways in RTT, the translatability of mouse models towards humans is not clear, especially regarding RTT females. Several mouse models for MDS have also been created [138,139]. Alternative models, such as primary cultures of patients' peripheral tissues, human embryonic stem cells (hESCs), or human-induced pluripotent stem cells (hiPSCs) reprogrammed from patients' somatic cells, have proven to be very useful [103,140]. Tang and colleagues found that in hiPSCs of a male RTT patient, elevating KCC2 levels could ameliorate the functional deficits caused by the absence of MeCP2, and showed that IGF1 treatment works in the mentioned tissue [141,142]. Kim et al., also found that in RTT hiPSC knockdown of LIN28 expression partially reversed the synaptic deficits [143]. Recently, 3D aggregates from hiPSCs have been developed in an attempt to mimic the complex architecture and functions of organs such as the brain [144]. Moreover, region-specific brain organoids have been generated. The created organoids have also proven to be mutation-dependent and different initial phenotypic alterations have been found in organoids with different backgrounds [145]. All these in vitro human-derived models seem truly promising, not only because of the molecular and genetic insight they are generating, but also because several drugs are being tested and these could, ultimately, undergo clinical trials.

Even though most of the clinical trials for RTT have female participants, according to the register of the U.S. National Library of Medicine a few clinical trials have incorporated male patients. Such is the case of NCT00593957, NTC01520363 with dextromethorphan, NCT02790034 with sarizotan, and NCT00299312, in which a phase of genetic and physical characterization of RTT patients has been done. On the other hand, there are no clinical trials registered yet for MDS, but some promising results have been obtained in the previous in vitro and animal models. Recently, Ash et al., found that the hyperactivity seen in ERK the pathway in MDS could, similar to other autism-associated disorders, be reversed with ERK-specific pharmacologic inhibitors [146,147]. Moreover, antisense oligonucleotide (ASO) therapies are showing promising results in mice, especially because of their ability to reduce MeCP2 expression in a dose-dependent manner [139]. These results show that CNS administration of *MECP2*-ASO is well tolerated and beneficial in a mouse model. Although these first studies do not include the *IRAK1* gene, they provide a translatable approach that could be feasible for treating MDS. The CRISPR-Cas system has been tested in animal models and human primary fibroblasts and has successfully corrected the duplication of *MECP2* including *IRAK1* [140]. However, there is not enough evidence so far to suggest possible approaches to therapy targeted the pathophysiology underlying these two diseases. Further work could bring deep brain stimulation, ASO, and gene therapy into the clinic within the coming decades [8].

5. Conclusions

RTT is a very well-defined syndrome in females and, when the required main and supportive criteria are met, the diagnosis becomes clear and accurate. In males, however, the clinical manifestations generated from variants and mutations in *MECP2* are so different that a diagnosis is not always reached. Classification of these male patients into the mentioned groups should help clinicians and geneticists to better understand the phenotypes that arise from alterations in *MECP2* and to establish the molecular basis for the genotype–phenotype correlation. In addition, technology is rapidly evolving and worldwide databases with detailed information are helping us understand and interpret the new or rare variants that are found in *MECP2.* It was proven that *MECP2* in males is also related to neurodevelopmental phenotypes, thus, we encourage geneticists not to exclude a variant in this gene without performing further studies, both molecular and functional.

Contrary to mutations in *MECP2*, duplication of the entire gene in males is associated with well-defined MDS. Even though there are several clinical and molecular aspects of it that are still unknown, further studies with large cohorts, such as the recent ones we have discussed here, will be promising. The platform Face2Gene will surely help decrease the time for an accurate diagnosis of MDS. Moreover, the implementation of a specific MDS scale will be of great value, not only to describe the status of each patient and to establish a genotype–phenotype correlation, but also to monitor the response and evolution of future clinical trials.

Author Contributions: A.P.-A. and J.A. conceptualized the review; A.P.-A. and A.F.M.-M. wrote the original draft; A.P.-A., A.F.M.-M., C.X. and J.A. reviewed and edited the draft; and J.A. supervised. All authors reviewed the article critically for intellectual content and agreed to the published version of the manuscript.

Funding: This research was funded with a grant from the Spanish Ministry of Health (Instituto de Salud Carlos III/FEDER, PI15/01159 and PI20/00389) and a doctoral grant from the Government of Catalonia (Secretaria d'Universitats I Recerca) and the European Social Fund with the grant number 2020 FI-B 00888 and the parent association "Síndrome duplicación MECP2: Miradas que hablan" (PFNR0085).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Xinhua, B.; Shengling, J.; Fuying, S.; Hong, P.; Meirong, L.; Wu, X.R. X chromosome inactivation in rett syndrome and Its correlations with mecp2 mutations and phenotype. *J. Child Neurol.* **2008**, *23*, 22–25. [CrossRef]
2. Mnatzakanian, G.N.; Lohi, H.; Munteanu, I.; Alfred, S.E.; Yamada, T.; MacLeod, P.J.M.; Jones, J.R.; Scherer, S.W.; Schanen, N.C.; Friez, M.J.; et al. A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat. Genet.* **2004**, *36*, 339–341. [CrossRef]
3. Dastidar, S.G.; Bardai, F.H.; Ma, C.; Price, V.; Rawat, V.; Verma, P.; Narayanan, V.; D'Mello, S.R. Isoform-specific toxicity of MeCP2 in postmitotic neurons: Suppression of neurotoxicity by FoxG1. *J. Neurosci.* **2012**, *32*, 2846–2855. [CrossRef]
4. Olson, C.O.; Zachariah, R.M.; Ezeonwuka, C.D.; Liyanage, V.R.B.; Rastegar, M. Brain region-specific expression of MeCP2 isoforms correlates with DNA methylation within Mecp2 regulatory elements. *PLoS ONE* **2014**, *9*, e90645. [CrossRef]
5. Martínez De Paz, A.; Khajavi, L.; Martin, H.; Claveria-Gimeno, R.; Tom Dieck, S.; Cheema, M.S.; Sanchez-Mut, J.V.; Moksa, M.M.; Carles, A.; Brodie, N.I.; et al. MeCP2-E1 isoform is a dynamically expressed, weakly DNA-bound protein with different protein and DNA interactions compared to MeCP2-E2. *Epigenetics Chromatin* **2019**, *12*, 1–16. [CrossRef] [PubMed]
6. Lyst, M.J.; Ekiert, R.; Ebert, D.H.; Merusi, C.; Nowak, J.; Selfridge, J.; Guy, J.; Kastan, N.R.; Robinson, N.D.; De Lima Alves, F.; et al. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **2013**, *16*, 898–902. [CrossRef]
7. Sharma, K.; Singh, J.; Frost, E.E.; Pillai, P.P. MeCP2 in central nervous system glial cells: Current updates. *Acta Neurobiol. Exp.* **2018**, *78*, 30–40. [CrossRef]

8. Sandweiss, A.J.; Brandt, V.L.; Zoghbi, H.Y. Advances in understanding of Rett syndrome and MECP2 duplication syndrome: Prospects for future therapies. *Lancet Neurol.* **2020**, *19*, 689–698. [CrossRef]
9. Tillotson, R.; Bird, A. The molecular basis of MeCP2 function in the brain. *J. Mol. Biol.* **2020**, *432*, 1602–1623. [CrossRef] [PubMed]
10. Bedogni, F.; Rossi, R.L.; Galli, F.; Cobolli Gigli, C.; Gandaglia, A.; Kilstrup-Nielsen, C.; Landsberger, N. Rett syndrome and the urge of novel approaches to study MeCP2 functions and mechanisms of action. *Neurosci. Biobehav. Rev.* **2014**, *46*, 187–201. [CrossRef] [PubMed]
11. Good, K.V.; Vincent, J.B.; Ausió, J. MeCP2: The genetic driver of Rett syndrome epigenetics. *Front. Genet.* **2021**, *12*, 620859. [CrossRef] [PubMed]
12. Shahbazian, M.D.; Antalffy, B.; Armstrong, D.L.; Zoghbi, H.Y. Insight into Rett syndrome: MeCP2 levels display tissue-and cell-specific differences and correlate with neuronal maturation. *Hum. Mol. Genet.* **2002**, *11*, 115–124. [CrossRef] [PubMed]
13. Kishi, N.; Macklis, J.D. MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Mol. Cell. Neurosci.* **2004**, *27*, 306–321. [CrossRef]
14. Amir, R.E.; Van den Veyver, I.B.; Wan, M.; Tran, C.Q.; Francke, U.; Zoghbi, H. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **1999**, *23*, 185–188. [CrossRef] [PubMed]
15. Wan, M.; Sung, S.; Lee, J.; Zhang, X.; Houwink-Manville, I.; Song, H.-R.; Amir, R.E.; Budden, S.; Naidu, S.; Luiz, J.; et al. Rett Syndrome and beyond: Recurrent spontaneous and familial MECP2 mutations at CpG hotspots. *Am. J. Hum. Genet* **1999**, *65*, 1520–1529. [CrossRef]
16. Armstrong, J.; Póo, P.; Pineda, M.; Aibar, E.; Geán, E.; Català, V.; Nia Monrós, E. *Classic Rett syndrome in a Boy as a Result of Somatic Mosaicism for a MECP2 Mutation*; Wiley-Liss: Hoboken, NJ, USA, 2001.
17. Maiwald, R.; Bönte, A.; Jung, H.; Bitter, P.; Storm, Z.; Laccone, F.; Herkenrath, P. De novo MECP2 mutation in a 46, XX male patient with Rett syndrome. *Neurogenetics* **2002**, *4*, 107–108. [CrossRef]
18. Lundvall, M.; Samuelsson, L.; Kyllerman, M. Male Rett phenotypes in T158M and R294X MeCP2-mutations. *Neuropediatrics* **2006**, *37*, 296–301. [CrossRef] [PubMed]
19. Psoni, S.; Sofocleous, C.; Traeger-Synodinos, J.; Kitsiou-Tzeli, S.; Kanavakis, E.; Fryssira-Kanioura, H. Phenotypic and genotypic variability in four males with MECP2 gene sequence aberrations including a novel deletion. *Pediatr. Res.* **2010**, *67*, 551–556. [CrossRef]
20. Neul, J.L.; Fang, P.; Barrish, J.; Lane, J.; Caeg, E.B.; Smith, E.O.; Zoghbi, H.; Percy, A.; Glaze, D.G. Specific mutations in Methyl-CpG-Binding Protein 2 confer different severity in Rett syndrome. *Neurology* **2008**, *70*, 1313–1321. [CrossRef]
21. Neul, J.L.; Lane, J.B.; Lee, H.S.; Geerts, S.; Barrish, J.O.; Annese, F.; Baggett, L.M.N.; Barnes, K.; Skinner, S.A.; Motil, K.J.; et al. Developmental delay in Rett syndrome: Data from the natural history study. *J. Neurodev. Disord.* **2014**, *6*, 20. [CrossRef]
22. Couvert, P.; Bienvenu, T.; Aquaviva, C.; Poirier, K.; Moraine, C.; Gendrot, C.; Verloes, A.; Andrès, C.; Le Fevre, A.C.; Souville, I.; et al. MECP2 is highly mutated in X-linked mental retardation. *Hum. Mol. Genet.* **2001**, *10*, 941–946. [CrossRef]
23. Yntema, H.G.; Kleesfstra, T.; Oudakker, A.R.; Romein, T.; de Vries, B.B.A.; Nillesen, W.; Sistermans, E.A.; Brunner, H.G.; Hamel, B.C.J.; van Bokhoven, H. Low frequency of MECP2 mutations in mentally retarded males. *Eur. J. Hum. Genet.* **2002**, *10*, 487–490. [CrossRef]
24. Bourdon, V.; Philippe, C.; Martin, D.; Verlò, A.; Grandemenge, A.; Jonveaux, P. MECP2 mutations or polymorphisms in mentally retarded boys diagnostic implications. *Mol Diagn* **2003**, *7*, 3–7. [PubMed]
25. Moog, U.; Van Roozendaal, K.; Smeets, E.; Tserpelis, D.; Devriendt, K.; Van Buggenhout, G.; Frijns, J.P.; Schrander-Stumpel, C. MECP2 mutations are an infrequent cause of mental retardation associated with neurological problems in male patients. *Brain Dev.* **2006**, *28*, 305–310. [CrossRef] [PubMed]
26. Ravn, K.; Nielsen, J.B.; Uldall, P.; Hansen, F.J. No correlation between phenotype and genotype in boys with a truncating MECP2 mutation. *J Med Genet* **2003**, *40*, e5. [CrossRef]
27. Yntema, H.G.; Oudakker, A.R.; Kleefstra, T.; Hamel, B.C.J.; van Bokhoven, H.; Chelly, J.; Kalscheuer, V.M.; Fryns, J.P.; Raynaud, M.; Moizard, M.P.; et al. In-frame deletion in MECP2 causes mild nonspecific mental retardation. *Am. J. Med. Genet.* **2002**, *107*, 81–83. [CrossRef] [PubMed]
28. World Health Organization. *International Statistical Classification of Disease and Related Health Problems*, 10th ed. World Health Organization: Geneva, Switzerland, 1992.
29. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed.; Barber, J.P., Connolly, M.B., Crits-Christoph, P., Gladis, L., Siqueland, L., Eds.; American Psychiatric Pub: Washington, DC, USA, 1994; Volume 68, pp. 1027–1032.
30. Hagberg, B.; Hanefeld, F.; Percy, A.; Skjeldal, O. 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.* **2002**, *6*, 293–297. [CrossRef]
31. Neul, J.L.; Kaufmann, W.E.; Glaze, D.G.; Christodoulou, J.; Clarke, A.J.; Bahi-Buisson, N.; Leonard, H.; Bailey, M.E.S.; Schanen, N.C.; Zappella, M.; et al. Rett syndrome: Revised diagnostic criteria and nomenclature. *Ann. Neurol.* **2010**, *68*, 944–950. [CrossRef]
32. Monrós, E.; Armstrong, J.; Aibar, E.; Poo, P.; Canós, I.; Pineda, M. Rett syndrome in Spain: Mutation analysis and clinical correlations. *Brain Dev.* **2001**, *23*, S251–S253. [CrossRef]
33. Kerr, A.M.; Nomura, Y.; Armstrong, D.; Anvret, M.; Belichenko, P.V.; Budden, S.; Cass, H.; Christodoulou, J.; Clarke, A.; Ellaway, C.; et al. Guidelines for reporting clinical features in cases with MECP2 mutations. *Brain Dev.* **2001**, *23*, 208–211. [CrossRef]

34. Colvin, L.; Fyfe, S.; Leonard, S.; Schiavello, T.; Ellaway, C.; De Klerk, N.; Christodoulou, J.; Msall, M. Describing the phenotype in Rett syndrome using a population database. *Arch. Dis. Child.* **2003**, *88*, 38–43. [CrossRef] [PubMed]
35. Huppke, P.; Gärtner, J. Molecular diagnosis of Rett Syndrome. *J. Child Neurol.* **2005**, *20*, 732–736. [CrossRef]
36. Moretti, P.; Zoghbi, H.Y. MeCP2 dysfunction in Rett syndrome and related disorders. *Curr. Opin. Genet. Dev.* **2006**, *16*, 276–281. [CrossRef] [PubMed]
37. Villard, L. MECP2 mutations in males. *J. Med. Genet.* **2007**, *44*, 417–423. [CrossRef] [PubMed]
38. Neul, J.L. The relationship of Rett syndrome and MECP2 disorders to autism. *Dialogues Clin. Neurosci.* **2012**, *14*, 253–262. [CrossRef] [PubMed]
39. Masuyama, T.; Matsuo, M.; Jing, J.J.; Tabara, Y.; Kitsuki, K.; Yamagata, H.; Kan, Y.; Miki, T.; Ishii, K.; Kondo, I. Classic Rett syndrome in a boy with R133C mutation of MECP2. *Brain Dev.* **2005**, *27*, 439–442. [CrossRef] [PubMed]
40. Neul, J.L.; Benke, T.A.; Marsh, E.D.; Skinner, S.A.; Merritt, J.; Lieberman, D.N.; Standridge, S.; Feyma, T.; Heydemann, P.; Peters, S.; et al. The array of clinical phenotypes of males with mutations in Methyl-CpG binding protein 2. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **2019**, *180*, 55–67. [CrossRef] [PubMed]
41. Buyse, I.M.; Fang, P.; Hoon, K.T.; Amir, R.E.; Zoghbi, H.Y.; Roa, B.B. Diagnostic testing for Rett Syndrome by DHPLC and direct sequencing analysis of the MECP2 Gene: Identification of several novel mutations and polymorphisms. *Am. J. Hum. Genet* **2000**, *67*, 1428–1436. [CrossRef]
42. Bianciardi, L.; Fichera, M.; Failla, P.; Di Marco, C.; Grozeva, D.; Mencarelli, M.A.; Spiga, O.; Mari, F.; Meloni, I.; Raymond, L.; et al. MECP2 missense mutations outside the canonical MBD and TRD domains in males with intellectual disability. *J. Hum. Genet.* **2016**, *61*, 95–101. [CrossRef]
43. Grozeva, D.; Carss, K.; Spasic-Boskovic, O.; Parker, M.J.; Archer, H.; Firth, H.V.; Park, S.M.; Canham, N.; Holder, S.E.; Wilson, M.; et al. De novo loss-of-function mutations in SETD5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. *Am. J. Hum. Genet.* **2014**, *94*, 618–624. [CrossRef]
44. Vidal, S.; Brandi, N.; Pacheco, P.; Maynou, J.; Fernandez, G.; Xiol, C.; Pascual-Alonso, A.; Pineda, M.; O'Callaghan, M.; Garcia-Cazorla, À.; et al. The most recurrent monogenic disorders that overlap with the phenotype of Rett syndrome. *Eur. J. Paediatr. Neurol.* **2019**, *23*, 609–620. [CrossRef] [PubMed]
45. Schönewolf-Greulich, B.; Bisgaard, A.M.; Dunø, M.; Jespersgaard, C.; Rokkjær, M.; Hansen, L.K.; Tsoutsou, E.; Sofokleous, C.; Topcu, M.; Kaur, S.; et al. Mosaic MECP2 variants in males with classical Rett syndrome features, including stereotypical hand movements. *Clin. Genet.* **2019**, *95*, 403–408. [CrossRef]
46. Gomot, M.; Gendrot, C.; Verloes, A.; Raynaud, M.; David, A.; Yntema, H.G.; Dessay, S.; Kalscheuer, V.; Frints, S.; Couvert, P.; et al. MECP2 gene mutations in non-syndromic X-linked mental retardation: Phenotype-genotype correlation. *Am. J. Med. Genet.* **2003**, *123*, 129–139. [CrossRef]
47. Sanlaville, D.; Prieur, M.; de Blois, M.C.; Genevieve, D.; Lapierre, J.M.; Ozilou, C.; Picq, M.; Gosset, P.; Morichon-Delvallez, N.; Munnich, A.; et al. Functional disomy of the Xq28 chromosome region. *Eur. J. Hum. Genet.* **2005**, *13*, 579–585. [CrossRef] [PubMed]
48. Lubs, H.; Abidi, F.; Blaymore Bier, J.-A.; Abuelo, D.; Ouzts, L.; Voeller, K.; Fennell, E.; Stevenson, R.E.; Schwartz, C.E.; Arena, F. XLMR syndrome characterized by multiple respiratory infections, hypertelorism, severe CNS deterioration and early death localizes to distal Xq28. *J. Med. Genet* **1999**, *85*, 243–248. [CrossRef]
49. Friez, M.J.; Jones, J.R.; Clarkson, K.; Lubs, H.; Abuelo, D.; Bier, J.A.B.; Pai, S.; Simensen, R.; Williams, C.; Giampietro, P.F.; et al. Recurrent infections, hypotonia, and mental retardation caused by duplication of MECP2 and adjacent region in Xq28. *Pediatrics* **2006**, *118*, e1687–e1695. [CrossRef] [PubMed]
50. Meins, M.; Lehmann, J.; Gerresheim, F.; Herchenbach, J.; Hagedorn, M.; Hameister, K.; Epplen, J.T. 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.* **2005**, *42*, e12. [CrossRef] [PubMed]
51. Van Esch, H.; Bauters, M.; Ignatius, J.; Jansen, M.; Raynaud, M.; Hollanders, K.; Lugtenberg, D.; Bienvenu, T.; Jensen, L.R.; Gecz, J.; et al. Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am. J. Hum. Genet.* **2005**, *77*, 442–453. [CrossRef]
52. Shao, L.; Shaw, C.A.; Lu, X.Y.; Sahoo, T.; Bacino, C.A.; Lalani, S.R.; Stankiewicz, P.; Yatsenko, S.A.; Li, Y.; Neill, S.; et al. Identification of chromosome abnormalities in subtelomeric regions by microarray analysis: A study of 5380 cases. *Am. J. Med. Genet. Part A* **2008**, *146A*, 2242–2251. [CrossRef]
53. Ramocki, M.B.; Tavyev, Y.J.; Peters, S.U. The MECP2 duplication syndrome. *Am. J. Med. Genet. Part A* **2010**, *152A*, 1079–1088. [CrossRef]
54. de Brouwer, A.P.M.; Yntema, H.G.; Kleefstra, T.; Lugtenberg, D.; Oudakker, A.R.; de Vries, B.B.A.; van Bokhoven, H.; Van Esch, H.; Frints, S.G.M.; Froyen, G.; et al. Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium. *Hum. Mutat.* **2007**, *28*, 207–208. [CrossRef] [PubMed]
55. Lugtenberg, D.; Kleefstra, T.; Oudakker, A.R.; Nillesen, W.M.; Yntema, H.G.; Tzschach, A.; Raynaud, M.; Rating, D.; Journel, H.; Chelly, J.; et al. Structural variation in Xq28: MECP2 duplications in 1% of patients with unexplained XLMR and in 2% of male patients with severe encephalopathy. *Eur. J. Hum. Genet.* **2009**, *17*, 444–453. [CrossRef] [PubMed]

56. Honda, S.; Hayashi, S.; Nakane, T.; Imoto, I.; Kurosawa, K.; Mizuno, S.; Okamoto, N.; Kato, M.; Yoshihashi, H.; Kubota, T.; et al. The incidence of hypoplasia of the corpus callosum in patients with dup (X)(q28) involving MECP2 is associated with the location of distal breakpoints. *Am. J. Med. Genet. Part A* **2012**, *158A*, 1292–1303. [CrossRef] [PubMed]
57. 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. Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet. Med.* **2006**, *8*, 784–792. [CrossRef]
58. Campos, M.; Churchman, S.M.; Santos-Rebouças, C.B.; Ponchel, F.; Pimentel, M.M.G. High frequency of nonrecurrent MECP2 duplications among Brazilian males with mental retardation. *J. Mol. Neurosci.* **2010**, *41*, 105–109. [CrossRef] [PubMed]
59. Lugtenberg, D.; De Brouwer, A.P.M.; Kleefstra, T.; Oudakker, A.R.; Frints, S.G.M.; Schrander-Stumpel, C.T.R.M.; Fryns, J.P.; Jensen, L.R.; Chelly, J.; Moraine, C.; et al. Chromosomal copy number changes in patients with non-syndromic X linked mental retardation detected by array CGH. *J. Med. Genet.* **2006**, *43*, 362–370. [CrossRef] [PubMed]
60. Rosenberg, C.; Knijnenburg, J.; Bakker, E.; Vianna-Morgante, A.M.; Sloos, W.; Otto, P.A.; Kriek, M.; Hansson, K.; Krepischi-Santos, A.C.V.; Fiegler, H.; et al. Array-CGH detection of micro rearrangements in mentally retarded individuals: Clinical significance of imbalances present both in affected children and normal parents. *J. Med. Genet.* **2006**, *43*, 180–186. [CrossRef] [PubMed]
61. Madrigal, I.; Rodríguez-Revenga, L.; Armengol, L.; González, E.; Rodriguez, B.; Badenas, C.; Sánchez, A.; Martínez, F.; Guitart, M.; Fernández, I.; et al. X-chromosome tiling path array detection of copy number variants in patients with chromosome X-linked mental retardation. *BMC Genom.* **2007**, *8*, 443. [CrossRef] [PubMed]
62. Bauters, M.; Van Esch, H.; Friez, M.J.; Boespflug-Tanguy, O.; Zenker, M.; Vianna-Morgante, A.M.; Rosenberg, C.; Ignatius, J.; Raynaud, M.; Hollanders, K.; et al. Nonrecurrent MECP2 duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair. *Genome Res.* **2008**, *18*, 847–858. [CrossRef]
63. Smyk, M.; Obersztyn, E.; Nowakowska, B.; Nawara, M.; Cheung, S.W.; Mazurczak, T.; Stankiewicz, P.; Bocian, E. Different-sized duplications of Xq28, including MECP2, in three males with mental retardation, absent or delayed speech, and recurrent infections. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **2008**, *147B*, 799–806. [CrossRef]
64. Velinov, M.; Novelli, A.; Gu, H.; Fenko, M.; Dolzhanskaya, N.; Bernardini, L.; Capalbo, A.; Dallapiccola, B.; Jenkins, E.C.; Brown, W.T. De-novo 2.15 Mb terminal Xq duplication involving MECP2 but not L1CAM gene in a male patient with mental retardation. *Clin. Dysmorphol.* **2009**, *18*, 9–12. [CrossRef]
65. Kirk, E.P.; Malaty-Brevaud, V.; Martini, N.; Lacoste, C.; Levy, N.; Maclean, K.; Davies, L.; Philip, N.; Badens, C. The clinical variability of the MECP2 duplication syndrome: Description of two families with duplications excluding L1CAM and FLNA. *Clin. Genet.* **2009**, *75*, 301–303. [CrossRef]
66. Clayton-Smith, J.; Walters, S.; Hobson, E.; Burkitt-Wright, E.; Smith, R.; Toutain, A.; Amiel, J.; Lyonnet, S.; Mansour, S.; Fitzpatrick, D.; et al. Xq28 duplication presenting with intestinal and bladder dysfunction and a distinctive facial appearance. *Eur. J. Hum. Genet.* **2009**, *17*, 434–443. [CrossRef] [PubMed]
67. Prescott, T.E.; Rødningen, O.K.; Bjørnstad, A.; Stray-Pedersen, A. Two brothers with a microduplication including the MECP2 gene: Rapid head growth in infancy and resolution of susceptibility to infection. *Clin. Dysmorphol.* **2009**, *18*, 78–82. [CrossRef]
68. Carvalho, C.M.B.B.; Zhang, F.; Liu, P.; Patel, A.; Sahoo, T.; Bacino, C.A.; Shaw, C.; Peacock, S.; Pursley, A.; Tavyev, J.Y.; et al. Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching. *Hum. Mol. Genet.* **2009**, *18*, 2188–2203. [CrossRef] [PubMed]
69. Echenne, B.; Roubertie, A.; Lugtenberg, D.; Kleefstra, T.; Hamel, B.C.J.; Van Bokhoven, H.; Lacombe, D.; Philippe, C.; Jonveaux, P.; de Brouwer, A.P.M. Neurologic aspects of MECP2 gene duplication in male patients. *Pediatr. Neurol.* **2009**, *41*, 187–191. [CrossRef]
70. Ramocki, M.B.; Peters, S.U.; Tavyev, Y.J.; Zhang, F.; Carvalho, C.M.B.; Schaaf, C.P.; Richman, R.; Fang, P.; Glaze, D.G.; Lupski, J.R.; et al. Autism and other neuropsychiatric symptoms are prevalent in individuals with MECP2 duplication syndrome. *Ann. Neurol.* **2009**, *66*, 771–782. [CrossRef] [PubMed]
71. Belligni, E.F.; Palmer, R.W.; Hennekam, R.C.M. MECP2 duplication in a patient with congenital central hypoventilation. *Am. J. Med. Genet. Part A* **2010**, *152*, 1591–1593. [CrossRef]
72. Bartsch, O.; Gebauer, K.; Lechno, S.; Van Esch, H.; Froyen, G.; Bonin, M.; Seidel, J.; Thamm-Mücke, B.; Horn, D.; Klopocki, E.; et al. Four unrelated patients with Lubs X-linked mental retardation syndrome and different Xq28 duplications. *Am. J. Med. Genet. Part A* **2010**, *152*, 305–312. [CrossRef]
73. Reardon, W.; Donoghue, V.; Murphy, A.M.; King, M.D.; Mayne, P.D.; Horn, N.; Birk Møller, L. Progressive cerebellar degenerative changes in the severe mental retardation syndrome caused by duplication of MECP2 and adjacent loci on Xq28. *Eur. J. Pediatr.* **2010**, *169*, 941–949. [CrossRef] [PubMed]
74. Makrythanasis, P.; Moix, I.; Gimelli, S.; Fluss, J.; Aliferis, K.; Antonarakis, S.E.; Morris, M.A.; Béna, F.; Bottani, A. De novo duplication of MECP2 in a girl with mental retardation and no obvious dysmorphic features. *Clin. Genet.* **2010**, *78*, 175–180. [CrossRef]
75. Honda, S.; Hayashi, S.; Imoto, I.; Toyama, J.; Okazawa, H.; Nakagawa, E.; Goto, Y.I.; Inazawa, J. Copy-number variations on the X chromosome in Japanese patients with mental retardation detected by array-based comparative genomic hybridization analysis. *J. Hum. Genet.* **2010**, *55*, 590–599. [CrossRef]
76. Fernández, R.M.; Núñez-Torres, R.; González-Meneses, A.; Antiñolo, G.; Borrego, S. Novel association of severe neonatal encephalopathy and Hirschsprung disease in a male with a duplication at the Xq28 region. *BMC Med. Genet.* **2010**, *11*, 137. [CrossRef]

77. Jezela-Stanek, A.; Ciara, E.; Juszczak, M.; Pelc, M.; Materna-Kiryluk, A.; Krajewska-Walasek, M.; Cryptic, X. Autosome translocation in a boy—Delineation of the phenotype. *Pediatr. Neurol.* **2011**, *44*, 221–224. [CrossRef]
78. Breman, A.M.; Ramocki, M.B.; Kang, S.H.L.; Williams, M.; Freedenberg, D.; Patel, A.; Bader, P.I.; Cheung, S.W. MECP2 duplications in six patients with complex sex chromosome rearrangements. *Eur. J. Hum. Genet.* **2011**, *19*, 409–415. [CrossRef] [PubMed]
79. Grasshoff, U.; Bonin, M.; Goehring, I.; Ekici, A.; Dufke, A.; Cremer, K.; Wagner, N.; Rossier, E.; Jauch, A.; Walter, M.; et al. De novo MECP2 duplication in two females with random X-inactivation and moderate mental retardation. *Eur. J. Hum. Genet.* **2011**, *19*, 507–512. [CrossRef]
80. Budisteanu, M.; Papuc, S.M.; Tutulan-Cunita, A.; Budisteanu, B.; Arghir, A. Novel clinical finding in MECP2 duplication syndrome. *Eur. Child Adolesc. Psychiatry* **2011**, *20*, 373–375. [CrossRef]
81. Mayo, S.; Monfort, S.; Roselló, M.; Orellana, C.; Oltra, S.; Armstrong, J.; Català, V.; Martínez, F. De novo interstitial triplication of MECP2 in a girl with neurodevelopmental disorder and random X chromosome inactivation. *Cytogenet. Genome Res.* **2011**, *135*, 93–101. [CrossRef] [PubMed]
82. Carvalho, C.M.B.; Ramocki, M.B.; Pehlivan, D.; Franco, L.M.; Gonzaga-Jauregui, C.; Fang, P.; McCall, A.; Pivnick, E.K.; Hines-Dowell, S.; Seaver, L.H.; et al. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat. Genet.* **2011**, *43*, 1074–1081. [CrossRef] [PubMed]
83. Utine, G.E.; Kiper, P.Ö.; Alanay, Y.; Haliloğlu, G.; Aktaş, D.; Boduroğlu, K.; Tunçbilek, E.; Alikaşifoğlu, M. Searching for copy number changes in nonsyndromic X-linked intellectual disability. *Mol. Syndromol.* **2012**, *2*, 64–71. [CrossRef]
84. Tang, S.S.; Fernandez, D.; Lazarou, L.P.; Singh, R.; Fallon, P. MECP2 triplication in 3 brothers—A rarely described cause of familial neurological regression in boys. *Eur. J. Paediatr. Neurol.* **2012**, *16*, 209–212. [CrossRef] [PubMed]
85. Honda, S.; Satomura, S.; Hayashi, S.; Imoto, I.; Nakagawa, E.; Goto, Y.I.; Inazawa, J. Concomitant microduplications of MECP2 and ATRX in male patients with severe mental retardation. *J. Hum. Genet.* **2012**, *57*, 73–77. [CrossRef]
86. Bijlsma, E.K.; Collins, A.; Papa, F.T.; Tejada, M.I.; Wheeler, P.; Peeters, E.A.J.; Gijsbers, A.C.J.; van de Kamp, J.M.; Kriek, M.; Losekoot, M.; et al. Xq28 duplications including MECP2 in five females: Expanding the phenotype to severe mental retardation. *Eur. J. Med. Genet.* **2012**, *55*, 404–413. [CrossRef]
87. Sanmann, J.N.; Bishay, D.L.; Starr, L.J.; Bell, C.A.; Pickering, D.L.; Stevens, J.M.; Kahler, S.G.; Olney, A.H.; Schaefer, G.B.; Sanger, W.G. Characterization of six novel patients with MECP2 duplications due to unbalanced rearrangements of the X chromosome. *Am. J. Med. Genet. Part A* **2012**, *158A*, 1285–1291. [CrossRef]
88. Vignoli, A.; Borgatti, R.; Peron, A.; Zucca, C.; Ballarati, L.; Bonaglia, C.; Bellini, M.; Giordano, L.; Romaniello, R.; Bedeschi, M.F.; et al. Electroclinical pattern in MECP2 duplication syndrome: Eight new reported cases and review of literature. *Epilepsia* **2012**, *53*, 1146–1155. [CrossRef]
89. Xu, X.; Xu, Q.; Zhang, Y.; Zhang, X.; Cheng, T.; Wu, B.; Ding, Y.; Lu, P.; Zheng, J.; Zhang, M.; et al. A case report of Chinese brothers with inherited MECP2-containing duplication: Autism and intellectual disability, but not seizures or respiratory infections. *BMC Med. Genet.* **2012**, *13*, 75. [CrossRef]
90. Yang, T.; Ramocki, M.B.; Neul, J.L.; Lu, W.; Roberts, L.; Knight, J.; Ward, C.S.; Zoghbi, H.Y.; Kheradmand, F.; Corry, D.B. Overexpression of methyl-CpG binding protein 2 impairs TH1 responses. *Sci. Transl. Med.* **2012**, *4*, 163ra158. [CrossRef] [PubMed]
91. Shimada, S.; Okamoto, N.; Ito, M.; Arai, Y.; Momosaki, K.; Togawa, M.; Maegaki, Y.; Sugawara, M.; Shimojima, K.; Osawa, M.; et al. MECP2 duplication syndrome in both genders. *Brain Dev.* **2013**, *35*, 411–419. [CrossRef] [PubMed]
92. Shimada, S.; Okamoto, N.; Hirasawa, K.; Yoshii, K.; Tani, Y.; Sugawara, M.; Shimojima, K.; Osawa, M.; Yamamoto, T. Clinical manifestations of Xq28 functional disomy involving MECP2 in one female and two male patients. *Am. J. Med. Genet. Part A* **2013**, *161A*, 1779–1785. [CrossRef]
93. Wax, J.R.; Pinette, M.G.; Smith, R.; Chard, R.; Cartin, A. Second-trimester prenasal and prefrontal skin thickening-Association with MECP2 triplication syndrome. *J. Clin. Ultrasound* **2013**, *41*, 434–437. [CrossRef] [PubMed]
94. Peters, S.U.; Hundley, R.J.; Wilson, A.K.; Carvalho, C.M.B.; Lupski, J.R.; Ramocki, M.B. Brief report: Regression timing and associated features in MECP2 duplication syndrome. *J. Autism Dev. Disord.* **2013**, *43*, 2484–2490. [CrossRef]
95. Scott Schwoerer, J.; Laffin, J.; Haun, J.; Raca, G.; Friez, M.J.; Giampietro, P.F. MECP2 duplication: Possible cause of severe phenotype in females. *Am. J. Med. Genet. Part A* **2014**, *164A*, 1029–1034. [CrossRef] [PubMed]
96. Novara, F.; Simonati, A.; Sicca, F.; Battini, R.; Fiori, S.; Contaldo, A.; Criscuolo, L.; Zuffardi, O.; Ciccone, R. MECP2 duplication phenotype in symptomatic females: Report of three further cases. *Mol. Cytogenet.* **2014**, *7*, 10. [CrossRef]
97. Fukushi, D.; Yamada, K.; Nomura, N.; Naiki, M.; Kimura, R.; Yamada, Y.; Kumagai, T.; Yamaguchi, K.; Miyake, Y.; Wakamatsu, N. Clinical characterization and identification of duplication breakpoints in a Japanese family with Xq28 duplication syndrome including MECP2. *Am. J. Med. Genet. Part A* **2014**, *164A*, 924–933. [CrossRef] [PubMed]
98. Bauer, M.; Kölsch, U.; Krüger, R.; Unterwalder, N.; Hameister, K.; Kaiser, F.M.; Vignoli, A.; Rossi, R.; Botella, M.P.; Budisteanu, M.; et al. Infectious and immunologic phenotype of MECP2 Duplication syndrome. *J. Clin. Immunol.* **2015**, *35*, 168–181. [CrossRef]
99. Miyatake, S.; Koshimizu, E.; Fujita, A.; Fukai, R.; Imagawa, E.; Ohba, C.; Kuki, I.; Nukui, M.; Araki, A.; Makita, Y.; et al. Detecting copy-number variations in whole-exome sequencing data using the exome hidden markov model: An "exome-first" approach. *J. Hum. Genet.* **2015**, *60*, 175–182. [CrossRef]
100. Trobaugh-Lotrario, A.; Martin, J.; López-Terrada, D. Hepatoblastoma in a male with MECP2 duplication syndrome. *Am. J. Med. Genet. Part A* **2016**, *170A*, 790–791. [CrossRef] [PubMed]
101. Zhang, Q.; Zhao, Y.; Yang, Y.; Bao, X. MECP2 duplication syndrome in a Chinese family. *BMC Med. Genet.* **2015**, *16*, 112. [CrossRef]

102. El Chehadeh, S.; Faivre, L.; Mosca-Boidron, A.L.; Malan, V.; Amiel, J.; Nizon, M.; Touraine, R.; Prieur, F.; Pasquier, L.; Callier, P.; et al. Large national series of patients with Xq28 duplication involving MECP2: Delineation of brain MRI abnormalities in 30 affected patients. *Am. J. Med. Genet. Part A* **2016**, *170A*, 116–129. [CrossRef]
103. Nageshappa, S.; Carromeu, C.; Trujillo, C.A.; Mesci, P.; Espuny-Camacho, I.; Pasciuto, E.; Vanderhaeghen, P.; Verfaillie, C.M.; Raitano, S.; Kumar, A.; et al. Altered neuronal network and rescue in a human MECP2 duplication model. *Mol. Psychiatry* **2016**, *21*, 178–188. [CrossRef]
104. Signorini, C.; De Felice, C.; Leoncini, S.; Møller, R.S.; Zollo, G.; Buoni, S.; Cortelazzo, A.; Guerranti, R.; Durand, T.; Ciccoli, L.; et al. MECP2 duplication syndrome: Evidence of enhanced oxidative stress. A comparison with Rett syndrome. *PLoS ONE* **2016**, *11*, e0150101. [CrossRef] [PubMed]
105. Yi, Z.; Pan, H.; Li, L.; Wu, H.; Wang, S.; Ma, Y.; Qi, Y. Chromosome Xq28 duplication encompassing MECP2: Clinical and molecular analysis of 16 new patients from 10 families in China. *Eur. J. Med. Genet.* **2016**, *59*, 347–353. [CrossRef] [PubMed]
106. San Antonio-Arce, V.; Fenollar-Cortés, M.; Ionescu, R.O.; DeSantos-Moreno, T.; Gallego-Merlo, J.; Cámara, F.J.I.; Pérez, M.C.O. MECP2 Duplications in symptomatic females. *Child Neurol. Open* **2016**, *3*. [CrossRef] [PubMed]
107. Tsuji-Hosokawa, A.; Matsuda, N.; Kurosawa, K.; Kashimada, K.; Morio, T. A case of MECP2 duplication syndrome with gonadotropin-dependent precocious puberty. *Horm. Res. Paediatr.* **2017**, *87*, 271–276. [CrossRef]
108. Ha, K.; Shen, Y.; Graves, T.; Kim, C.H.; Kim, H.G. The presence of two rare genomic syndromes, 1q21 deletion and Xq28 duplication, segregating independently in a family with intellectual disability. *Mol. Cytogenet.* **2016**, *9*, 74. [CrossRef]
109. Lim, Z.; Downs, J.; Wong, K.; Ellaway, C.; Leonard, H. Expanding the clinical picture of the MECP2 Duplication syndrome. *Clin. Genet.* **2017**, *91*, 557–563. [CrossRef]
110. El Chehadeh, S.; Touraine, R.; Prieur, F.; Reardon, W.; Bienvenu, T.; Chantot-Bastaraud, S.; Doco-Fenzy, M.; Landais, E.; Philippe, C.; Marle, N.; et al. Xq28 duplication including MECP2 in six unreported affected females: What can we learn for diagnosis and genetic counselling? *Clin. Genet.* **2017**, *91*, 576–588. [CrossRef]
111. Moirangthem, A.; Tuteja Bhatia, M.; Srivastava, P.; Mandal, K.; Rai, A.; Phadke, S.R. Expansion of the phenotypic spectrum in three families of methyl CpG-binding protein 2 duplication syndrome. *Clin. Dysmorphol.* **2017**, *26*, 73–77. [CrossRef]
112. Yon, D.K.; Park, J.E.; Kim, S.J.; Shim, S.H.; Chae, K.Y. A sibship with duplication of Xq28 inherited from the mother; genomic characterization and clinical outcomes. *BMC Med. Genet.* **2017**, *18*, 1–9. [CrossRef]
113. Li, X.; Xie, H.; Chen, Q.; Yu, X.; Yi, Z.; Li, E.; Zhang, T.; Wang, J.; Zhong, J.; Chen, X. Clinical and molecular genetic characterization of familial MECP2 duplication syndrome in a Chinese family. *BMC Med. Genet.* **2017**, *18*, 131. [CrossRef]
114. Deshwar, A.R.; Dupuis, L.; Bergmann, C.; Stavropoulos, J.; Mendoza-Londono, R. Severe rhizomelic shortening in a child with a complex duplication/deletion rearrangement of chromosome X. *Am. J. Med. Genet. Part A* **2018**, *176A*, 450–454. [CrossRef] [PubMed]
115. Bauer, M.; Krüger, R.; Kölsch, U.; Unterwalder, N.; Meisel, C.; Wahn, V.; Von Bernuth, H. Antibiotic prophylaxis, immunoglobulin substitution and supportive measures prevent infections in MECP2 duplication syndrome. *Pediatr. Infect. Dis. J.* **2018**, *37*, 466–468. [CrossRef]
116. Miguet, M.; Faivre, L.; Amiel, J.; Nizon, M.; Touraine, R.; Prieur, F.; Pasquier, L.; Lefebvre, M.; Thevenon, J.; Dubourg, C.; et al. Further delineation of the MECP2 duplication syndrome phenotype in 59 French male patients, with a particular focus on morphological and neurological features. *J. Med. Genet.* **2018**, *55*, 359–371. [CrossRef] [PubMed]
117. Pitzianti, M.B.; Palombo, A.S.; Esposito, S.; Pasini, A. Rett syndrome in males: The different clinical course in two brothers with the same microduplication MECP2 Xq28. *Int. J. Environ. Res. Public Health* **2019**, *16*, 3075. [CrossRef]
118. Kanai, S.; Okanishi, T.; Fujimoto, A.; Itamura, S.; Baba, S.; Nishimura, M.; Itomi, K.; Enoki, H. Successful corpus callosotomy for post-encephalopathic refractory epilepsy in a patient with MECP2 duplication syndrome. *Brain Dev.* **2019**, *41*, 296–300. [CrossRef] [PubMed]
119. Marafi, D.; Suter, B.; Schultz, R.; Glaze, D.; Pavlik, V.N.; Goldman, A.M. Spectrum and time course of epilepsy and the associated cognitive decline in MECP2 duplication syndrome. *Neurology* **2019**, *92*, E108–E114. [CrossRef]
120. Lotti, F.; Geronzi, U.; Grosso, S. Electroencephalographic and epilepsy findings in mecp2 duplication syndrome. A family study. *Brain Dev.* **2019**, *41*, 456–459. [CrossRef]
121. Giudice-Nairn, P.; Downs, J.; Wong, K.; Wilson, D.; Ta, D.; Gattas, M.; Amor, D.; Thompson, E.; Kirrali-Borri, C.; Ellaway, C.; et al. The incidence, prevalence and clinical features of MECP2 duplication syndrome in Australian children. *J. Paediatr. Child Health* **2019**, *55*, 1315–1322. [CrossRef] [PubMed]
122. Peters, S.U.; Fu, C.; Suter, B.; Marsh, E.; Benke, T.A.; Skinner, S.A.; Lieberman, D.N.; Standridge, S.; Jones, M.; Beisang, A.; et al. Characterizing the phenotypic effect of Xq28 duplication size in MECP2 duplication syndrome. *Clin. Genet.* **2019**, *95*, 575–581. [CrossRef]
123. Pascual-Alonso, A.; Blasco, L.; Vidal, S.; Gean, E.; Rubio, P.; O'Callaghan, M.; Martínez-Monseny, A.F.; Castells, A.A.; Xiol, C.; Català, V.; et al. Molecular characterization of Spanish patients with MECP2 duplication syndrome. *Clin. Genet.* **2020**, *97*, 610–620. [CrossRef]
124. Gutiérrez-Sánchez, A.M.; Marín-Andrés, M.; López-Lafuente, A.; Monge-Galindo, L.; López-Pisón, J.; Peña-Segura, J.L. Síndrome de duplicación MECP2 familiar. *Rev. Neurol.* **2020**, *70*, 309–310. [CrossRef] [PubMed]

Publication 4: Further study of males carrying variants in *MECP2*: beyond Rett syndrome.

Authors: Ainhoa Pascual-Alonso[1,2], Noelia Rodríguez[1,2], Mar O'Callaghan[3], Juan Darío Ortigoza[3] and Judith Armstrong[2,4,5].

Reference: Manuscript in preparation.

Abstract:

Variants in *MECP2* can cause a wide variety of phenotypes from the well known RTT or MDS to more subtle X-linked intellectual disabilities. Since the phenotypes that variants in *MECP2* can cause are not always precise, classifying a variant in that gene is not always straightforward. A combination of the ACMG guidelines, *in silico* prediction tools and molecular and functional studies is often needed to assign a causality to a variant. The implementation of NGS derived methodologies, such as clinical exome sequencing, has enabled the finding of variants in genes that otherwise might not have been interrogated for sequencing due to the presenting clinical traits.

For the present manuscript we gathered seven males with variants in *MECP2* and different degrees of neurodevelopmental delay and ID. In four cases the variants were inherited from their asymptomatic mothers. DNA was extracted from blood leukocytes to confirm the variants. Besides, fibroblast cell lines were established from skin biopsies of five probands, four carrier mothers and one affected sister. DNA, RNA and protein were extracted from the fibroblasts to perform the molecular studies. We confirmed all the variants in DNA, measured the expression levels of the two isoforms of *MECP2*, checked the exon junctions sequence and the generation of new isoforms and quantified MeCP2 protein levels. XCI was tested in female samples. The obtained data aided us with the reclassification of six out of the seven variants. In fact, studying the carrier mothers proved to be useful for the variant classification and it should be recommended for genes located in the X chromosome, since XCI could be the reason for the carrier's apparently asymptomatic phenotype.

Further study of males carrying variants in *MECP2*: beyond Rett syndrome.

A. Pascual-Alonso[1,2], N. Rodriguez[1,2], M. O'Callaghan[3], J.D. Ortigoza[3] J. Armstrong[2,4,5]

1. Sant Joan de Déu Research Foundation, Esplugues de Lobregat (Barcelona), Spain.
2. Sant Joan de Déu Research Institute (IRSJD), Hospital Sant Joan de Déu, Esplugues de Lobregat (Barcelona), Spain.
3. Neurology Section, Hospital Sant Joan de Déu, Esplugues de Lobregat (Barcelona), Spain.
4. Molecular and Genetics Medicine Section, Hospital Sant Joan de Déu, Esplugues de Lobregat (Barcelona), Spain.
5. CIBER-ER (Biomedical Network Research Center for Rare Diseases), Instituto de Salud Carlos III, Madrid, Spain.

0. Abstract

Methyl-CpG binding protein 2 (*MECP2)* is a multifunctional gene located in the region Xq28 that undergoes X chromosome inactivation (XCI) in females. Variants in *MECP2* in females usually cause Rett syndrome (RTT; OMIM#312750) whereas in males trigger a broad spectrum of phenotypes from congenital encephalopathies to mild intellectual disability (ID), and sometimes, even RTT. Since the phenotypes that can cause *MECP2* variants in males are not specific, the correlation of a variant with a diagnosis is not always straightforward. Variant classification using the ACMG guidelines, *in silico* prediction tools and molecular and functional studies is advised. Here, we present seven male cases with seven different variants in *MECP2*. In four cases, the variants are inherited from apparently asymptomatic mothers. Skin biopsies were obtained from five probands, four carrier mothers and an affected sister and primary fibroblast cell lines were established. DNA, RNA and proteins were extracted from the fibroblasts, and DNA was also extracted from blood leukocytes. All the variants were validated in DNA and RNA samples. mRNA and protein levels of *MECP2* have been quantified and the exon junctions have been checked. In addition, XCI was performed in the carrier mothers. As a result, new evidence related to those variants' effect has been obtained and we could reclassify six out of the seven variants. Molecular studies of the probands and the carrier mothers

are recommended for genes located in the X chromosome, since XCI could be explaining the asymptomatic phenotype of the female carriers.

Keywords: *MECP2*; variant classification; Rett syndrome.

Abbreviations: X chromosome inactivation (XCI), Rett syndrome (RTT), methyl-binding domain (MBD), transcriptional repression domain (TRD), intellectual disability (ID), quantitative reverse transcription PCR (qRT-PCR), genome Aggregation Database (gnomAD), variant of uncertain significance (VUS), autism spectrum disorder (ASD).

1. Introduction

Methyl CpG binding protein 2 (*MECP2)* gene (OMIM*300005) is a multifunctional gene located in the X chromosome (Xq28). *MECP2* has four exons and generates two isoforms, *MECP2_e1* (that contains exons 1, 3 and 4) and *MECP2_e2* (containing exons 2, 3 and 4) (1). The coding protein, MeCP2, possesses two main functional domains, the methyl-binding domain (MBD) and the transcriptional repression domain (TRD). MBD and TRD are the regions where the majority of the disease-causing mutations are located (2). MeCP2 acts as both transcription activator and repressor (3). Besides, it is involved in microRNA processing and in splicing regulation (4,5).

The malfunction of MeCP2 is related to several disorders such as X-linked mental retardation syndrome (OMIM #300055), severe neonatal encephalopathy (OMIM #300673), autism susceptibility (OMIM #300496), *MECP2* duplication syndrome (OMIM #300260) and, especially, to Rett syndrome (RTT; OMIM #312750). RTT is a neurodevelopmental disorder with an incidence of 1:10.000-15.000 that occurs mostly in women, although few male cases have been reported. RTT is characterised by a period of apparent normal development until 6-18 months of age followed by a regression in which arrest of psychomotor development happens. Consequently, partial or complete loss of previously acquired skills such as speech or ambulation occur and stereotypic movements and seizures appear (6).

Less than 10% of the registered entries in the RettBASE database are male cases (2). The disparity of cases for each sex together with the difficulties in creating the first *Mecp2* male mouse model suggested that mutations in *MECP2* in males were lethal. However, several publications of male cases have revealed that

males with mutations in *MECP2* show a wide variety of phenotypes, from severe encephalopathies leading to premature deaths to patients with mild intellectual disability (ID) and autistic features (7,8). The thorough characterisation of these boys revealed that a few cases of males with RTT exist, but that other disorders should be considered.

Males with variants in *MECP2* can be classified in four groups: i) classic or atypical RTT (whenever consensus criteria for each RTT type are fulfilled) (9); ii) neonatal encephalopathy or iii) progressive encephalopathy (depending on the moment the clinical features emerge) and iv) cognitive impairment (when no worsening of the phenotype is seen) (10).

The implementation of next generation sequencing as a diagnostic tool has enabled the finding of candidate variants in *MECP2* for children that otherwise would not have been screened for the gene and would remain undiagnosed. Some of the detected variants had already been found in girls with RTT, others might look pathological due to the effect they generate in MeCP2. But there are other variants in which no straightforward relationship can be seen with the phenotype of the patients, and *in silico* studies and functional experiments should be performed in order to gain information about the variant's pathogenicity. There are more than 900 variants reported in *MECP2* in public databases (2) but, even though the ACMG guidelines have become a useful tool to aid with variant classification, still, some of *MECP2's* variants' significance remains unclear (11).

Here we present seven cases with seven different variants in the gene *MECP2*. In four out of seven cases variants were inherited from the mothers who are apparently asymptomatic. By combining the analysis of different molecular techniques and some *in silico* tools we tried to understand if those variations are the reason for the altered phenotype our probands show.

2. Material and Methods

2.1 Samples

Written informed consent was obtained from the patients and their legally responsible tutors following appropriate ethic protocols for the analysis of genes related to RTT.

DNA was extracted from peripheral blood leukocytes using the Puregene DNA Isolation kit (Gentra System, Minneapolis, USA). A skin biopsy was obtained from five probands, four mothers and an affected sister, and were cultivated to generate primary lines of fibroblast cells. DNA, RNA and proteins were extracted from the fibroblasts using DNeasy Blood & Tissue Kit, RNeasy Fibrous Tissue Mini Kit (both from Qiagen, Hilden, Germany) and RIPA buffer and protease inhibitors, always according to the manufacturer's guidelines. We got the cDNA using the SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR of Invitrogen™ (Thermo Fisher, California, USA) following the manufacturer's instructions.

2.2. Sanger sequencing

We validated the variants in DNA of the available tissues and in RNA.

PCR was performed using the Expand High Fidelity PCR System kit (Roche, Mannheim, Germany). This protocol was carried out on a SimpliAmp Thermal cycler in accordance with the manufacturer's instructions (Applied BioSystems, Massachusetts, USA). Primers for the different *MECP2* exons were designed for the genomic transcript NM_004992.3 using the Primer3 program [Table S1] (12). The PCR products were cleaned with Exonuclease I and Shrimp Alkaline Phosphatase of Illustra™ (GE Healthcare Life Sciences, UK) and sequenced using Big-Dye® Terminator version 3.1 Cycle Sequencing Kit in an Applied Biosystems 3730/DNA Analyzer (Applied BioSystems, Massachusetts, USA). The raw data was analysed with Chromas trace viewer (http://technelysium.com.au/wp/chromas/). Variants were annotated based on *MECP2* NM_004992.3.

2.3. X chromosome inactivation assay (XCI)

The XCI status of the females was determined by analysing the methylation status of the highly polymorphic trinucleotide X-linked androgen receptor (AR) locus as described by Allen et al. (13). For each subject, half of the genomic DNA sample was digested with HpaII restriction enzyme (New England Biolabs, Massachusetts, USA) in accordance with the manufacturer's instructions. A region of the AR locus was amplified by PCR from digested and undigested DNA using fluorochrome-labelled primers. Samples were electrophoresed on an ABI Prism Genetic Analyzer 3130 and the peak areas quantified using the Peak Scanner software (Thermo Fisher Scientific, California, USA). XCI was considered skewed when the inactivation ratio was 80:20 or higher.

2.4. Quantitative reverse transcription PCR (qRT-PCR)

The generated cDNA was amplified with qRT-PCR to measure the relative expression of the two isoforms of *MECP2*. Primers were designed using the software Primer3 avoiding intronic sequences [Table S1]. *ALAS1* was used as a reference amplicon. The qRT-PCR was performed with PowerUP SYBR Green Master Mix kit in a QuantStudio 6 Flex Real-Time PCR system (Applied Biosystems, Massachusetts, USA). Analysis was performed using the ΔΔCt relative quantification method. All reactions were conducted in triplicate and the average was used for quantitative analysis. Three control samples from individuals of the same gender were used to normalise the expression levels. Product specificity was assessed via melting curve analysis.

2.5. Western blot

Western blot analysis of protein samples from fibroblast cells was performed. Proteins were separated in the precast polyacrylamide gels Mini-PROTEAN®TGX™ 4-20% (BioRad Laboratories, USA) and transferred to a nitrocellulose membrane. Membranes were blocked with BSA (A7906, Merck KGaA, Darmstadt, Germany) and PBST buffer 5% for 1h at room temperature. Incubation with primary antibodies was against polyclonal anti-MeCP2 (rabbit; ab2828, Abcam, Spain) at a concentration of 1:500. As a loading control we used monoclonal anti-Vinculin (mouse; ab129002, Abcam, Spain) at a concentration of 1:500, both primary antibodies were incubated at 4°C overnight. The secondary antibodies used were the horseradish peroxidase-conjugated antibodies donkey anti-mouse and donkey anti-rabbit of Invitrogen™ (Thermo Fisher, California, USA) at a concentration of 1:5000. Secondary antibodies were incubated at room temperature for 1h. Membranes were washed three times with PBST and the signal was revealed using the Enhanced Chemiluminescence System (Pierce® ECL Western Blotting Substrate, Thermo Scientific). Quantification was measured with the programme ImageJ (https://imagej.net/Fiji).

2.6. Tools for variant classification

In order to study the variants and their possible effects we used Franklin (https://franklin.genoox.com/clinical-db/home), which follows the ACMG guidelines to classify the variants. In addition, the literature, the RettBASE:

MECP2 Variant Database, genome Aggregation Database (gnomAD) and UCSC Genome Browser platforms were revised.

3. Results

Seven patients were referred to us for further study after a variant in the gene *MECP2* had been found in a clinical exome [Table 1, Figure S1].

Table 1: The seven detected variants in the gene *MECP2* in the seven cases. NM_004992.3 and NP_004983.1 are used for the annotation. For case 6, where the variant causes a deletion that affects only one of the isoforms NP_001104262 was used. The origin of the variant was studied whenever the parent's sample was available.

Case	Variant (DNA)	Variant (Protein)	Inheritance
Patient 1	c.561C>G	p.G187G	Mother
Patient 2	c.602C>T	p.A201V	Mother
Patient 3	c.909C>G	p.I303M	*De novo*
Patient 4	c.595C>T	p.P199S	Mother
Patient 5	c.347C>G	p.S128C	Unknown
Patient 6	c.-113C>T	MeCP2_e1 p.Glu17_Leu21del MeCP2_e2 normal	Mother
Patient 7	c.753dupC	p.G252R*7	*De novo*

Case 1

P1 is an 18 years old boy born to non-consanguineous parents that presents chronic encephalopathy and moderate to severe ID. By 14 months of age, he had global psychomotor delay and by the age of two, generalised epilepsy. He has developed autistic traits, controlled epilepsy and stereotypic hand movements. The variant c.561C>G (p.Gly187Gly) was found in hemizygosity. We found that the mother, M1, carries the same variant in heterozygosity. The variant was confirmed in both proband and mother in leukocyte and fibroblast origin DNA [Figure S2]. We proceeded to confirm the variant in cDNA of fibroblast origin and discovered that the variant was present in P1 but was missing in M1's cDNA [Figure S2].

XCI assay was performed in the two DNA samples from M1 we had. The fibrous tissue showed a complete skewed pattern whereas the blood tissue tended to this skewed pattern [Table 2].

The variant is classified as a variant of uncertain significance (VUS) according to the ACMG classification (PM2, BP7). This variant has never been reported, not in gnomAD, not in affected individuals.

No new isoforms were detected by PCR and the splicing junctions were all okay. The mRNA levels of both *MECP2* isoforms were studied. P1 expresses 20% more of *MECP2_e1*, whereas M1 expresses 30% more of both isoforms [Figure S3a,b]. Western Blot assay was performed to see if there were differences at the protein level and both individuals have 20% less expression of MeCP2 [Figure S3c].

Table 2: XCI ratios of the studied females in leukocyte DNA and fibroblast DNA. 80:20 is the established threshold to consider the inactivation to be skewed.

ID	XCI leukocyte DNA	XCI fibroblast DNA
M1	70:30	98:2
M2	87:23	77:23
M4	86:14	85:15
M6	59:41	78:22
S6	82:18	69:31

Case 2

P2 is a boy born to non-consanguineous parents of Spanish European ancestry and was delivered at 35 weeks of gestation. He was large for gestational age and he had congenital macrocephaly. He developed infantile spasms that were controlled with antiepileptic drugs, and was diagnosed with autism spectrum disorder (ASD). He was found to be hemizygous for a maternal inherited c.602C>T (p.Ala201Val) variant. The variant was confirmed in P2 and his mother M2 in all the available DNA samples [Figure S2]. We confirmed the variant in fibroblast cDNA and discovered that, again, the variant was present in P2 but not in M2's cDNA [Figure S2].

XCI assay of M2 showed a skewed pattern in leukocytes and an almost skewed pattern in fibroblasts [Table 2].

This variant is described as benign (BA1 and BS2, BP6 edited due to special guidelines) and has the rs61748381. There are 291 alleles registered in gnomAD and one case of homozygosity has been reported.

The two isoforms of *MECP2* were amplified by PCR in the cDNA but no alternative product was seen. The splicing junctions were unaltered. qRT-PCR revealed a 30% less expression of *MECP2_e1* and *MECP2_e2* in P2, but a 50% of higher expression in both isoforms in M2 [Figure S3a,b]. Western blot assay showed that P2 has a 40% less expression of MeCP2 whereas M2 has 20% less [Figure S3c].

Case 3

P3 was delivered without incident after a spontaneous gestation. P3 is a 15 years old boy with global psychomotor development delay, ID, conduct and sleep disorder, unstable gait, generalized tremor, limited use of hands and poor language. He has the *de novo* variant c.909C>G (p.Ile303Met) and was referred because of suspicion of atypical RTT. The variant was detected in leukocytes DNA [Figure S2].

The variant is classified as VUS according to ACMG guidelines (PM2, PM5 and PM1; PP3 edited due to special guidelines) with no alleles registered in gnomAD. Since no fibroblast tissue could be obtained, no further tests were performed.

Case 4

P4 was born to non-consanguineous parents of central American ancestry in an uneventful delivery. He suffered meningitis shortly after being born. P4 is a 21 years old male with global neurodevelopmental delay, mild muscle hypotonia, hand flapping stereotypies, no speech and encephalopathy. He harbours the variant c.595C>T (p.Pro199Ser) which is inherited from his mother. The variant was detected in all the available DNA and RNA samples from P4 and M4 [Figure S2].

M4's XCI showed a skewed inactivation in leukocytes and in fibroblasts [Table 2]. In this case, M4 expresses almost exclusively the alternative allele in mRNA.

The variant has been classified as VUS (PM2 and BS2, BP6 edited due to special guidelines). There is an associated rs373329231 and there are 3 alleles registered in gnomAD (in the east Asian population).

No alternative isoforms were detected by PCR. *MECP2_e1* was expressed 30% more in P4 and M4, whereas *MECP2_e2* was expressed 15% more [Figure S3a,b]. Western blot analysis showed 45% less MeCP2 amounts for P4 and 30% less for M4 [Figure S3c].

Case 5

P5 was adopted so no perinatal nor parent information is available. P5 is a 16 years old male with mild ID, ASD, psychosis, mood disorder, general tremor, muscular rigidity and impairment of social abilities. He has a sister with ID and conduct disorder. P5 has the variant c.347C>G (p.Ser128Cys). The variant was seen in the leukocyte DNA sample of P5.

The variant has been classified as a VUS (PM2, PM1, PP3) and no allele has been registered in gnomAD. Leukocyte DNA was only available so no further experiments were performed [Figure S2].

Case 6

P6 was gestated by artificial insemination with gametes from both parents. Pregnancy was controlled and ended with eutocic delivery. P6 is a 2 years old boy with severe encephalopathy, psychomotor delay, generalised epilepsy and sleep disorder. He has an affected sister (S6) with the same variant and RTT phenotype. The variant c.-113C>T (NM_001110792.2 p.Glu17_Leu21del) is inherited from their mother (M6), who has a somatic mosaicism and mild ID. The variant was found in all three members of the family in hemizygosity or heterozygosity in all the DNA samples [Figure S2].

M6 has a random XCI in leukocytes but is slightly skewed in fibroblasts. S6 has a skewed XCI in leukocytes and a random XCI in fibroblasts [Table 2].

The variant is classified as VUS (PM2, PP5 edited due to special guidelines). No allele appears in gnomAD.

Study of the RNA revealed a 15bp deletion between the end of exon 1 and the beginning of exon 2 in all three members of the family. P6 is hemizygous for the deletion and M6 and S6 are heterozygous [Figure S2]. The expression of *MECP2_*

e1 is null in P6, is reduced to half in S6, whereas it is normal in the mother. The expression levels of *MECP2_e2* are normal in all three [Figure S3a,b]. Western blot analysis shows that P6 and S6 possess half of MeCP2 amounts while M6 expresses the same MeCP2 amounts as the healthy controls [Figure S3c].

Case 7

The seventh boy was referred to us with RTT suspicion. He had neonatal encephalopathy, psychomotor delay, severe mental disability, microcephaly and hearing impairment. He died at the age of 2 years and 6 months. He had a de novo c.753dupC. (p.G252R*7). The variant was confirmed in DNA [Figure S2].

The variant is classified as VUS (PVS1 and PM2, PP5 is edited due to special guidelines) and it is not registered in gnomAD.

mRNA expression of isoform *MECP2_e1* is 55% lower and expression of isoform 2 is 25% lower than controls' [Figure S3a,b]. MeCP2 quantification showed 48% of expression of the protein [Figure S3c].

4. Discussion

We here present seven cases of male probands with different degrees of developmental delay, autistic features and ID with variants in *MECP2*. Their clinical picture matches with the one that has been described in other males with variants in *MECP2* (10,14). We used a variant interpreter to evaluate the variants and see if they could be the cause of these boys' phenotype. According to Franklin, six variants (P1, P3, P4, P5, P6, P7) are VUS and one (P2) is benign. We got fibroblasts from cases 1, 2, 4, 6, 7, their mothers and a symptomatic sister to perform more studies at different molecular levels. We performed qRT-PCR and Western blot in all the cases in which fibroblasts were obtained, and found out that our indexes, the affected sister and three carrier mothers have impaired protein amounts, suggesting a contribution of the variants to, at least, MeCP2's lifespan. We admit the limitations of our experiments and see the need of performing further functional studies to properly reclassify the variants following ACMG guidelines. However, the obtained data enabled us to tilt the variant classification towards a likely benign or likely pathogenic and to give a diagnosis to six of these children.

During the study of case 1, several hypotheses were tested in order to find the reason why the variant might cause the abnormal phenotype. The variant of P1 does not lead to an amino acid change, so the most plausible explanation for a pathogenicity involved an altered splicing; event which was also predicted by some bioinformatics tools that claimed that an exonic splicing silencer was created. We could not detect an altered splicing event in our samples. M1 has a complete skewed XCI in fibroblasts and does not express the allele with the variant. That leads us to think that the cell could be inactivating an aberrant allele as a protective mechanism. However, we detected the same protein impairment in fibroblasts from P1 and M1, suggesting that the variant could not be the cause of the variations seen in *MECP2* and MeCP2. That variant might probably be benign.

The variant from case 2 is classified as benign mainly due to its abundance in healthy populations. Besides, the variant has been reported as a polymorphism in several publications (15–17). Seeing that at mRNA level M2 silences the variant's expression was the motor of this more exhaustive molecular study. No alterations were detected in the splicing but the mRNA quantities of both isoforms and protein abundances were impaired for P2. However, once more, M2 also expresses a slight reduction of MeCP2 levels without expressing the variant, which makes us think that another mechanism might be regulating *MECP2* in this family.

The lack of fibroblast tissue made it impossible to further study the variant from case 3. That variant is reported in RettBASE and in the literature in two females with variable degree of phenotypic severity (18,19). Kammoun et al.'s female presented mild ID and tremors, whereas Zvereff et al.'s female had developmental and speech delay, poor hand coordination and epilepsy. Our P3 presents all those phenotypic traits plus conduct and sleep disorder. More functional studies should be conducted but it looks like the p.I303M variant could be causing this specific phenotype.

The variant from case 4 does not generate alternative splicing but impairs MeCP2 amounts up to 40% in P4 and M4. In this case, M4 expresses almost exclusively the allele with the variant in fibroblast derived mRNA, and thus, the variant seems to be the cause of the protein impairment in these two individuals. Since the mother is asymptomatic, the XCI might be opposite in brain tissues, where the wt allele might be expressed, leading to a healthy phenotype. Therefore, the variant seems to be altering MeCP2 and might be related to the phenotype of P4.

The variant from case 5 was complicated to study since P5 was adopted and no parental information was available. The adoptive parents said that the proband's sister presents ID and a similar behavioural disorder, but since she was adopted by a different family the medical history record could not be completed. Definitely, studying the mRNA and protein levels of these two children would shed some light to the effect of the variant.

The variant seen in case 6 has been reported in one female with classic RTT (20). Sheikh et al. reported the splicing event that shortens the first exon and that alters the amino acid sequence from MeCP2_e1. We could only amplify the pre-mRNA molecule of *MECP2* and thus, we could not observe the whole specific resulting sequence of *MECP2_e1*, even though we detected the 15nt deletion causing the aberrant isoform. Besides, we discovered a null expression of *MECP2_e1* in P6, and half expression levels of the same isoform in his sister. MeCP2 levels were impaired to half as well in both siblings. S6 has been diagnosed with classic RTT as well, and the proband had severe congenital encephalopathy. Altogether, our data confirms the pathogenicity of the variant.

There are four entries for the variant of case 7 in RettBASE. Three of them are females and two are diagnosed with classic RTT (the third registry does not describe the phenotype). The fourth entry corresponds to a male with severe neonatal encephalopathy, like our P7 (21). We found that the expression levels of *MECP2_e1, MECP2_e2* and MecP2 were diminished in P7. Our molecular results and the fact that two more patients have been clinically diagnosed with classical RTT and neonatal encephalopathy makes it probable to be in front of a pathogenic variant.

5. Conclusion

Our molecular studies justify the need to perform functional studies with the variants found in P3, P4, P6 and P7 since those variants are accumulating evidence towards pathogenicity. On the contrary, variants from P1, and P2, seem not to be the causative variants of the MeCP2 impairment detected and thus, the phenotype of these children. Finally, the lack of further tissues made it impossible to collect information about the variant from P5, so it remains as a VUS with no suggestions.

As a conclusion, we encourage geneticists not to dismiss a variant located in a gene from the X chromosome, as *MECP2,* only because it is present in an apparently asymptomatic female. The XCI could explain the mild phenotype the carriers have, although it can be very difficult to determine its status in target tissues (22). In general, the study of different molecular levels should be taken into consideration before jumping into conclusions about the pathogenicity of a variant.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We thank all the patients and their families for contributing to the present study.

Funding

This work was funded by the grant from the Spanish Ministry of Health (Instituto de Carlos III/FEDER, PI15/01159 and PI20/00389) and a doctoral grant from the Government of Catalonia (Secretaria d'Universitats I Recerca) and the European Social Fund with the grant number 2020 FI-B 00888.

Bibliography

1. Mnatzakanian GN, Lohi H, Munteanu I, Alfred SE, Yamada T, MacLeod PJM, et al. A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. Nat Genet. 2004;36(4):339–41.
2. Krishnaraj R, Ho G, Christodoulou J. RettBASE: Rett syndrome database update. Hum Mutat. 2017 Aug 1;38:922–31.
3. Chahrour M, Sung YJ, Shaw C, Zhou X, Wong STC, Qin J, et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. Science (80-). 2008;320(5880):1224–9.

4. Wu H, Tao J, Chen PJ, Shahab A, Ge W, Hart RP, et al. Genome-wide analysis reveals methyl-CpG-binding protein 2-dependent regulation of microRNAs in a mouse model of Rett syndrome. PNAS. 2010 Oct 19;107(42):18161–6.
5. Young JI, Hong EP, Castle JC, Crespo-Barreto J, Bowman AB, Rose MF, et al. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. PNAS. 2005;102(49):17551–8.
6. Collins BE, Neul JL. Rett syndrome and MECP2 Duplication syndrome: disorders of MeCP2 dosage. Neuropsychiatr Dis Treat. 2022;18:2813–35.
7. Psoni S, Sofocleous C, Traeger-Synodinos J, Kitsiou-Tzeli S, Kanavakis E, Fryssira-Kanioura H. Phenotypic and genotypic variability in four males with MECP2 gene sequence aberrations including a novel deletion. Pediatr Res [Internet]. 2010;67(5):551–6. Available from: http://sift.jcvi.org/
8. Ravn K, Nielsen JB, Uldall P, Hansen FJ. No correlation between phenotype and genotype in boys with a truncating MECP2 mutation. J Med Genet [Internet]. 2003;40:e5. Available from: http://www.jmedgenet.com/cgi/content/full/40/1/e5
9. Neul JL, Kaufmann WE, Glaze DG, Christodoulou J, Clarke AJ, Bahi-Buisson N, et al. Rett syndrome: revised diagnostic criteria and nomenclature. Ann Neurol. 2010 Dec;68(6):944–50.
10. Neul JL, Benke TA, Marsh ED, Skinner SA, Merritt J, Lieberman DN, et al. The array of clinical phenotypes of males with mutations in Methyl-CpG binding protein 2. Am J Med Genet Part B Neuropsychiatr Genet. 2019 Jan 1;180(1):55–67.
11. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May 8;17(5):405–24.
12. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-new capabilities and interfaces. Nucleic Acids Res. 2012 Aug;40(15):e115.
13. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. Am J Hum Genet. 1992;51:1229–39.
14. Pascual-Alonso A, Martínez-Monseny AF, Xiol C, Armstrong J. MECP2-related disorders in males. Int J Mol Sci. 2021 Sep 1;22:9610.

15. Amano K, Nomura Y, Segawa M, Yamakawa K. Mutational analysis of the MECP2 gene in Japanese patients with Rett syndrome. J Hum Genet. 2000;45:231– 236.
16. Fukuda T, Yamashita Y, Nagamitsu S, Miyamoto K, Jin JJ, Ohmori I, et al. Methyl-CpG binding protein 2 gene (MECP2) variations in Japanese patients with Rett syndrome: pathological mutations and polymorphisms. Brain Dev. 2005;27:211–7.
17. Wong VCN, Li SYH. Rett syndrome: prevalence among Chinese and a comparison of MECP2 mutations of classic Rett syndrome with other neurodevelopmental disorders. J Child Neurol. 2007;22(12):1397–400.
18. Zvereff V, Carpenter L, Patton D, Cabral H, Rita D, Wilson A, et al. Molecular diagnostic dilemmas in Rett syndrome. Brain Dev [Internet]. 2012;34(9):750–5. Available from: http://dx.doi.org/10.1016/j.braindev.2011.12.012
19. Kammoun F, de Roux N, Boespflug-Tanguy O, Vallée L, Seng R, Tardieu M, et al. Screening of MECP2 coding sequence in patients with phenotypes of decreasing likelihood for Rett syndrome: a cohort of 171 cases. J Med Genet. 2004;41(6):1–7.
20. Sheikh TI, Mittal K, Willis MJ, Vincent JB. A synonymous change, p.Gly16Gly in MECP2 Exon 1, causes a cryptic splice event in a Rett syndrome patient. Orphanet J Rare Dis [Internet]. 2013;8(1):1. Available from: Orphanet Journal of Rare Diseases
21. Ben Zeev B, Yaron Y, Schanen NC, Wolf H, Brandt N, Ginot N, et al. Rett syndrome: clinical manifestations in males with MECP2 mutations. J Child Neurol. 2002;17(1):20–4.
22. Xiol C, Vidal S, Pascual-Alonso A, Blasco L, Brandi N, Pacheco P, et al. X chromosome inactivation does not necessarily determine the severity of the phenotype in Rett syndrome patients. Sci Rep [Internet]. 2019 Dec 19;9(1):11983. Available from: http://www.nature.com/articles/s41598-019-48385-w

Supplementary information

Table S1: List of the primers for the variant validation and *MECP2* isoform sequencing. The first 3 pairs of primers are designed to amplify DNA, while the next four primer pairs are designed for cDNA. The last 2 *MECP2* pairs of primers were used for the *MECP2_e1* and *MECP2_e2* relative quantification. *ALAS1* was used as a reference gene in the qRT-PCR.

Primer name	Forward sequence	Reverse sequence
*MECP2*_ex1	CAATTGACGGCATCGCCGCTGAGA	CATCCGCCAGCCGTGTCGTCCGAC
*MECP2*_ex3	TCTCTGTTGTTGTCCTGGGGAAG	CCCTGGGCACATACATTTTC
*MECP2*_ex4	GGCAGTGTGACTCTCGTTCA	GTCCTTTCCCGCTCTTCTC
*MECP2*_ex1_ ex2-3	CGCGCGCTCCCTCCT	TGACTTTTCTTCCCTGAGCCCTAA
*MECP2*_ex3_ex4	CAAAGCAGAGACATCAGAAGGG	CTTCCCAGGACTTCTCCA
*MECP2*_ex1-3_ex3 (isof1)	AGGAGAGACTGGAAGAAAAGT	CTTGAGGGGTTTGTCCTTGA
*MECP2*_ex2_ex3 (isof2)	CTCACCAGTTTCCTGCTTTGATGT	CTTGAGGGGTTTGTCCTTGA
ALAS1	CCTTTGGTTGTGTTGGAGGG	CCGGCATCCATTAGCATCTG

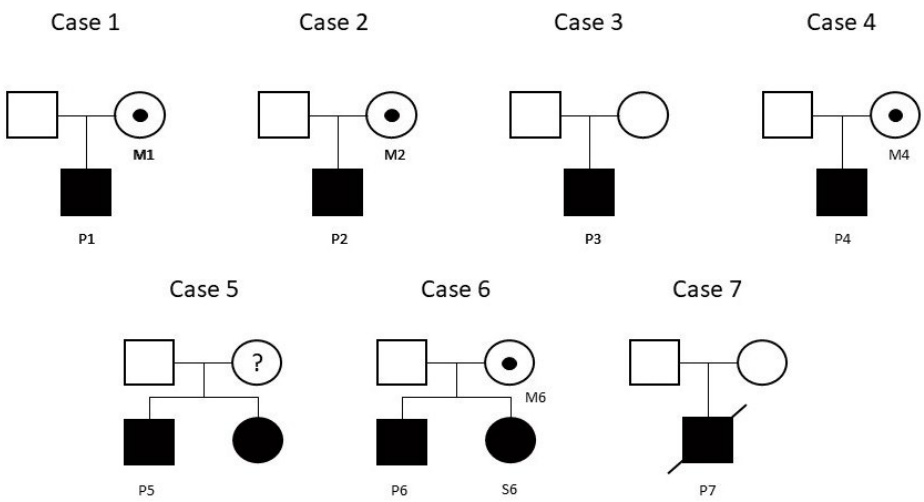

Figure S1: Familiar trees of the seven cases.

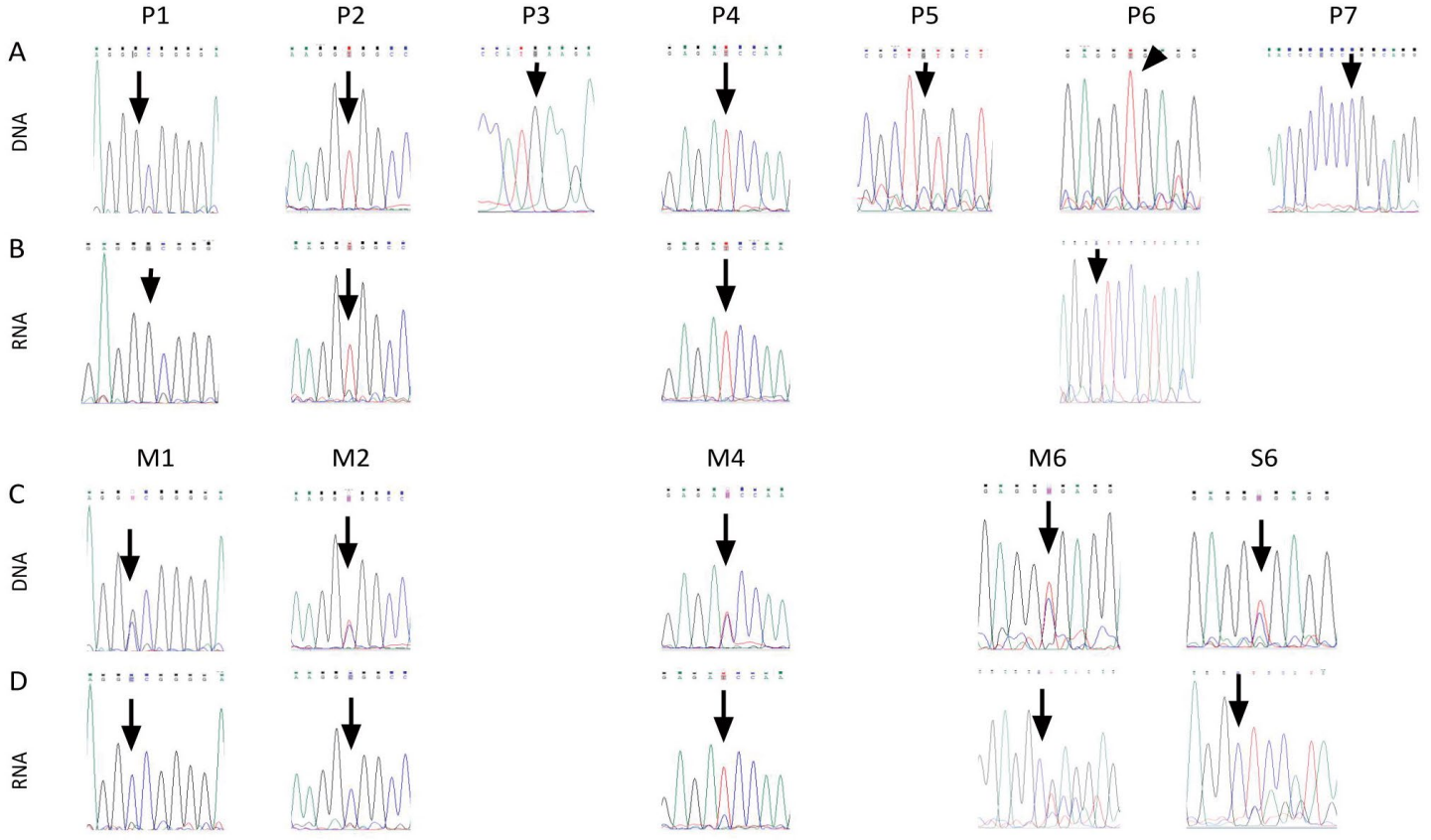

Figure S2: Sequencing results of *MECP2* gene in the seven study cases. A) Sequencing results of the patients' leukocyte DNA samples. Note that fibroblast DNA was also sequenced and the same result as with leukocytes were obtained. B) Sequencing results of the patients' fibroblasts RNA samples. C) Sequencing results of the available carrier mother's leukocyte DNA samples. As with their offspring, fibroblast DNA was also sequenced and the same variants were detected. D) Sequencing results of the available carrier mothers' fibroblasts RNA samples. All the detected SNVs have been marked with an arrow. The variants are in hemizygosity in male patients and in heterozygosity in females. In case 6, even though a SNV was found in the DNA, at

mRNA level a deletion of 15 nt spanning the end of exon 1 and the beginning of exon 2 was detected. That deletion has been seen in heterozygosis in M6 and S6 [...GAGG**C/T**(GAGGAGGAGAGACT G) CTCCATAAAAA...].

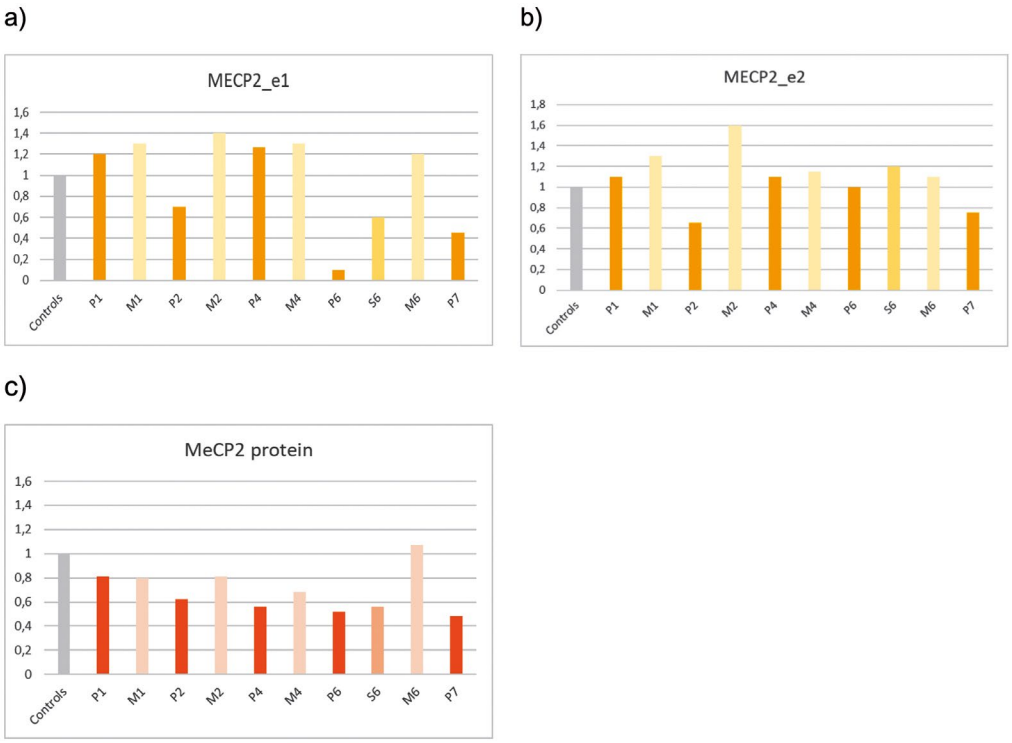

Figure S3: A) Relative quantifications of *MECP2_e1* and B) *MECP2_e2*. C) Relative quantification of MeCP2 protein.

Chapter 2

Chapter 2: Characterization of dysregulated molecular processes in RTT patients using transcriptomic and proteomic technologies.

In the second chapter we will unravel the dysregulated processes of a RTT cohort, which corresponds with the second objective of the thesis. The publication we are presenting in the current chapter will summarise for the first time the main findings of applying a multi-omics approach to a classic RTT cohort, and the similarities that exist between classic RTT, RTT-like and MDS patients at molecular level.

Performing experiments with high quality samples is crucial to avoid artefacts or dysregulations caused by external reasons. To ensure the maximum sample quality for the experiments we developed a protocol in which a quality control test was done to all the RNA samples. With it, the RNA samples that had suffered oxidative stress during cell culture were spotted and discarded before the RNAseq was conducted. Whenever proteins were also obtained from the same "stressed" cultures, that sample was also discarded to ensure the same quality for the proteomics experiments. By applying our protocol we increased the reproducibility of different replicas of the same patient, capturing the differential expression caused by the syndrome. The protocol is explained in the Material and Methods section of the publication.

Transcriptomics or proteomics of RTT have been studied before, but usually in mouse models and rarely with sample sizes bigger than 10. Also, the tissues studied in humans are not always stable. We gathered a cohort of 22 patients with classic RTT and performed RNAseq and TMT-MS experiments with mRNA and proteins from fibroblast primary cell lines derived from the patients' skin biopsies. Without a gold standard workflow to analyse the transcriptomics and proteomics data, several tools were tested until a robust way of normalisation was found (also explained in the Material and Methods section). The same workflow was applied to samples from 15 MDS and 12 RTT-like patients and the lists of DEGs and DEPs together with the enriched terms were compared. The thorough description of the findings is detailed in the following publication.

Publication 1: *Identification of molecular signatures and pathways involved in Rett syndrome using a multi-omics approach.*

Authors: Ainhoa Pascual-Alonso, Clara Xiol, Dmitrii Smirnov, Robert Kopajtich, Holger Prokisch and Judith Armstrong.

Reference: Hum Genomics. Under revision.

Abstract:

Transcriptomic and proteomic technologies have been successfully used to discover which molecular processes are altered in RTT. Unfortunately, those studies have usually been performed with few samples and frequently, in mouse models. We have been able to generate a collection of fibroblast primary cell lines from skin biopsies of patients with *MECP2* related disorders. In total, we have selected 22 classic RTT (21 female and 1 male), 15 MDS (all male), 12 RTT-like patients with mutations in *CDKL5* (1 female, 3 males), *FOXG1* (1 female, 1 male), *NR2F1* (1 female), *GRIN2B* (1 female), *AHDC1* (1 female) and 3 female patients without molecular diagnosis, and 13 healthy controls (7 female and 6 male) and we have performed RNAseq and TMT-MS experiments.

The obtained transcriptomic and proteomic data was analysed in a case-control approach, and a list of DEGs and DEPs was generated from each patient cohort. The DEGs and the DEPs underwent an enrichment analysis and an upstream transcription factor study in order to obtain groups of genes and proteins related to the same biological processes. We identified molecular alterations that could contribute to the phenotype, such as genes involved in cytoskeleton, vesicular transport or mRNA and rRNA processing machinery. We also compared whether DEGs and DEPs expressed in opposite directions exist between classic RTT and MDS cohorts. Interesting results were obtained from the transcriptomics experiments but the multi-omics approach showed little overlap. Besides, the RTT and RTT-like cohorts were also compared and several consistently deregulated genes and proteins were found. The integration of omics data makes it possible to decipher the molecular consequences of *MECP2* dysfunction, enlarging the picture of the molecular landscape of RTT patients and *MECP2*-related disorders.

Identification of molecular signatures and pathways involved in Rett syndrome using a multi-omics approach

Ainhoa Pascual-Alonso[*1,2], Clara Xiol[*1,2], Dmitrii Smirnov[3,4], Robert Kopajtich[3,4], Holger Prokisch[3,4], Judith Armstrong[#2,5,6]

1. Fundació per la Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain
2. Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain
3. Institute of Human Genetics, Technical University of Munich, Munich, Germany
4. Institute of Neurogenomics, Helmholtz Zentrum München, Munich, Germany
5. CIBER-ER (Biomedical Network Research Center for Rare Diseases), Instituto de Salud Carlos III (ISCIII), Madrid, Spain
6. Genomic Unit, Molecular and Genetic Medicine Section, Hospital Sant Joan de Déu, Barcelona, Spain

*Both authors contributed equally to this work.

#Corresponding author, judith.armstrong@sjd.es

0. Abstract

Background: Rett syndrome (RTT) is a neurodevelopmental disorder mainly caused by mutations in the methyl-CpG-binding protein 2 gene (*MECP2*). MeCP2 is a multifunctional protein involved in many cellular processes, but the mechanisms by which its dysfunction causes disease are not fully understood. The duplication of the *MECP2* gene causes a distinct disorder called MECP2 duplication syndrome (MDS), highlighting the importance of tightly regulating its dosage for proper cellular function. Additionally, some patients with mutations in genes other than *MECP2* exhibit phenotypic similarities with RTT, indicating that these genes may also play a role in similar cellular functions. The purpose of this study was to characterise the molecular alterations in patients with RTT in order to identify potential biomarkers or therapeutic targets for this disorder.

Methods: We used a combination of transcriptomics (RNAseq) and proteomics (TMT-mass spectrometry) to characterise the expression patterns in fibroblast cell lines from 22 patients with RTT and detected mutation in *MECP2*, 15 patients with MDS, 12 patients with RTT-like phenotypes and 13 healthy controls. Transcriptomics and proteomics data were used to identify differentially expressed genes both at RNA and protein levels, which were further inspected via enrichment and upstream regulator analyses and compared to find shared features in patients with RTT.

Results: We identified molecular alterations in cellular functions and pathways that may contribute to the disease phenotype in patients with RTT, such as deregulated cytoskeletal components, vesicular transport elements, ribosomal subunits and mRNA processing machinery. We also compared RTT expression profiles with those of MDS seeking changes in opposite directions that could lead to the identification of MeCP2 direct targets. Some of the deregulated transcripts and proteins were consistently affected in patients with RTT-like phenotypes, revealing potentially relevant molecular processes in patients with overlapping traits and different genetic aetiology.

Conclusions: The integration of data in a multi-omic analysis has helped to interpret the molecular consequences of *MECP2* dysfunction, contributing to the characterisation of the molecular landscape in patients with RTT. The comparison with MDS provides knowledge of MeCP2 direct targets, whilst the correlation with RTT-like phenotypes highlights processes potentially contributing to the pathomechanism leading these disorders.

Keywords: Rett syndrome, MECP2 duplication syndrome, Rett-like phenotypes, Multi-omics, Transcriptomics, Proteomics, Differential expression.

1. Introduction

Rett syndrome (RTT, OMIM#312750) is a severe neurodevelopmental disorder characterised by psychomotor regression after a period of normal early development. It mainly affects girls, who typically present with loss of purposeful hand use and expressive language, gait abnormalities and stereotypic hand movements. In addition, the main symptoms can be accompanied by a variety of other dysfunctions, such as seizures, breathing disturbances, scoliosis, impaired sleep patterns, and abnormal muscle tone (1). The diagnosis of RTT is

mainly clinical and is based on a set of criteria that differentiate between typical RTT and three atypical variants with distinctive features: the preserved speech variant, the congenital variant, and the early-onset seizure variant (2).

Most typical and atypical RTT cases are caused by loss-of-function mutations in the methyl-CpG-binding protein 2 (*MECP2*, OMIM*300005) gene, located on the X chromosome (3). MeCP2 is a chromatin-associated protein that acts as a transcriptional regulator, both repressing and activating transcription, and is also involved in maintaining heterochromatin structure, regulating splicing through interaction with splicing factors and miRNA processing by binding to microprocessor components (1). MeCP2 is expressed ubiquitously but is especially abundant in mature neurons. MeCP2 has proved to be crucial for neuronal maturation, dendritic arborisation, and synaptic plasticity (1).

Mutations in other genes have also been identified in patients with RTT. Pathogenic variants in cyclin-dependent kinase-like 5 (*CDKL5*, OMIM*300203) and forkhead box G1 (*FOXG1*, OMIM*164874) have been detected in a substantial number of cases with early-onset seizure and congenital RTT variants, respectively (4,5). Moreover, with the generalisation of next-generation sequencing (NGS), the number of genes associated with RTT has increased remarkably (6–8). Some of these are novel findings whilst others have already been associated with different neurodevelopmental disorders or epileptic encephalopathies. Patients with overlapping phenotypes with RTT but who do not fulfil established clinical criteria are termed 'RTT-like'. Therefore, any patient with a combination of RTT features can be described as RTT-like (6,7).

MeCP2 levels are tightly regulated in the cells and not only a loss-of-function can have pathogenic effects. Chromosomal duplications at Xq28 encompassing the *MECP2* and *IRAK1* genes, leading to their gain-of-function, cause *MECP2* duplication syndrome (MDS), a neurological disorder characterised by intellectual disability (ID), infantile hypotonia, seizures, speech impairment, and recurrent respiratory infections (9). It mainly affects males, whilst penetrance in females is highly dependent on X-chromosome inactivation (XCI). Phenotypic variability is high in patients with MDS and potentially related to the size and content of the duplication, which is unique for each family (10). However, a clear genotype-phenotype correlation has not yet been found.

One of the drawbacks in studying the downstream molecular effects of MeCP2 dysfunction is the lack of accessibility to samples of the primarily affected tissue, the brain. In the search for new tissues, skin fibroblasts have

demonstrated greater consistency in gene expression and include more OMIM and neurologically relevant genes compared with whole blood (11,12).

Around 70 experimental and repurposed drugs have been investigated for RTT but there is no approved treatment yet (13). In RTT clinical trials, the success of the tested drugs is evaluated by measuring the improvement in the symptomatology and quality of life of the patients. The lack of a biomarker complicates an objective quantification of the improvements derived from drug treatments. An efficient way to extract huge amounts of molecular data in order to find biomarkers could be by analysing the RNA profiles and proteome of the patients using multi-omics technology.

To date, no multi-omics analysis has been performed with RTT human samples and only one has been published with 4 RTT mice samples (14). Here, we aim to fill that knowledge gap by studying a cohort of 22 patients with RTT, 12 patients with RTT-like and 15 patients with MDS. Integration of transcriptomics and proteomics data could be a promising approach to find new potential therapeutic targets and biomarkers.

2. Material and methods

2.1. Clinical and molecular characterisation

The study was approved by the Hospital Sant Joan de Déu (HSJD) ethical committee, Comitè d'Ètica d'Investigació Clínica-Fundació Sant Joan de Déu (CEIC; internal code: PIC-219-20). Sixty-two individuals (49 patients and 13 healthy age-matched controls) participated in this study and provided written informed consent. Patients were recruited after clinical and genetic confirmation of their pathology as described elsewhere (15). Eleven out of the fifteen MDS patients were described in Pascual-Alonso et al. (16) and the four new patients were characterised in the same way. We studied 22 patients with RTT and mutations in *MECP2* (21 females, 1 male); 15 male patients with MDS; 12 patients with RTT-like phenotypes and mutations in *CDKL5* (1 female, 3 males), *FOXG1* (1 female, 1 male), *NR2F1* (1 female), *GRIN2B* (1 female), and *AHDC1* (1 female), and 3 female patients without molecular diagnosis; as well as 13 healthy controls (7 females, 6 males) [Table 1]. Clinical severity of patients with RTT and RTT-like phenotypes was measured using the clinical severity score designed by Dr Pineda (17).

Skin biopsies from the 62 individuals were obtained and primary fibroblast cell lines were established. Fibroblast lines were grown on plates with Dulbecco's Modified Eagle's Medium high glucose with glutamine, supplemented with 10% heat-inactivated foetal bovine serum and 1% penicillin, streptomycin, and B amphotericin (all from Thermo Fisher, Waltham, MA, USA). Cultures were kept at 37°C with 5% CO2 in a humidified atmosphere. When the cells reached 70%-80% confluence, they were trypsinised and either re-sowed on new plates or harvested for subsequent RNA or protein extraction. Frozen vials from all the fibroblast lines were entrusted to the Biobanc 'Hospital Infantil Sant Joan de Déu per a la Investigació', which is integrated into the Spanish Biobank Network of ISCIII for the sample and data procurement.

DNA was extracted from fibroblast cell lines using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. XCI was performed in all female samples as described by Allen et al. (18). XCI was considered skewed with an allele ratio of 80:20 or greater [Table S1].

Table 1: The composition of the studied cohort, which consists of individuals with Rett syndrome (RTT) with mutations in *MECP2*, *MECP2* duplication syndrome (MDS), Rett-like (RTT-like) with mutations in different genes that are not *MECP2* and healthy controls.

Disease cohort	Female	Male	Total
RTT	21	1	**22**
MDS	-	15	**15**
RTT-like	8	4	**12**
CDKL5	1	3	**4**
FOXG1	1	1	**2**
NR2F1	1	-	**1**
GRIN2B	1	-	**1**
AHDC1	1	-	**1**
Unknown mutation	3	-	**3**
Healthy controls	7	6	**13**

2.2. RNA sequencing

RNA was extracted from cultured fibroblast pellets using the RNeasy Fibrous Tissue Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The obtained RNA was then measured with a Nanodrop spectrophotometer and examined in an agarose gel to check its purity and integrity.

To further confirm the quality of the isolated RNA and to diminish undesirable gene alterations due to cell stress conditions (19), we performed reverse transcription-quantitative polymerase chain reaction (RT-qPCR) of five genes that are part of the oxidative respiratory chain: *MT-CO1*, *MT-CO2*, *MT-CYB*, *MT-ND4*, and *MT-ATP6*. First, 500 ng of total RNA was processed according to the manufacturer's instructions, and double-stranded complementary DNA (cDNA) was generated in the presence of random hexamers using the SuperScript III First-Strand Synthesis SuperMix for qRT-PCR kit (Thermo Fisher, Waltham, MA, USA). Primers for the five mitochondrial genes and two additional housekeeping genes (*RPLP0* and *ALAS1*) were designed with Primer3 software (20) [Table S2]. The qRT-PCR was performed with PowerUp SYBR Green Master Mix (Thermo Fisher, Waltham, MA, USA) in an QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems, Waltham, MA, USA). All reactions were conducted in triplicate and the average of each triplicate group was used for the analysis, which is based on the ΔΔCt relative quantification method. Three non-stressed control samples were added to the experiment to get the normalised values. Amplified product specificity was assessed via melting curve. All samples that overexpressed two or more genes more than 1.5-fold the values of non-stressed controls were discarded [Figure S2].

For each sample, 2500 ng of RNA was used for library preparation. Illumina's TruSeq Stranded mRNA kit (Illumina, San Diego, CA, USA) was used following the manufacturer's protocol. Libraries were quantified in a 4200 TapeStation (Agilent Technologies, Santa Clara, CA, USA) and their integrity was checked. Sequencing was performed on an Illumina NextSeq 500 sequencer and 75 base pair (bp) paired-end reads with around 40 million paired reads per sample successfully mapped to the reference genome. At least two healthy controls of the same sex as the patients were included in all the runs to enable normalisation and to control the batch effect.

To validate the RNAseq experiment, we chose 22 significantly differentially expressed genes (DEGs) and performed RT-qPCR as explained above with a total of 23 different samples [Figure S2].

2.3. Differential expression (DE) analysis

RNAseq reads were aligned to the human reference genome (GRCh37/hg19) using STAR (v2.4.2a) in a strand-specific manner. Uniquely mapped reads were counted for each gene using the HTSeq package (v2.0.2) (21) with gene models from GENCODE release 29. A final count matrix for analysis was generated by averaging the values of raw counts from different replicates of the same sample. Counts per million mapped reads (CPM) were computed and only genes where more than 50% of samples had at least 1 CPM were kept.

We first inspected age, sex, biopsy origin and batch as possible covariates in the differential expression study by principal component analysis (PCA) and cluster analysis, but found no clear patterns in our samples (data not shown). PCA identified the primary sources of variation in our data. The first three principal components, explaining 18.8%, 16.1%, and 7.4% of the variance, were subsequently used in the model construction for differential expression analysis with DESeq2 (v1.34.0) (22). We used a Benjamini-Hochberg (BH) corrected p-value of 0.05 to consider significant differences.

2.4. Enrichement and upstrem regulation analysis

Enrichment analysis was performed using the clusterProfiler (v4.2.2) (23) and ReactomePA (v1.38.0) (24) R packages. Both overrepresentation analysis (ORA) and gene set enrichment analysis (GSEA) were carried out, using only significant DEGs and all expressed genes, respectively. Potentially enriched terms were searched in Gene Ontology (GO) (25), the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (26), WikiPathways (WP) (27), and the Reactome pathway database (RP) (28). All genes with CPM greater than 1 in at least 50% of samples and with an existing EntrezID were used as background (a total of 11904 genes). The cut-off value for considering a significantly enriched term was 0.05 in BH-corrected p-value.

We considered upstream transcription factors (TFs) responsible for some of the differential expression changes observed in our data, and used the ChIP-X Enrichment Analysis 3 (ChEA3) tool to identify them (29). ChEA3 contains information about TF gene co-expression, association in ChIP-seq studies and co-occurrence in gene lists, which is used to predict upstream TFs involved in the regulation of the user inputted gene lists. The lists of DEGs resulting

from differential expression analysis were fed to ChEA3 to predict the possible involvement of TFs in their dysregulation.

2.5. Proteomics

Proteomics experiments were performed at the BayBioMS core facility at the Technical University of Munich (TUM) in Germany. Fibroblast cell pellets containing around 0.5 million cells were sent frozen. These cells were thawed and lysed with urea containing buffer and quantified using BCA Protein Assay Kit (Thermo Scientific, Waltham, MA, USA).

For the proteomics experiment, 15 µg of protein extract was reduced, alkylated, and digested using Trypsin Gold (Promega, Madison, WI, USA). Digests were acidified, desalted, and TMT-labelled following the protocol described by Zecha et al. (30) using the TMT 11-plex labelling reagent (Thermo Fisher, Waltham, MA, USA). TMT batches were organised to always include one reference sample that is common to all batches in order to enable normalisation. Liquid chromatography-mass spectrometry (LC-MS) measurements were run on a Fusion Lumos Tribrid mass spectrometer (Thermo Fisher, Waltham, MA, USA) operated in data-dependent acquisition mode and multi-notch MS3 mode. MaxQuant version 1.6.3.4 (31) was used for peptide identification, and protein groups were obtained. Missing values were imputed with the minimal value across the dataset (32).

2.6. Differential expression in proteomics data

Prior to any analysis, MS data were adjusted with respect to one identical control sample that was present in each MS batch as described previously (32). Recalibrated intensities were log transformed for normalisation and proteins that were not detected in all samples were removed. An initial exploratory inspection of data by PCA and cluster analysis revealed that samples were grouped by MS batch (data not shown). Therefore, we carried out the differential expression analysis using the limma (v3.50.3) package (33) in R, including the MS batch as a covariate in the model to adjust for this confounding factor. We used the removeBatchEffect function from limma to recapitulate the exploratory analysis after batch correction and we observed no more clustering by MS batch. Finally, we took a nominal p-value of 0.05 as a threshold to define differentially expressed proteins (DEPs).

We carried out enrichment analysis just like we did for transcriptomics data. As a background, we considered the proteins that we detected in all samples with a valid EntrezID (a total of 5894 genes).

3. Results

3.1. Transcriptomic profiles in primary fibroblast cell cultures

First of all, we examined the similarity between the transcriptomic profiles obtained from primary fibroblast cell cultures and those from several brain areas, in order to understand how many of the molecular alterations that we identify could be extrapolated to neural tissues. We used publicly available data from the Genotype-Tissue Expression (GTEx) project and compared mean TPM (Transcripts per Kilobase Million) in fibroblast cultured cells and 11 brain areas: amygdala, anterior cingulate cortex, caudate basal ganglia, frontal cortex, cerebellar hemisphere, substantia nigra, hippocampus, hypothalamus, nucleus accumbens basal ganglia, putamen basal ganglia and spinal cord. 98.5% of detected transcripts (TPM>0.5) in GTEx cultured fibroblasts RNAseq samples correspond to genes with some degree of expression in at least one neural tissue [Figure S3]. More than 99% of the transcripts detected in our analysis are also reliably detected in GTEx cultured fibroblasts samples, indicating that the vast majority of the data that we are analyzing may be extrapolated to biological processes occurring in the brain and therefore may impact neurological phenotypes.

3.2. Characterisation of RTT-*MECP2* versus controls

There were similar *MECP2* mRNA amounts in patients with RTT and controls, whereas MeCP2 protein amount was significantly reduced in patients with RTT [Figure 1A, B]. We found a significant correlation between MeCP2 levels and the Pineda clinical severity score of our patients with RTT, indicating that more severely affected patients present lower amounts of MeCP2 protein [Figure 1C].

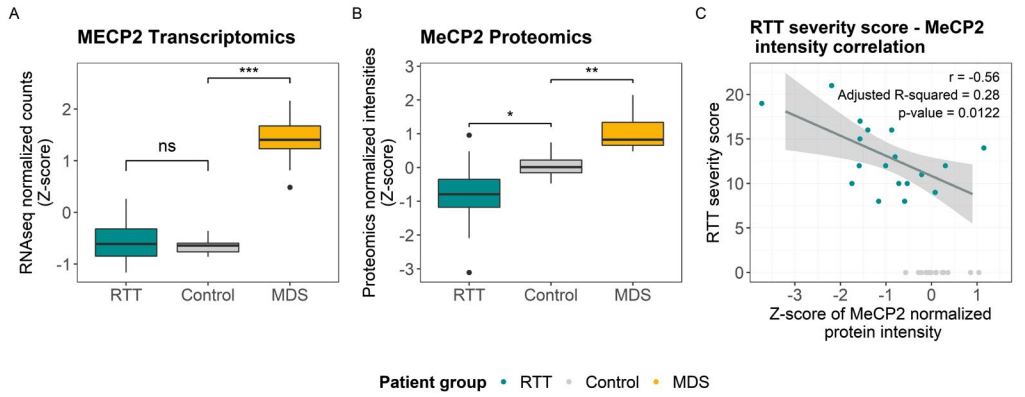

Figure 1: A) *MECP2* expression levels for RTT, MDS and control individuals obtained by RNAseq. B) MeCP2 intensity levels for RTT, MDS and control individuals obtained by proteomics. C) Pearson's correlation between the severity score of RTT patients and MeCP2 levels.

Transcriptomics and upstream regulator analysis

Differential expression analysis of patients with RTT carrying *MECP2* mutations versus healthy controls showed 3446 DEGs, 1713 upregulated and 1733 downregulated [Figure 2A]. We subsequently used these DEGs as input for upstream regulator analysis with ChEA3. We inspected the top 40 ranked TFs searching for proteins that regulate a large number of the identified DEGs, since they would potentially be driving some of these transcriptomic alterations. The list of DEGs was significantly enriched in *CREB1* and *SRF* targets (Fisher's exact test $p < 0.05$ in 5 of the 6 primary libraries in ChEA3). These two TFs have remarkable functions in neural tissues and could regulate the expression of 1253 and 1017 of the identified DEGs, respectively [Figure2B, Table S5a]. More than 98% of these potential targets have some degree of expression in at least one region of the nervous system, indicating that the alterations in transcriptomic networks identified in primary fibroblast cell cultures may affect the nervous system as well.

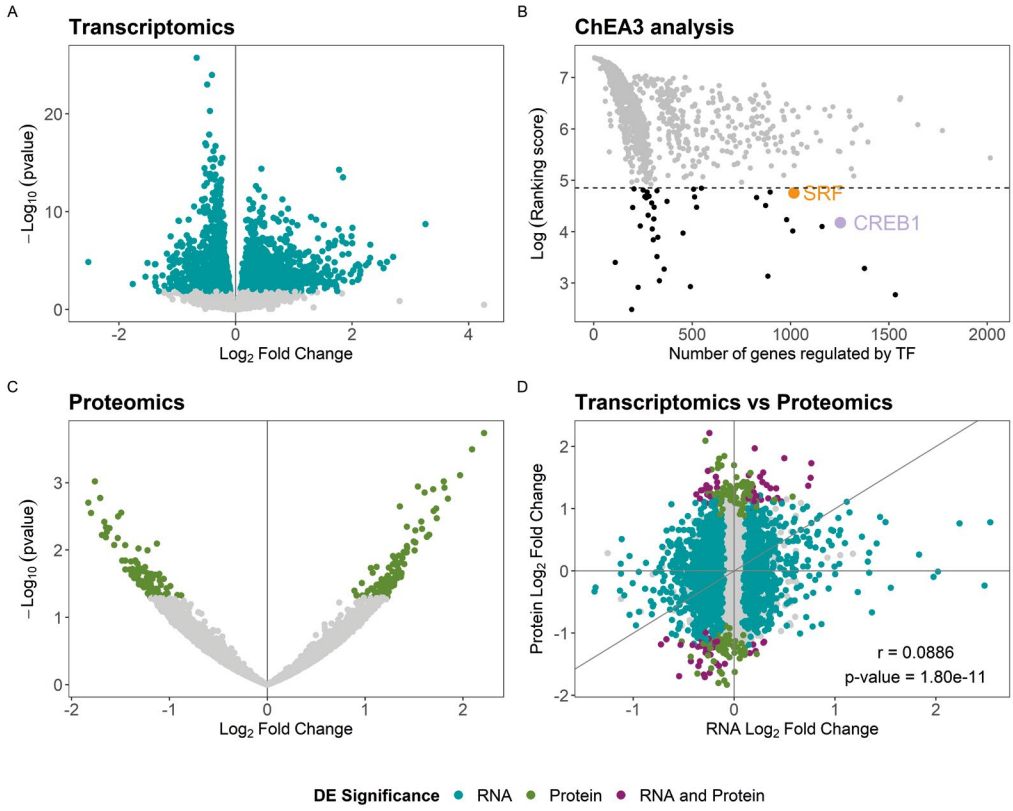

Figure 2: Summary of the results of the RTT-*MECP2* versus healthy controls analysis. A) RNAseq DE analysis results. The coloured genes are considered differentially expressed, passing a threshold of BH<0.05. B) Upstream TF ChEA3 analysis for the DEGs. The 40 TFs that were further studied are colored in black. C) Proteomics DE analysis results. The coloured proteins are considered differentially expressed, passing a threshold of nominal p.value <0.05. D) An integrated view of the transcriptomics and proteomics results. The genes that are significant at both analysis are coloured in purple.

Multi-omics: integrating transcriptomics and proteomics data

Proteomics differential expression analysis revealed 224 DEPs, 123 upregulated and 101 upregulated [Figure 2C, Table S3a,b]. 33 and 28 of these are CREB1 and SRF targets, respectively. The number of DEPs is markedly lower than the number of DEGs identified in transcriptomics, probably in part due to the fact that mass spectrometry identified roughly half (5918) of the number of genes mapped in the RNAseq experiment (12448). Almost 97% of the genes detected via mass spectrometry were also identified in RNAseq. Although the correlation between transcriptome and proteome differential expression

findings was not high [Pearson correlation coefficient = 0.09, p = 1.8e-11, Figure 2D], we found 75 genes deregulated at both the RNA and protein levels in patients with RTT [Supplementary Table 3C]. The overlap between DEGs and DEPs is not significantly higher than expected by chance (Fisher's exact test p = 0.1397, OR = 1.18), but some of the concordant genes constitute strong candidates for deciphering some of the pathomechanisms behind RTT, as well as for establishing biomarkers of this disorder [Table 2].

Enrichment analysis uncovered significant overrepresentation of genes and proteins involved in several cellular functions and processes, some of which may be extrapolated to neuronal tissues and thus are especially interesting when trying to elucidate the pathomechanisms underlying RTT [Table S4a,b] The most remarkable pathways that repeatedly appeared significantly enriched with DEGs and DEPs were cytoskeletal processes, vesicular activity, rRNA processing and mRNA splicing [Figure 3, Table 2]. The vast majority of the consistent DEGs and DEPs driving this enrichment have some degree of expression in at least one brain area according to GTEx publicly available data.

Table 2: Genes with concordant differential expression in transcriptomics and proteomics, that are involved in the main biological processes identified via enrichment analysis.

Gene	Direction	Biological process	Potential TF
AFAP1	Upregulated	Cytoskeletal processes	SRF
FMNL2	Upregulated	Cytoskeletal processes	CREB1
FNBP1L	Upregulated	Cytoskeletal processes	CREB1
KIF3A	Upregulated	Cytoskeletal processes	-
MARCKSL1	Upregulated	Cytoskeletal processes	-
PLS3	Upregulated	Cytoskeletal processes	SRF
ARMC9	Downregulated	Cytoskeletal processes	SRF
ARHGEF1	Downregulated	Cytoskeletal processes	-
CDC42EP1	Downregulated	Cytoskeletal processes	-
IQGAP3	Downregulated	Cytoskeletal processes	SRF
PLXNB2	Downregulated	Cytoskeletal processes	CREB1, SRF

Gene	Direction	Biological process	Potential TF
EIF4G3	Upregulated	RNA processing	-
NUDT12	Upregulated	RNA processing	-
SART1	Downregulated	RNA processing	CREB1, SRF
DDX31	Downregulated	RNA processing	CREB1, SRF
DDX54	Downregulated	RNA processing	SRF
MYBBP1A	Downregulated	RNA processing	SRF
NCALD	Upregulated	Vesicular activity	-
PREPL	Upregulated	Vesicular activity	CREB1
TMED1	Downregulated	Vesicular activity	SRF
ZFPL1	Downregulated	Vesicular activity	CREB1, SRF
AGPAT3	Downregulated	Metabolism	-
AACS	Downregulated	Metabolism	CREB1
CTBS	Upregulated	Metabolism	-
DCAKD	Downregulated	Metabolism	CREB1
HS2ST1	Upregulated	Metabolism	-
ORMDL2	Downregulated	Metabolism	-
PCK2	Downregulated	Metabolism	-
PI4KB	Downregulated	Metabolism	-
UAP1L1	Downregulated	Metabolism	CREB1
COMT	Downregulated	Metabolism	CREB1

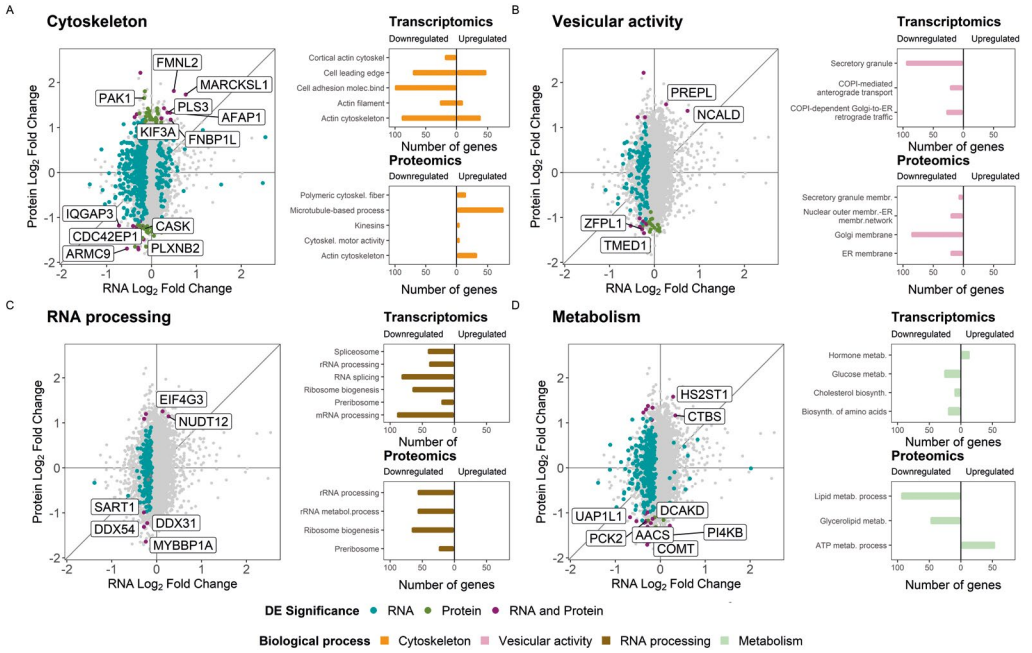

Figure 3: Summary of the enriched Biological Processes found in RTT-*MECP2* versus healthy control analysis. A) Significant DEGs and DEPs related to cytoskeleton (coloured in the dotplot) and the enriched terms found for those DEGs (upper barplot) and DEPs (lower barplot). B) Significant DEGs and DEPs related to vesicular activity and the enriched terms found for them. C) Significant DEGs and DEPs related to RNA processing and the enriched terms found for them. D) Significant DEGs and DEPs related to metabolism and the enriched terms found for them.

3.3. Patients with RTT versus patients with MDS

We compared the results of the DE analysis performed with patients with classical RTT and patients with MDS to identify common gene expression dysregulations that could shed some light into the pathomechanism underlying both syndromes.

Transcriptomics DE analysis of male patients with MDS versus male controls, and female patients with RTT versus female controls revealed 2465 and 3716 DEGs, respectively. Proteomics DE analysis returned 300 and 238 DEPs, respectively. Of these, 721 DEGs and 12 DEPs are shared between both groups, but only 2 genes are dysregulated with both omics in both syndromes [Figure

4F]. Those common DE genes are *MYO1C* and *HARS2*. *MYO1C* (OMIM*606538) is a myosin involved in cytoskeletal organisation and vesicle trafficking to the plasma membrane and is consistently downregulated. *HARS2* (OMIM*600783) is a mitochondrial histidyl-tRNA synthetase 2. At the RNA levels, it is consistently upregulated in patients with MDS and downregulated in patients with RTT. At the protein level, however, it is upregulated in both sets of patients.

Because *MECP2* expression is decreased in RTT and increased in MDS, we wondered whether they share DEGs that are expressed in opposite directions. In our cohort, 82 DEGs were positively correlated with MeCP2 expression levels (hence, upregulated in MDS and downregulated in RTT), and 100 DEGs were negatively correlated with MeCP2 expression (upregulated in RTT and downregulated in MDS) [Table S6a,b]. Enrichment analysis of those two gene sets revealed that pathways related to cytoskeleton and mRNA processing are altered. In addition, we found other molecular functions and pathways commonly altered between the RTT and MDS cohorts, some of which could help to understand why these two syndromes share clinical traits [Figure 4A,B; Table S4c].

The 82 DEGs downregulated in patients with RTT and upregulated in patients with MDS are overrepresented in terms related to mRNA processing and cell cycle [Figure 4A, Table S4c]. mRNA-splicing related genes appear dysregulated in both analyses. Interestingly, 8 of the 82 DEGs are part of spliceosome complexes and another four are related to mRNA stability, processing, and maturation functions. When looking at the ChEA3 TF enrichment analysis that regulates the same 82 DEGs, we found several TFs, most of them zinc finger proteins, described as cell cycle regulators and also *SRF*, which we found in the RTT ChEA3 analysis [Figure 4C, Table S5b]. These results are consistent with our findings in transcriptomics enrichment.

The 100 DEGs upregulated in RTT and downregulated in MDS enrich processes related to neurogenesis regulation; signalling cascades, such as Wnt, BMP, and TGFß; and the cytoskeleton [Figure 4B, Table S4c]. TF analysis with ChEA3 for the 100 DEGs revealed that *CREB1* (BH < 0.05 in DE analysis) is upregulated and *SRF* is downregulated, and that they regulate 39% and 22% of the shared 100 DEGs, respectively. Moreover, the following TFs related to neuronal function are also enriched in the ChEA3 analysis: *HEYL, GLIS2, NFATC4* and *JUN* [Figure 4C, Table S5c].

Among the shared 12 DEPs, three, *APPL2*, *CNPY4*, and *CTSC*, regulate immune response and are downregulated in RTT and upregulated in MDS [Figure 4E, Table

S6c]. Two other DEPs are related to cytoskeleton functions: *REPS1* and *CNN1*. *REPS1* (OMIM*614825) is a signalling adaptor protein that mediates cytoskeletal changes as endocytosis, and the protein is upregulated in both syndromes. *CNN1* (OMIM*600806) can bind to the cytoskeleton and produce smooth muscle contractions and is upregulated in RTT and downregulated in MDS.

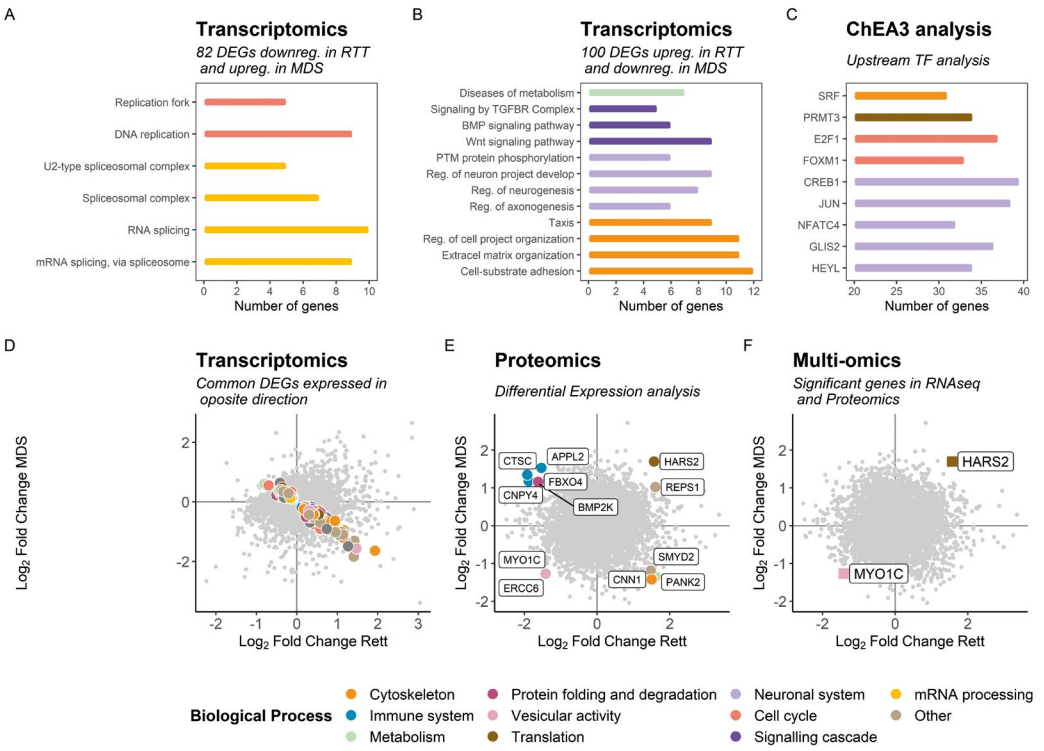

Figure 4: Summary of the common findings between the analysis of female RTT-*MECP2* patients versus female controls and male MDS patients versus male controls. A) Enrichment analysis results for the shared 82 DEGs downregulated in RTT and upregulated in MDS, coloured by Biological Process (BP). B) Enrichment analysis results for the shared 100 DEGs upregulated in RTT and downregulated in MDS coloured by BP. C) Relevant TF from ChEA3 analysis for the 82 and 100 DEGs dysregulated at transcriptomic level coloured by BP. D) Transcriptomics DE analysis results. The common 182 DEGs expressed in opposite directions are coloured by BP. E) Common 12 DEPs coloured by BP. F) Common DEGs and DEPs obtained from a multi-omics approach.

3.4. Patients with RTT versus patients with RTT-like phenotypes

Our RTT-like cohort was recruited considering their resemblance to the RTT phenotype. It encompassed nine patients with mutations in five different genes plus three patients without an established molecular diagnosis. The greater heterogeneity of this group complicated the identification of DEGs, as well as the interpretation of the differential expression results. Therefore, we established a significance threshold of BH < 0.1 for transcriptomics to be able to call DEGs despite the data heterogeneity. We interpreted these results in comparison with those obtained in typical patients with RTT, searching for shared molecular alterations that could constitute common grounds in the pathogenesis of overlapping disorders of diverse genetic nature.

DE analysis of transcriptomics data revealed 63 genes consistently altered in patients with RTT and RTT-like phenotypes (25 upregulated and 38 downregulated) [Figure 5A, Table S6d]. SRF targets were significantly overrepresented in these common DEGs, with 31 putative targets out of 63 common DEGs). This could implicate SRF transcriptional regulation as a common mechanism linking the molecular phenotypes in RTT-spectrum disorders.

Proteomics data showed 81 proteins consistently dysregulated (39 upregulated and 42 downregulated) [Figure 5C; Table S6e], but no gene was altered both in transcriptomics and proteomics reaching significance. Nevertheless, some of the candidate genes identified in the multi-omics approach in RTT patients maintained a consistent dysregulation at the protein level in patients with RTT-like phenotypes [Table 3].

Table 3: Candidate genes identified in multi-omics analysis of patients with RTT and with a concordant alteration at the protein level in patients with RTT-like phenotypes.

Gene	Direction	Biological process	Potential TF
ARMC9	Upregulated	Cytoskeletal processes	SRF
DDX31	Downregulated	RNA processing	CREB1, SRF
DDX54	Downregulated	RNA processing	SRF
MYBBP1A	Downregulated	RNA processing	SRF
COMT	Downregulated	Metabolism	CREB1

Transcriptomic and proteomic profiles of patients with RTT-like phenotypes are significantly correlated to those of patients with typical RTT ($r = 0.69$, adj-$R^2 = 0.47$, $p < 0.001$ in transcriptomics; $r = 0.75$, adj-$R^2 = 0.56$, $p < 0.001$ in proteomics) [Figure 5A and C]. Enrichment analysis of common DEGs and DEPs revealed terms related to cytoskeletal organisation, RNA processing, vesicular activity and metabolism, which constitute shared molecular alterations shared in patients with typical RTT and RTT-like phenotypes and could explain phenotypic overlap to some extent [Figure 5B and D; Table S4d-e].

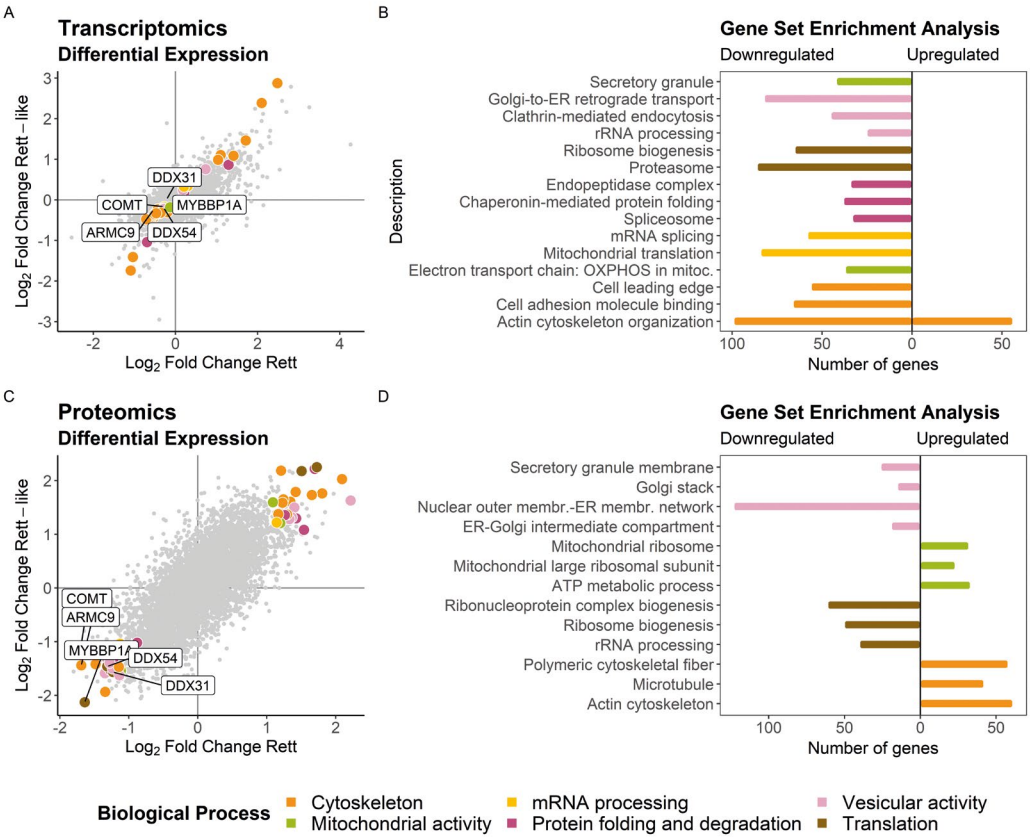

Figure 5: Summary of the common findings between the analysis of RTT-*MECP2* patients versus healthy controls and RTT-like patients versus healthy controls. A) Common 63 DEGs (25 upregulated and 38 downregulated), colored by Biological Process (BP). B) Gene Set Enrichment Analysis results for the shared 63 DEGs colored by BP. C) Common 81 DEPs (39 upregulated and 42 downregulated), colored by BP. D) Gene Set Enrichment Analysis results for the shared 81 DEGs colored by BP.

4. Discussion

4.1. Multi-omics expression in RTT patients with *MECP2* mutations

Cytoskeletal actin-filament-based processes play a crucial role in neuronal development, and their dysregulation is associated with cognitive disorders like RTT (34–37). We discovered a significant enrichment of cytoskeleton-related DEGs and DEPs in RTT patients compared to healthy controls.

Our study found that *ARMC9* (OMIM*617612), a gene involved in cilium assembly, signaling and transport, was significantly downregulated in both mRNA and protein levels in patients with RTT. Its implication in cytoskeletal dynamics and the cytoskeletal abnormalities found in patients with RTT suggest a potential link between ARMC9 and RTT pathogenesis (38). Scaffolding proteins, actin monomers, and regulatory proteins were upregulated in RTT patients. We found an upregulation of p21-activated kinase 1 (*PAK1*, OMIM*602590), essential for regulation of the actin cytoskeleton and controls dendritic spine morphogenesis and excitatory synapse formation (39). Moreover, Roux et al. (40) found an upregulation of proteins related to cytoskeletal motor activities, such as tubulin monomers and kinesins, that could be implicated in axonal transport to the neuronal growth cone. Our study also showed a downregulation of protein levels of Ca2+/calmodulin-activated Ser-Thr kinase (*CASK*, OMIM*300172), a scaffolding protein that is involved in synaptic transmembrane protein anchoring in the brain (41). CASK dysfunction is a promising route towards understanding some of the pathomechanisms behind RTT since it has been linked to neurodevelopmental disorders with overlapping phenotypes with RTT (42).

Another consistently downregulated mRNA and protein was *COMT* (OMIM+116790), a methyltransferase required for the metabolism and degradation of catecholamine neurotransmitters, including epinephrine, norepinephrine, and dopamine (43). Patients with RTT and RTT mouse models have shown low levels of these biogenic amines, and alteration in dopaminergic metabolism has been associated with the characteristic motor deficits of RTT (44).

We also found a significant enrichment in genes and proteins related to **vesicular activity** located in the Golgi apparatus and the nuclear outer membrane-endoplasmic reticulum membrane network, as well as secretory

vesicles. We found a significant upregulation of vesicular proteins located in neuronal axons. Prolyl endopeptidase like (*PREPL*, OMIM*609557) is a cytoplasmic protein with high expression in neuronal tissues. *PREPL* interacts with adaptor complex 1 (AP-1), a protein complex that plays an essential role in vesicular trafficking (45). NCALD is a neuronal calcium sensor protein that is involved in calcium signalling. It interacts with clathrin and actin and is involved in the modulation of endocytosis and synaptic vesicle recycling. NCALD was found to bind clathrin only at low calcium levels, resulting in inhibition and modulation of synaptic vesicle recycling (46). Our study also found a significant downregulation of ZFPL1, a cis-Golgi membrane protein that regulates trafficking from the endoplasmic reticulum to the Golgi apparatus and maintains cis-Golgi structural and functional integrity (47).

We identified a downregulation of genes and proteins involved in **rRNA processing** and ribosome biogenesis in patients with RTT (48–50). This could affect general protein translation in affected cells, possibly due to a reduction in mTORC1 activity (48,49). The downregulation of three proteins that interact with MeCP2, *MYBBP1A*, *DDX31*, and *DDX54*, could explain alterations in rRNA processing and mRNA splicing (51). The exact nature of the interaction between MeCP2 and these proteins is still unknown. We also observed downregulation of DEGs involved in **mRNA splicing** and spliceosomal complexes in patients with RTT. *MECP2* is known to interact with splicing factors (37,52,53), but a recent publication questions its role as a global regulator of splicing (54). Additional studies are needed to clarify MECP2's role in splicing since many genes involved in mRNA splicing are repeatedly dysregulated in different transcriptomics experiments.

CREB1 (OMIM*123810), which is a known MeCP2 interactor, regulates transcription in processes of relevance for neuronal survival and memory consolidation, among others (55,56). In astrocytes, it even regulates genes related to mitochondrial function, vesicle dynamics, and the cytoskeleton (57). Besides, one third of our DEGs are regulated by *CREB1* and *CREB1* itself was significantly upregulated in our cohort at the mRNA level. *SRF* (OMIM*600589), which is an integrator of mitogen-activated protein kinase (MAPK) and Rho-GTPase-mediated signalling, regulates cytoskeletal dynamics. SRF binds to the serum response element (SRE) sequence, present in a subset of cytoskeletal genes such as *ACTB* and immediate early genes (IEGs) (58). Besides, *SRF* regulates neuronal morphology and activity-dependent transcription (59) and suppression of *SRF*-mediated transcriptional responses has been found to produce a reduction in dendritic complexity in cortical neurons, which could

contribute to the neuronal spine dysgenesis phenotype observed in patients with an RTT-spectrum phenotype (60).

None of the genes regulated in opposite directions in transcriptomics and proteomics were known MeCP2 partners. We analysed the functional relationships between them, but no clear biological processes were identified. We hypothesise that the discordance in transcriptomics and proteomics may be due to cellular compensatory processes.

4.2. RTT and MDS: *MECP2* gene, two syndromes

Our study found that there are common genes between RTT and MDS. We found two shared genes, *MYO1C* and *HARS2*, which are a cytoskeletal component and a tRNA synthetase, respectively, in common significant DEGs and DEPs. The AKT/mTOR signaling pathway is downregulated in Mecp2 null models, indicating a deregulation of transcription followed by a limited ability to generate functional proteins (48). Both syndromes seem to have a deregulation of the correct protein synthesis.

82 DEGs were downregulated in patients with RTT and upregulated in patients with MDS, indicating that *MECP2* is important for mRNA processing. Moran-Salvador et al. found a downregulated group of genes involved in DNA replication and cell proliferation in hepatic stellate cells of Mecp2-null mice and suggested inhibition of Mecp2 phosphorylation as a liver fibrosis treatment (61).

The 100 DEGs upregulated in RTT and downregulated in MDS revealed processes related to neurogenesis regulation, cytoskeleton, and Wnt, BMP, and TGFß signaling cascades. The Wnt, BMP, and TGFß signaling pathways are also involved in osteoblast activity and maintenance of cartilage (62–64). Patients with RTT suffer from scoliosis, low bone mass density, and a higher bone fracture rate than the general population (65,66). Scoliosis is the most commonly reported orthopedic issue in patients with MDS, and osteopenia, contractures of joints, and fractures have also been reported (67).

Our results detected four transcription factors related to neuronal function, *HEYL, GLIS2, NFATC4,* and *JUN*. *HEYL* and *GLIS2* promote neuronal differentiation (68,69), while *NFATC4* regulates adult hippocampal neurogenesis and shares a common signaling process with *BDNF* for neuron maturation (70,71). *BDNF* modulates many aspects of neuronal development, synaptic

transmission and plasticity, and its dysregulation is found in RTT (72). *JUN* plays a role in neuronal migration and axon-dendritic architecture, and its inhibition reduces breathing abnormalities in RTT mice and induced Pluripotent Stem Cells neuronal models, and rescues the dendritic spine alterations (73).

Our findings indicate a resemblance of both syndromes at a molecular level, with several TFs involved in neural processes and dendritic complexity. Therapeutic strategies that seem promising for one syndrome could also benefit the other if the correct gene dosages are reached.

4.3. RTT-spectrum: one common clinical presentation, different mutated genes

The results of our study found that patients with RTT-spectrum disorders share common molecular alterations that could impact neuronal phenotypes. Almost one third of the common DEGs are involved in cytoskeleton organisation and regulation, and some of these have important functions in neurons. The malfunctioning of cytoskeletal genes with prominent functions in neurite development could lead to neuronal spine dysgenesis and, consequently, to the emergence of disorders with common traits derived from this structural synaptic dysfunction (74,75). The enrichment in putative SRF targets amongst shared DEGs highlights the potential implication of SRF transcriptional regulation in RTT-spectrum common molecular alterations leading to overlapping phenotypes.

We also detected an overrepresentation of several terms related to nervous system development and structure, supporting that common molecular alterations found in patients with RTT-spectrum phenotypes can impact neuronal phenotypes. The downregulation of *ARMC9* observed in patients with typical RTT can also be observed in patients with RTT-like phenotypes, constituting a link between RTT-spectrum disorders and the overlapping phenotype caused by loss-of-function variants in this gene.

The patients with RTT-spectrum phenotypes in our study shared a downregulation of *SNRPC* expression at the RNA level that was not replicated in proteomics. This transcriptional alteration was also previously found in post-mortem brain tissue and embryonic stem cells of patients with RTT (49,76). *SNRPC* is a spliceosome component involved in 5' splice-site recognition, so it may affect the splicing of many different targets and could constitute a shared

mechanism of splicing dysregulation of patients with RTT-spectrum phenotypes. The dysregulation of splicing factors and regulators has been described in RTT as well as in other monogenic intellectual disabilities and in autism spectrum disorders (ASD) (77).

Protein translation may be affected in all patients with RTT-spectrum phenotypes. Several rRNA processing and ribosome-biogenesis-related proteins found altered in patients with RTT were also consistently dysregulated in patients with RTT-like phenotypes, indicating this commonality. DDX54, DDX31, and MYBBP1A are MeCP2 partners and are linked to rRNA expression and pre-processing and could explain, at least to some extent, the shared dysregulation of ribosome biogenesis.

5. Conclusions

Numerous studies have investigated the transcriptomes of individuals with RTT, resulting in over 60 published articles. Our study found that studying other human tissues, such as fibroblasts, can reflect the same dysregulations caused by loss of function of *MECP2*. However, integrating all knowledge is complicated by the heterogeneity in experiments and tissue-specific effects of *MECP2*. Dysregulation of various cellular functions was identified, including cytoskeletal organization, vesicular activity, translation and mRNA processing, which are altered in patients with RTT, RTT-like phenotypes, and MDS. *ARMC9* could be a potential biomarker for RTT and RTT-spectrum disorders. TF analysis supports *CREB1* and *SRF* as potential therapeutic targets. Shared dysregulated biological processes and cellular functions were found between patients with RTT, MDS, and RTT-like phenotypes, with RTT and RTT-like being more similar than MDS. Further studies are necessary to validate these findings.

List of abbreviations

ASD: autism spectrum disorder; BH: Benjamini-Hochberg corrected p-value; bp; base pair; *CDKL5*: Cyclin-dependent kinase-like 5; cDNA: double-stranded complementary DNA; CEIC : Comitè d'Ètica d'Investigació Clínica-Fundació Sant Joan de Déu; ChIP-X Enrichment Analysis 3; CPM: Counts per million mapped reads; DEG: Differential expression genes; DEP: Differential expression proteins;

EEGs :Electroencephalograms; *FOXG1*: Forkhead box G1; GRCh37 (hg19): Homo sapiens (human) genome assembly GRCh37 (hg19) from Genome Reference Consortium; GSEA: GO: Gene Ontology; HSJD: Hospital Sant Joan de Déu; ID: Intelectual Disability; IEGs : Immediate early genes; iPSCs: induced Pluripotent Stem Cells.KEGG: Kyoto Encyclopedia of Genes and Genomes; LC-MS: Liquid chromatography-mass spectrometry; MDS: MECP2-Duplication syndrome; *MECP2:* Methyl-CpG-binding protein 2; NGS: Next Generation Sequencing; NF-κB : nuclear factor kappa B; ORA: Overrepresentation analysis; PCA : Principal component analysis; RNA-seq: RNA-sequencing; RP: Reactome pathway; RT-qPCR : Reverse transcription-quantitative polymerase chain reaction; RTT: Rett syndrome; RTT-like: Rett-like; TF: Transcription factors; ChEA3: WP: WikiPathways; XCI: X-chromosome inactivation;

Supplementary information

Additional File 1. Table S1: Sequence of the designed primers. Table S2: Detailed description of the study cohort. Table S3: RNAseq and proteomics differential expression results for RTT - MECP2 versus healthy controls. Table S4: Enrichment analysis results for the RTT-MECP2 versus healthy controls, RTT versus MDS, RTT versus RTT-like, transcriptomic and proteomic data. Table S5: ChEA3 upstream analysis results for RTT patients versus healthy controls, RTT versus MDS, RTT versus RTT-like. Table S6: Summary for the DEG and DEP from RTT-MECP2 vs MDS and RTT-MECP2 versus RTT-like.

Additional File 2. Figure S1: Detection of fibroblast samples under oxidative stress conditions. Figure S2: Comparison of the gene expression results between RNAseq and RT-qPCR.

Declarations

Ethics approval and consent to participate

All individuals included or their legal guardians provided written informed consent before evaluation. The study was approved by the ethical committee of the Fundació Sant Joan de Déu (#PIC-219-20). The research conformed to the principles of the Declaration of Helsinki.

Consent for publication

All individuals included or their legal guardians provided written consent to share pseudonymized patient data and analysis data.

Availability of data and methods

Our ethics approval and consent agreements allow us to share nonidentifable patient data and analysis data only, as such, we cannot provide BAM or VCF fles. The analysis data provided are the gene expression count matrices, as well as the privacy-preserving count matrices of split and unsplit reads overlapping annotated splice sites from RNA-seq. They will be available for download without restriction when this article will be published.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded with a grant from the Spanish Ministry of Health (Instituto de Salud Carlos III/FEDER, PI20/00389), the parent association 'Síndrome duplicación MECP2: Miradas que hablan' (PFNR0085), funding from Muévete por los que no Pueden (PCP/00282), the doctoral grant FI from the Government of Catalonia (Secretaria d'Universitats I Recerca) and the European Social Fund (2020 FI-B 00888) to Ainhoa Pascual-Alonso, and the FPU (Formación del Profesorado Universitario) doctoral grant from the Spanish Ministry of Science, Innovation and Universities (FPU18/02152) to Clara Xiol.

Authors' contributions

Conceptualization: APA, CX, JA. Design of the work: JA. Acquisition: CX, APA. Data Analysis: APA, CX, DS, RK. Data Interpretation: APA, CX, JA. Supervision: JA, HP. Writing—original draft preparation: CX, APA, JA. Writing—review & editing: all authors. All authors read and approved the final manuscript.

Acknowledgements

We thank all patients and their families for their collaboration and support in this study. We would like to thank the Cytogenetics Department of the Hospital Sant Joan de Déu, especially Lluisa, Teresa, Carlos, Rosa, and Esther for starting some of the fibroblast primary cell cultures. We would like to thank the Bioinformatics department of the Hospital Sant Joan de Déu for their help in optimising the RNAseq data obtention protocol. We would like to thank the Proteomics Core Facility at TUM for performing the proteomics experiments. We thank the 'Biobanc de l'Hospital Infantil Sant Joan de Déu per a la Investigació' integrated in the Spanish Biobank Network of Instituto de Salud Carlos III (ISCIII) for sample storage.

Bibliography

1. Ehrhart F, Coort SLM, Cirillo E, Smeets E, Evelo CT, Curfs LMG. Rett syndrome - Biological pathways leading from MECP2 to disorder phenotypes. Orphanet J Rare Dis. 2016 Nov 25;11(1):158.
2. Neul JL, Kaufmann WE, Glaze DG, Christodoulou J, Clarke AJ, Bahi-Buisson N, et al. Rett syndrome: revised diagnostic criteria and nomenclature. Ann Neurol. 2010 Dec;68(6):944–50.
3. Neul JL, Fang P, Barrish J, Lane J, Caeg EB, Smith EO, et al. Specific mutations in Methyl-CpG-Binding Protein 2 confer different severity in Rett syndrome. Neurology. 2008;70(16):1313–21.
4. Bahi-Buisson N, Nectoux J, Rosas-Vargas H, Milh M, Boddaert N, Girard B, et al. Key clinical features to identify girls with CDKL5 mutations. Brain. 2008;131(10):2647–61.
5. Ariani F, Hayek G, Rondinella D, Artuso R, Mencarelli MA, Spanhol-Rosseto A, et al. FOXG1 is responsible for the congenital variant of Rett syndrome. Am J Hum Genet. 2008;83(1):89–93.
6. Vidal S, Xiol C, Pascual-Alonso A, O'Callaghan M, Pineda M, Armstrong J. Genetic landscape of Rett syndrome spectrum: improvements and challenges. Int J Mol Sci. 2019 Aug 12;20(16):3925.
7. Schönewolf-Greulich B, Bisgaard AM, Møller RS, Dunø M, Brøndum-Nielsen K, Kaur S, et al. Clinician's guide to genes associated with Rett-like phenotypes–Investigation of a Danish cohort and review of the literature. Clin Genet. 2019 Feb 1;95:221–30.

8. Ehrhart F, Sangani NB, Curfs LMG. Current developments in the genetics of Rett and Rett-like syndrome. Curr Opin Psychiatry. 2018 Mar 1;31(2):103–8.
9. Van Esch H. MECP2 duplication syndrome. Mol Syndromol. 2011 Apr;2:128–36.
10. Peters SU, Fu C, Suter B, Marsh E, Benke TA, Skinner SA, et al. Characterizing the phenotypic effect of Xq28 duplication size in MECP2 duplication syndrome. Clin Genet. 2019 May;95:575–81.
11. Murdock DR, Dai H, Burrage LC, Rosenfeld JA, Ketkar S, Müller MF, et al. Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. J Clin Invest. 2021 Jan 4;131(1):e141500.
12. Yépez VA, Gusic M, Kopajtich R, Mertes C, Smith NH, Alston CL, et al. Clinical implementation of RNA sequencing for Mendelian disease diagnostics. Genome Med. 2022 Dec 1;14:38.
13. Gomathi M, Padmapriya S, Balachandar V. Drug studies on Rett syndrome: from bench to bedside. J Autism Dev Disord. 2020 Aug 1;50(8):2740–64.
14. Pacheco NL, Heaven MR, Holt LM, Crossman DK, Boggio KJ, Shaffer SA, et al. RNA sequencing and proteomics approaches reveal novel deficits in the cortex of Mecp2-deficient mice, a model for Rett syndrome. Mol Autism. 2017;8(56).
15. Vidal S, Brandi N, Pacheco P, Gerotina E, Blasco L, Trotta JR, et al. The utility of Next Generation Sequencing for molecular diagnostics in Rett syndrome. Sci Rep. 2017;7:12288.
16. Pascual-Alonso A, Blasco L, Vidal S, Gean E, Rubio P, O'Callaghan M, et al. Molecular characterization of Spanish patients with MECP2 duplication syndrome. Clin Genet. 2020 Apr 1;97:610–20.
17. Monrós E, Armstrong J, Aibar E, Poo P, Canós I, Pineda M. Rett syndrome in Spain: mutation analysis and clinical correlations. Brain Dev. 2001;23:S251–3.
18. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. Am J Hum Genet. 1992;51:1229–39.
19. Al-Kafaji G, Sabry MA, Bakhiet M. Increased expression of mitochondrial DNA-encoded genes in human renal mesangial cells in response to high glucose-induced reactive oxygen species. Mol Med Rep. 2016;13(2):1774–80.
20. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-new capabilities and interfaces. Nucleic Acids Res. 2012 Aug;40(15):e115.

21. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.
22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15:550.
23. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innov. 2021 Aug 28;2(3):100141.
24. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016;12(2):477–9.
25. Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, et al. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2021 Jan 8;49(D1):D325–34.
26. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021 Jan 8;49(D1):D545–51.
27. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, et al. WikiPathways: connecting communities. Nucleic Acids Res. 2021 Jan 8;49(D1):D613–21.
28. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022 Jan 7;50(D1):D687–92.
29. Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res. 2019 Jul 1;47(W1):W212–24.
30. Zecha J, Satpathy S, Kanashova T, Avanessian SC, Kane MH, Clauser KR, et al. TMT labeling for the masses: a robust and cost-efficient, in-solution labeling approach. Mol Cell Proteomics. 2019;18(7):1468–78.
31. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016 Dec 1;11(12):2301–19.
32. Kopajtich R, Smirnov D, Stenton SL, Loipfinger S, Meng C, Scheller IF, et al. Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders. medRxiv
33. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Jan 6;43(7):e47.

34. Pecorelli A, Cervellati C, Cortelazzo A, Cervellati F, Sticozzi C, Mirasole C, et al. Proteomic analysis of 4-hydroxynonenal and nitrotyrosine modified proteins in RTT fibroblasts. Int J Biochem Cell Biol. 2016 Dec 1;81:236–45.
35. Cortelazzo A, De Felice C, Pecorelli A, Belmonte G, Signorini C, Leoncini S, et al. Beta-actin deficiency with oxidative posttranslational modifications in Rett syndrome erythrocytes: insights into an altered cytoskeletal organization. PLoS One. 2014;9(3).
36. Varderidou-Minasian S, Hinz L, Hagemans D, Posthuma D, Altelaar M, Heine VM. Quantitative proteomic analysis of Rett iPSC-derived neuronal progenitors. Mol Autism. 2020 May 27;11:38.
37. Ehrhart F, Coort SL, Eijssen L, Cirillo E, Smeets EE, Bahram Sangani N, et al. Integrated analysis of human transcriptome data for Rett syndrome finds a network of involved genes. World J Biol Psychiatry. 2019;21(10):712–25.
38. Latour BL, van de Weghe JC, Rusterholz TDS, Letteboer SJF, Gomez A, Shaheen R, et al. Dysfunction of the ciliary ARMC9/TOGARAM1 protein module causes Joubert syndrome. J Clin Invest. 2020 Aug 3;140(8):4423–39.
39. Zhang H, Webb DJ, Asmussen H, Niu S, Horwitz AF. A GIT1/PIX/Rac/PAK signaling module regulates spine morphogenesis and synapse formation through MLC. J Neurosci. 2005 Mar 30;25(13):3379–88.
40. Roux JC, Zala D, Panayotis N, Borges-Correia A, Saudou F, Villard L. Modification of Mecp2 dosage alters axonal transport through the Huntingtin/Hap1 pathway. Neurobiol Dis. 2012 Feb;45(2):786–95.
41. Setou M, Nakagawa T, Seog D-H, Hirokawa N. Kinesin superfamily motor protein KIF17 and mLin-10 in NMDA receptor–containing vesicle transport. Science (80-). 2000;288:1796–802.
42. Patel PA, Hegert J V., Cristian I, Kerr A, LaConte LEW, Fox MA, et al. Complete loss of the X-linked gene CASK causes severe cerebellar degeneration. J Med Genet. 2022 Nov 1;59(11):1044–57.
43. Chen J, Song J, Yuan P, Tian Q, Ji Y, Ren-Patterson R, et al. Orientation and cellular distribution of membrane-bound catechol-O-methyltransferase in cortical neurons: implications for drug development. J Biol Chem. 2011;286(40):34752–60.
44. Szczesna K, De La Caridad O, Petazzi P, Soler M, Roa L, Saez MA, et al. Improvement of the rett syndrome phenotype in a mecp2 mouse model upon treatment with levodopa and a dopa-decarboxylase inhibitor. Neuropsychopharmacology. 2014;39(12):2846–56.

45. Radhakrishnan K, Baltes J, Creemers JWM, Schu P. Trans-Golgi network morphology and sorting is regulated by prolyl-oligopeptidase-like protein PREPL and the AP-1 complex subunit μ1A. J Cell Sci. 2013 Mar 1;126(5):1155–63.
46. Riessland M, Kaczmarek A, Schneider S, Swoboda KJ, Löhr H, Bradler C, et al. Neurocalcin delta suppression protects against spinal muscular atrophy in humans and across species by restoring impaired endocytosis. Am J Hum Genet. 2017 Feb 2;100(2):297–315.
47. Chiu CF, Ghanekar Y, Frost L, Diao A, Morrison D, McKenzie E, et al. ZFPL1, a novel ring finger protein required for cis-Golgi integrity and efficient ER-to-Golgi transport. EMBO J. 2008;27(7):934–47.
48. Ricciardi S, Boggio EM, Grosso S, Lonetti G, Forlani G, Stefanelli G, et al. Reduced AKT/mTOR signaling and protein synthesis dysregulation in a Rett syndrome animal model. Hum Mol Genet. 2011 Mar;20(6):1182–96.
49. Li Y, Wang H, Muffat J, Cheng AW, Orlando DA, Lovén J, et al. Global transcriptional and translational repression in human-embryonic- stem-cell-derived rett syndrome neurons. Cell Stem Cell. 2013 Oct 3;13:446–58.
50. Rodrigues DC, Mufteev M, Weatheritt RJ, Djuric U, Ha KCH, Ross PJ, et al. Shifts in ribosome engagement impact key gene sets in neurodevelopment and ubiquitination in Rett syndrome. Cell Rep. 2020 Mar 24;30(12):4179–96.
51. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, et al. Architecture of the human interactome defines protein communities and disease networks. Nature. 2017 May 25;545(7655):505–9.
52. Cheng TL, Chen J, Wan H, Tang B, Tian W, Liao L, et al. Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain. Sci Rep. 2017;7:1–12.
53. Long SW, Ooi JYY, Yau PM, Jones PL. A brain-derived MeCP2 complex supports a role for MeCP2 in RNA processing. Biosci Rep. 2011 Oct;31(5):333–43.
54. Chhatbar K, Cholewa-Waclaw J, Shah R, Bird A, Sanguinetti G. Quantitative analysis questions the role of MeCP2 as a global regulator of alternative splicing. PLoS Genet. 2020;16(10):1–14.
55. Chen Y, Shin BC, Thamotharan S, Devaskar SU. Creb1-Mecp2-mCpG complex transactivates postnatal murine neuronal glucose transporter isoform 3 expression. Endocrinology. 2013;154(4):1598–611.
56. Kaldun JC, Sprecher SG. Initiated by CREB: resolving gene regulatory programs in learning and memory. BioEssays. 2019 Aug 1;41(8).

57. Pardo L, Valor LM, Eraso-Pichot A, Barco A, Golbano A, Hardingham GE, et al. CREB regulates distinct adaptive transcriptional programs in astrocytes and neurons. Sci Rep. 2017 Dec 1;7:6390.
58. Tabuchi A, Ihara D. Regulation of dendritic synaptic morphology and transcription by the SRF cofactor MKL/MRTF. Front Mol Neurosci. 2021 Nov 2;14:7678.
59. Knöll B, Nordheim A. Functional versatility of transcription factors in the nervous system: the SRF paradigm. Trends Neurosci. 2009 Aug;32(8):432 42.
60. Shiota J, Ishikawa M, Sakagami H, Tsuda M, Baraban JM, Tabuchi A. Developmental expression of the SRF co-activator MAL in brain: role in regulating dendritic morphology. J Neurochem. 2006 Sep;98(6):1778–88.
61. Moran-Salvador E, Garcia-Macia M, Sivaharan A, Sabater L, Zaki MYW, Oakley F, et al. Fibrogenic activity of MECP2 is regulated by phosphorylation in hepatic stellate cells. Gastroenterology. 2019 Nov 1;157:1398-1412.e9.
62. Liu J, Xiao Q, Xiao J, Niu C, Li Y, Zhang X, et al. Wnt/β-catenin signalling: function, biological mechanisms, and therapeutic opportunities. Signal Transduct Target Ther. 2022 Dec 1;7(3).
63. Lowery JW, Rosen V. The BMP pathway and its inhibitors in the skeleton. Physiol Rev. 2018;98:2431–52.
64. Finnson KW, Chi Y, Bou-Gharios G, Leask A, Philip A. TGF-beta signaling in cartilage homeostasis and osteoarthritis. Front Biosci. 2012;S4:251–68.
65. Downs J, Bebbington A, Woodhead H, Jacoby P, Jian L, Jefferson A, et al. Early determinants of fractures in Rett syndrome. Pediatrics. 2008 Mar 1;121(3):540–6.
66. Pecorelli A, Cordone V, Schiavone ML, Caffarelli C, Cervellati C, Cerbone G, et al. Altered bone status in Rett syndrome. Life. 2021 Jun 1;11:521.
67. Ta D, Downs J, Baynam G, Wilson A, Richmond P, Leonard H. A brief history of MECP2 duplication syndrome: 20-years of clinical understanding. Orphanet J Rare Dis. 2022 Dec 1;17:131.
68. Jalali A, Bassuk AG, Kan L, Israsena N, Mukhopadhyay A, McGuire T, et al. HeyL promotes neuronal differentiation of neural progenitor cells. J Neurosci Res. 2011 Mar;89(3):299–309.
69. Lamar E, Kintner C, Goulding M. Identification of NKL, a novel Gli-Kruppel zinc-finger protein that promotes neuronal differentiation. Development. 2001;128:1335–46.

70. Li L, Ke K, Tan X, Xu W, Shen J, Zhai T, et al. Up-regulation of NFATc4 involves in neuronal apoptosis following intracerebral hemorrhage. Cell Mol Neurobiol. 2013 Oct;33(7):893–905.
71. Ding B, Dobner PR, Mullikin-Kilpatrick D, Wang W, Zhu H, Chow CW, et al. BDNF activates an NFI-dependent neurodevelopmental timing program by sequestering NFATc4. Mol Biol Cell. 2018 Apr 15;29(8):975–87.
72. Li W, Pozzo-Miller L. BDNF deregulation in Rett syndrome. Neuropharmacology. 2014;76.
73. Musi CA, Castaldo AM, Valsecchi AE, Cimini S, Morello N, Pizzo R, et al. JNK signaling provides a novel therapeutic target for Rett syndrome. BMC Biol. 2021 Dec 1;19:256.
74. Xu X, Miller EC, Pozzo-Miller L. Dendritic spine dysgenesis in Rett syndrome. Front Neuroanat. 2014 Sep 10;8:1–8.
75. Phillips M, Pozzo-Miller L. Dendritic spine dysgenesis in autism related disorders. Neurosci Lett. 2015 Oct 3;601:30–40.
76. Colantuoni C, Jeon OH, Hyder K, Chenchik A, Khimani AH, Narayanan V, et al. Gene expression profiling in postmortem Rett syndrome brain: differential gene expression and patient classification. Neurobiol Dis. 2001;8:847–65.
77. Shah S, Richter JD. Do fragile X syndrome and other intellectual disorders converge at aberrant pre-mRNA splicing? Front Psychiatry. 2021 Sep 10;12:715346.

Chapter 3

Chapter 3: Characterization of dysregulated molecular processes in MDS patients and carrier mothers using transcriptomic and proteomic technologies.

In the third and last chapter we will discover which biological processes are altered in MDS patients and in the *MECP2* duplication carrier mothers. This chapter corresponds to the third objective of the thesis.

Omics technology has been used to decipher the insights of several disorders due to its capacity to obtain a huge amount of information from a unique sample with one experiment. Transcriptomic and proteomic of MDS patients have never been studied, but they have successfully been applied to study RTT samples. We have carried out the same workflow we optimised to study the RTT patients in fibroblast primary cell lines from 17 MDS patients and 10 carrier mothers, and performed RNAseq and TMT-MS. The following publication will show, for the first time, the significantly dysregulated processes in a male and a female MDS cohort and in a *MECP2* duplication asymptomatic carrier cohort when multi-omics (transcriptomic and proteomic) approach is conducted.

Publication 1: Molecular profiles of *MECP2* duplication syndrome patients and carriers.

Authors: Ainhoa Pascual-Alonso, Clara Xiol, Dmitrii Smirnov, Robert Kopajtich, Holger Prokisch and Judith Armstrong.

Reference: Author's manuscript. Submitted to Experimental and Molecular Medicine.

Abstract:

Two of the most common omics, transcriptomic and proteomic, have been used to study the effect of the dysregulation of *MECP2* in RTT samples, but not in MDS. The fibroblast primary cell lines that we established from skin biopsies of 17 MDS patients (15 males and 2 females) and 10 *MECP2* duplication carrier mothers, together with 13 healthy controls (6 males and 7 females) were selected to perform RNAseq and TMT-MS. In addition, the results from the male MDS patient cohort were compared to 21 classic RTT patient-derived fibroblast samples.

The first detailed description of the altered processes found in transcriptomic and proteomic studies in male MDS patients, female MDS patients and carrier mothers is presented in this manuscript. We found that in male MDS patients DEGs and DEPs related to cytoskeleton, neuronal system, vesicular transport and immune system are mainly downregulated. In female MDS patients, however, the DEGs and DEPs mainly participate in translation and mRNA processing. Despite the high number of DEGs and DEPs obtained separately in the *MECP2* duplication carrier mothers' study, a multi-omics analysis revealed few enriched terms, and no specific biological process seems to be altered. In addition, similarities and differences between the cohorts are also discussed. With this multi-omics approach, some genes implicated in neuronal functions that are expressed in the brain and that are not associated with any other disorder presenting ID or neurodevelopmental delay have been highlighted as biomarker candidates for MDS. These candidate genes are *TMOD2, SRGAP1, COPS2*, *CNPY2*, *IGF2BP1, MOB2, VASP, FZD7, ECSIT* and *KIF3B.* In spite of the shared phenotypes of the studied cohorts, defining the RNA and protein profiles have shown that our four cohorts are less alike than expected.

Molecular profiles of *MECP2* duplication syndrome patients and carriers.

Ainhoa Pascual-Alonso[1,2], Clara Xiol1,[2], Dmitrii Smirnov[3,4], Robert Kopajtich[3,4], Holger Prokisch[3,4], Judith Armstrong[#2,5,6]

1. Fundació per la Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain.
2. Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain.
3. Institute of Human Genetics, Technical University of Munich, Munich, Germany.
4. Institute of Neurogenomics, Helmholtz Zentrum München, Munich, Germany.
5. CIBER-ER (Biomedical Network Research Center for Rare Diseases), Instituto de Salud Carlos III (ISCIII), Madrid, Spain.
6. Genomic Unit, Molecular and Genetic Medicine Section, Hospital Sant Joan de Déu, Esplugues de Llobregat, Spain

#Corresponding author, judith.armstrong@sjd.es

0. Abstract

MECP2 duplication syndrome (MDS) is an X-linked neurodevelopmental disorder caused by the gain of dose of at least the genes *MECP2* and *IRAK1* and is characterised by intellectual disability (ID), developmental delay, early onset hypotonia, epilepsy and recurrent infections. It mainly affects males, although females can be mildly affected or asymptomatic carriers. Rett syndrome (RTT) is mainly triggered by loss of function mutations in *MECP2* and is a well described syndrome that presents ID, epilepsy, lack of purposeful hand use and impaired speech, among others. As a result of implementing omics technology, altered biological pathways in human RTT samples have been reported, but such molecular characterisation has not been performed in MDS patients. We gathered human skin fibroblasts from 17 patients with MDS, 10 *MECP2*

duplication carrier mothers and 21 RTT patients, and performed multi-omics (RNAseq and proteomics) analysis. Here, we provide a thorough description and compare the shared and specific dysregulated biological processes between the cohorts. We also highlight the genes *TMOD2, SRGAP1, COPS2*, *CNPY2*, *IGF2BP1, MOB2, VASP, FZD7, ECSIT*, *KIF3B and NR1H2* as biomarker and therapeutic target candidates due to their implication in neuronal functions. Defining the RNA and protein profiles has shown that our four cohorts are less alike than expected by their shared phenotypes.

Keywords: *MECP2*, *MECP2* duplication, Rett syndrome, RNAseq, TMT-mass spectrometry.

1. Introduction

MECP2 duplication syndrome (MDS; OMIM#300260) is an X-linked neurodevelopmental disorder that mainly affects males. It is a rare syndrome with just over 600 cases reported worldwide [1]. MDS is characterised by intellectual disability (ID), developmental delay, infantile hypotonia, epilepsy, poor or absent speech, progressive lower extremity spasticity, recurrent infections, gastrointestinal problems, autistic features and mild dysmorphic features [2]. Females harbouring the *MECP2* duplication present a variable phenotype ranging from asymptomatic carriers to mildly affected girls, depending on the X chromosome inactivation (XCI) status. *MECP2* duplications can be *de novo* or inherited from carrier mothers who are often asymptomatic, but some cases with neuropsychiatric symptoms or learning difficulties have been reported [3].

MDS is caused by the duplication of, at least, the genes *MECP2* (OMIM*300005) and *IRAK1* (OMIM*300283). Duplication locations, sizes, gene contents and dosages are specific to each family, and sometimes even for each individual [1]. Unfortunately, no clear genotype-phenotype correlation has been found to date.

Methyl-CpG-binding protein 2 (*MECP2*) is located on the long arm of the X chromosome (Xq28) and undergoes XCI in females. *MECP2* has four exons and produces two well-known transcripts that differ in the 5' region [4]. The resulting protein, MeCP2, is ubiquitously expressed, with the MeCP2_e1 isoform being predominant in the brain. Besides, MeCP2 levels positively correlate with the maturation of the neurons, highlighting the importance of MeCP2 in

neuronal maintenance and maturation [5]. *MECP2* is a transcriptional regulator that represses or activates transcription depending on its molecular context [6]. *MECP2* also functions as a chromatin remodeller; it interacts with RNA splicing machinery and microRNA processing machinery and participates in neuronal development, maturation and synapse formation [7,8].

Interleukin-1 receptor-associated kinase 1 (*IRAK1*) is located downstream of *MECP2* and is always duplicated and overexpressed in MDS. *IRAK1* is ubiquitously expressed and is part of the Toll-like receptor (TLR) and interleukin-1 receptor (IL-1R) signalling cascades, which are involved in pathogen recognition and the modulation of inflammatory and immune responses [9]. *IRAK1*'s role in the immune system makes it a candidate to explain the severe recurrent infections that MDS patients suffer.

Apart from causing MDS, mutations in *MECP2* trigger Rett syndrome (RTT; OMIM#312750), severe neonatal encephalopathy (OMIM#300673) or X-linked mental retardation syndrome (OMIM#300055). RTT is a neurodevelopmental disorder caused by loss of function mutations in *MECP2*. It is characterised by a normal early development followed by a psychomotor regression that occurs between the first 6 to 18 months of life which includes the appearance of stereotypic hand movements, ID, seizures, breathing disturbances and loss of speech, among others [10].

The main affected organ in MDS and RTT is the brain. The impossibility of obtaining human samples from the primarily altered tissue made us focus on a different target tissue. Skin fibroblasts have proven to be a useful resource; the gene expression is less variable than in whole blood and more disease-related genes are expressed [11].

Studying the global transcriptome and proteome is becoming incredibly useful for diagnostics and for research [12]. Few transcriptomics studies performed in MDS have been published: five studies in mice models [6,13–16], one study using modified cell lines [17] and one study with two MDS patients [18]. No proteomics experiments have been published to date. For RTT, however, there are around 70 published transcriptomic studies and around 30 proteomic studies performed in mice and human samples. Those publications have broadened the knowledge about the biological pathways dysregulated by *MECP2,* but the syndromes causing pathomechanisms remain unknown.

Promising strategies are being developed to restore the normal MeCP2 dosage, although there is no registered clinical trial for MDS in the U.S National Library of Medicine. CRISPR-Cas9 technology has been able to remove the duplication of *MECP2* in fibroblast cells from one MDS patient [19]. Antisense oligonucleotide (ASO) therapy has proven to downregulate MeCP2 expression and restore the expression levels of other dysregulated genes without inducing toxicity in mice carrying two human *MECP2* alleles [20]. They have also detected that each gene needed a different amount of time (days and even weeks) to respond. Multi-omics could help to decipher the dynamics of all of the implicated genes and aid with dosing timings. In addition, obtaining the molecular signatures of the patient groups could reveal biomarkers that can be used to measure treatment response or even as a prognosis tool.

Here, we present the results of multi-omics (transcriptomics and proteomics) experiments performed in skin fibroblasts of 17 MDS patients and 10 carrier mothers and compared them to the results of 21 RTT patients. Delineating the molecular signature of the different cohorts and knowing which pathways are altered is crucial not only to gain insights into the pathomechanism of the syndrome, but also to find biomarkers and therapeutic targets that could be used in upcoming clinical trials.

2. Material and Methods

2.1. Clinical and molecular characterization

Seventeen patients with MDS (15 males and 2 females), 10 *MECP2* duplication carrier mothers, 21 patients with RTT and 13 age-matched healthy controls (6 males and 7 females) were enrolled in this study; written informed consent was obtained from them all [Table 1; SI 1].

Table1: Characterization of the MDS families with fibroblast cell lines. The carrier mothers' duplications coincide with their offsprings' duplications. M10 was included in the carrier's study but her newborn daughter (P18) was excluded from the female MDS cohort. XCI was studied in fibroblast DNA or in whole blood when fibroblasts were not available. More detailed information can be found in the SI 1. n.a. stands for "not available". n.i. stands for "not informative".

Patient ID	Gender	Duplication origin	Duplication location	Duplication size (Mb)	Genomic coordinates (GRCh37/hg19)	Patient's XCI	Carrier mother's ID	Carrier mother's XCI	Family relations
P1	M	maternal	tandem	0.465	Xq28(153101077-153565901)x2	-	M1	99:1	-
P2	M	maternal	tandem	0.637	Xq28(152957295-153594098)x2; 1q25.3(183996990-184265525)x3	-	M2	95:5	-
P3	M	maternal	tandem	0.74	Xq28(152832700-153576940)x2	-	M3	ni	-
P4	M	maternal	tandem	0.416	Xq28(153194797-153611490)x2	-	M4	76:24	Cousins
P5	M	maternal	tandem	0.416	Xq28(153194797-153611490)x2	-	M5	70:30	
P6	M	maternal	tandem	0.314	Xq28(153287517-153601836)x2	-	M6	100:0	-
P7	M	maternal	tandem	0.75	Xq28(153120541-153870079)x2	-	M7	87:13	-
P8	M	maternal	chr Xp	14.3	Xq27.2q28(140928466-155232885)x2	-	M8	95:5	-
P9	M	maternal	tandem	0.511	Xq28(152962751-153473892)x2	-	M9	ni	Brothers
P10	M	maternal	tandem	0.524	Xq28(152949788-153473892)x2	-	M9	ni	
P11	M	maternal	tandem	0.221	Xq28(153194797-153406233)x2	-	-	99:1(blood)	-
P12	M	*de novo*	tandem	0.764	Xq28(153028550-153792888)x2	-	-	-	-
P13	M	maternal	tandem	0.386	Xq28(153155029-153541192)x2	-	-	n.a.	-
P14	M	*de novo*	chr 18	2.7	Xq28(152111224-154841455)x2; 15q13.2(30783615-31089985)x3; 18p11.32(159550-496915)x1	-	-	-	-
P15	M	*de novo*	chr Y	5.8	Xq28(149116213-154929279)x2; Yq11.2 22q11.23(20826207-28629893)x0	-	-	-	-
P16	F	*de novo*	tandem	0.346	Xq28(153194797-153541289)x3	51:49	-	-	-
P17	F	*de novo*	n.a.	0.638	Xq28(153023556-153661653)x3	91:9	-	-	-
P18	F	maternal	tandem	7.94	Xq27.3q28(146407370-154351599)x3	-	M10	94:6	

The study was approved by the Sant Joan de Déu Hospital's ethical committee, CEIC: Comitè d'Ètica d'Investigació Clínica-Fundació Sant Joan de Déu (internal code: PIC-219-20). Twelve patients with MDS were molecularly characterised in Pascual-Alonso et al. [21]. Six new families were recruited and studied following the same approach. Carrier mothers were also molecularly characterised in order to confirm the location and size of the duplication. The duplication in carrier mother number 10 of our cohort was detected in a prenatal test. We considered it to be too early to include her newborn daughter in our cohort since we do not know whether she will develop MDS clinical traits or not. The 21 RTT patients have a mutation in *MECP2* and present the necessary criteria for RTT diagnosis [10]. The clinical severity of RTT patients was evaluated according to Dr Pineda's clinical severity score [22].

Skin biopsies were obtained from the 61 individuals and primary fibroblast cell lines were established. Fibroblast cell lines were grown as explained elsewhere [23]. When a confluence of 70–80% was reached, cells were trypsinised and either re-sown in new plates or harvested for subsequent RNA or protein extraction. All fibroblast lines were frozen and entrusted to the Biobanc "Hospital Infantil Sant Joan de Déu per a la Investigació", which is integrated in the Spanish Biobank Network of ISCIII for sample and data procurement.

DNA was extracted from fibroblast cell lines using DNeasy Blood & Tissue kit (Qiagen, Germany) following the manufacturer's instructions. XCI was studied in all female samples, as described elsewhere [23]. XCI was considered skewed with an allele ratio of 80:20 or greater. RNA was extracted from the fibroblast cell lines using the RNeasy Fibrous Tissue Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. Proteins were extracted at the BayBioMS core facility at the Technical University of Munich (TUM) in Germany, as described elsewhere [24].

2.2. Transcriptomics

To ensure the quality of the RNAs and discard any undesirable effects due to cell stress, we performed RT-qPCR of five genes of the oxidative respiratory chain (*MT-CO1, MT-CO2, MT-CYB, MT-ND4* and *MT-ATP6)* and two housekeeping genes (*RPLP0* and *ALAS1*), as shown by Pascual-Alonso et al. [23]. All samples that overexpressed two or more mitochondrial genes by more than 1.5-fold of the values of non-stressed controls were discarded.

Illumina's TruSeq Stranded mRNA kit was used for RNAseq library preparation (Illumina, California, USA) and the manufacturer's protocol was followed. A 4200 TapeStation (Agilent Technologies, California, USA) was used to quantify and check library integrity. Sequencing was performed on an Illumina NextSeq 500 sequencer. Paired-end reads of 75 bp were obtained and around 40 million paired reads per sample successfully mapped to the reference genome. Two healthy controls of the same sex as the patients were at least included in all the runs to enable normalisation and control the batch effect.

For RNAseq validation, 22 significantly differentially expressed genes (DEGs) were selected and RT-qPCR was performed with 23 different samples, as explained in Pascual-Alonso et al. [23].

As described by Pascual-Alonso et al. reads were aligned with STAR (v2.4.2a) to the human reference genome (GRCh37/hg19) in a strand-specific manner [23]. Uniquely mapped reads were counted for each gene using the HTSeq package (v2.0.2) with gene models from GENCODE release 29. For the final count matrix, the raw counts from the replicates of the same individual were averaged. Counts per million mapped reads (CPM) were computed and only genes where more than 50% of samples had at least 1 CPM were kept. The first three principal components were introduced in the model construction for DE analysis with DESeq2 (v1.34.0). A cut-off value of 0.05 in Benjamini-Hochberg (BH) corrected p-value was established for considering differences significant.

2.3. Proteomics

All proteomics experiments were performed at the BayBioMS core facility at the TUM in Germany, as described by Kopajtich et al. [24]. The same reference sample was included in all TMT batches in order to enable normalisation. LC-MS measurements were run on a Fusion Lumos Tribrid mass spectrometer (Thermo Fisher, Massachusetts, USA) operated in the data-dependent acquisition mode and multi-notch MS3 mode. MaxQuant version 1.6.3.4 was used for peptide identification. Missing values were imputed with the minimal value across the dataset.

Mass-spectrometry (MS) data was adjusted with respect to one common control sample that was present in each MS batch [24]. Intensities were recalibrated and log transformed for normalization. Proteins undetected in at least one sample were removed. DE analysis was performed using the limma (v3.50.3)

package in R and we included the MS batch as a covariate in the model. We set a nominal p-value of 0.05 as a threshold to define differentially expressed proteins (DEPs).

2.4. Enrichment analysis and upstream regulator analysis (ChEA3)

Enrichment analysis was performed with overrepresentation analysis (ORA) and gene set enrichment analysis (GSEA) approaches, using only significant DEGs or DEPs and all expressed genes, respectively. As explained elsewhere, clusterProfiler (v4.2.2) and ReactomePA (v1.38.0) R packages were used for the enrichment analysis [23]. Enriched terms were searched in Gene Ontology (GO), Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway database, WikiPathways (WP) and Reactome pathway database (RP). For transcriptomics results, all genes with CPM greater than 1 in at least 50% of samples and with an existing EntrezID were used as background (a total of 11904 genes). For proteomics results, proteins that were detected in all samples and with a valid EntrezID (a total of 5894 proteins) were used. The cut-off value for considering a significantly enriched term was 0.05 in BH-corrected p-value.

To study whether upstream transcription factors (TFs) were responsible for some of the differentially expressed changes observed in our data, the ChIP-X Enrichment Analysis 3 (ChEA3) tool was used. The lists of DEGs and DEPs resulting from DE analysis were given to ChEA3 to predict the possible involvement of TFs in their dysregulation.

3. Results

3.1. Male MDS patients vs controls

Biased results due to a non-homogeneous distribution of females and males in the cohort were avoided by performing separately differential expression (DE) analysis for MDS males and females. DE analysis of male MDS patients against healthy controls showed 2465 differentially expressed genes (DEGs)

and 300 differentially expressed proteins (DEPs) [Figure 1A, B; SI 2a,b]. From those DEGs and DEPs, 103 genes are significant in both transcriptomics and proteomics analysis [Figure 1C; SI 2c].

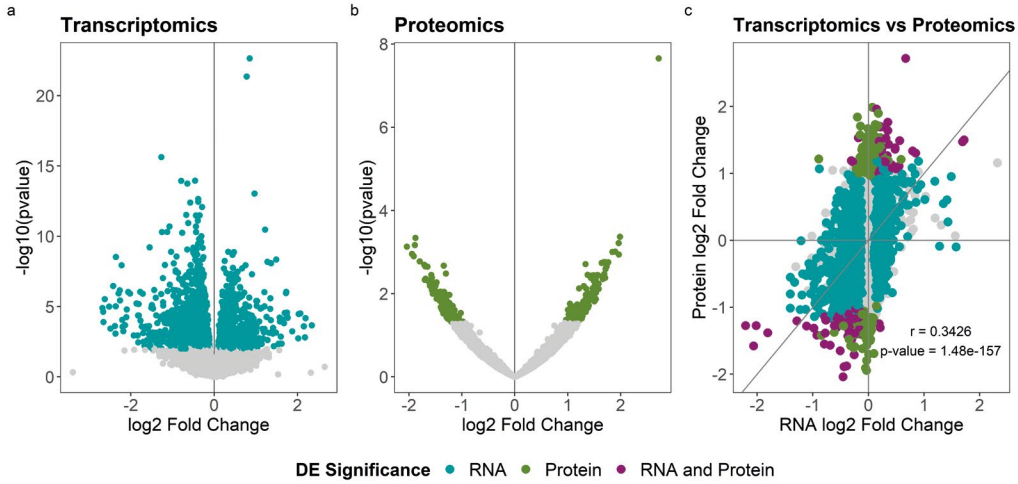

Figure 1: Summary of the results of the male MDS patients versus male healthy controls analysis. a) RNAseq DE analysis results. The coloured genes are considered differentially expressed, passing a threshold of BH < 0.05. b) Proteomics DE analysis results. The coloured proteins are considered differentially expressed, passing a threshold of nominal p.value < 0.05. c) An integrated view of the transcriptomics and proteomics results. The genes that are significant at both analyses are coloured in purple.

Enrichment analysis of transcriptomics and proteomics results revealed several significantly dysregulated biological processes such as the cytoskeleton, neuronal system, vesicular activity, immune system and Wnt and NF-kB signalling cascades [Figure 2A, Figure 3; SI 3a,b,c]. Upstream TF analysis revealed three relevant TFs: *CREB1, NFATC4* and *NR1H2*. In our cohort, more than a third of the DEGs and DEPs (38% and 37%) are regulated by CREB1, around 9% are regulated by *NFATC4,* and 17% by *NR1H2* [SI 4a,b]. Also, *NR1H2* mRNA is significantly downregulated in our MDS male cohort.

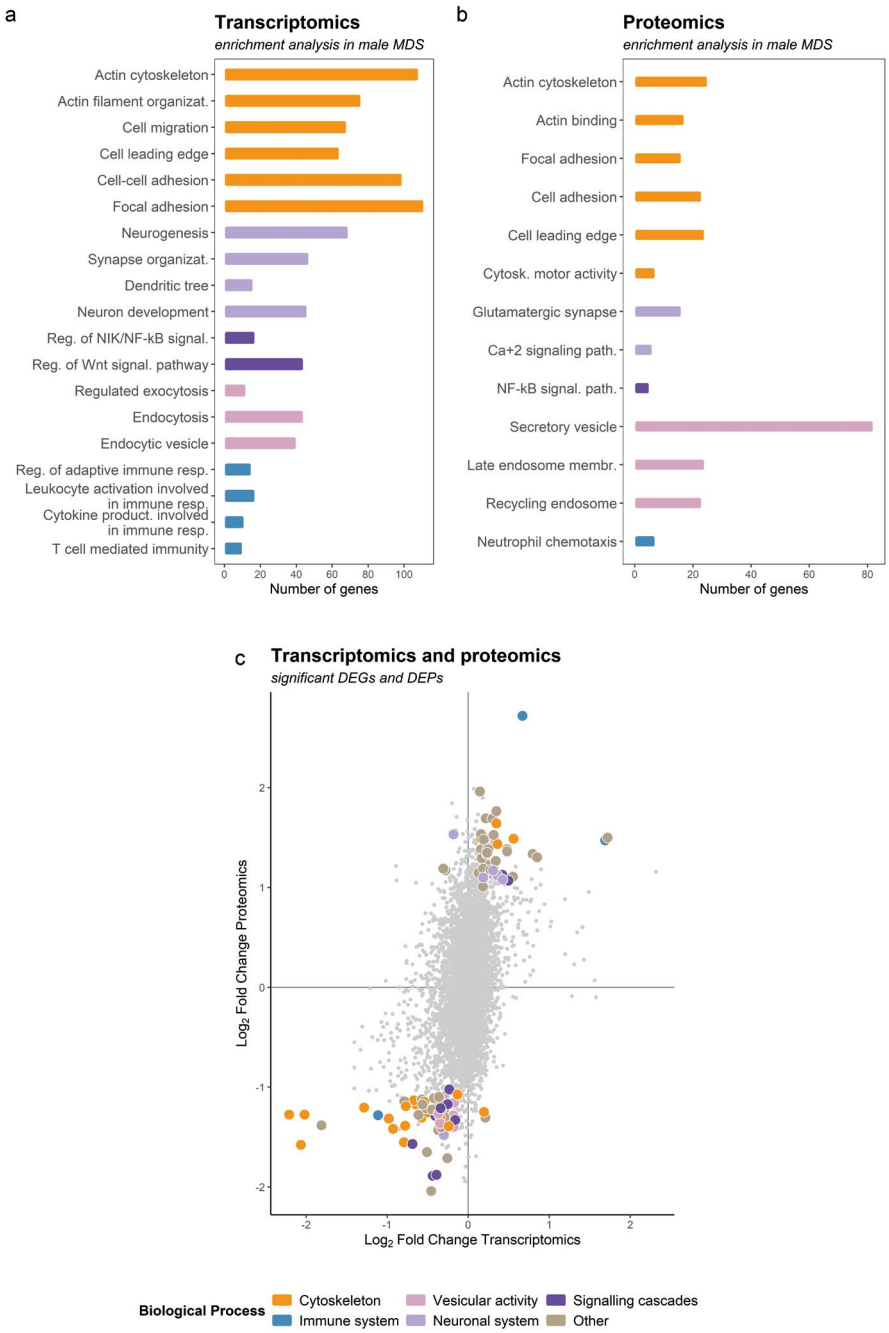

Figure 2: Summary of the common findings of the analysis of male MDS patients versus male healthy controls. a) Enrichment analysis results for the transcriptomics experiment, coloured by Biological Process (BP). b) Enrichment analysis results for the proteomics experiments, coloured by BP. c) Common significant genes for transcriptomics and proteomics DE analysis results. The common 103 DEGs and DEPs are coloured by BP.

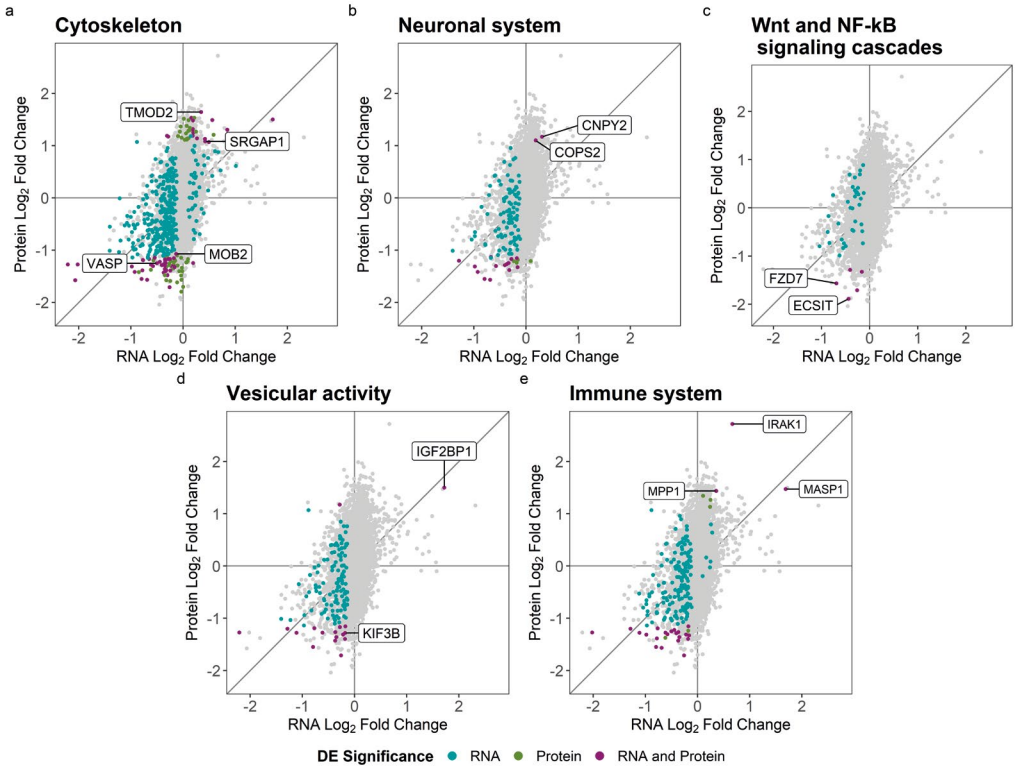

Figure 3: Summary of the main biological processes found enriched in the analysis of male MDS patients versus male healthy control analysis. a) Significant DEGs and DEPs related to cytoskeleton (coloured in the dotplot). b) Significant DEGs and DEPs related to the neuronal system. c) Significant DEGs and DEPs related to Wnt and NF-kB signalling cascades. d) Significant DEGs and DEPs related to vesicular activity. e) Significant DEGs and DEPs related to the immune system.

DE analysis was also performed, stratifying the patients according to their duplications' size and location to see whether a correlation with patient severity or prognosis could be found [SI 2d]. No specific dysregulated molecular pathways were found in any group.

3.2. Female MDS patients vs controls

We analysed the two girls with MDS that we characterised. One of the girls has a skewed XCI while the other is random. DE analysis of the female MDS patients against female controls revealed 5720 DEGs and 493 DEPs [SI 2e,f]. Two hundred and forty-nine genes are significantly differentially expressed at the mRNA and protein levels [SI 2g]. An enrichment analysis with the 249 common genes showed that 26% are related mainly to translation [SI 3f]. Translation related DEGs are dysregulated in both directions but DEPs are predominantly upregulated. In the upstream analysis with ChEA3, the TFs *PA2G4* and *PRMT3* were found, both of which are related to translation machinery [SI 4c,d]. ChEA3 analysis also shows *CREB1* in the female cohort, which regulates 34% of the DEGs and 46% of the DEPs found in female MDS patients. In addition, splicing-related terms are enriched with downregulated DEGs but upregulated DEPs.

A comparison between the MDS male and female cohorts showed that there are only three genes significantly dysregulated in both cohorts and in both omics that are not present in the carrier cohort: *ABCC4*, *STK17B* and *MYO1C*. The enrichment analysis showed some shared biological pathways between MDS male and female cohorts, such as cellular adhesion, vesicular activity and synapses, all of which are downregulated [Figure 4A,B,E,F,I,J; SI 3a-f].

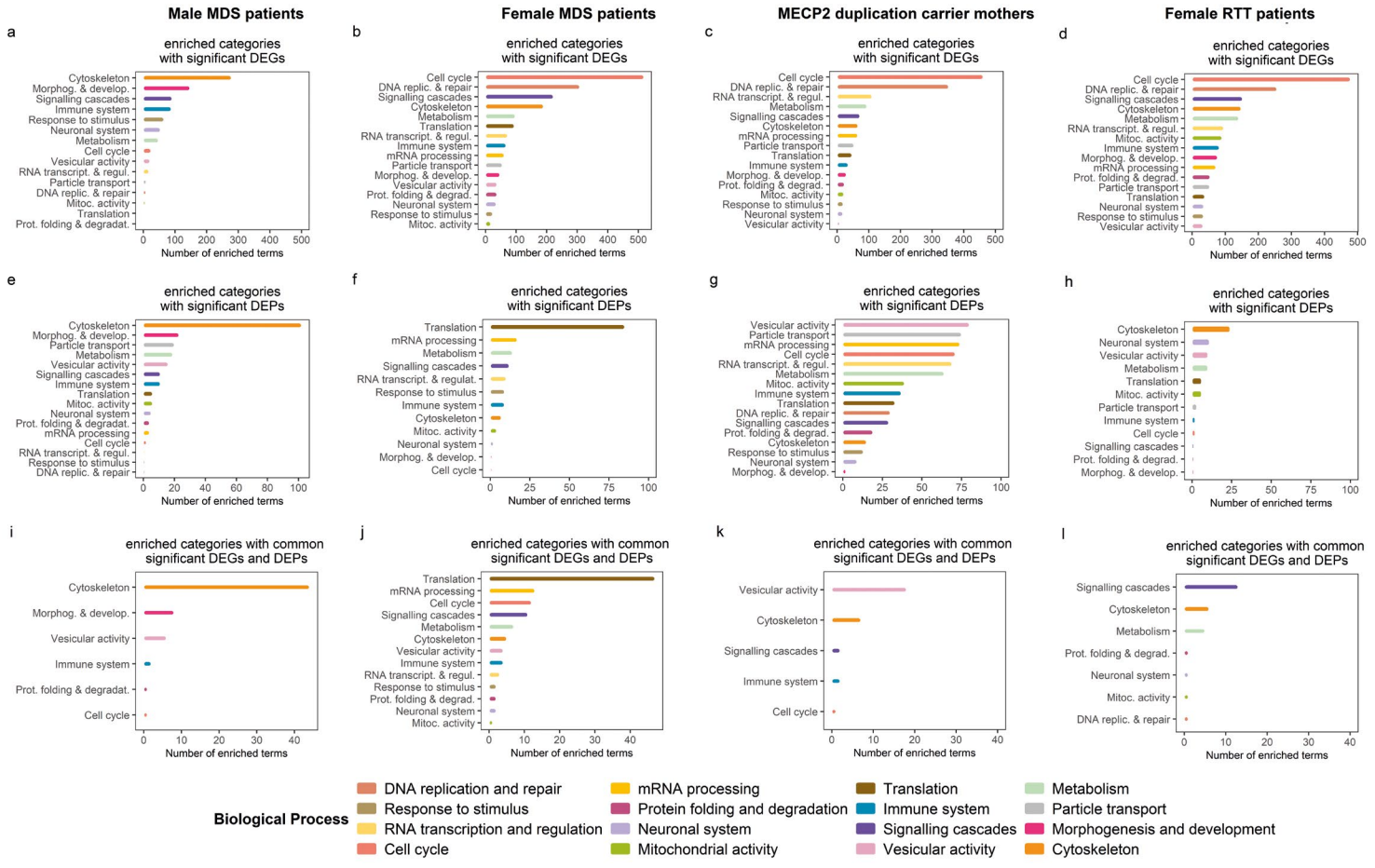

Figure 4: Summary of the main biological processes (BP) found enriched in the analysis of male and female MDS patients, *MECP2* duplication carrier mothers and female RTT patients. Upper row: a, b, c, d) Enriched BP found when DEGs from RNAseq were analysed. Middle row: e, f, g, h) Enriched BP found when DEPs from proteomics were analysed. Lower row: i, j, k, l) Enriched BP found when the common DEGs and DEPs were analysed.

3.3. *MECP2* duplication carrier mothers vs controls

MECP2 duplication carrier mothers underwent the XCI assay and 6/10 had a skewed XCI [SI 1]. The XCI of two carrier mothers was close to skewed (70:30 or above) while the remaining two were not informative. The DE analysis of the 10 carrier mothers against female healthy controls showed 2888 DEGs and 635 DEPs [SI 2h,i]. In total, 177 genes are dysregulated at the mRNA and protein levels [SI 2j] and enrichment analysis of those common genes showed few terms [SI 3i]. Upstream analysis with ChEA3 again showed *CREB1* to be an important TF that regulates 37% and 42% of the DEGs and DEPs, respectively [SI 4e,f].

When considering the DEGs and DEPs that are not shared between MDS males and carrier mothers, terms related to the cytoskeleton were still present for males, while ribosome binding and phagocytic activity-related terms appeared for the carrier mothers. General enrichment analysis of carrier mothers has shown multiple dysregulated genes related to the cell cycle, splicing, nuclear transport, protein folding and degradation and phagocytosis at the mRNA or protein levels, with most of them being downregulated [SI 3g,h].

Our female MDS patients and carrier mothers share a high percentage of DEGs, but we can see no common enriched terms when considering the genes that are differentially expressed in both omics and are shared between the two cohorts [Figure 4B,C,F,G,J,K; SI 3d-i].

3.4. Male MDS vs female RTT patients

Since *MECP2* is altered in both RTT and MDS, we compared the DE analysis results of the male MDS cohort and a female RTT cohort [Figure 4A,D,E,H; SI 2k-m, SI 3k-m]. At the transcriptomics level, 721 significant DEGs were shared between the two analyses [SI 2n]. At the proteomics level, only 12 DEPs were common [SI 2o] and only two genes were significantly altered in both of the omics and in both syndromes. We specifically queried the DEGs that are dysregulated in the same or opposite directions [SI 3j]. The 82 DEGs that are downregulated in RTT and upregulated in MDS (which corresponds to MeCP2 expression levels in these patients) are significantly enriched with processes involved in mRNA processing, mRNA splicing and DNA replication. ChEA3 revealed several TFs that regulate the cell cycle and DNA replication [SI 4g]. As in the other three cohorts, ChEA3 also returned the TF *CREB1*.

The 100 DEGs upregulated in RTT and downregulated in MDS are related to signalling cascades such as Wnt, BMP and TGFß. Enrichment analysis also revealed terms related to cell adhesion, cell projection organisation and cell motility. ChEA3 analysis showed TFs involved in neuronal function: *HEYL, GLIS2, NFATC4* and *JUN* [23] [SI 4h]. Enrichment analysis for the commonly upregulated DEGs both syndromes showed no enriched terms, but the analysis for the downregulated DEGs showed cytoskeleton and some synapse-related terms [SI 3j].

4. Discussion

4.1.1. Male MDS patients: altered pathways

Cytoskeletal functions

The most recurrently enriched terms in male MDS patients are related to actin cytoskeletal functions, specifically to processes involved in cell migration, cell adhesion and filament organisation. From these 103 significant DEGs and DEPs, almost a third are related to the cytoskeleton and most of them are downregulated [Figure 3A]. The function of some of those genes occurs at the neuronal level [Table 2]; for example, *TMOD2* (OMIM*602928) regulates dendritic arborisation and *SRGAP1* (OMIM*606523) regulates neuronal differentiation; both of these are upregulated. *MOB2* (OMIM*611969) regulates neurite formation and *VASP* (OMIM*601703) regulates dendritic spine morphology, axon guidance and neuronal migration, and both of these are downregulated. It has been thoroughly described that cytoskeletal alterations lead to neuronal malfunction in cognitive and neurodevelopmental disorders such as RTT, autism or Fragile X [25–27]. The dysregulation of genes and proteins related to actin dynamics and cell adhesion can alter the dendritic spine morphology and synapse plasticity needed for proper synapse function [25,26,28].

Neuronal system

Even though we have performed our studies in fibroblast tissue, we also identified some terms related directly to the neuronal system in the enrichment analysis, which are mainly downregulated [Figure 3B]. Two of the genes that fall into that category are *COPS2* (OMIM*604508), which participates in neuronal

differentiation in the early stages and *CNPY2* (OMIM*605861), which regulates neurite outgrowth; both are upregulated. In MDS mouse models, the dendritic spine density is initially higher but falls below normal levels around the same age as the neurobehavioral traits emerge [29]. The genes that we have found to be significantly differentially expressed could be interesting candidates for direct study in neuronal tissues of MDS patients.

In addition, three upstream TFs (*CREB1, NFATC4* and *NR1H2)* are involved in neuronal activity. *CREB1* (OMIM*123810) is a transcriptional activator that participates in neurogenesis, neuronal plasticity and survival [30]. Interestingly, it has been reported that MeCP2 promotes gene transcription via an interaction with CREB1 [6]. *NFATC4* (OMIM*602699) is a TF that regulates adult hippocampal neurogenesis and, together with *BDNF,* mediates synaptic plasticity [31]. *BDNF* (OMIM*113505) is regulated by *MECP2* and is needed by GABAergic interneurons for their maturation. Quadrato et al. reported that GABAaR stimulation via *NFATC4* activation decreases innate anxiety in mice [31]. Since around half of MDS patients present anxiety, this could be an interesting pathway to explore [2]. *NR1H2* (OMIM*600380) is a TF that is involved in the layering of brain structures [32]. The *NR1H2* agonist GW3965 rescues synaptic formation and function in hippocampus neurons [32]. *NR1H2* also regulates cholesterol trafficking and the same GW3965 agonist rescued impaired cholesterol levels and axonal defects in cortical projection neurons [33]. A correct neurotransmission depends on cholesterol [34]; therefore, trying to restore NR1H2 neuronal levels could be beneficial for MDS patients in multiple ways.

Wnt and NF-κB signalling cascades

Wnt and NF-κB signalling cascades have also been shown to be enriched with downregulated DEGs and DEPs [Figure 3C]. The Wnt signalling cascade takes part in neuronal development, axon and dendrite branching and synapse formation [35,36], whereas the NF-κB pathway regulates neurogenesis, dendritic complexity, axon guidance and peripheral nerve myelination, among others [37]. NF-kB has been reported to be upregulated in Mecp2-null mice and in Irak1-overexpressing neurons [38]. Also, partial silencing of the aberrant NF-kB signalling improves the dendritic phenotype of the RTT mouse model and expands its lifespan. *FZD7* (OMIM*603410), a Wnt protein receptor required for spine plasticity and the migration of glutamatergic neurons, as well as *ECSIT* (OMIM*608388), which regulates the NF-κB pathway, are significantly downregulated at the mRNA and protein levels in our cohort. These signalling cascades might be impaired and could compromise proper synapse function and axon formation.

Vesicular transport

Terms related to vesicular transport are also enriched with downregulated DEGs and DEPs [Figure 3D]. These genes transport different kinds of cargos. *KIF3B* (OMIM*603754) is downregulated and transports vesicles containing the NMDAR subunit 2A (NR2A), which is required for adequate synaptic plasticity needed for learning and memory [39]. In addition, KIF3B inhibition results in an increased spine density and *Kif3b* haploinsufficient models have an upregulated long-term potentiation (LTP) [39,40]. Both features have been reported in MDS mouse models [29,41]. *IGF2BP1* (OMIM*608288) is upregulated, while *IGF2BP1* is involved in growth cone migration, dendritic branching and synapse formation in neurons, and also transports transcripts required for axonal regeneration on adult sensory neurons [42]. Altogether, our data emphasise a disruption in vesicle formation, cargo uptake and secretion. Specifically, *KIF3B* could explain two of the traits that have been found in MDS mouse model brains.

Immune system

Several processes related to the immune system are also significantly enriched with mainly downregulated DEGs and DEPs [Figure 3E]. MDS patients suffer from recurrent infections and elements of their immune system have been reported to be impaired [43]. Recurrent infections, together with epilepsy, are some of the major factors that trigger regression and cause death in MDS patients. Among our consistently dysregulated genes and proteins, *IRAK1, MASP1* and *MPP1* are upregulated; the first two are involved in innate immunity function and *MPP1* in neutrophil polarity. It has been hypothesised that the overexpression of *IRAK1* could be key to understanding the malfunction of these children's immune systems. Recently, Gottschalk et al. evaluated four immortalised fibroblasts and six PBMCs, alongside whole blood samples of MDS patients, showing no significant difference in IL-6 and IL-8 production [44].The NF-kB pathway by TLR and IL-1R signalling should be further studied in several tissues of MDS patients to determine whether it is implicated at the neuronal and immunological level in MDS.

Table 2: Interesting significant differentially expressed genes and proteins from the analysis of male MDS patients against healthy male controls.

Gene	Expression	Biological Process
TMOD	Upregulated	Cytoskeleton
SRGAP1	Upregulated	Cytoskeleton
MOB2	Downregulated	Cytoskeleton
VASP	Downregulated	Cytoskeleton
COPS2	Upregulated	Neuronal system
CNPY2	Upregulated	Neuronal system
FZD7	Downregulated	Signalling cascade
ECSIT	Downregulated	Signalling cascade
KIF3B	Downregulated	Vesicular transport
IGF2BP1	Upregulated	Vesicular transport
IRAK1	Upregulated	Immune system
MASP1	Upregulated	Immune system
MPP1	Upregulated	Immune system

4.1.2. Male MDS patients: grouped by duplication size and location

It is thought that part of the phenotypic variability of MDS patients can be explained by the size, location and content of the duplication, characteristics that are exclusive to each family. No clear genotype-phenotype correlation has been found to date, although it has been reported that larger duplications and triplications tend to lead to a more severe phenotype [45]. In our MDS cohort, we have no triplications and all of the duplications in tandem have a size smaller than 1 Mb, whereas the duplications outside the Xq28 region are larger than 1 Mb. Besides, 11/13 tandem duplications are inherited from a carrier mother, while 2/3 duplications located elsewhere are *de novo*. No significant enriched terms or dysregulated molecular pathways were found when analysing MDS patients grouped by duplication size and location. Definitely, larger cohorts, especially those with large and translocated duplications, should be considered for this approach.

4.2.1. Female MDS patients: altered pathways

Translation

Around a quarter of the shared DEGs and DEPs in female MDS patients are involved in translation. Also, two upstream TFs from the ChEA3 analysis participate in the formation of translation machinery. *PA2G4* (OMIM*602145), which is significantly downregulated at the mRNA level, is involved in ribosome assembly and participates in rRNA processing, while *PRMT3* (OMIM*603190) participates in the maturation of the 80S ribosome and its coding protein is upregulated in our cohort. Buist et al. recently reported an impairment in protein translation and mTOR signalling in RTT human brains but not in a cell model overexpressing *MECP2* isoforms [17]. Those findings are consistent with our data, since no relevant impairment has been detected in our male MDS patients, but females, whose phenotype is milder, present dysregulation of the genes related to translation, which are mainly upregulated. All of the DEGs, DEPs and TFs related to translation, together with those found in the enriched terms involved in proteasome regulation and nonsense mediated decay, could reflect cellular compensation as a consequence of the duplication, explaining why females have a milder phenotype and a different molecular signature than males with MDS. To further confirm the status of protein translation machinery in MDS, other MDS models and tissues should also be tested.

mRNA processing

Splicing-related terms have downregulated DEGs but upregulated DEPs. The tight regulation of alternative splicing is crucial for neurodevelopment and also for synaptic plasticity [46]. *MECP2* interacts with splicing factors and regulates splicing [47]. However, it has recently been stated that *MECP2* can only regulate the alternative splicing of specific genes rather than doing it in a global way [48]. Splicing dysregulations have been found in RTT and in other monogenic intellectual disabilities [46], so the affected transcripts might be implicated in similar biological functions, explaining why those splicing defects are recurrently detected in disorders with overlapping traits.

4.2.2. Comparison between female and male MDS patients

Only three genes were differentially expressed in male and female MDS patients in both omics: *ABCC4*, *STK17B* and *MYO1C*. *ABCC4* (OMIM*605250)

is needed for dendritic cell migration, which are key initiators of the immune response [49]. *ABCC4* is downregulated in MDS males but upregulated in females. Those results are consistent with the immune system dysfunction reported in males but not in females with MDS. *STK17B* (OMIM*604727) is a downstream effector of protein kinase C (PKC) in the immune system. PKC is also involved in dendrite development and synapse plasticity in Purkinje cells. Reduced levels of STK17B in mouse models have been shown to protect Purkinje cell dendrites from the negative impact of the activation of PKC signalling. STK17B is upregulated in MDS males but downregulated in females. Finally, *MYO1C* (OMIM*606538), which is downregulated in male and female MDS cohorts, stabilises actin in the Golgi complex and regulates cargo transport. However, when depleted, *MYOC1* induces fragmentation of the Golgi complex, the loss of cellular F-actin and a delay in transport [50]. *MYO1C* is also involved in the recycling of glucose transporters in response to insulin [51]. In addition, the ChEA3 study returned *CREB1* in both cohorts.

Also, MDS male and female cohorts share some biological pathways, such as cellular adhesion, vesicular activity and synapses; all of these are downregulated. These commonly altered biological processes can contribute to the dysregulation in neuronal architecture and synapse function that occurs in MDS patients, even if it is to a different degree. Therefore, a larger MDS female cohort should be analysed. On the other hand, no enriched terms related to immune system deficiencies were found in females. These data reflect that recurrent infections are a problem in MDS male patients but not in females.

4.3.1. *MECP2* duplication carrier mothers: altered pathways

The enrichment of the 177 commonly dysregulated DEGs and DEPs shows few terms. The terms have been classified as cytoskeleton and vesicular organisation-related, but they are annotated based on the Cellular Component GO domain; thus, no specific information about any altered biological processes is observed. These results show that even though multiple DEGs and DEPs exist at the mRNA and protein levels, no evident dysregulation is detected in the carriers from a multi-omics perspective. This could explain the normal and healthy phenotype that these carriers present.

4.3.2. Comparison between *MECP2* duplication carrier mothers and male MDS patients

Enrichment analysis in *MECP2* duplication carrier mothers reveals differentially expressed genes related to cell cycle, splicing, nuclear transport, protein folding, protein degradation and phagocytosis at the mRNA or protein levels, which are mainly downregulated and missing in male MDS patients. Those genes that are not present in males with MDS could be implicated in cellular compensation against the effects of the duplication.

4.3.3. Comparison between *MECP2* duplication carrier mothers and female MDS patients

Despite the high amount of initially shared DEGs, no common enriched terms have been found between female cohorts at both omics levels. Those results suggest that *MECP2* duplication generates gene and protein dysregulation, so that some female cells are able to compensate at the transcriptomics and proteomics levels, living healthily as a carrier. As mentioned above, the DEGs and DEPs that are related to RNA transcription and regulation, translation, protein folding and degradation in the carrier cohort could be crucial for this purpose. A skewed XCI may contribute to a lower number of cells with the duplication and therefore, to the asymptomatic phenotype of the carriers. However, the DEGs and DEPs found in our carriers suggest that XCI is not sufficient to ensure correct molecular regulation. It is unclear how the cells trigger the skewed XCI and any other molecular compensatory mechanism in some girls but not in all. DE analysis of a larger cohort of females with the *MECP2* duplication classified by their phenotypes would be crucial to characterise their molecular signatures and answer these new questions.

4.4. Comparison between male MDS patients and RTT patients

The opposing expression levels of *MECP2* between RTT and MDS made it interesting to compare the DE results obtained from both omics studies. However, only two genes, *MYO1C* and *HARS2,* were commonly dysregulated in both syndromes and omics. *MYO1C* (OMIM*606538) is a myosin involved in cytoskeletal organisation and vesicle trafficking, being consistently downregulated. *HARS2* (OMIM*600783) is a mitochondrial histidyl-tRNA

synthetase 2. At the RNA level, it is upregulated in patients with MDS and downregulated in patients with RTT, whereas it is upregulated in both sets of patients at the protein level. Having only two common genes demonstrates that these patients are not as comparable as thought at the molecular level, at least in fibroblast tissue. The downstream effect of *MECP2* seems to be quite different depending on its dose. It should be noted that some shared DEGs have been detected at the mRNA level which highlight biological processes that could be involved in the pathomechanism causing at least some of the features of RTT and MDS.

We have detected 100 DEGs that are upregulated in RTT and downregulated in MDS mRNA, which present enriched terms related to the Wnt, BMP and TGFß signalling cascades. These signalling pathways participate in neurogenesis regulation, myelin synthesis and synapse formation [23,35,36,52], functions that, when altered, could explain the malfunction of the neuronal tissues in MDS and RTT. Wnt, BMP and TGFß are also implicated in osteoblast regulation and the maintenance of cartilage [53–55], and scoliosis and bone fractures have been reported in these syndromes [2,23,56]. The dysregulation found in the Wnt, BMP and TGFß signalling pathways could be related to the problems found in the skeletal system of RTT and MDS patients since childhood. Enrichment analysis in those 100 DEGs also revealed terms related to cell adhesion, cell projection organisation and cell motility. An altered cytoskeleton seems to be implicated in the deficient synaptic activity of MDS and RTT patients.

Zoghbi's group reported that the majority of the dysregulated genes found in male *MECP2*-Tg and *Mecp2*-null mice hypothalami and cerebella were shared between both models and that most of them were upregulated [6,13]. In our cohort, however, less than 30% of the DEGs are common between MDS and RTT patients and the gene dysregulation occurs evenly in both directions. The integrative transcriptomics analysis of 43 mouse studies performed by Trostle et al. also showed an even distribution of the dysregulated genes [57]. The limited correlation of the findings between both species and tissues makes it necessary to confirm the results in patient-derived samples.

Getting upstream TFs related to neuronal processes regulating the common DEGs between MDS and RTT patients emphasises the resemblance of both syndromes at the transcriptomics level. A commonly altered molecular signature could lead to a shared therapeutic strategy benefiting both syndromes. Unfortunately, the 100 DEGs are not significant in the female MDS cohort, even

though a larger cohort of affected girls will determine how alike all of these children are.

The commonly upregulated DEGs in both MDS and RTT showed no enriched terms, indicating that all of those DEGs are involved in different biological functions. Enrichment analysis for the consistently downregulated DEGs showed cytoskeleton- and synapse-related terms. The common DEGs that do not follow the expected behaviour for a gene regulated by MeCP2 in these two syndromes could be secondary or indirect effects that cannot be traced with the RNAseq technique.

At the proteome level, only 12 proteins are shared, as described in Pascual-Alonso et al. [23]. Overall, a limited correlation between transcriptomics and proteomics experiments, in part due to the producing and degrading rates of the molecules and the different analysis tools and resolution of the techniques themself, complicates the generation of more comparable results between these two omics [58]. Also, it becomes much more challenging to compare the outputs from two different cohorts. Improvements in the omics technologies will solve these complications in the next few years.

5. Conclusions

Our male MDS cohort shows the significant upregulation of the genes *TMOD2, SRGAP1, COPS2*, *CNPY2* and *IGF2BP1* and the significant downregulation of *MOB2, VASP, FZD7, ECSIT* and *KIF3B* at the mRNA and protein levels. They have been published to be involved in neurite formation, dendritic arborisation, synaptic plasticity and neuronal differentiation and migration, thus being implicated in the neuronal dysfunctions reported in MDS. Those genes are expressed in brain tissue and are not associated with any other disorder presenting ID or neurodevelopmental delay, making them candidates for therapeutic targets and diagnostic biomarkers of MDS. In particular, *KIF3B* seems to be a promising candidate as its inhibition leads to increased spine density and upregulated LTP, two features that have been seen in the brains of MDS mouse models.

We have demonstrated that the transcriptomics and proteomics profiles of males and females with MDS, carrier mothers and RTT patients are specific for each cohort. Females with MDS present milder forms of the syndrome, most

frequently being referred for diagnosis due to mild to moderate intellectual disabilities or familiar male MDS cases. The little overlap of DEGs and DEPs found between male and female MDS cohorts also corroborates the phenotypic disparity related to sex in this syndrome. Also, carrier mothers and MDS females present a specific molecular signature, which is similar at the mRNA level, but differs at the protein level. We hypothesised that the dysregulated DEGs and DEPs related to transcription regulation, mRNA processing, translation activity and protein degradation could be compensating for the effects of the duplication in carrier mothers but not fully in MDS females. Nevertheless, more female MDS patients should be added to the DE analysis since two samples are far away from an ideal sample size. Obtaining optimal sample sizes is a particular challenge when working with rare diseases as reaching desirable sample numbers is not always feasible. Performing the integrative analysis of several already published data sets will become a solution to the current sample size problem in rare diseases. The limited amount of overlap that we have found between the two omics experiments coincides with the results of other groups [58,59].

Definitively, multi-omics has proven to be a useful technique to gain insight into the altered genes and molecular processes of a syndrome, as well as to find biomarkers and therapeutic targets.

Acknowledgements

We thank all patients and their families for their collaboration and support in this study. We would like to thank the Cytogenetics Department of the Hospital Sant Joan de Déu for starting some of the fibroblast primary cell cultures. We would like to thank the Proteomics Core Facility at TUM for performing the proteomics experiments. We thank the 'Biobanc de l'Hospital Infantil Sant Joan de Déu per a la Investigació' integrated in the Spanish Biobank Network of Instituto de Salud Carlos III (ISCIII) for sample storage. We thank the Spanish Ministry of Health (Instituto de Salud Carlos III/FEDER, PI20/00389) and the parent association 'Síndrome duplicación MECP2: Miradas que hablan' (PFNR0085) for their funding; the European Social Fund and the Government of Catalonia (Secretaria d'Universitats I Recerca) for the doctoral grant FI (2020 FI-B 00888) given to Ainhoa Pascual-Alonso, and the Spanish Ministry of Science, Innovation and Universities for the doctoral grant FPU (FPU18/02152) given to Clara Xiol.

Conflict of interest

The authors declare that they have no competing interests.

Supplementary information

Supplementary information is available at Experimental & Molecular Medicine's website.

Supplementary information accompanies the manuscript on the Experimental & Molecular Medicine` website (http://www.nature.com/emm/).

References

1. Pascual-Alonso, A., Martínez-Monseny, A. F., Xiol, C. & Armstrong, J. MECP2-related disorders in males. *Int. J. Mol. Sci.* **22**, 9610 (2021).
2. Ta, D. *et al.* A brief history of MECP2 duplication syndrome: 20-years of clinical understanding. *Orphanet J. Rare Dis.* **17**, 131 (2022).
3. Bijlsma, E. K. *et al.* Xq28 duplications including MECP2 in five females: expanding the phenotype to severe mental retardation. *Eur. J. Med. Genet.* **55**, 404–413 (2012).
4. Mnatzakanian, G. N. *et al.* A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat. Genet.* **36**, 339–341 (2004).
5. Kishi, N. & Macklis, J. D. MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Mol. Cell. Neurosci.* **27**, 306–321 (2004).
6. Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science (80-.).* **320**, 1224–1229 (2008).
7. Gulmez Karaca, K., Brito, D. V. C. & Oliveira, A. M. M. MeCP2: a critical regulator of chromatin in neurodevelopment and adult brain function. *Int. J. Mol. Sci.* **20**, 4577 (2019).
8. Sandweiss, A. J., Brandt, V. L. & Zoghbi, H. Y. Advances in understanding of Rett syndrome and MECP2 duplication syndrome: prospects for future therapies. *Lancet Neurol.* **19**, 689–698 (2020).

9. Gottipati, S., Rao, N. L. & Fung-Leung, W. P. IRAK1: a critical signaling mediator of innate immunity. *Cell. Signal.* **20**, 269–276 (2008).
10. Neul, J. L. *et al.* Rett syndrome: revised diagnostic criteria and nomenclature. *Ann. Neurol.* **68**, 944–950 (2010).
11. Murdock, D. R. *et al.* Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *J. Clin. Invest.* **131**, e141500 (2021).
12. Stenton, S. L., Kremer, L. S., Kopajtich, R., Ludwig, C. & Prokisch, H. The diagnosis of inborn errors of metabolism by an integrative "multi-omics" approach: a perspective encompassing genomics, transcriptomics, and proteomics. *J. Inherit. Metab. Dis.* 1–11 (2019) doi:10.1002/jimd.12130.
13. Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C. A. & Zoghbi, H. Y. Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Hum. Mol. Genet.* **18**, 2431–2442 (2009).
14. Samaco, R. C. *et al.* Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2 duplication syndrome. *Nat. Genet.* **44**, 206–221 (2012).
15. Orlic-Milacic, M. *et al.* Over-expression of either MECP2-e1 or MECP2-e2 in neuronally differentiated cells results in different patterns of gene expression. *PLoS One* **9**, e91742 (2014).
16. Chen, L. *et al.* MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5509–5514 (2015).
17. Buist, M. *et al.* Differential sensitivity of the protein translation initiation machinery and mTOR Signaling to MECP2 gain-and loss-of-function involves MeCP2 isoform-specific homeostasis in the brain. *Cells* **11**, (2022).
18. Sun, Y. *et al.* Lack of MECP2 gene transcription on the duplicated alleles of two related asymptomatic females with Xq28 duplications and opposite X-chromosome inactivation skewing. *Hum. Mutat.* **42**, 1429–1442 (2021).
19. Wojtal, D. *et al.* Spell checking nature: versatility of CRISPR/Cas9 for developing treatments for inherited disorders. *Am. J. Hum. Genet.* **98**, 90–101 (2016).
20. Shao, Y. *et al.* Antisense oligonucleotide therapy in a humanized mouse model of MECP2 duplication syndrome. *Sci. Transl. Med* **13**, 7785 (2021).
21. Pascual-Alonso, A. *et al.* Molecular characterization of Spanish patients with MECP2 duplication syndrome. *Clin. Genet.* **97**, 610–620 (2020).
22. Monrós, E. *et al.* Rett syndrome in Spain: mutation analysis and clinical correlations. *Brain Dev.* **23**, S251–S253 (2001).

23. Pascual-Alonso, A. *et al.* Identication of molecular signatures and pathways involved in Rett syndrome using a multi-omics approach. *Hum. Genomics* (2023) doi:10.21203/rs.3.rs-2492515/v1. Preprint at https://www.researchsquare.com/article/rs-2492515/v1
24. Kopajtich, R. *et al.* Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders. *medRxiv* (2021) doi:10.1101/2021.03.09.21253187.
25. Fortin, D. A., Srivastava, T. & Soderling, T. R. Structural modulation of dendritic spines during synaptic plasticity. *Neuroscientist* **18**, 326–341 (2012).
26. Verpelli, C. & Sala, C. Molecular and synaptic defects in intellectual disability syndromes. *Curr. Opin. Neurobiol.* **22**, 530–536 (2012).
27. Liaci, C. *et al.* Neuronal cytoskeleton in intellectual disability: from systems biology and modeling to therapeutic opportunities. *Int. J. Mol. Sci.* **22**, 6167 (2021).
28. Betancur, C., Sakurai, T. & Buxbaum, J. D. The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci.* **32**, 402–412 (2009).
29. Jiang, M. *et al.* Dendritic arborization and spine dynamics are abnormal in the mouse model of MECP2 duplication syndrome. *J. Neurosci.* **33**, 19518–19533 (2013).
30. Pardo, L. *et al.* CREB regulates distinct adaptive transcriptional programs in astrocytes and neurons. *Sci. Rep.* **7**, 6390 (2017).
31. Quadrato, G. *et al.* Modulation of GABAA receptor signaling increases neurogenesis and suppresses anxiety through NFATc4. *J. Neurosci.* **34**, 8630–8645 (2014).
32. Li, X. *et al.* Loss of liver X receptor β in astrocytes leads to anxiety-like behaviors via regulating synaptic transmission in the medial prefrontal cortex in mice. *Mol. Psychiatry* **26**, 6380–6393 (2021).
33. Mou, Y. *et al.* Impaired lipid metabolism in astrocytes underlies degeneration of cortical projection neurons in hereditary spastic paraplegia. *Acta Neuropathol. Commun.* **8**, (2020).
34. Lin, J., de Rezende, V. L., de Aguiar da Costa, M., de Oliveira, J. & Gonçalves, C. L. Cholesterol metabolism pathway in autism spectrum disorder: from animal models to clinical observations. *Pharmacol. Biochem. Behav.* 173522 (2023) doi:10.1016/j.pbb.2023.173522.
35. He, C. W., Liao, C. P. & Pan, C. L. Wnt signalling in the development of axon, dendrites and synapses. *Open Biol.* **8**, (2018).

36. Armenteros, T., Andreu, Z., Hortigüela, R., Lie, D. C. & Mira, H. BMP and WNT signalling cooperate through LEF1 in the neuronal specification of adult hippocampal neural stem and progenitor cells. *Sci. Rep.* **8**, 1–14 (2018).
37. Gutierrez, H. & Davies, A. M. Regulation of neural process growth, elaboration and structural plasticity by NF-κB. *Trends Neurosci.* **34**, 316–325 (2011).
38. Kishi, N. *et al.* Reduction of aberrant NF-κB signalling ameliorates Rett syndrome phenotypes in Mecp2-null mice. *Nat. Commun.* **7**, (2016).
39. Alsabban, A. H., Morikawa, M., Tanaka, Y., Takei, Y. & Hirokawa, N. Kinesin Kif3b mutation reduces NMDAR subunit NR 2A trafficking and causes schizophrenia-like phenotypes in mice . *EMBO J.* **39**, (2020).
40. Joseph, N. F., Grinman, E., Swarnkar, S. & Puthanveettil, S. V. Molecular motor KIF3B acts as a key regulator of dendritic architecture in cortical neurons. *Front. Cell. Neurosci.* **14**, (2020).
41. Na, E. S. *et al.* A mouse model for MeCP2 duplication syndrome: MeCP2 overexpression impairs learning and memory and synaptic transmission. *J. Neurosci.* **32**, 3109–3117 (2012).
42. Núñez, L. *et al.* Tagged actin mRNA dysregulation in IGF2BP1 −/− mice. *PNAS* **119**, (2022).
43. Bauer, M. *et al.* Antibiotic prophylaxis, immunoglobulin substitution and supportive measures prevent infections in MECP2 duplication syndrome. *Pediatr. Infect. Dis. J.* **37**, 466–468 (2018).
44. Gottschalk, I. *et al.* IRAK1 duplication in MECP2 duplication syndrome does not increase canonical NF-κB–induced inflammation. *J. Clin. Immunol.* 1–19 (2022) doi:10.1007/s10875-022-01390-7.
45. Del Gaudio, D. *et al.* Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet. Med.* **8**, 784–792 (2006).
46. Shah, S. & Richter, J. D. Do fragile X syndrome and other intellectual disorders converge at aberrant pre-mRNA splicing? *Front. Psychiatry* **12**, 715346 (2021).
47. Li, R. *et al.* Misregulation of alternative splicing in a mouse model of Rett syndrome. *PLoS Genet.* **12**, e1006129 (2016).
48. Chhatbar, K., Cholewa-Waclaw, J., Shah, R., Bird, A. & Sanguinetti, G. Quantitative analysis questions the role of MeCP2 as a global regulator of alternative splicing. *PLoS Genet.* **16**, 1–14 (2020).
49. Van De Ven, R. *et al.* A role for multidrug resistance protein 4 (MRP4; ABCC4) in human dendritic cell migration. *Blood* **112**, 2353–2359 (2008).

50. Capmany, A. *et al.* MYO1C stabilizes actin and facilitates the arrival of transport carriers at the Golgi complex. *J. Cell Sci.* **132**, jcs225029 (2019).
51. Åslund, A. *et al.* Myosin 1c: A novel regulator of glucose uptake in brown adipocytes. *Mol. Metab.* **53**, (2021).
52. Fjodorova, M., Noakes, Z. & Li, M. A role for TGFβ signalling in medium spiny neuron differentiation of human pluripotent stem cells. *Neuronal Signal.* **4**, (2020).
53. Liu, J. *et al.* Wnt/β-catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduct. Target. Ther.* **7**, (2022).
54. Lowery, J. W. & Rosen, V. The BMP pathway and its inhibitors in the skeleton. *Physiol Rev* **98**, 2431–2452 (2018).
55. Finnson, K. W., Chi, Y., Bou-Gharios, G., Leask, A. & Philip, A. TGF-beta signaling in cartilage homeostasis and osteoarthritis. *Front. Biosci.* **S4**, 251–268 (2012).
56. Pecorelli, A. *et al.* Altered bone status in Rett syndrome. *Life* **11**, 521 (2021).
57. Trostle, A. J. *et al.* A comprehensive and integrative approach to MeCP2 disease transcriptomics. *Int. J. Mol. Sci.* **24**, 5122 (2023).
58. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
59. Pacheco, N. L. *et al.* RNA sequencing and proteomics approaches reveal novel deficits in the cortex of Mecp2-deficient mice, a model for Rett syndrome. *Mol. Autism* **8**, (2017).

DISCUSSION

RTT and MDS are two rare disorders with the same altered gene: *MECP2*. RTT is diagnosed based on the phenotype according to established diagnostic criteria, whereas MDS's diagnosis is mainly genetic. Both syndromes are phenotypically characterized and the genetic alteration triggering the syndromes is known in the majority of the patients. However, one of the current limitations is that the existing genotype-phenotype correlations for RTT and MDS can only foresee the prognosis of a few individual patients. For instance, in MDS, harbouring a triplication has been associated with a more severe phenotype, but patients with a duplication, which are the majority, can develop different degrees of severity and no further guidelines regarding the patient's prognosis can be given.

Despite the efforts of several groups, the underlying pathomechanism that leads from the altered gene to the clinical features is still unknown. Molecular data generated over the last two decades suggest that *MECP2* exerts subtle but ubiquitous effects rather than dysregulating a single specific pathway. The use of high-throughput technologies has made it possible to study the global state of these patients' tissues and to delineate some of the molecular effects of *MECP2*. Some breakthroughs have been made, but several questions remain unanswered.

To address these challenges, we opted to first, molecularly characterize patients with alterations involving *MECP2* using a range of traditional tests, and then, to study our RTT and MDS patients by applying transcriptomic and proteomic technologies, to obtain a complete picture of specific and common dysregulated pathways in these syndromes. As a result, we found that cytoskeleton-related genes are mainly downregulated, together with genes involved in the neuronal system, vesicular activity or immune system. In addition, the application of a multi-omics approach has highlighted some significantly dysregulated genes that could be candidates for therapeutic targets of the syndromes due to their neuronal functions and expression in brain tissues. Altogether, the generation of multi-omics data, especially when using adequate sample sizes, has proven to be a reliable and useful source of information to study RTT and MDS.

1. What we have learnt from our RTT patients

1.1. Regarding the DNA sequence composition within and surrounding *MECP2*

The study of large deletions within *MECP2* in classic RTT patients showed that breakpoints are not randomly distributed across the gene, but tend to be located in two areas: intron 2 and the DPR in exon 4. Both regions contain repetitive elements, including direct and inverted small repeats or Alu elements. Repetitive elements have been associated with DNA breakpoints and can cause all kinds of genomic rearrangements, including deletions, inversions, translocations and duplications. It is possible that the seven RTT patients, whose deletion breakpoints could not be sequenced, have complex rearrangements with inverted segments that escaped the experimental approach we used. Using a long-range polymerase, we detected an inversion within a deletion in patient P19, proving that deletions within *MECP2* may not be as simple as expected. In addition, to detect complex rearrangements, the primers should be designed considering the different orientations of the target segments, which was not the case for our patients. A tiling path oligonucleotide aCGH or whole genome sequencing approach may be appropriate to characterize the remaining RTT cases.

Carvalho et al. have studied the breakpoints of the duplications causing MDS using a tiling path oligonucleotide aCGH and discovered that a fair proportion of MDS patients have complex rearrangements with duplicated, triplicated and inverted segments within the original "duplication" [165]. They also discovered a region where distal breakpoints are clustered, located about 47kb upstream of *MECP2*. That region has direct and inverted LCR sequences, and similar to what happens with the deletions of RTT, generates an unstable DNA structure and induce DNA strand lesions that can become complex rearrangements. Therefore, repetitive elements located within and around *MECP2*, facilitate breakpoints and rearrangements. The resulting CNVs generate different pathologies, demonstrating once again that the correct dosage and functioning of the gene *MECP2* is vital.

1.2. Looking for a genotype-phenotype correlation in RTT

The variability that exists between RTT patients, even among those with the same mutation, makes it difficult to find a genotype-phenotype correlation. We checked the expression levels of *MECP2* and MeCP2 in our classic RTT patients and compared it to healthy controls. No significant difference in mRNA levels was found, but a significant impairment in protein amounts was detected. Furthermore, when we looked at the correlation between MeCP2 levels and the Pineda severity scores, we found a significant correlation (pval=0.012), showing that the lower the MeCP2 levels, the more affected the patient. These findings did not correlate with the XCI status of the fibroblasts, the age or the type of mutation. These results suggest that the contribution of genetic modifiers or environmental factors to RTT may be greater than expected. It would be interesting to see whether MeCP2 levels in fibroblasts remain stable with age. If so, MeCP2 itself could be a prognostic biomarker for these patients. A thorough examination of the exact phenotypic characteristics of patients with higher severity scores would enable paediatricians to provide more accurate counselling.

1.3. The importance of choosing an informative model for the experiments

There is a long list of RTT and MDS experimental models available to researchers, and choosing the most appropriate one depends on several aspects. The most studied RTT mouse model is Bird's *Mecp2*-null male mouse. More than 30 transcriptomic experiments have been performed in male RTT mice, while only 4 experiments have been reported in female mice despite being a more clinically relevant model [160]. As the male RTT mouse model lacks the XCI effect, it has been considered a "simplified" version of RTT. As a consequence, male mice are more similar to each other and experimental results are more homogeneous. These advantages have made it the animal model of choice, allowing research to progress more quickly. However, researchers should not forget that RTT occurs predominantly in females, which do indeed present a heterogeneous phenotype. Poor overlap has been found between DEGs from male and female mouse models [166]. Although the male RTT mouse models were engineered to gain insight into RTT, the results obtained could benefit the males presenting neonatal encephalopathies due to variants in *MECP2*.

Undoubtedly, mouse models have been a key element in RTT research over the past decades. However, it has been proven that translatability from mouse to human is not straightforward [166]. For example, the Sarizotan clinical trial was designed only based on RTT mouse model results. This trial failed to show efficacy in phases 2 and 3 and was cancelled. It has also been proven that the glutamatergic, serotonergic and cholinergic systems, which are altered in RTT patients, differ between humans and rodents [76]. The gathered evidence suggests that the human-derived models seem to be the recommended strategy we should all follow from now on.

Access to patients and families has led us to study human-derived samples. By studying humans directly, we can focus on the target population and determine their true genetic status without having to worry about extrapolating to another species.

The most affected organ in RTT and MDS is the brain. Human brain tissue is generally available post-mortem and the process of dying is mirrored in the gene expression profile [167]. Due to the difficulty of obtaining brain tissue samples of human origin, alternatives have been sought in more accessible tissues. A minimally invasive tissue is whole blood, but unfortunately, we found low reproducibility when measuring the mRNA levels of the lymphocytes. The dynamics of the blood, which is a constantly renewing tissue that also adapts in response to diet, drugs or infection, may affect the results of the experiments. We looked for a more stable tissue and opted for the fibroblasts derived from skin biopsies.

Literature supports fibroblasts as a reliable and robust tissue. Compared to whole blood, skin fibroblasts have demonstrated a higher consistency in gene expression, facilitating the detection of biologically relevant variations [168]. In addition, fibroblasts express a more comprehensive set of transcripts than whole blood, including more OMIM and neurologically relevant genes [168,169]. Skin fibroblasts are becoming widespread in rare disease investigation [170] and have already been successfully used in several publications to characterize molecular alterations in RTT and RTT-like patients [161,162,171–175]. In our personal experience, fibroblasts have proven to be stable and provide reproducible data, which we consider a prerequisite for any kind of experiment. We have also chosen to use patient-derived and unmodified cell lines in order to capture the background of these cells, even if this implies further genetic modifiers that we cannot identify yet. Using too simplified models will ultimately fail to capture the molecular environment of these patient's cells and will not lead to successful clinical trials.

1.4. Multi-omics results: altered biological processes in RTT

The cytoskeleton in RTT

Transcriptomic and proteomic analysis of the RTT and MDS cohorts revealed hundreds of dysregulated DEGs and DEPs but undoubtedly, cytoskeleton related genes are among the most promising ones.

Genes involved in cytoskeletal functions are fundamental to all cells because they are essential for cell shape, migration and proliferation, among other things. These processes are also needed for neuronal development and the proper functioning of the entire nervous system. Actin and microtubule cytoskeleton participate in neuronal migration, neurite initiation and outgrowth, axon differentiation, elongation and regeneration, dendritic spine morphodynamics and synapse functioning [157,158]. Furthermore, defects in actin and microtubules cause a wide range of nervous system abnormalities and neurodevelopmental disorders. Alterations in cytoskeleton-related proteins have been previously described in RTT patients and mouse models [163,164,167,172,175,176], and have been linked to the alterations in dendritic spine morphology and axonal organisation observed in MeCP2-deficient neurons [79,177–181]. The current evidence suggests that modulating cytoskeleton-based processes could offer new therapeutic paradigms to treat neurological defects. An example is the microtubule stabilising agent epothilone D, that impairs axonal dysfunctioning and Alzheimer-like pathology in aged transgenic mice [182].

Our multi-omic studies have revealed several significantly dysregulated genes and proteins that participate in cytoskeleton processes. The majority of these genes are downregulated. The RTT multi-omic analysis showed some cytoskeleton related genes with known functions in neuronal tissues. *ARMC9* is a known RTT spectrum gene involved in cilium assembly, signalling and transport, and is significantly downregulated. Bi-allelic *ARMC9* deficiency causes Joubert syndrome 30 (OMIM #617622), which presents some phenotypic overlap with classical RTT [183]. *CASK* phosphorylates synaptic cell-adhesion molecules [185], and belongs to the protein complex involved in the trafficking of glutamate ionotropic receptor NMDA type subunit 2B (*GRIN2B*) to the plasma membrane [186]. We found it downregulated and, in fact, *CASK* deficiency in both humans and mice has been shown to cause a malfunctioning of the brainstem leading to hypoventilation, a common feature among RTT patients [187].

COMT is a methyltransferase required for the metabolism and degradation of catecholamine neurotransmitters, including epinephrine, norepinephrine and dopamine [188], and we found it downregulated. RTT patients and mouse models have shown low levels of these biogenic amines and the alteration in dopaminergic metabolism has been associated with the characteristic motor deficits of RTT [189,190]. A combined treatment with levodopa and a dopa-decarboxylase inhibitor (Ddci) in *Mecp2* KO mice proved to rise their dopamine levels as well as to ameliorate their mobility deficits, tremor and breathing disturbances [189]. Since dopamine signalling promotes dendritic spine formation, the low levels of this neurotransmitter could be related to the defective dendritic spine morphogenesis observed in RTT neurons. Following the combined treatment with levodopa and Ddci, hippocampal neurons showed a significant increase in both the total number of spines and the number of mature spines. *Mecp2*-null mice also exhibit low levels of TH, the rate-limiting enzyme in dopamine synthesis [189]. Hence, the downregulation in *COMT* in RTT patients could be another sign of the altered dopaminergic metabolism, either as a direct effect of MeCP2 dysfunction or as a secondary alteration due to the malfunctioning of the pathway.

PREPL interacts with AP-1, a protein complex that organises membrane proteins in clathrin-coated vesicles at the trans Golgi network and endosomes [191]. Loss-of-function mutations in *PREPL* cause congenital myasthenic syndrome-22 (OMIM #616224) in a recessive manner, which presents with delayed motor development, waddling gait and hypotonia with onset at birth [192]. The interaction of *PREPL* with AP-1 promotes the dissociation of the protein complex from the plasma membrane [191]. Hence, the overexpression of *PREPL* that we observe in RTT patients' fibroblasts could produce a lack of membrane-bound AP-1, which is necessary for the correct membrane positioning of SNAREs [193]. This alteration could play a role in the altered vesicular activity that we identified through enrichment analysis. Moreover, the correct positioning of AP-1 is also needed for the trafficking of vesicular acetylcholine transporters, which could explain the motor deficiencies observed in both RTT and myasthenic syndrome patients [194]. Therefore, the defective functioning of the five highlighted genes that we have found consistently significantly dysregulated may be contributing to the neurological alterations seen in RTT mice and patients.

Other enriched processes found in RTT patients

Enrichment analysis revealed a downregulation of DEGs and DEPs involved in **rRNA processing** and ribosome biogenesis, which would impact general

protein translation in the affected cells. Previous studies have identified a reduction in mTORC1 activity [83,195], leading to a dysregulation of ribosomal proteins and translational initiation factors [83,196] and resulting in a global reduction in protein synthesis in RTT neurons [83,195,197].

Several **immune system** related DEGs and DEPs have been found dysregulated in our cohort. In particular, a significant enrichment of upregulated DEGs involved in the complement system cascade has been detected. The complement system is part of the innate immune response and is needed for recognition and clearing of pathogens, apoptotic cells or cellular debris. Some of its components are also involved in synapse pruning [198]. Recent evidence in animal models of Alzheimer disease and other disorders with impaired cognitive phenotypes showed an association between excessive complement-mediated synapse elimination and neuronal system dysregulation [199,200]. Immune system in general has also been found altered in several tissues of both RTT mouse models and patient cell lines, and it is thought that it could participate in the syndrome's deterioration [201]. All the gathered evidence leads to the thinking of a possible defective response and defence against external stimuli and antigens in RTT [201,202].

Before the causal gene for RTT was identified, some authors considered RTT to be a mitochondrial disorder due to the phenotypic overlap with mitochondrial diseases [203]. At the molecular level, **mitochondrial dysfunction** has been repeatedly reported in RTT mice and human samples [161,204–207]. Some of the reported alterations included increased production of reactive oxygen species (ROS), reduced levels of antioxidants and decreased activity of the mitochondrial respiratory chain [203,208]. Impairment of mitochondrial respiratory chain function leads to increased free radical production and thus oxidative stress, which may provoke protein dysregulation, oxidative tissue damage and ultimately contribute to neuronal network dysfunction in RTT [205,206,209]. We found that at the mRNA level, terms involved in respiratory electron transport, ATP metabolism, TCA cycle and mitochondrial translation were enriched with downregulated DEGs, whereas at the protein level, these terms were enriched with upregulated DEPs. Some authors propose a model in which MeCP2 modulates the expression of certain mitochondrial genes, which are key to maintaining mitochondrial function. As a result, respiratory function is impaired and mitochondria must increase their respiratory rates to achieve normal levels of ATP production. Over time, this leads to increased ROS production and oxidative stress, which impairs cellular energetic functions and contributes to the pathogenesis of RTT [208].

2. What we have learnt from our MDS patients

2.1. About the implication of *IRAK1* in MDS

It is a fact that duplication of at least *MECP2* and *IRAK1* is needed to present MDS. A case of a patient with a duplication encompassing *MECP2* but not *IRAK1* was presented to our centre. Evaluation by a paediatrician experienced in MDS clearly showed that the patient lacked the core clinical features of MDS, thus, proving the need for both genes to develop the syndrome. This requirement has been accepted worldwide since early 2000, but the development of a mouse model with duplicated *Mecp2* and *Irak1* only happened this year. The new mouse model presents the phenotype of previous MDS mouse models, but it also has an abnormal immune response, which is a core characteristic of MDS patients that was missing in previous models.

Recently, Gottschalk et al. evaluated *IRAK1*'s role in the canonical NF-κB signalling pathway by measuring cytokine production, IκBα degradation and *IRAK1* phosphorylation in 4 immortalised fibroblasts and 6 peripheral blood mononuclear cells and whole blood samples of MDS patients [210]. The canonical NF-κB signalling pathway mediates inflammatory response and they hypothesised that overexpression of *IRAK1* could over-activate the pathway causing excessive inflammation in MDS patients. Overall, they saw no significant difference between IL-6 and IL-8 production under different stimulators and surprisingly the cytokine production was impaired in the fibroblasts compared to healthy controls. *IRAK1* phosphorylation and IκBα degradation had a similar timing pattern compared to the control samples when activated with IL-1β. They concluded that since NF-κB pathway was not increased, inflammation in MDS is caused by the duplication of *MECP2* itself rather than by *IRAK1*.

Out of curiosity, we checked the expression levels of members of the canonical NF-κB pathway in our omics experiments. In the MDS cohort, mRNAs of *IL6* and *IL8* were significantly downregulated (BH<0.01) but no protein of those cytokines was detected in the proteomics experiment. *IRAK1* and *IRAK4* are significantly upregulated (BH<0.001), whereas towards the end of the NF-κB pathway, the following mRNAs are downregulated: *IkBα*, *NFKBIE, REL* (all three with pval<0.05), *RELA, RELB, NFKBIB, NFKBIL1* (all four with BH<0.05). Although at protein levels all these genes but *IRAK1* have recovered the expression levels of controls, that might not be the case in other tissues such as lung epithelia.

Having several *NFKB1* inhibitors downregulated might lead to an overexpression of the genes transcribed by NFKB1, while downregulation of RELA could alter the transcription of the genes regulated by RELA homo- and heterodimers. It would be interesting to investigate whether the NF-kB pathway's response to stimulus is also altered in our MDS fibroblasts as a result of the dysregulated genes we have found, and to see whether this is consistent with the findings of Gottschalk et al. Definitely, NF-kB pathway should be further studied in several tissues of MDS patients to discover whether it participates at the neuronal and immunological level in these individuals.

Taken together, the evidence from patients and animal models clearly indicates that in addition to *MECP2, IRAK1* also appears to be involved in the pathophysiology of MDS, although the mechanism by which it contributes to the syndrome remains unknown.

2.2. MDS diagnosis in females with the duplication of *MECP2*

The case of females in MDS is extremely interesting because of the role of XCI. XCI is usually studied with the AR method, which is not allele specific. The skewed results that are found in the carrier mothers are assumed to reflect the silencing of the allele with the duplication, explaining why carriers are asymptomatic. The XCI pattern is specific to each cell population and extrapolation from one tissue to another is not always possible, as proven with RTT [59]. However, the recurrent skewed pattern reported in MDS carriers suggest that the cell may be able to detect large rearrangements and silence them ubiquitously by XCI, whereas the point mutations or small indels causing RTT escape cellular control more easily and are inactivated differently in each tissue. Unfortunately, not all the alleles with the duplication are silenced in all the tissues and some girls develop MDS. Depending on the resulting XCI patterns, female MDS patients will present a phenotype that can be as severe as in males or much milder, with preserved ambulation or communication, for example.

At this point, a question could be raised about the girls who have been molecularly diagnosed as MDS patients, but who do not have any other MDS core features other than mild ID. This is the case of a girl who came to our hospital. She was referred to the clinicians regarding learning difficulties and attention deficits and the *MECP2* duplication was found by aCGH. She was

diagnosed with MDS, but now she is a 13-year-old teenager with dyslexia who walks, talks, attends high school and interacts with her classmates. We also have another female case who was diagnosed with MDS during a prenatal test. She is now two years old and has not yet developed any clinical features associated with MDS. The cases of these two girls raise the question of where (and when) to draw the line and consider a girl with the *MECP2* duplication to be a carrier or an affected patient. Both girls have been enrolled in the research projects of our department, but the older girl's sample is still being processed so could not be included in any omic experiment. The newborn girl's sample underwent both omic analyses, but the high impact on the list of DEGs and DEPs when adding one sample to a group of two samples, together with the lack of a healthy age- and sex-matched control sample, made it difficult to analyse and obtain reliable conclusions about the similarity between these MDS females. The study of these two girls, and others diagnosed in the coming years, will certainly help to expand the clinical and molecular picture of females with *MECP2* duplication.

2.3. Looking for a genotype-phenotype correlation in MDS

By the characteristics of the duplication

Despite attempts to establish a genotype-phenotype correlation for MDS, only a correlation based on the number of segments gained has been demonstrated (i.e. triplications result in a more severe phenotype than duplications).

We looked for a correlation by grouping our patients according to their duplication size and location (i.e. in tandem in the same Xq28 region or outside Xq28). In our cohort, all duplications larger than 1Mb were also located outside the Xq28 region. In our initial characterization, we found a trend between duplication size and location and clinical severity in our MDS patients, but without an appropriate severity scale, no significance could be achieved.

We also performed a DE analysis by grouping MDS patients according to these criteria but the enrichment studies performed with the multi-omics approach did not reveal any significant term. We only had three samples for the group with large and translocated duplications and our results may not be as reliable as wanted due to the effect of the small sample size. Definitely, larger cohorts of big and translocated duplications should be considered for this

approach in order to confirm whether the size and location of the duplication influences the severity of the syndrome and by which molecular mechanisms.

By the gene content of the duplication

It has been hypothesised that the duplication of certain genes is the cause of particular clinical features. *FLNA* is one of these candidate genes and it has been suggested to contribute to gastrointestinal problems [129]. However, patients without *FLNA* duplication and with gastrointestinal problems have been widely reported. Whether *FLNA* only causes a specific type of bowel dysfunction or whether other genetic causes also contribute to this clinical trait remains unstudied. The importance of conducting a detailed patient characterization is evident for the success of these association studies.

Another candidate gene that may influence the severity of MDS is *RAB39B* [132]. However, the publication suggesting this has two limitations. The first is that all the large duplications but one encompassed *RAB39B,* so they could not be certain about being the gene the cause of the more severe phenotype or the large size of the duplication and its full gene content. The second limitation is that the CSS they used was designed for RTT and, even though MDS and RTT share some clinical features, various core traits of MDS are missed by that scale. A bespoke MDS severity score would be needed in order to confirm these findings and the ones from other genotype-phenotype correlation studies. Besides, some of the core phenotype features of MDS appear with age, such as seizures. That implies a necessity to annually update the clinical score of the patients. In addition, a MDS severity scale would be essential to describe the natural history of the disorder and to monitor the evolution and response of the patients in future clinical trials. Clinicians at our hospital have been developing a severity scale for MDS patients and it will be published soon. Eight traits (sitting, ambulation, hand use, language, nonverbal communication, epilepsy, infections and gastrointestinal problems) are measured in it. Hopefully, its implementation will shed some light into the validity of the genotype-phenotype correlations that are being studied.

It sounds logical that by focusing on gene content a correlation between some of these genes and clinical traits will emerge. On the other hand, the persistent failures of this approach may suggest that instead of a single causative gene, the existing phenotype could be the result of small additive effects of multiple genes. If this were true, perhaps artificial intelligence could help us find a genotype-phenotype correlation for MDS. Artificial intelligence has been

applied to improve diagnosis, prognosis or drug repositioning in rare diseases [211], so perhaps it would be possible for it to interpret multi-omics data to detect patterns and biomarkers and help with prognosis in MDS patients.

2.4. Multi-omics results: altered biological processes in male MDS patients

The cytoskeleton in MDS

The MDS multi-omic analysis showed that almost a third of the consistent DEGs and DEPs are related to cytoskeleton. Similar to what happens in RTT, some of these genes' function occurs at neuronal level. *TMOD2* regulates dendritic arborization, *SRGAP1* regulates neuronal differentiation, *MOB2* regulates neurite formation and *VASP* regulates dendritic spine morphology, axon guidance and neuronal migration. The first two genes are consistently upregulated and the second two are downregulated.

Nine genes involved in the vesicular transport of different kinds of cargos are among the 103 consistently dysregulated DEGs and DEPs. We highlighted two of them: *KIF3B* and *IGF2BP1*. *KIF3B* transports vesicles containing NMDAR subunit 2A required for an adequate synaptic plasticity needed for learning and memory, and we found it downregulated [212,213]. It has been reported that KIF3B inhibition results in an increased spine density and *Kif3b* haploinsufficient models have an upregulated long term potentiation (LTP) [212,213]. These two features have been seen in MDS mouse models [136,142]. *IGF2BP1* allows mRNA transport and transient storage, modulating the location and rate at which those transcripts bind to the translational apparatus and protecting them from degradation [214,215]. Besides, in neurons, *IGF2BP1* is involved in neuronal development, synapse formation and it also transports transcripts required for axonal regeneration on adult sensory neurons [215]. Altogether, our data demonstrate an alteration in neuronal activity, particularly in synaptic function, that could be partly explained by the impairment of genes related to the cytoskeleton and vesicular activity.

Other enriched processes in male MDS patients

Immune system related terms are also significantly enriched with mainly downregulated DEGs and DEPs. Dysregulation of the immune system was

the most expected process, given the serious health problem that infections represent for MDS patients. The recurrent infections, together with epilepsy, are one of the major factors that triggers regression and causes death in MDS patients. It has been reported that elements of their immune system are impaired [104,108,109,111,216]. Among our consistently dysregulated genes and proteins we have *IRAK1, MASP1* and *MPP1*, the first two are involved in innate immunity function and *MPP1* in neutrophil polarity.

Terms involved in **metabolism** have also been found in our analysis, specially related with lipids and fatty acids. Neuronal membranes contain polyunsaturated fatty acids that are susceptible to oxidative stress. Oxidative stress has been associated with neurological disorders and diseases presenting cognitive impairment [217]. Oxidised products derived from polyunsaturated fatty acids could reflect an ongoing oxidative brain damage in MDS, similar to what has been reported in RTT [123].

ERO1A, FKBP7, *NGLY1* and *PCMT1*, which are consistently upregulated, intervene in **protein folding** status, both helping during the folding step or recognizing misfolded proteins and initiating their repair or exporting them to be eliminated. Their upregulation may reflect a cellular need to eliminate the overproduction of proteins due to *MECP2*'s overexpression.

2.5. Multi-omics analysis results of female MDS patients

In the female MDS multi-omic analysis, we found that 66 of the DEGs and DEPs (26% of all common genes) are related to translation activity. Buist et al. recently reported an impairment of protein translation and mTOR signalling in RTT human brains, but not in a cell model overexpressing *MECP2* isoforms [218]. These findings are consistent with our data, since no relevant impairment was found in our male MDS patients, but in the females, whose phenotype is milder, there is a dysregulation of the translation-related genes, which are mainly upregulated. All the differentially expressed genes and TFs related to translation found in the female MDS patients, together with those found in the enriched pathways involved in proteasome regulation and nonsense-mediated decay, could reflect a cellular compensation as a consequence of the duplication, explaining why females have a milder phenotype and a different molecular signature than males with MDS.

Alterations in cytoskeletal functions and vesicle trafficking were also found. These may contribute to the dysregulation of neuronal architecture and synaptic function mentioned above. The low overlap of DEGs and especially DEPs found between male and female MDS cohorts confirms the gender-related phenotypic disparity in this syndrome.

3. What we have learnt from our *MECP2* duplication carrier mothers

3.1. Asymptomatic or symptomatic?

MECP2 duplication carrier mothers are described as "apparently asymptomatic". Ramocki et al. examined 8 female carriers and found that the majority of the women exhibited neuropsychiatric abnormalities such as anxiety, depression or compulsions, as well as features of autism phenotype prior to the birth of a child with MDS [219]. One of the MDS familiar cases we have characterized in the Spanish cohort (I7) was published by Bijlsma et al. and they reported that his carrier mother (M7) had learning difficulties at school and that she presented depression and compulsions before having her son [127]. Two of the mothers from the Spanish MDS cohort (M20 and M21), the two whose XCI does not reach the skewed threshold, are two sisters that had learning difficulties at school but no IQ test or further tests were done [220]. Although Ramocki and Bijlsma's carrier mothers were carefully studied, generalisation of these results to the rest of the carrier population should be made with caution. The three traits that are most recurrently referred to are anxiety, depression and compulsions (e.g. need for routine, structure, cleaning and ordering); the first two behaviours have a strong environmental component and the third could be influenced by the social environment in which the women were raised. In addition, the same autism phenotype questionnaire used by Ramocki et al. could have been applied to non-carrier mothers of children with and without MDS. A statistical difference between these groups of mothers could add value to these statements.

At transcriptomics level, we found an extensive set of genes differentially expressed in our 10 carrier mothers, that when checked at protein level, recovered normality. These results indicate that these females are able to

compensate for the vast majority of the dysregulations that have occurred at mRNA level and that the final picture does not have altered gene sets involved in specific biological processes. However, it is true that some consistently altered genes and proteins have been found, and that we cannot rule out the possibility that those genes, for example the ones involved in cytoskeleton functions, may generate some subtle neuropsychiatric aspects in the carrier females, even though no significant enriched terms related to it have been found in our analysis.

To answer the question of whether carrier mothers have a specific phenotype, it would be ideal to evaluate them extensively. A questionnaire that looks at the autism phenotype from an adult perspective, combined with an assessment of the individual's social environment prior to the birth of the affected children, might shed some light on this issue. An immunological assessment could also be added, as some of Ramocki's carriers had autoimmune diseases and two of our carrier mothers, who are sisters, also mentioned that they needed recurrent vaccine boosters. We did not find enriched terms suggesting a dysregulated immune system either. However, a single gene may be enough to cause this immune problem, and it could be that the gene responsible is duplicated in only a few women. As suggested earlier, the questionnaire and the studies should also be applied to non-carrier mothers to confirm the statistical significance of the results.

3.2. Multi-omics analysis results from *MECP2* duplication carrier mothers

It is thought that a skewed XCI may contribute to a lower number of cells expressing the molecular alterations caused by the duplication and therefore, to the asymptomatic phenotype of the carriers. However, we found multiple significant DEGs and DEPs in our carriers, which suggests that a skewed XCI is not sufficient to ensure a correct molecular profile. Enrichment analysis performed in our carrier mothers revealed a few terms, but no specific altered biological process. Hence, despite the presence of hundreds of significant DEGs and DEPs, from a multi-omic perspective, no apparent dysregulation is detected in the carriers, explaining their asymptomatic phenotype. Our data indicate that some women are able to molecularly compensate for the dysregulations caused by the duplication and become asymptomatic. DEGs and DEPs associated with **RNA transcription and regulation, translation, protein folding**

and degradation from the carrier cohort, which are not significant in the male and female cohorts, may be essential for this. How the cells trigger molecular compensatory mechanisms in some but not all females remains unknown. The use of single cell technology may be appropriate to investigate how each cell functions in these women and what processes are active in the cells according to the allele they express.

4. Multi-omics approach to analyse RTT and MDS cohorts

4.1. Multi-omics data analysis and limitations

Zoghbi's group performed one of the first high throughput transcriptomics experiments in two brain areas of a RTT and a MDS mouse model. They found that the majority of the genes were downregulated in the *Mecp2*-null mice and upregulated in the *MECP2*-Tg mice [15,16]. Contrary to what they found in mice hypothalami and cerebella, in our human derived fibroblasts we saw that less than 30% of the DEGs were common between RTT and MDS and that the DEGs were evenly dysregulated in both directions. The integrative transcriptomics analysis of 43 mouse studies performed by Trostle et al. showed an even distribution of the dysregulated genes [166]. Zoghbi's group also found that neuronal and synaptic activities were upregulated in MDS transgenic mice [16], but the terms we found related to the neuronal system were more often downregulated. These data suggest that despite being a global regulator, *MECP2* might have specific functions, and thereby downstream effects, in each tissue. Nevertheless, the DE results in mice may not reflect the dysregulations we found in human fibroblasts, and thus, further studies are recommended.

Around 70 transcriptomic studies have been carried out in RTT and, contrary to what one might expect, each experiment reports a specific set of genes that barely overlap with the rest, even within a species [166,167]. First of all, each experiment uses a different species, RTT model, tissue, *MECP2* mutation or disease time-point, introducing specific expression patterns linked to those aspects. For example, Tanaka et al. demonstrated that each *MECP2* mutation affects a partly different set of genes [221]. Even studies performed with similar conditions present minimally overlapping lists of DEGs [176]. Secondly, the

technology used to measure mRNA expression and to identify proteins has evolved and become more sensitive. In addition, the lack of a gold standard pipeline to analyse each omics data has resulted in that each group has developed its own pipeline and set of tools to normalise and study their data. As a result, there is little overlap when comparing DEGs and DEPs from different studies.

Out of curiosity, we collected the DEG lists of 50 published RTT transcriptomic experiments and saw that 75% of the significant DEGs identified in our RTT RNAseq were reported as significant elsewhere. On the contrary, at least 9200 significant DEGs were reported by others but were missing from our experiment. Obtaining consistent gene sets across studies reinforces the idea that some of the effects of *MECP2* are truly global, and that studying tissues other than the human brain is a good approach in *MECP2*-related disorders. However, the disparity seen between experiments warns us to be careful about the interpretations that are done from the data of a single experiment; especially when sample sizes are small, which occurs frequently. To overcome the current limitations we waited until we gathered cohorts of six samples or more (22 classic RTT, 17 MDS, 10 carrier mothers and 13 healthy controls), a condition that is recommended and most of the experiments fail to meet [166,222].

4.2. The challenge of the sample sizes in rare diseases

The majority of transcriptomic experiments performed in RTT ranged from three to six samples, with the largest human cohort consisting of 12 RTT patients [204]. Only two mouse studies were conducted with larger cohorts of up to 50 mice [223,224]. The largest RTT cohort that has been studied with proteomics experiments consists of 25 patients [225]. Up to date, no multi-omics analysis has been performed with RTT human samples and only one has been published in which 4 samples of RTT mouse models have been studied [176]. For MDS, 5 transcriptomics studies in mouse models [15,16,226–228] and one study with two MDS patients [229] have been published. No proteomics experiments have been published so far.

It has been demonstrated that small sample sizes lead to low statistical power, and their robustness and validity may not be as reliable as desired. A drawback that must be taken into consideration when working with small sample sizes is that one sample can have a decisive impact in the output of a statistical comparison. That means that when a study group consists of few samples, sometimes the extremely high or low expression of a gene in a single sample is enough to change the group's

arithmetic mean and to obtain a significant (or no significant) result that might not be representative of the group when more samples are considered. Unfortunately, when studying rare disorders gathering large cohorts of patients is complicated and it is part of our duty to interpret the results we have generated with caution, being aware that we will probably have false positive and negative results.

While analysing the MDS cohort we found that some X chromosome related genes were differentially expressed due to the unbalanced sex distribution of the cohort. Therefore, we separated the males and females for the DE analysis. The male MDS cohort consisted of 15 patients and the results were reproducible and reliable. We did not want to overlook the females, that were only two, so we decided to publish their data, but always using restrictive thresholds. As has been acknowledged several times, the female MDS patient's data should be interpreted with caution. However, those two datasets (as well as the rest) are now available and can be re-studied together with other datasets in an integrative approach. Integrative approaches address the limitations of small sample sizes increasing the statistical power of the analysis and improving the outcome by reducing the false positives. Therefore, they can capture the common transcriptomic changes that are likely to cause the *MECP2*-related disorders [166,167].

Having a global idea of what is dysregulated in all these children will generate new research projects that may eventually lead to treatments or therapies. Once the common dysregulated processes have been elucidated and some drug candidates have been tested, a more individualised experimental approach may be beneficial. For example, scRNAseq could be used to measure each patient's gene dysregulation in order to administer a tailored dose of a drug or to follow its effect over time. This could be particularly important in females to see how cells expressing the wt and affected alleles respond.

4.3. Common multi-omics results between RTT and MDS patients

RTT and MDS share some common phenotypic features such as ID, neurodevelopmental delay, hypotonia, epilepsy and gastrointestinal problems. We wanted to compare the results of DE analysis performed on classic RTT patients and MDS patients to identify common gene expression dysregulations that might shed light on the pathomechanism underlying both syndromes. 721

DEGs and 12 DEPs are common to both cohorts, but only 2 genes (*MYO1C* and *HARS2*) are dysregulated with both omics in both syndromes.

MYO1C is a myosin implicated in cytoskeletal organisation and vesicle trafficking. It stabilises actin in the Golgi complex and regulates cargo transport. When depleted, MYOC1 induces fragmentation of the Golgi complex, loss of cellular F-actin and a delay in vesicle transport [230]. *MYO1C* is also involved in the recycling of glucose transporters in response to insulin and we found it consistently downregulated [231]. Although to date, only four cases of girls with RTT and diabetes have been published, abnormal glucose levels and insulin resistance have been reported in RTT mice and patients [232]. Glucose is essential in the brain cells as a source of energy and is also necessary for neurotransmitter metabolism such as GABA and acetylcholine. The downregulation of *MYO1C* suggests that glucose reabsorption is partially impaired, so higher blood glucose levels would also be expected in our cohort. Metformin, an anti-diabetic drug that acts as a blood glucose lowering agent, was administered to female RTT mice and the mitochondrial function was rescued [233]. No data have been collected on glucose metabolism in children with MDS, but the molecular similarities between these two syndromes lead us to believe that MDS may benefit from some of the therapies developed for RTT.

HARS2 is a mitochondrial histidyl-tRNA synthetase that charges tRNAs. It is consistently upregulated in MDS patients while in RTT, the mRNA is downregulated and the protein is upregulated. The amino acid that charges, histidine, is an essential amino acid. It is the precursor of histamine, a substance that immune system cells release during an allergy. Histamine is also needed for myelination and can act as a neurotransmitter interacting with other neurotransmitters such as GABA. Histamine increases neuronal excitability and maybe, it could contribute to the excitability/inhibitory imbalance reported in RTT. However, the effect of having an increased *HARS2*, which is mitochondrial, but not the cytosolic synthetase *HARS1* needs to be further explored.

Common transcriptomics results between RTT and MDS patients

The expression of *MECP2* is decreased in RTT and increased in MDS, so we wondered whether they share DEGs that are expressed in opposite directions. The 82 DEGs downregulated in RTT and upregulated in MDS patients are overrepresented in terms related to DNA replication, cell cycle and mRNA splicing, terms that have been described by us and by other groups [167,234]. Eight

of the 82 DEGs are part of spliceosome complexes and another four are related to mRNA stability, processing and maturation functions.

The 100 DEGs upregulated in RTT and downregulated in MDS enrich processes related to neurogenesis regulation, cytoskeleton and Wnt, BMP and TGFß signalling cascades. Wnt, BMP and TGFß signalling pathways play a role in osteoblast activity and maintenance of cartilage, among others [235–237]. RTT girls suffer from scoliosis, their bone mass density is low, even lower in RTT patients with the so-called "severe" typical MECP2 mutations (R106T, R168X, R255X, R270X), and their bone fracture rate is four times that of the general population [238–240]. Scoliosis is the most commonly reported orthopaedic issue in MDS patients, and osteopenia, contractures of joints and fractures have also been reported [106]. The dysregulation found in Wnt, BMP and TGFß signalling pathways could be related to the problems found in the skeletal system of RTT and MDS patients since childhood. However, it should be noted that valproate, a drug used to control epilepsy in RTT and some MDS, may also increase the risk of bone fractures [45]. Besides, BMP signalling cascade has recently been found upregulated in neural stem/precursor cells from RTT patients and treatment with BMP inhibitors partially rescued the abnormal development detected in RTT brain organoids [241]. The fine-tuning of BMP signalling cascade may also benefit the downregulation seen in MDS.

TGFß signalling pathway is also implicated in innate and adaptive immunity [242]. In addition, three out of the 12 DEPs (*APPL2, CNPY4* and *CTSC*) regulate immune response and are downregulated in RTT and upregulated in MDS. An impaired ability to eliminate foreign antigens has been suggested in RTT patients [202]. RTT mice also have an altered cytokine profile, and inflammasomes, protein complexes with a role in innate immune responses, were also found to be dysregulated in RTT fibroblasts [243]. Finding altered pathways related to immunity in both RTT and MDS patients highlights again the resemblance of the two syndromes at the molecular level.

Common proteomics results between RTT and MDS patients

Surprisingly, only 12 DEPs are shared between the RTT and MDS cohorts, and these proteins are involved in quite varied activities. Transcriptomic and proteomic profiles have shown limited correlation across previously published studies in different species and cell types [244–246]. The squared Pearson correlation coefficient between these two omics is of about 0.40. This means that about 40% of the protein variation can be explained by mRNA abundances. The remaining 60% is thought to

be explained by the different rates of mRNA and protein degradation and by post-transcriptional and translational regulatory mechanisms. In addition, miRNAs have been reported to fine-tune protein abundance and could also contribute to the difference in rates between mRNA and protein production and degradation. Furthermore, the different levels of sensitivity and error measurement that transcriptomic and proteomic technologies have may also influence the different outcomes that both experiments generate [246].

In concordance with previously published results, the correlation between transcriptome and proteome DE findings in our study was low (Pearson correlation coefficient = 0.09 for all RTT; 0.17 for female RTT; 0.34 for male MDS; 0.12 for female MDS and 0.09 for carrier mothers). Interestingly, the higher correlation coefficient corresponds to the male MDS cohort, perhaps because it is a large and homogeneous group. The impairment in protein synthesis found in the RTT cohort may help to explain, at least in part, some of the differences observed between the differentially expressed gene sets in transcriptomic and proteomic.

The different technologies behind RNAseq and TMT-MS allow the measurement of different amounts of genes and proteins; in our case, TMT-MS identified about half of the genes detected by RNAseq. In addition to different sensitivities and measurement limitations of the technologies, the intrinsic properties of the molecules also contribute to the lack of correlation. The life-time of the molecules, understood as a combination of post-transcriptional, translational and degradation regulation, also contributes to the final mRNA and protein abundance. The different rates at which these processes occur in each molecule may be reflected in the low correlation we observe. Future improvements in the resolution of current omic technologies and their combination with other methods, such as high-throughput cell imaging or ribosome footprinting, will allow the collection of new data that will resolve these discrepancies.

4.4. Transcription factor analysis

Transcription factors (TFs) could also be responsible for some of the differentially expressed changes detected in our experiments. Therefore, we used the ChEA3 tool to identify them. The advantage of this approach is that by targeting a single TF several dysregulated genes can be compensated for. The limitation is that TFs are associated with more genes than the subset we need

to correct. Thus, modifying a TF could lead to the deregulation of other genes, with consequences for the phenotype. Several TFs that regulate genes involved in neuronal activity were identified in our cohorts, and some of them even target the DEGs that are common to the RTT and MDS cohorts and are expressed in opposite directions. Despite the difficulties, fine-tuning these TFs could be a potential therapeutic approach.

Two potential TFs we have found are *NFATC4* and *NR1H2*. *NFATC4* is the TF of 24% of the genes that are significantly upregulated in RTT and downregulated in MDS, and it also regulates about 9% of the DEGs and DEPs found in the male MDS cohort. *NFATC4* regulates adult neurogenesis and mediates synaptic plasticity with *BDNF* [247]. *BDNF* is regulated by *MECP2* and is needed for GABAergic interneuron maturation. GABAaR stimulation via *NFATC4* activation decreases innate anxiety in mice [247], a trait that some RTT and MDS patients present [248–250].

NR1H2 is the TF of approximately 17% of the DEGs and DEPs found in the male MDS cohort and is involved in the layering of brain structures and regulates cholesterol trafficking [251]. GW3965, an *NR1H2* agonist, rescues synapse formation and function in hippocampal neurons and ameliorates cholesterol levels and axonal defects in cortical projection neurons [251,252]. A correct neurotransmission depends on cholesterol [253], and therefore, trying to restore NR1H2 neuronal levels could be beneficial for MDS patients in multiple ways.

The results of these experiments showed that regulation of *NFATC4* and *NR1H2* could improve the anxiety and some of the synaptic impairments seen in RTT and MDS patients.

5. Future prospects

5.1. Expanding the knowledge and integrating analyses

The present work has provided a number of avenues for future research. We have published robust and reliable RNA and protein profiles of classic RTT patients, male MDS patients and *MECP2* duplication carrier mothers, the latter two cohorts being reported for the first time. In total, we have generated 71

RNAseq and 73 TMT-MS datasets that will be publicly available. Therefore, other research groups may benefit from these data and could search for specific processes or pathways in which they are experts to obtain new findings.

Knowing that fibroblasts are a reliable tissue to study *MECP2*-related disorders, our datasets could be integrated with other datasets to increase the sample size and mitigate the drawbacks caused by the use of reduced sample sizes. Integrative analysis allows the finding of core DEGs and DEPs and could eliminate the individual variability captured by existing experiments and focus on the disease-causing alterations. The establishment of collaborative networks or the creation of high throughput databases with datasets derived from individuals with *MECP2*-related disorders, similar to what Trostle et al. have initiated with mouse transcriptomics data, could also increase the speed of research in this big data era [166]. In addition, by integrating new patients with our own, it will be possible to determine whether the trends we have found in the various attempts to establish a genotype-phenotype correlation are significant or not. Using large cohorts, other subgroups, based on molecular or clinical features, may also reveal significant correlations that have not yet been studied. In any case, and most importantly, to successfully establish a correlation with MDS patients, an MDS severity scale should be implemented.

5.2. Further validation of the candidate genes

As a result of our multi-omic experiments, we highlighted five significantly dysregulated genes from the RTT cohort (*ARMC9, CASK, COMT* and *PREPL*), six genes and one TF from the MDS cohort (*TMOD2, SRGAP1, MOB2, VASP, KIF3B, IGF2BP1* and *NR1H2*), and two genes and one TF shared between the RTT and MDS cohorts (*MYO1C, HARS2* and *NFATC4*) that are worthy of further investigation because of their known involvement in neuronal functions. For some of these genes, there are even some experiments demonstrating amelioration of known altered processes in mouse models. Restoring the expression levels of the first 10 candidate genes could improve synaptic function in these children. In particular, *KIF3B* seems to be a promising candidate because its inhibition leads to increased spine density and upregulated LTP, two features that have been seen in the brains of MDS mouse models. *NR1H2* may also have a neuronal effect by balancing brain cholesterol levels. Activation of *NFATC4* reduces innate anxiety in mice and may do the same in RTT and MDS patients. A functional study upregulating *MYO1C* and seeing whether glucose metabolism changes, could

also lead to an improvement in the mitochondrial dysfunction as reported by Urbinati et al. in RTT mice. Finally, studying the effect of *HARS2* in histaminergic neurons of RTT and MDS cell models could be an interesting approach to see if there is excessive histamine production due to *HARS2* upregulation and if this alteration could be reversed by downregulating *HARS2*. Hopefully, some of the functional studies done with these candidate genes will lead to the development of treatments that will improve the quality of life and lifespan of these children.

5.3. Expanding the clinical and molecular picture of females with *MECP2* duplication

With the standardisation of NGS as a diagnostic tool, an increase in the detection of women carrying a *MECP2* duplication is expected. A thorough molecular and clinical characterization of all women carrying the duplication, regardless of affected status, will also uncover a whole new picture, ranging from asymptomatic to severely affected. The study of their molecular profiles will show which processes are silenced by the XCI effect in asymptomatic carriers and which cause the phenotype. scRNAseq could also reveal the precise transcriptomics effects of expressing the allele with and without the duplication in females. This information could reveal potential processes that need to be compensated for to achieve asymptomatic carrier status and live normally.

5.4. More applications for multi-omics in RTT and MDS

The utility of a multi-omic approach to find biomarkers has been proven, which are still missing in these disorders and that will be essential to monitor the effects of drugs during clinical trials. In RTT clinical trials, the effect of the drugs is evaluated by measuring the improvement in the symptomatology and quality of life of the patients. Those are usually measured via clinical severity scores, checklists, records of breathing patterns or EEGs [254]. The lack of a biomarker complicates an objective quantification of the improvement. A comprehensive knowledge of the molecular dysregulations caused by the malfunction of *MECP2* could lead us to the finding of biomarkers that could be applied in clinical trials, both as response measuring tools or as a prognosis tool. For example, Shovlin et al. detected RNA profiles in patients undergoing Mecasermin clinical trials that matched with the clinical phenotypes and that changed as the trial went on [255].

The disease stages that a classic RTT patient goes through are well defined, but no longitudinal study has been done in MDS patients. Apart from studying MDS patients from a clinical point of view, a multi-omic based longitudinal study in female and male cohorts would also be an interesting approach. It has been reported that the transcriptome changes with age [256], so obtaining the RNA and protein profiles at different ages and disease stages would be interesting from a research point of view and perhaps even helpful from a medical point of view. For example, in the aforementioned cases of females carrying a *MECP2* duplication but without a fully penetrant MDS phenotype, performing multi-omics at the age when the RNA and protein profiles acquire the "definitive" profile of an asymptomatic carrier or an affected girl would help in the diagnosis and genetic counselling of the girl.

CONCLUSIONS

The present doctoral thesis has identified which genes, proteins and molecular processes are significantly altered in patients with Rett syndrome (RTT) and *MECP2* duplication syndrome (MDS), and has examined what is common and what is specific to each cohort. The generation of reliable datasets has also allowed us to find candidate genes for potential therapeutic targets and biomarkers.

Specifically, we can conclude the following from the thesis project:

- Deletions involving only exon 4 of *MECP2* correlate with milder RTT phenotypes and deletions involving *IRAK1* or exons 1 and 2 are associated with more severe phenotypes.

- A significant genotype-phenotype correlation has been found in RTT patients between the MeCP2 levels and the Pineda severity score, with lower protein levels associated with a more severe phenotype.

- In patients with MDS, we found that those with duplications outside the Xq28 region tended to have a more severe phenotype than those with duplications in tandem.

- A variant located in a gene from the X chromosome should not be discarded without further evaluation simply because it is present in an apparently asymptomatic female. XCI could explain the absence of clinical features in the carriers.

- Genes involved in neuronal function and expressed in brain tissue were found to be dysregulated in the patients' fibroblasts, proving the utility of fibroblast tissue in the study of *MECP2*-related disorders.

- Patients with RTT and MDS have dysregulated processes involved in the cytoskeleton, vesicular activity and immune system.

- *ARMC9, CASK, COMT* and *PREPL* are candidate biomarkers for RTT patients.

- *TMOD, SRGAP1, MOB2, VASP, KIF3B* and *IGFBP1* are candidate biomarkers for male MDS patients. In particular, *KIF3B* is a promising candidate for MDS, as its inhibition leads to two features reported in the brains of the MDS mouse model.

- Transcriptomic and proteomic profiles of males and females with MDS, carrier mothers and RTT patients are specific to each cohort.

- The dysregulated DEGs and DEPs related to transcriptional regulation, mRNA processing, translational activity and protein degradation present in the carrier mothers' cohort may compensate for the effects of the duplication in them, but not in MDS females.

BIBLIOGRAPHY

1. Xinhua, B. *et al.* X chromosome inactivation in Rett syndrome and its correlations with MECP2 mutations and phenotype. *J. Child Neurol.* **23**, 22–25 (2008).
2. Mnatzakanian, G. N. *et al.* A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat. Genet.* **36**, 339–341 (2004).
3. Martínez De Paz, A. *et al.* MeCP2-E1 isoform is a dynamically expressed, weakly DNA-bound protein with different protein and DNA interactions compared to MeCP2-E2. *Epigenetics and Chromatin* **12**, (2019).
4. Dastidar, S. G. *et al.* Isoform-specific toxicity of MeCP2 in postmitotic neurons: suppression of neurotoxicity by FoxG1. *J. Neurosci.* **32**, 2846–2855 (2012).
5. Olson, C. O., Zachariah, R. M., Ezeonwuka, C. D., Liyanage, V. R. B. & Rastegar, M. Brain region-specific expression of MeCP2 isoforms correlates with DNA methylation within Mecp2 regulatory elements. *PLoS One* **9**, (2014).
6. Krishnaraj, R., Ho, G. & Christodoulou, J. RettBASE: Rett syndrome database update. *Hum. Mutat.* **38**, 922–931 (2017).
7. Lagger, S. *et al.* MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. *PLoS Genet.* **13**, e1006793 (2017).
8. Lei, M., Tempel, W., Chen, S., Liu, K. & Min, J. Plasticity at the DNA recognition site of the MeCP2 mCG-binding domain. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1862**, 194409 (2019).
9. Kumar, A. *et al.* Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. *J. Cell Sci.* **121**, 1128–1137 (2008).
10. Ip, J. P. K., Mellios, N. & Sur, M. Rett syndrome: insights into genetic, molecular and circuit mechanisms. *Nat. Rev. Neurosci.* **19**, 368–382 (2018).
11. Adams, V. H., McBryant, S. J., Wade, P. A., Woodcock, C. L. & Hansen, J. C. Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J. Biol. Chem.* **282**, 15057–15064 (2007).
12. Sharma, K., Singh, J., Frost, E. E. & Pillai, P. P. MeCP2 in central nervous system glial cells: current updates. *Acta Neurobiol. Exp. (Wars).* **78**, 30–40 (2018).
13. Sandweiss, A. J., Brandt, V. L. & Zoghbi, H. Y. Advances in understanding of Rett syndrome and MECP2 duplication syndrome: prospects for future therapies. *Lancet Neurol.* **19**, 689–698 (2020).
14. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185–188 (1999).

15. Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science (80-.).* **320**, 1224–1229 (2008).
16. Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C. A. & Zoghbi, H. Y. Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Hum. Mol. Genet.* **18**, 2431–2442 (2009).
17. Wu, H. *et al.* Genome-wide analysis reveals methyl-CpG-binding protein 2-dependent regulation of microRNAs in a mouse model of Rett syndrome. *PNAS* **107**, 18161–18166 (2010).
18. Young, J. I. *et al.* Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *PNAS* **102**, 17551–17558 (2005).
19. Chhatbar, K., Cholewa-Waclaw, J., Shah, R., Bird, A. & Sanguinetti, G. Quantitative analysis questions the role of MeCP2 as a global regulator of alternative splicing. *PLoS Genet.* **16**, 1–14 (2020).
20. Bedogni, F. *et al.* Rett syndrome and the urge of novel approaches to study MeCP2 functions and mechanisms of action. *Neurosci. Biobehav. Rev.* **46**, 187–201 (2014).
21. Cheng, T. L. & Qiu, Z. MeCP2: multifaceted roles in gene regulation and neural development. *Neurosci. Bull.* **30**, 601–609 (2014).
22. Good, K. V., Vincent, J. B. & Ausió, J. MeCP2: the genetic driver of Rett syndrome epigenetics. *Front. Genet.* **12**, 620859 (2021).
23. Tao, J. *et al.* Phosphorylation of MeCP2 at Serine 80 regulates its chromatin association and neurological function. *PNAS March* **24**, 4882–4887 (2009).
24. Cohen, S. *et al.* Genome-wide activity-dependent MeCP2 phosphorylation regulates nervous system development and function. *Neuron* **72**, 72–85 (2011).
25. Bellini, E. *et al.* MeCP2 post-translational modifications: a mechanism to control its involvement in synaptic plasticity and homeostasis? *Front. Cell. Neurosci.* **8**, 1–15 (2014).
26. Gulmez Karaca, K., Brito, D. V. C. & Oliveira, A. M. M. MeCP2: a critical regulator of chromatin in neurodevelopment and adult brain function. *Int. J. Mol. Sci.* **20**, 4577 (2019).
27. Shahbazian, M. D., Antalffy, B., Armstrong, D. L. & Zoghbi, H. Y. Insight into Rett syndrome: MeCP2 levels display tissue-and cell-specific differences and correlate with neuronal maturation. *Hum. Mol. Genet.* **11**, 115–124 (2002).
28. Kishi, N. & Macklis, J. D. MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Mol. Cell. Neurosci.* **27**, 306–321 (2004).

29. Ross, P. D. *et al.* Exclusive expression of MeCP2 in the nervous system distinguishes between brain and peripheral Rett syndrome-like phenotypes. *Hum. Mol. Genet.* **25**, 4389–4404 (2016).
30. Hagberg, B., Aicardi J, Dias, K. & Ramos, O. 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 (1983).
31. World Health Organization. International statistical classification of diseases and related health problems 10th revision (ICD-10). *Geneva: WHO* vol. 10 (1992).
32. Association, A. P. Diagnostic and statistical manual of mental disorders (4th ed.). in *Barber, J. P., Connolly, M. B., Crits-Christoph, P., Gladis, L., & Siqueland, L* vol. 68 1027–1032 (1994).
33. Hagberg, B., Hanefeld, F., Percy, A. & Skjeldal, O. 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. in *European Journal of Paediatric Neurology* vol. 6 293–297 (W.B. Saunders Ltd, 2002).
34. Neul, J. L. *et al.* Rett syndrome: revised diagnostic criteria and nomenclature. *Ann. Neurol.* **68**, 944–950 (2010).
35. Kirby, R. S. *et al.* Longevity in Rett syndrome: analysis of the North American database. *J. Pediatr.* **156**, 135–138 (2010).
36. Tarquinio, D. C. *et al.* The changing face of survival in Rett syndrome and MECP2-related disorders. *Pediatr. Neurol.* **53**, 402–411 (2015).
37. Anderson, A., Wong, K., Jacoby, P., Downs, J. & Leonard, H. Twenty years of surveillance in Rett syndrome: what does this tell us? *Orphanet J. Rare Dis.* **9**, 87 (2014).
38. Singh, J., Lanzarini, E. & Santosh, P. Autonomic dysfunction and sudden death in patients with Rett syndrome: a systematic review. *J. Psychiatry Neurosci.* **45**, 150–181 (2020).
39. Monrós, E. *et al.* Rett syndrome in Spain: mutation analysis and clinical correlations. *Brain Dev.* **23**, S251–S253 (2001).
40. Kerr, A. M. *et al.* Guidelines for reporting clinical features in cases with MECP2 mutations. *Brain Dev.* **23**, 208–211 (2001).
41. Colvin, L. *et al.* Describing the phenotype in Rett syndrome using a population database. *Arch. Dis. Child.* **88**, 38–43 (2003).
42. Cosentino, L., Vigli, D., Franchi, F., Laviola, G. & De Filippis, B. Rett syndrome before regression: a time window of overlooked opportunities for diagnosis and intervention. *Neurosci. Biobehav. Rev.* **107**, 115–135 (2019).

43. Collins, B. E. & Neul, J. L. Rett syndrome and MECP2 Duplication syndrome: disorders of MeCP2 dosage. *Neuropsychiatr. Dis. Treat.* **18**, 2813–2835 (2022).
44. Henriksen, M. W. *et al.* Epilepsy in classic Rett syndrome: course and characteristics in adult age. *Epilepsy Res.* **145**, 134–139 (2018).
45. Leonard, H. *et al.* Valproate and risk of fracture in Rett syndrome. *Arch. Dis. Child.* **95**, 444–448 (2010).
46. Kyle, S. M., Vashi, N. & Justice, M. J. Rett syndrome: a neurological disorder with metabolic components. *Open Biol.* **8**, 170216 (2018).
47. Fehr, S. *et al.* The CDKL5 disorder is an independent clinical entity associated with early-onset encephalopathy. *Eur. J. Hum. Genet.* **21**, 266–273 (2013).
48. Kortüm, F. *et al.* The core FOXG1 syndrome phenotype consists of postnatal microcephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis. *J. Med. Genet.* **48**, 396–406 (2011).
49. Vidal, S. *et al.* Genetic landscape of Rett syndrome spectrum: improvements and challenges. *Int. J. Mol. Sci.* **20**, 3925 (2019).
50. Ehrhart, F., Sangani, N. B. & Curfs, L. M. G. Current developments in the genetics of Rett and Rett-like syndrome. *Curr. Opin. Psychiatry* **31**, 103–108 (2018).
51. Neul, J. L. *et al.* Developmental delay in Rett syndrome: data from the natural history study. *J. Neurodev. Disord.* **6**, 20 (2014).
52. Wan, M. *et al.* Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. *Am. J. Hum. Genet* **65**, 1520–1529 (1999).
53. Hardwick, S. A. *et al.* Delineation of large deletions of the MECP2 gene in Rett syndrome patients, including a familial case with a male proband. *Eur. J. Hum. Genet.* **15**, 1218–1229 (2007).
54. Laccone, F. *et al.* Large deletions of the MECP2 gene detected by gene dosage analysis in patients with Rett syndrome. *Hum. Mutat.* **23**, 234–244 (2004).
55. Armstrong, J. *et al.* Classic Rett syndrome in a boy as a result of somatic mosaicism for a MECP2 mutation. *Wiley-Liss* (2001).
56. Topçu, M. *et al.* Somatic mosaicism for a MECP2 mutation associated with classic Rett syndrome in a boy. *Eur. J. Hum. Genet.* **10**, 77–81 (2002).
57. Huppke, P. & Gärtner, J. Molecular diagnosis of Rett syndrome. *J. Child Neurol.* **20**, 732–736 (2005).
58. Villard, L. MECP2 mutations in males. *J. Med. Genet.* **44**, 417–423 (2007).
59. Xiol, C. *et al.* X chromosome inactivation does not necessarily determine the severity of the phenotype in Rett syndrome patients. *Sci. Rep.* **9**, 11983 (2019).

60. Cuddapah, V. A. *et al.* Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome. *J. Med. Genet.* **51**, 152–158 (2014).
61. Shovlin, S. & Tropea, D. Transcriptome level analysis in Rett syndrome using human samples from different tissues. *Orphanet J. Rare Dis.* **13**, 113 (2018).
62. Ravn, K. *et al.* Two new Rett syndrome families and review of the literature: expanding the knowledge of MECP2 frameshift mutations. *Orphanet J. Rare Dis.* **6**, 58 (2011).
63. Enikanolaiye, A. *et al.* Suppressor mutations in Mecp2-null mice implicate the DNA damage response in Rett syndrome pathology. *Genome Res.* **30**, 540–552 (2020).
64. Downs, J. *et al.* Environmental enrichment intervention for Rett syndrome: an individually randomised stepped wedge trial. *Orphanet J. Rare Dis.* **13**, 3 (2018).
65. Achilly, N. P., Wang, W. & Zoghbi, H. Y. Presymptomatic training mitigates functional deficits in a mouse model of Rett syndrome. *Nature* **592**, 596–600 (2021).
66. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **27**, 322–326 (2001).
67. Chen, R. Z., Akbarian, S., Tudor, M. & Jaenisch, R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.* **27**, 327–331 (2001).
68. McGraw, C. M., Samaco, R. C. & Zoghbi, H. Y. Adult neural function requires MeCP2. *Science (80-.).* **333**, 186 (2011).
69. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of neurological defects in a mouse model of Rett syndrome. *Science (80-.).* **315**, 1143–1147 (2007).
70. Robinson, L. *et al.* Morphological and functional reversal of phenotypes in a mouse model of Rett syndrome. *Brain* **135**, 2699–2710 (2012).
71. Lioy, D. T. *et al.* A role for glia in the progression of Rett-syndrome. *Nature* **475**, 497–500 (2011).
72. Jin, X. R., Chen, X. S. & Xiao, L. MeCP2 deficiency in neuroglia: new progress in the pathogenesis of Rett syndrome. *Front. Mol. Neurosci.* **10**, 316 (2017).
73. Tang, X. *et al.* KCC2 rescues functional deficits in human neurons derived from patients with Rett syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 751–756 (2016).
74. Tang, X. *et al.* Pharmacological enhancement of KCC2 gene expression exerts therapeutic effects on human Rett syndrome neurons and Mecp2 mutant mice. *Sci. Transl. Med* **11**, (2019).

75. Kim, J. J. *et al.* Proteomic analyses reveal misregulation of LIN28 expression and delayed timing of glial differentiation in human iPS cells with MECP2 loss-of-function. *PLoS One* **14**, (2019).
76. Gomes, A. R., Fernandes, T. G., Cabral, J. M. S. & Diogo, M. M. Modeling Rett syndrome with human pluripotent stem cells: mechanistic outcomes and future clinical perspectives. *Int. J. Mol. Sci.* **22**, (2021).
77. Trujillo, C. A. *et al.* Pharmacological reversal of synaptic and network pathology in human MECP2 -KO neurons and cortical organoids. *EMBO Mol. Med.* **13**, e12523 (2021).
78. Asaka, Y., Jugloff, D. G. M., Zhang, L., Eubanks, J. H. & Fitzsimonds, R. M. Hippocampal synaptic plasticity is impaired in the Mecp2-null mouse model of Rett syndrome. *Neurobiol. Dis.* **21**, 217–227 (2006).
79. Degano, A. L., Pasterkamp, R. J. & Ronnett, G. V. MeCP2 deficiency disrupts axonal guidance, fasciculation, and targeting by altering Semaphorin 3F function. *Mol. Cell. Neurosci.* **42**, 243–254 (2009).
80. Lombardi, L. M., Baker, S. A. & Zoghbi, H. Y. MECP2 disorders: from the clinic to mice and back. *J. Clin. Invest.* **125**, 2914–2923 (2015).
81. Banerjee, A., Miller, M. T., Li, K., Sur, M. & Kaufmann, W. E. Towards a better diagnosis and treatment of Rett syndrome: a model synaptic disorder. *Brain* **142**, 239–248 (2019).
82. Mouro, F. M., Miranda-Lourenço, C., Sebastião, A. M. & Diógenes, M. J. From cannabinoids and neurosteroids to statins and the ketogenic diet: new therapeutic avenues in Rett syndrome? *Front. Neurosci.* **13**, 680 (2019).
83. Li, Y. *et al.* Global transcriptional and translational repression in human-embryonic- stem-cell-derived rett syndrome neurons. *Cell Stem Cell* **13**, 446–458 (2013).
84. O'Leary, H. M. *et al.* Placebo-controlled crossover assessment of mecasermin for the treatment of Rett syndrome. *Ann. Clin. Transl. Neurol.* **5**, 323–332 (2018).
85. Glaze, D. G. *et al.* Double-blind, randomized, placebo-controlled study of trofinetide in pediatric Rett syndrome. *Neurology* **92**, E1912–E1925 (2019).
86. Neul, J. L. *et al.* Design and outcome measures of LAVENDER, a phase 3 study of trofinetide for Rett syndrome. *Contemp. Clin. Trials* **114**, 106704 (2022).
87. Li, W. & Pozzo-Miller, L. BDNF deregulation in Rett syndrome. *Neuropharmacology* **76**, (2014).
88. Naegelin, Y. *et al.* Fingolimod in children with Rett syndrome: the FINGORETT study. *Orphanet J. Rare Dis.* **16**, 19 (2021).

89. Djukic, A. *et al.* Pharmacologic treatment of Rett syndrome with glatiramer acetate. *Pediatr. Neurol.* **61**, 51–57 (2016).
90. Nissenkorn, A., Kidon, M. & Ben-Zeev, B. A potential life-threatening reaction to glatiramer acetate in Rett syndrome. *Pediatr. Neurol.* **68**, 40–43 (2017).
91. Hurley, E. N. *et al.* Efficacy and safety of cannabidivarin treatment of epilepsy in girls with Rett syndrome: a phase 1 clinical trial. *Epilepsia* **63**, 1736–1747 (2022).
92. Garg, S. K. *et al.* Systemic delivery of MeCP2 rescues behavioral and cellular deficits in female mouse models of Rett syndrome. *J. Neurosci.* **33**, 13612–13620 (2013).
93. Grimm, N. B. & Lee, J. T. Selective Xi reactivation and alternative methods to restore MECP2 function in Rett syndrome. *Trends Genet.* **38**, 920–943 (2022).
94. Huong Le, T. T. *et al.* Efficient and precise CRISPR/Cas9-mediated MECP2 modifications in human-induced pluripotent stem cells. *Front. Genet.* **10**, 625 (2019).
95. Przanowski, P. *et al.* Pharmacological reactivation of inactive X-linked Mecp2 in cerebral cortical neurons of living mice. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7991–7996 (2018).
96. Sinnamon, J. R. *et al.* In vivo repair of a protein underlying a neurological disorder by programmable RNA editing. *Cell Rep.* **32**, 107878 (2020).
97. Psoni, S. *et al.* Phenotypic and genotypic variability in four males with MECP2 gene sequence aberrations including a novel deletion. *Pediatr. Res.* **67**, 551–556 (2010).
98. Ravn, K., Nielsen, J. B., Uldall, P. & Hansen, F. J. No correlation between phenotype and genotype in boys with a truncating MECP2 mutation. *J Med Genet* **40**, e5 (2003).
99. Neul, J. L. *et al.* The array of clinical phenotypes of males with mutations in Methyl-CpG binding protein 2. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **180**, 55–67 (2019).
100. Sanfeliu, A., Hokamp, K., Gill, M. & Tropea, D. Transcriptomic analysis of Mecp2 mutant mice reveals differentially expressed genes and altered mechanisms in both blood and brain. *Front. Psychiatry* **10**, 278 (2019).
101. Lugtenberg, D. *et al.* Structural variation in Xq28: MECP2 duplications in 1% of patients with unexplained XLMR and in 2% of male patients with severe encephalopathy. *Eur. J. Hum. Genet.* **17**, 444–453 (2009).

102. Honda, S. *et al.* The incidence of hypoplasia of the corpus callosum in patients with dup (X)(q28) involving MECP2 is associated with the location of distal breakpoints. *Am. J. Med. Genet. Part A* **158A**, 1292–1303 (2012).

103. Van Esch, H. MECP2 duplication syndrome. *Mol. Syndromol.* **2**, 128–136 (2011).

104. Miguet, M. *et al.* Further delineation of the MECP2 duplication syndrome phenotype in 59 French male patients, with a particular focus on morphological and neurological features. *J. Med. Genet.* **55**, 359–371 (2018).

105. Giudice-Nairn, P. *et al.* The incidence, prevalence and clinical features of MECP2 duplication syndrome in Australian children. *J. Paediatr. Child Health* **55**, 1315–1322 (2019).

106. Ta, D. *et al.* Medical comorbidities in MECP2 duplication syndrome: results from the international MECP2 duplication database. *Children* **9**, 633 (2022).

107. Ak, M. *et al.* Exploring the characteristics and most bothersome symptoms in MECP2 duplication syndrome to pave the path toward developing parent-oriented outcome measures. *Mol. Genet. Genomic Med.* **10**, e1989 (2022).

108. Bauer, M. *et al.* Infectious and immunologic phenotype of MECP2 duplication syndrome. *J. Clin. Immunol.* **35**, 168–181 (2015).

109. Bauer, M. *et al.* Antibiotic prophylaxis, immunoglobulin substitution and supportive measures prevent infections in MECP2 duplication syndrome. *Pediatr. Infect. Dis. J.* **37**, 466–468 (2018).

110. Prescott, T. E., Rødningen, O. K., Bjørnstad, A. & Stray-Pedersen, A. 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 (2009).

111. Lim, Z., Downs, J., Wong, K., Ellaway, C. & Leonard, H. Expanding the clinical picture of the MECP2 duplication syndrome. *Clin. Genet.* **91**, 557–563 (2017).

112. Takeguchi, R. *et al.* Early diagnosis of MECP2 duplication syndrome: insights from a nationwide survey in Japan. *J. Neurol. Sci.* **422**, (2021).

113. Peters, S. U. *et al.* Phenotypic features in MECP2 duplication syndrome: effects of age. *Am. J. Med. Genet. Part A* **185A**, 362–369 (2021).

114. Ramocki, M. B., Tavyev, Y. J. & Peters, S. U. The MECP2 duplication syndrome. *Am. J. Med. Genet. Part A* **152A**, 1079–1088 (2010).

115. Sanmann, J. N. *et al.* Characterization of six novel patients with MECP2 duplications due to unbalanced rearrangements of the X chromosome. *Am. J. Med. Genet. Part A* **158A**, 1285–1291 (2012).

116. Shimada, S. *et al.* Clinical manifestations of Xq28 functional disomy involving MECP2 in one female and two male patients. *Am. J. Med. Genet. Part A* **161A**, 1779–1785 (2013).
117. Yi, Z. *et al.* Chromosome Xq28 duplication encompassing MECP2: clinical and molecular analysis of 16 new patients from 10 families in China. *Eur. J. Med. Genet.* **59**, 347–353 (2016).
118. Bauters, M. *et al.* Nonrecurrent MECP2 duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair. *Genome Res.* **18**, 847–858 (2008).
119. Del Gaudio, D. *et al.* Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet. Med.* **8**, 784–792 (2006).
120. Carvalho, C. M. B. *et al.* Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat. Genet.* **43**, 1074–1081 (2011).
121. Tang, S. S., Fernandez, D., Lazarou, L. P., Singh, R. & Fallon, P. MECP2 triplication in 3 brothers - A rarely described cause of familial neurological regression in boys. *Eur. J. Paediatr. Neurol.* **16**, 209–212 (2012).
122. Wax, J. R., Pinette, M. G., Smith, R., Chard, R. & Cartin, A. Second-trimester prenasal and prefrontal skin thickening-Association with MECP2 triplication syndrome. *J. Clin. Ultrasound* **41**, 434–437 (2013).
123. Signorini, C. *et al.* MECP2 duplication syndrome: evidence of enhanced oxidative stress. A comparison with Rett syndrome. *PLoS One* **11**, (2016).
124. Smyk, M. *et al.* Different-sized duplications of Xq28, including MECP2, in three males with mental retardation, absent or delayed speech, and recurrent infections. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **147B**, 799–806 (2008).
125. Mayo, S. *et al.* De novo interstitial triplication of MECP2 in a girl with neurodevelopmental disorder and random X chromosome inactivation. *Cytogenet. Genome Res.* **135**, 93–101 (2011).
126. Peters, S. U. *et al.* The behavioral phenotype in MECP2 duplication syndrome: a comparison with idiopathic autism. *Autism Res.* **6**, 42–50 (2013).
127. Bijlsma, E. K. *et al.* Xq28 duplications including MECP2 in five females: expanding the phenotype to severe mental retardation. *Eur. J. Med. Genet.* **55**, 404–413 (2012).
128. Gottipati, S., Rao, N. L. & Fung-Leung, W. P. IRAK1: a critical signaling mediator of innate immunity. *Cell. Signal.* **20**, 269–276 (2008).

129. Clayton-Smith, J. *et al.* Xq28 duplication presenting with intestinal and bladder dysfunction and a distinctive facial appearance. *Eur. J. Hum. Genet.* **17**, 434–443 (2009).

130. El Chehadeh, S. *et al.* Large national series of patients with Xq28 duplication involving MECP2: delineation of brain MRI abnormalities in 30 affected patients. *Am. J. Med. Genet. Part A* **170A**, 116–129 (2016).

131. Weller, S. & Gärtner, J. Genetic and clinical aspects of X-linked hydrocephalus (L1 disease): mutations in the L1CAM gene. *Hum. Mutat.* **18**, 1–12 (2001).

132. Peters, S. U. *et al.* Characterizing the phenotypic effect of Xq28 duplication size in MECP2 duplication syndrome. *Clin. Genet.* **95**, 575–581 (2019).

133. Pitzianti, M. B., Palombo, A. S., Esposito, S. & Pasini, A. Rett syndrome in males: the different clinical course in two brothers with the same microduplication MECP2 Xq28. *Int. J. Environ. Res. Public Health* **16**, (2019).

134. Collins, A. L. *et al.* Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* **13**, 2679–2689 (2004).

135. Shao, Y. *et al.* Antisense oligonucleotide therapy in a humanized mouse model of MECP2 duplication syndrome. *Sci. Transl. Med* **13**, 7785 (2021).

136. Na, E. S. *et al.* A mouse model for MeCP2 duplication syndrome: MeCP2 overexpression impairs learning and memory and synaptic transmission. *J. Neurosci.* **32**, 3109–3117 (2012).

137. Maino, E. *et al.* A Cas9-fusion proximity-based approach generates an Irak1-Mecp2 tandem duplication mouse model for the study of MeCP2 duplication syndrome. *bioRxiv* (2023) doi:10.1101/2023.02.07.527511.

138. Li, X., Nie, Y., Qiu, Z. & Wang, S. Human MECP2 transgenic rats show increased anxiety, severe social deficits, and abnormal prefrontal neural oscillation stability. *Biochem. Biophys. Res. Commun.* **648**, 28–35 (2023).

139. Liu, Z. *et al.* Autism-like behaviours and germline transmission in transgenic monkeys overexpressing MeCP2. *Nature* **530**, 98–102 (2016).

140. Nageshappa, S. *et al.* Altered neuronal network and rescue in a human MECP2 duplication model. *Mol. Psychiatry* **21**, 178–188 (2016).

141. Wojtal, D. *et al.* Spell checking nature: versatility of CRISPR/Cas9 for developing treatments for inherited disorders. *Am. J. Hum. Genet.* **98**, 90–101 (2016).

142. Jiang, M. *et al.* Dendritic arborization and spine dynamics are abnormal in the mouse model of MECP2 duplication syndrome. *J. Neurosci.* **33**, 19518–19533 (2013).

143. Ash, R. T. *et al.* Inhibition of elevated ras-mapk signaling normalizes enhanced motor learning and excessive clustered dendritic spine stabilization in the mecp2-duplication syndrome mouse model of autism. *eNeuro* **8**, (2021).
144. Sztainberg, Y. *et al.* Reversal of phenotypes in MECP2 duplication mice using genetic rescue or antisense oligonucleotides. *Nature* **528**, 123–126 (2015).
145. Tromp, T. R., Stroes, E. S. G. & Hovingh, G. K. Gene-based therapy in lipid management: the winding road from promise to practice. *Expert Opin. Investig. Drugs* **29**, 483–493 (2020).
146. Na, E. S., Morris, M. J., Nelson, E. D. & Monteggia, L. M. GABAA a receptor antagonism ameliorates behavioral and synaptic impairments associated with MeCP2 overexpression. *Neuropsychopharmacology* **39**, 1946–1954 (2014).
147. Yu, B. *et al.* Reversal of social recognition deficit in adult mice with MECP2 duplication via normalization of MeCP2 in the medial prefrontal cortex. *Neurosci. Bull.* **36**, 570–584 (2020).
148. Sun, L. *et al.* Visualization and correction of social abnormalities-associated neural ensembles in adult MECP2 duplication mice. *Sci. Bull.* **65**, 1192–1202 (2020).
149. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).
150. Rozanova, S. *et al.* Quantitative mass spectrometry-based proteomics: an overview. in *Methods in Molecular Biology* vol. 2228 85–116 (Humana Press Inc., 2021).
151. Stenton, S. L., Kremer, L. S., Kopajtich, R., Ludwig, C. & Prokisch, H. The diagnosis of inborn errors of metabolism by an integrative "multi-omics" approach: a perspective encompassing genomics, transcriptomics, and proteomics. *J. Inherit. Metab. Dis.* 1–11 (2019) doi:10.1002/jimd.12130.
152. Kelly, R. T. Single-cell proteomics: progress and prospects. *Mol. Cell. Proteomics* **19**, 1739–1748 (2020).
153. Wu, D. C., Yao, J., Ho, K. S., Lambowitz, A. M. & Wilke, C. O. Limitations of alignment-free tools in total RNA-seq quantification. *BMC Genomics* **19**, 510 (2018).
154. Chen, C., Hou, J., Tanner, J. J. & Cheng, J. Bioinformatics methods for mass spectrometry-based proteomics data analysis. *Int. J. Mol. Sci.* **21**, 2873 (2020).
155. Kammers, K., Cole, R. N., Tiengwe, C. & Ruczinski, I. Detecting significant changes in protein abundance. *EuPA Open Proteomics* **7**, 11–19 (2015).
156. Marano, D., Fioriniello, S., D'esposito, M. & Della Ragione, F. Transcriptomic and epigenomic landscape in rett syndrome. *Biomolecules* **11**, 967 (2021).

157. Fortin, D. A., Srivastava, T. & Soderling, T. R. Structural modulation of dendritic spines during synaptic plasticity. *Neuroscientist* **18**, 326–341 (2012).
158. Kapitein, L. C. & Hoogenraad, C. C. Building the neuronal microtubule cytoskeleton. *Neuron* **87**, 492–506 (2015).
159. Krishnaraj, R. *et al.* Genome-wide transcriptomic and proteomic studies of Rett syndrome mouse models identify common signaling pathways and cellular functions as potential therapeutic targets. *Hum. Mutat.* **40**, 2184–2196 (2019).
160. Marballi, K. & MacDonald, J. L. Proteomic and transcriptional changes associated with MeCP2 dysfunction reveal nodes for therapeutic intervention in Rett syndrome. *Neurochem. Int.* **148**, 105076 (2021).
161. Cicaloni, V. *et al.* Proteomic profiling reveals mitochondrial alterations in Rett syndrome. *Free Radic. Biol. Med.* **155**, 37–48 (2020).
162. Segatto, M. *et al.* Cholesterol metabolism is altered in Rett syndrome: a study on plasma and primary cultured fibroblasts derived from patients. *PLoS One* **9**, e104834 (2014).
163. Varderidou-Minasian, S. *et al.* Quantitative proteomic analysis of Rett iPSC-derived neuronal progenitors. *Mol. Autism* **11**, 38 (2020).
164. Cortelazzo, A. *et al.* Beta-actin deficiency with oxidative posttranslational modifications in Rett syndrome erythrocytes: insights into an altered cytoskeletal organization. *PLoS One* **9**, (2014).
165. Carvalho, C. M. B. B. *et al.* Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching. *Hum. Mol. Genet.* **18**, 2188–2203 (2009).
166. Trostle, A. J. *et al.* A comprehensive and integrative approach to MeCP2 disease transcriptomics. *Int. J. Mol. Sci.* **24**, 5122 (2023).
167. Ehrhart, F. *et al.* Integrated analysis of human transcriptome data for Rett syndrome finds a network of involved genes. *World J. Biol. Psychiatry* **21**, 712–725 (2019).
168. Murdock, D. R. *et al.* Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *J. Clin. Invest.* **131**, e141500 (2021).
169. Yépez, V. A. *et al.* Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med.* **14**, 38 (2022).
170. Montgomery, S. B., Bernstein, J. A. & Wheeler, M. T. Toward transcriptomics as a primary tool for rare disease investigation. *Cold Spring Harb. Mol. Case Stud.* **8**, (2022).

171. Pecorelli, A. *et al.* Alteration of serum lipid profile, SRB1 loss, and impaired Nrf2 activation in CDKL5 disorder. *Free Radic. Biol. Med.* **86**, 156–165 (2015).
172. Pecorelli, A. *et al.* Proteomic analysis of 4-hydroxynonenal and nitrotyrosine modified proteins in RTT fibroblasts. *Int. J. Biochem. Cell Biol.* **81**, 236–245 (2016).
173. Sbardella, D. *et al.* Defective proteasome biogenesis into skin fibroblasts isolated from Rett syndrome subjects with MeCP2 non-sense mutations. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1866**, 165793 (2020).
174. Sticozzi, C. *et al.* Scavenger receptor B1 post-translational modifications in Rett syndrome. *FEBS Lett.* **587**, 2199–2204 (2013).
175. Gold, W. A., Lacina, T. A., Cantrill, L. C. & Christodoulou, J. MeCP2 deficiency is associated with reduced levels of tubulin acetylation and can be restored using HDAC6 inhibitors. *J. Mol. Med.* **93**, 63–72 (2015).
176. Pacheco, N. L. *et al.* RNA sequencing and proteomics approaches reveal novel deficits in the cortex of Mecp2-deficient mice, a model for Rett syndrome. *Mol. Autism* **8**, (2017).
177. Belichenko, P. V. *et al.* Widespread changes in dendritic and axonal morphology in Mecp2-mutant mouse models of Rett syndrome: evidence for disruption of neuronal networks. *J. Comp. Neurol.* **514**, 240–258 (2009).
178. Ronnett, G. V. *et al.* Olfactory biopsies demonstrate a defect in neuronal development in Rett's syndrome. *Ann. Neurol.* **54**, 206–218 (2003).
179. Matarazzo, V. *et al.* The transcriptional repressor Mecp2 regulates terminal neuronal differentiation. *Mol. Cell. Neurosci.* **27**, 44–58 (2004).
180. Matarazzo, V. & Ronnett, G. V. Temporal and regional differences in the olfactory proteome as a consequence of MeCP2 deficiency. *PNAS* **101**, 7763–7768 (2004).
181. Kishi, N. & Macklis, J. D. Dissecting MECP2 function in the central nervous system. *J. Child Neurol.* **20**, 753–759 (2005).
182. Zhang, B. *et al.* The microtubule-stabilizing agent, epothilone D, reduces axonal dysfunction, neurotoxicity, cognitive deficits, and Alzheimer-like pathology in an interventional study with aged tau transgenic mice. *J. Neurosci.* **32**, 3601–3611 (2012).
183. Latour, B. L. *et al.* Dysfunction of the ciliary ARMC9/TOGARAM1 protein module causes Joubert syndrome. *J. Clin. Invest.* **140**, 4423–4439 (2020).
184. Zhang, H., Webb, D. J., Asmussen, H., Niu, S. & Horwitz, A. F. A GIT1/PIX/Rac/PAK signaling module regulates spine morphogenesis and synapse formation through MLC. *J. Neurosci.* **25**, 3379–3388 (2005).

185. Mukherjee, K. *et al.* CASK functions as a Mg2+-independent neurexin kinase. *Cell* **133**, 328–339 (2008).

186. Setou, M., Nakagawa, T., Seog, D.-H. & Hirokawa, N. Kinesin superfamily motor protein KIF17 and mLin-10 in NMDA receptor–containing vesicle transport. *Science (80-.).* **288**, 1796–1802 (2000).

187. Patel, P. A. *et al.* Complete loss of the X-linked gene CASK causes severe cerebellar degeneration. *J. Med. Genet.* **59**, 1044–1057 (2022).

188. Chen, J. *et al.* Orientation and cellular distribution of membrane-bound catechol-O-methyltransferase in cortical neurons: implications for drug development. *J. Biol. Chem.* **286**, 34752–34760 (2011).

189. Szczesna, K. *et al.* Improvement of the Rett syndrome phenotype in a Mecp2 mouse model upon treatment with levodopa and a dopa-decarboxylase inhibitor. *Neuropsychopharmacology* **39**, 2846–2856 (2014).

190. Ide, S., Itoh, M. & Goto, Y. I. Defect in normal developmental increase of the brain biogenic amine concentrations in the mecp2-null mouse. *Neurosci. Lett.* **386**, 14–17 (2005).

191. Radhakrishnan, K., Baltes, J., Creemers, J. W. M. & Schu, P. Trans-Golgi network morphology and sorting is regulated by prolyl-oligopeptidase-like protein PREPL and the AP-1 complex subunit μ1A. *J. Cell Sci.* **126**, 1155–1163 (2013).

192. Régal, L. *et al.* PREPL deficiency with or without cystinuria causes a novel myasthenic syndrome. *Neurology* **82**, 1254–1260 (2014).

193. Hirst, J. *et al.* Distinct and overlapping roles for AP-1 and GGAs revealed by the 'knocksideways' system. *Curr. Biol.* **22**, 1711–1716 (2012).

194. Kim, M. H. & Hersh, L. B. The vesicular acetylcholine transporter interacts with clathrin-associated adaptor complexes AP-1 and AP-2. *J. Biol. Chem.* **279**, 12580–12587 (2004).

195. Ricciardi, S. *et al.* Reduced AKT/mTOR signaling and protein synthesis dysregulation in a Rett syndrome animal model. *Hum. Mol. Genet.* **20**, 1182–1196 (2011).

196. Thoreen, C. C. *et al.* A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109–113 (2012).

197. Rodrigues, D. C. *et al.* Shifts in ribosome engagement impact key gene sets in neurodevelopment and ubiquitination in Rett syndrome. *Cell Rep.* **30**, 4179–4196 (2020).

198. Stephan, A. H., Barres, B. A. & Stevens, B. The complement system: an unexpected role in synaptic pruning during development and disease. *Annu. Rev. Neurosci.* **35**, 369–389 (2012).

199. Schartz, N. D. & Tenner, A. J. The good, the bad, and the opportunities of the complement system in neurodegenerative disease. *J. Neuroinflammation* **17**, 354 (2020).

200. Ziabska, K., Ziemka-Nalecz, M., Pawelec, P., Sypecka, J. & Zalewska, T. Aberrant complement system activation in neurological disorders. *Int. J. Mol. Sci.* **22**, 4675 (2021).

201. Pecorelli, A., Cervellati, C., Cordone, V., Hayek, J. & Valacchi, G. Compromised immune/inflammatory responses in Rett syndrome. *Free Radic. Biol. Med.* **152**, 100–106 (2020).

202. Fiumara, A. *et al.* Peripheral lymphocyte subsets and other immune aspects in Rett syndrome. *Pediatr. Neurol.* **21**, 619–621 (1999).

203. Gold, W. A. *et al.* Mitochondrial dysfunction in the skeletal muscle of a mouse model of Rett syndrome (RTT): implications for the disease phenotype. *Mitochondrion* **15**, 10–17 (2014).

204. Pecorelli, A. *et al.* Genes related to mitochondrial functions, protein degradation, and chromatin folding are differentially expressed in lymphomonocytes of Rett syndrome patients. *Mediators Inflamm.* **2013**, (2013).

205. Filosa, S., Pecorelli, A., D'Espositoc, M., Valacchi, G. & Hajek, J. Exploring the possible link between MeCP2 and oxidative stress in Rett syndrome. *Free Radic. Biol. Med.* **88**, 81–90 (2015).

206. Can, K. *et al.* Neuronal redox-imbalance in Rett syndrome affects mitochondria as well as cytosol, and is accompanied by intensified mitochondrial O2 consumption and ROS release. *Front. Physiol.* **10**, 1–18 (2019).

207. Aldosary, M. *et al.* Rett syndrome, a neurodevelopmental disorder, whole-transcriptome, and mitochondrial genome multiomics analyses identify novel variations and disease pathways. *Omi. A J. Integr. Biol.* **24**, 160–171 (2020).

208. Shulyakova, N., Andreazza, A. C., Mills, L. R. & Eubanks, J. H. Mitochondrial dysfunction in the pathogenesis of Rett syndrome: implications for mitochondria-targeted therapies. *Front. Cell. Neurosci.* **11**, 58 (2017).

209. Rojas-Charry, L., Nardi, L., Methner, A. & Schmeisser, M. J. Abnormalities of synaptic mitochondria in autism spectrum disorder and related neurodevelopmental disorders. *J. Mol. Med.* **99**, 161–178 (2021).

210. Gottschalk, I. *et al.* IRAK1 duplication in MECP2 duplication syndrome does not increase canonical NF-κB–induced inflammation. *J. Clin. Immunol.* 1–19 (2022) doi:10.1007/s10875-022-01390-7.

211. Brasil, S. *et al.* Artificial Intelligence (AI) in rare diseases : is the future brighter? *Genes (Basel).* 978 (2019) doi:10.3390/genes10120978.

212. Alsabban, A. H., Morikawa, M., Tanaka, Y., Takei, Y. & Hirokawa, N. Kinesin Kif3b mutation reduces NMDAR subunit NR 2A trafficking and causes schizophrenia-like phenotypes in mice . *EMBO J.* **39**, (2020).

213. Joseph, N. F., Grinman, E., Swarnkar, S. & Puthanveettil, S. V. Molecular motor KIF3B acts as a key regulator of dendritic architecture in cortical neurons. *Front. Cell. Neurosci.* **14**, (2020).

214. Huang, H. *et al.* Recognition of RNA N 6 -methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat. Cell Biol.* **20**, 285–295 (2018).

215. Núñez, L. *et al.* Tagged actin mRNA dysregulation in IGF2BP1 −/− mice. *PNAS* **119**, (2022).

216. Yang, T. *et al.* Overexpression of methyl-CpG binding protein 2 impairs TH1 responses. *Sci. Transl. Med.* **4**, (2012).

217. Shichiri, M. The role of lipid peroxidation in neurological disorders. *J. Clin. Biochem. Nutr* **54**, 151–160 (2014).

218. Buist, M. *et al.* Differential sensitivity of the protein translation initiation machinery and mTOR Signaling to MECP2 gain-and loss-of-function involves MeCP2 isoform-specific homeostasis in the brain. *Cells* **11**, (2022).

219. Ramocki, M. B. *et al.* Autism and other neuropsychiatric symptoms are prevalent in individuals with MECP2 duplication syndrome. *Ann. Neurol.* **66**, 771–782 (2009).

220. Pascual-Alonso, A. *et al.* Molecular characterization of Spanish patients with MECP2 duplication syndrome. *Clin. Genet.* **97**, 610–620 (2020).

221. Tanaka, Y. *et al.* Transcriptional regulation in pluripotent stem cells by methyl CpG-binding protein 2 (MeCP2). *Hum. Mol. Genet.* **23**, 1045–1055 (2014).

222. Schurch, N. J. *et al.* How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *Rna* **22**, 839–851 (2016).

223. Tudor, M., Akbarian, S., Chen, R. Z. & Jaenisch, R. Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *PNAS* **99**, 15536–15541 (2002).

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

225. Cortelazzo, A. *et al.* Subclinical inflammatory status in Rett syndrome. *Mediators Inflamm.* **2014**, (2014).

226. Samaco, R. C. *et al.* Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2 duplication syndrome. *Nat. Genet.* **44**, 206–221 (2012).

227. Orlic-Milacic, M. *et al.* Over-expression of either MECP2-e1 or MECP2-e2 in neuronally differentiated cells results in different patterns of gene expression. *PLoS One* **9**, e91742 (2014).

228. Chen, L. *et al.* MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5509–5514 (2015).

229. Sun, Y. *et al.* Lack of MECP2 gene transcription on the duplicated alleles of two related asymptomatic females with Xq28 duplications and opposite X-chromosome inactivation skewing. *Hum. Mutat.* **42**, 1429–1442 (2021).

230. Capmany, A. *et al.* MYO1C stabilizes actin and facilitates the arrival of transport carriers at the Golgi complex. *J. Cell Sci.* **132**, jcs225029 (2019).

231. Åslund, A. *et al.* Myosin 1c: A novel regulator of glucose uptake in brown adipocytes. *Mol. Metab.* **53**, (2021).

232. Vuu, Y. M., Roberts, C. T. & Rastegar, M. MeCP2 is an epigenetic factor that links DNA methylation with brain metabolism. *Int. J. Mol. Sci.* **24**, 4218 (2023).

233. Urbinati, C. *et al.* Chronic treatment with the anti-diabetic drug metformin rescues impaired brain mitochondrial activity and selectively ameliorates defective cognitive flexibility in a female mouse model of Rett syndrome. *Neuropharmacology* **224**, 109350 (2023).

234. Miller, R. A. *et al.* Beyond pathway analysis: identification of active subnetworks in Rett syndrome. *Front. Genet.* **10**, 59 (2019).

235. Liu, J. *et al.* Wnt/β-catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduct. Target. Ther.* **7**, (2022).

236. Lowery, J. W. & Rosen, V. The BMP pathway and its inhibitors in the skeleton. *Physiol Rev* **98**, 2431–2452 (2018).

237. Finnson, K. W., Chi, Y., Bou-Gharios, G., Leask, A. & Philip, A. TGF-beta signaling in cartilage homeostasis and osteoarthritis. *Front. Biosci.* **S4**, 251–268 (2012).

238. Downs, J. *et al.* Early determinants of fractures in Rett syndrome. *Pediatrics* **121**, 540–546 (2008).

239. Caffarelli, C. *et al.* Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with bone disease severity in Rett syndrome. *BMC Med. Genet.* **21**, 1–9 (2020).

240. Pecorelli, A. *et al.* Altered bone status in Rett syndrome. *Life* **11**, 521 (2021).

241. Nakashima, H. *et al.* MeCP2 controls neural stem cell fate specification through miR-199a-mediated inhibition of BMP-Smad signaling. *Cell Rep.* **35**, (2021).

242. Lai, L. Y. S., Gracie, N. P., Gowripalan, A., Howell, L. M. & Newsome, T. P. SMAD proteins: mediators of diverse outcomes during infection. *Eur. J. Cell Biol.* **101**, 151204 (2022).

243. Pecorelli, A. *et al.* Altered inflammasome machinery as a key player in the perpetuation of Rett syndrome oxinflammation. *Redox Biol.* **28**, 101334 (2020).

244. Lundberg, E. *et al.* Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* **6**, 450 (2010).

245. Haider, S. & Pal, R. Integrated analysis of transcriptomic and proteomic data. *Curr. Genomics* **14**, 91–110 (2013).

246. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).

247. Quadrato, G. *et al.* Modulation of GABAA receptor signaling increases neurogenesis and suppresses anxiety through NFATc4. *J. Neurosci.* **34**, 8630–8645 (2014).

248. Ta, D. *et al.* A brief history of MECP2 duplication syndrome: 20-years of clinical understanding. *Orphanet J. Rare Dis.* **17**, 131 (2022).

249. Buchanan, C. B. *et al.* Anxiety-like behavior and anxiolytic treatment in the Rett syndrome natural history study. *J. Neurodev. Disord.* **14**, 1–11 (2022).

250. Edwards, G. *et al.* Prevalence of anxiety symptomatology and diagnosis in syndromic intellectual disability: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* **138**, (2022).

251. Mou, Y. *et al.* Impaired lipid metabolism in astrocytes underlies degeneration of cortical projection neurons in hereditary spastic paraplegia. *Acta Neuropathol. Commun.* **8**, (2020).

252. Li, X. *et al.* Loss of liver X receptor β in astrocytes leads to anxiety-like behaviors via regulating synaptic transmission in the medial prefrontal cortex in mice. *Mol. Psychiatry* **26**, 6380–6393 (2021).

253. Lin, J., de Rezende, V. L., de Aguiar da Costa, M., de Oliveira, J. & Gonçalves, C. L. Cholesterol metabolism pathway in autism spectrum disorder: from animal models to clinical observations. *Pharmacol. Biochem. Behav.* 173522 (2023) doi:10.1016/j.pbb.2023.173522.

254. Gomathi, M., Padmapriya, S. & Balachandar, V. Drug studies on Rett syndrome: from bench to bedside. *J. Autism Dev. Disord.* **50**, 2740–2764 (2020).

255. Shovlin, S. *et al.* Molecular signatures of response to Mecasermin in children with Rett syndrome. *Front. Neurosci.* **16**, 868008 (2022).

256. Gonorazky, H. D. *et al.* Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease. *Am. J. Hum. Genet.* **104**, 466–483 (2019).

ANNEXES

Annex 1:

Table A1: Pineda's score for classic and atypical RTT severity evaluation. Modified from Monrós et al. [39].

Feature	Score	Definition
Age of onset	3 2 1	0-12 months 12-24 months >24 months
Microcephaly	0 1	Absent Present
Sits unsupported	0 1 2 3 +1	Acquired <8 months Acquired 8-16 months Acquired >16 months Never acquired Lost acquisition
Ambulation	0 1 2 3 4	Acquired <18 months Acquired 18-30 months Acquired >30 months Lost acquisition Never acquired
Language	0 1 2	Preserved and propositive Lost Never acquired
Epilepsy	0 1 2	Absent Present and controlled Uncontrolled or early epilepsy
Respiratory function	0 1	No dysfunction Hyperventilation and/or apnea
Hand use	0 1 2 3 4	Acquired and conserved Lost purposefulness: 2-6 years Lost purposefulness: <2 years Lost all acquisitions Never acquired
Onset of stereotypies	0 1 2 3	>10 years >36 months 18-36 months <18 months

Annex 2:

Other publications generated due to collaborations with other research groups during the PhD.

Publication 1: Unravelling molecular pathways altered in *MECP2*-related syndromes, in the search for new potential avenues for therapy.

Authors: Alba-Aina Castells, Rafel Balada, Alba Tristán-Noguero, Mar O'Callaghan, Elisenda Cortès-Saladelafont, Ainhoa Pascual-Alonso, Àngels Garcia-Cazorla, Judith Armstrong and Soledad Alcántara.

Reference: Biomedicines 2021; 9, 148. https://doi.org/10.3390/biomedicines9020148

Publication 2: Analysis of the circRNA and T-UCR populations identifies convergent pathways in mouse and human models of Rett syndrome.

Authors: Edilene Siqueira, Aida Obiols-Guardia, Olga C. Jorge-Torres, Cristina Oliveira-Mateos, Marta Soler, Deepthi Ramesh-Kumar, Fernando Setién, Daniëlle van Rossum, Ainhoa Pascual-Alonso, Clara Xiol, Cristina Ivan, Masayoshi Shimizu, Judith Armstrong, George A. Calin, R. Jeroen Pasterkamp, Manel Esteller and Sonia Guil.

Reference: Molecular Therapy: Nucleic Acids 2021; 27:621-644. https://doi.org/10.1016/j.omtn.2021.12.030.