UNIVERSITAT DE BARCELONA

Comprehensive analysis of diagnostic approaches and molecular landscape in Rett syndrome spectrum disorders

Clara Xiol Viñas

Comprehensive analysis of diagnostic approaches and molecular landscape of Rett syndrome spectrum disorders

CLARA XIOL VIÑAS

PhD Thesis 2023

Universitat de Barcelona

Tesi doctoral

Comprehensive analysis of diagnostic approaches and molecular landscape in Rett syndrome spectrum disorders

Memòria presentada per **Clara Xiol Viñas**

Per optar al grau de Doctora per la Universitat de Barcelona

Programa de Doctorat en Genètica

Tesi realitzada a l'Institut de Recerca Sant Joan de Déu

Directors

Judith Armstrong Morón i Alfonso Luis de Oyarzábal Sanz

Tutora

Raquel Rabionet Janssen

Doctoranda

Clara Xiol Viñas

Barcelona, octubre de 2023

Agraïments

Després de més de set anys implicada en aquest gran projecte, que va començar com un treball de final de grau i va anar evolucionant a tesi, estic a punt d'acabar (o això em diuen, perquè sembla que no s'acabi mai, això!). Arribats a aquest punt, només em queda donar les gràcies a totes les persones que m'han acompanyat en aquest camí i que han fet possible d'una manera o altra que aquest treball arribés a bon port.

En primer lloc, gràcies als pares, mares i pacients de les associacions catalana i espanyola de la síndrome de Rett, per la seva sempre incansable implicació en tots els projectes que hem dut a terme. Sou el motor que ens fa voler avançar una mica cada dia. Espero que aquesta tesi serveixi de trampolí cap a nous projectes que encara estan per venir.

Als meus directors de tesi, que sent ben diferents m'han aportat sempre el millor de cadascun. Judith, gràcies per ensenyar-me cada dia que l'èxit és la suma dels intents. Per deixar-me equivocar-me i aprendre dels errors i dels èxits (esperats i inesperats), i per desbloquejar-me quan no era capaç d'avançar sola. Alfonso, gracias por enseñarme a confiar más en mí. Por tu entusiasmo, por ayudarme a ver las cosas desde otra perspectiva y por estar siempre operativo para hacer fácil lo difícil.

A tots els membres del laboratori de genètica i del grup de recerca en síndrome de Rett, que he vist evolucionar a *Genòmica pel diagnòstic de malalties rares*. A les del principi, que em van ensenyar bona part del que sé: Silvia, Nuri, Paola i Laura, com hem avançat des del 2016! Em vau acollir i des de llavors hem crescut, hem après, ens hem traslladat... i el nou lab ja se'ns ha quedat petit! Als que han compartit el camí: Ainhoa, Uliana, Gonzalo, Carlota, María i Noelia, han estat uns anys ben intensos, gràcies per compartir els alts i els baixos! I a les noves (i joves!) adquisicions: Marta, David i Nídia, i també especialment al Roger: no se m'acut ningú millor a qui passar-li el testimoni d'aquest projecte. *Calienta, ¡que sales!*

And of course, these acknowledgements would not be complete without mentioning all the wonderful people in Prokisch Lab. Aiman, Beryll, Catalina, Caterina, Christiane, Dewi, Dima, Fatemeh, Masaru, Nikita, Riccardo and Robert. I will always

feel Munich as my second home, and that is thanks to you. Thank you for those awesome four months! And thank you very much, Holger, for sharing as much of your knowledge as of your kindness. I will be forever grateful to you all.

Es mereix també una menció especial el Guerau, el tercer director a l'ombra. Gràcies per endinsar-me al món de les pantalles negres, per confiar sempre en mi i ser un suport incondicional. També vull agrair a la Laura i la Dèlia per acollir-me tan bé i per fer-me sentir una més durant un últim any de tesi tan intens. I a en Joe: sense el *tech support* aquesta tesi no s'hagués acabat mai. Em temo que et deuré permanentment un cafè i un dònut.

I no podien faltar els *et al.* que no surten als articles:

Els biòlegs: Ana, David, Elisa, Laia, Marc, Marta, Nico, Ramon, Ricard i Taïme. Fa 10 anys començàvem aquest camí junts i m'heu fet gaudir-lo cada dia. No m'imagino una millor companyia amb qui compartir classes (els que veníeu), cafès, biblios, *Bangs*, excursions i viatges. Sou els millors.

Les de sempre: Aida, Ari, Cris, Elena, Gemma, Ivet, Júlia, Lara, Laura, Maria B., Maria C., Marta C. i Marta P. Heu estat sempre presents per fer-me costat en aquest camí. Us prometo que ara surto de la cova, gràcies per esperar.

Al Joan i a la Mireia, que són els que més passes han caminat al meu costat: De 1998 al 2017 pràcticament compartint aula, alegries i penes. Poc ens ho pensàvem a P3 que faríem un camí tan llarg plegats!

Als meus pares i al Ferran, per ser-hi sempre. Sou pilars incondicionals. Em sento molt afortunada de tenir-vos.

Aquesta tesi també és una mica de tots vosaltres.

“It’s a dangerous business, Frodo, going out your door. You step onto the door and if you don’t keep your feet, there’s no knowing where you might be swept off to.”

– Bilbo Baggins, The Lord of The Rings (by JRR Tolkien)

Abstract

Rett syndrome (RTT) is a severe neurodevelopmental disorder characterized by a regression in acquired skills, such as purposeful hand use and language, after an apparently normal early development. RTT affects almost exclusively females and is mainly caused by mutations in the X-linked *MECP2* gene, encoding methyl-CpG-binding protein 2 (MeCP2). MeCP2 is a global regulator of gene expression that operates through different mechanisms, including transcriptional regulation, chromatin architecture, splicing modulation, and miRNA processing. Nevertheless, the precise pathomechanisms by which MeCP2 deficiency leads to RTT remain elusive.

MeCP2 plays a pivotal role in neuronal maturation and maintenance in the post-natal brain, and its deficiency causes severe defects in dendritic arborization and synaptogenesis. Currently, RTT has no cure or any effective pharmacological treatment, but the delineation of the downstream effects of MeCP2 deficiency could lead to the identification of biomarkers and potential therapeutic targets for RTT. The reversibility of RTT-like features in *Mecp2*-null mouse models upon *Mecp2* reactivation strongly suggests that symptomatic patients could benefit from counteracting the effects of MeCP2 deficiency. This doctoral thesis aims to profile the molecular landscape of RTT in different ways to contribute to the understanding of the pathomechanisms behind this disorder.

MECP2 being an X-linked gene, it has been long hypothesized that X chromosome inactivation (XCI) patterns may influence the phenotype of RTT patients. Therefore, this thesis has studied XCI patterns in blood and brain samples of RTT patients with different recurrent *MECP2* mutations to investigate their potential correlation with the severity of the clinical phenotype.

Although the main features of RTT are neurologic in nature, MeCP2 is a ubiquitously expressed protein. In this thesis, we have characterized gene expression levels in primary fibroblast cell cultures directly derived from RTT patients using an integrative multi-omics approach that combines transcriptomic and proteomic data to identify the most robust gene expression changes. We have identified an enrichment in cellular processes such as cytoskeletal activity, vesicular transport, energy

metabolism and RNA processing, with important implications for neurological phenotypes despite having studied an extraneurological tissue. Moreover, we have investigated the effects of MeCP2 deficiency on the expression of GABAergic synapse proteins, and identified a developmental stage-dependent positive regulation of their expression by MeCP2, linking GABAergic neurotransmission defects with early events in RTT pathophysiology.

With the advent of next-generation sequencing, many patients with a clinical diagnosis of RTT have been found to have mutations in genes other than *MECP2*. Understanding the relationships and interactions between these genes may help identifying common pathomechanisms leading to the overlapping phenotypes and pinpoint common therapeutic targets. In this thesis, we have used comprehensive multi-omics genomic testing to solve cases with no molecular diagnosis, and we have studied the molecular alterations in RTT-spectrum patients fibroblasts searching for common gene expression changes also found in RTT patients.

Contents

List of tables

List of figures

INTRODUCTION

1. Rett syndrome

Rett syndrome (RTT, OMIM#312750) is a severe neurodevelopmental disorder which mainly affects young females. It is considered a rare disease, but with an estimated incidence of approximately 1:10 000 girls by age 12, it is the second most common cause of intellectual disability in females, only after Down syndrome[1,2].

The first report of this clinical entity dates from 1966, when the Austrian paediatrician Andreas Rett noted the same distinctive hand stereotypies in two girls, leading to the publication of the first series of 22 patients with the same features[3,4]. In 1983, a group of paediatricians led by Bengt Hagberg further described the disorder in a larger cohort, and for the first time named the disorder “Rett syndrome”[5].

1.1. Clinical perspective

RTT is characterized by a regression in acquired skills after apparently normal early psychomotor development, particularly with the loss of purposeful hand use and language[4]. After the regression, affected females can develop a range of secondary traits such as acquired microcephaly, gait ataxia, seizures, disturbed breathing patterns with periods of hyperventilation and apnoea, and scoliosis[1].

The classical description of the natural history of RTT starts with a normal early development phase that usually lasts for the first 6 to 18 months, during which no gross psychomotor abnormalities are observed in the child[2,6]. Nevertheless, recent studies have shown that certain subtle, almost subclinical developmental defects are detectable even in this early period, which could be accompanied by specific molecular signatures[7,8]. These initial signs constitute bases for early detection and potential therapeutic action once effective therapies and treatments are developed.

After this early period of normal psychomotor development, patients develop symptoms following four stages of the disease: stagnation, regression, pseudostationary period and late motor deterioration[1,6]. The fact that the regression period is followed by a recovery phase differentiates RTT from neurodegenerative conditions[6].

- **Stage I - Developmental stagnation** (ages 6mo – 1.5y)

At age 6-18 months, the progression of developmental milestones is modestly delayed. Most patients experience growth retardation and deceleration of head growth, which results in reduced size of the head and brain (acquired microcephaly)[2,9]. These features may be accompanied by the first signs of muscle hypotonia, evidenced in the form of postural weakness.

- **Stage II - Developmental regression** (ages 1.5y – 4y)

After a few weeks or months with no significant improvement in developmental milestones, a period of loss of acquired skills begins, which can last up to a year. During this stage, purposeful hand use is replaced by hand stereotypies, which may involve clapping, flapping, mouthing, wringing or the characteristic washing movements that drew the attention of the first paediatricians to describe the disease. In this period, acquired language, such as babbling or first words, is also lost, and RTT patients appear unresponsive to the environment and social interactions with a remarkable lack of eye contact. The autistic features displayed during this phase also include sound hypersensitivity, irritability, and self-abusive behaviour. Gross motor abnormalities also become apparent during this stage, displaying gait dysfunctions involving a lack of coordination, ataxia, and apraxia. In many patients, breathing disturbances involving hyperventilation and apnoea also appear during this stage.

- **Stage III - Pseudostationary period** (ages 3y – 10y)

Later, patients enter a stabilization phase that can last from years to decades. There is a partial recovery in communication and a higher degree of connection with the world, with intense eye contact. Gross motor skills are also improved, with some patients recovering or developing ambulation, although gait ataxia and hand apraxia and stereotypies are prominent. At this point, most patients develop seizures. In some patients, seizures may be manageable with antiepileptic drugs, but in others, they can be completely refractory to treatment, although in most cases they will tend to decrease in severity after adolescence. Patients show a range of behavioural abnormalities including bruxism, inappropriate laughing, crying, and screaming spells and a high degree of anxiety. During this period, osteopenia, scoliosis, and

rigidity become apparent, while there is an unapparent progression of neurologic deterioration that ultimately leads to the final disease stage.

- **Stage IV - Late motor deterioration** (ages 10y – after 40y)

The last stage of the disease, which can last for decades, begins with the loss of ambulation and the patient becomes wheelchair-bound. The loss of mobility worsens scoliosis and rigidity, and patients develop dystonia and Parkinsonian features. Additional autonomic dysfunctions arise, such as vasomotor disturbances leading to cold, blue feet, dysphagia, constipation, and cardiac dysfunction. Some patients survive after age 40y in a severely impaired physical condition.

1.1.1. Diagnostic criteria for typical RTT

The diagnosis of RTT is mainly clinical and follows a set of criteria proposed for the first time in 2001 and periodically revised and updated since[10,11]. To consider a diagnosis of typical RTT, there must be a period of regression followed by a phase of recovery or stabilization, and the patient must fulfil all main and exclusion criteria (Table 1).

Main criteria	Exclusion criteria	Supportive 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	1. Brain injury secondary to trauma (peri- or postnatally), neurometabolic disease, or severe infection that causes neurological problems 2. Grossly abnormal psychomotor development in first 6 months of life	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 pain response 11. Intense eye communication

Table 1. Revised diagnostic criteria for RTT[11].

Even though supportive criteria are usually present in typical RTT patients, they are not necessary for the diagnosis[11].

1.1.2. Atypical forms of RTT

After the standardisation of RTT diagnostic criteria, it was recommended that patients who deviate from the characteristic RTT pattern (in the age of onset, natural history, severity and the presence or absence of certain distinctive features), and therefore do not fulfil all the necessary criteria for typical RTT, should be designated as "atypical"[12]. In accordance with the revised criteria, atypical RTT patients must also have experienced a regression period followed by recovery or stabilization and must fulfil at least 2 of the 4 main criteria and 5 out of 11 supportive criteria[11].

There are several specifically documented variant forms of atypical RTT:

- **Preserved speech (Zappella) variant**: patients with this atypical RTT variant present an overall milder phenotype. Although they go through all disease stages, both regression and late motor deterioration are significantly delayed compared to patients with typical RTT[11,13]. Hand stereotypies are also present but are less prominent and allow for better retained hand use. The denomination of this variant stems from the marked recovery of language in the pseudostationary period. Intellectual disability tends to be milder (with an IQ of up to 50), and they can properly use single words or even phrases. Moreover, features such as epilepsy, autonomic dysfunction and scoliosis are less pronounced[13,14].
- **Early onset seizures (Hanefeld) variant**: patients with the Hanefeld variant develop seizures remarkably early, before regression[15,16]. These patients present a phenotype marked by severe infantile spasms and refractory myoclonic or tonic-clonic epilepsy[11,17], while stereotypies, breathing dysfunctions and other signs appear later in the disease course.
- **Congenital (Rolando) variant**: patients with the Rolando variant of RTT present a severe postnatal microcephaly before age 4 months, an earlier regression period (by age 5 months) and marked hypotonia[18]. In some cases, it is difficult to appreciate regression, and therefore it was termed "congenital", although a regression period is required for a correct diagnosis[15]. Early psychomotor development is remarkably abnormal, with severely delayed milestones and

inability to walk. Autonomic dysfunctions, like peripheral vasomotor disturbances and breathing abnormalities, are prominent, and patients display distinctive tongue stereotypies and jerky movements of the limbs. The severity of seizures impedes communication, and patients lack the distinctive intense eye contact observed in typical RTT[11].

Less delineated atypical forms include the *forme fruste* of RTT, with a discrete pattern of neurodevelopmental abnormalities that only become compatible with RTT as patients age; and late regression, where the characteristic regression period starts much later, after 3 years old[6].

1.1.3. Phenotypic heterogeneity

With the delineation of all these variants, it became patent that the phenotypic spectrum of patients with RTT is wide, ranging from a severe new-born encephalopathy to much milder cases with a degree of ambulation and language[4,19]. Phenotypic variability has been observed even amongst patients comprised within the same family, such as the cases of two couples of sisters where one was classic RTT and the other had the preserved speech variant[13,14].

The range of variability encompasses the presence or absence of specific traits and comorbidities, as well as variations in severity and age of onset. Therefore, several clinical scales that consider these features have been developed to assess and quantify the severity of the clinical presentation in RTT[4]. Although different scoring systems have been published, they all assign higher scores for the earlier debut or higher severity of symptoms[20–22] (Table 2). Therefore, severity scores tend to become higher as patients grow up and new comorbidities appear[4].

Feature	Score	Definition
Age of onset	3	0-12 months
	2	12-24 months
	1	> 24 months
Microcephaly	0	Absent
	1	Present
Sits unsupported	0	Acquired <8 months
	1	Acquired 8-16 months
	2	Acquired >16 months
	3	Never acquired
	+1	Lose acquisition
Ambulation	0	Acquired <18 months
	1	Acquired <30 months
	2	Acquired >30 months
	3	Lose acquisition
	4	Never acquired
Language	0	Preserved and propositive
	1	Lost
	2	Never acquired
Epilepsy	0	Absent
	1	Present and controlled
	2	Uncontrolled or early epilepsy
Respiratory function	0	No dysfunction
	1	Hyperventilation and/or apnoea
Hands use	0	Acquired and conserved
	1	Lose purposefulness: 2-6 years
	2	Lose purposefulness: <2 years
	3	Lose all acquisitions
	4	Never acquired
Onset of stereotypies	0	> 10 years
	1	> 36 months
	2	18-36 months
	3	< 18 months

Table 2. Pineda clinical severity score for RTT[21]. The range of scoring comprises values from 1 (the mildest case) to 23 (the most severe).

1.2. Genetic perspective

Before the identification of the genetic cause of RTT, many theories arose about which was the underlying nature of the disease. Initially, it was thought to be a metabolic disorder, since the first report by Andreas Rett found an apparent association with hyperammonaemia[3], but this finding was later attributed to a laboratory error. Later, it was proposed that it might be a mitochondrial disease based on mitochondrial morphology and metabolism abnormalities identified in RTT patients[23].

Since patients with RTT are almost exclusively female, one hypothesis was that it could be an X-linked dominant disorder associated with lethality in males[5,24–28]. Based on this assumption, exclusion mapping studies in families with more than one RTT case limited the candidate region to Xq28, and in 1999 the causal gene in this region was identified as *MECP2,* the gene encoding methyl-CpG-binding protein 2 (MeCP2)[29].

1.2.1. MeCP2: Methyl-CpG-binding protein 2

Structure

MECP2 (OMIM*300005) is an X-linked gene, located in the long arm of the X-chromosome (Xq28), between the Interleukin-1 receptor associated kinase gene (*IRAK1*) and the red opsin gene (*RCP*). It is encompassed by four major exons that are alternatively spliced into two different isoforms (Figure 1A and B). Isoform e1 (NM_001110792) comprises exons 1, 3 and 4, and produces a 498 amino acids protein from the translation start codon in exon 1. Isoform e2 (NM_004992) comprises all four exons and does not use the translation start codon in exon 1, but rather an alternative translation initiation site towards the end of exon 2, producing a shorter 486 amino acids protein[1,30,31].

The proteins produced from both isoforms are identical except for a few amino acids in the N-terminal end and share the structure of five main domains[30] (Figure 1C): the N-terminal Domain (NTD), the methyl binding domain (MBD), the interdomain (ID), the transcription repressor domain (TRD) and the C-terminal domain (CTD).

The majority of the MeCP2 protein is unstructured, and therefore it is considered an intrinsically disordered protein[32].

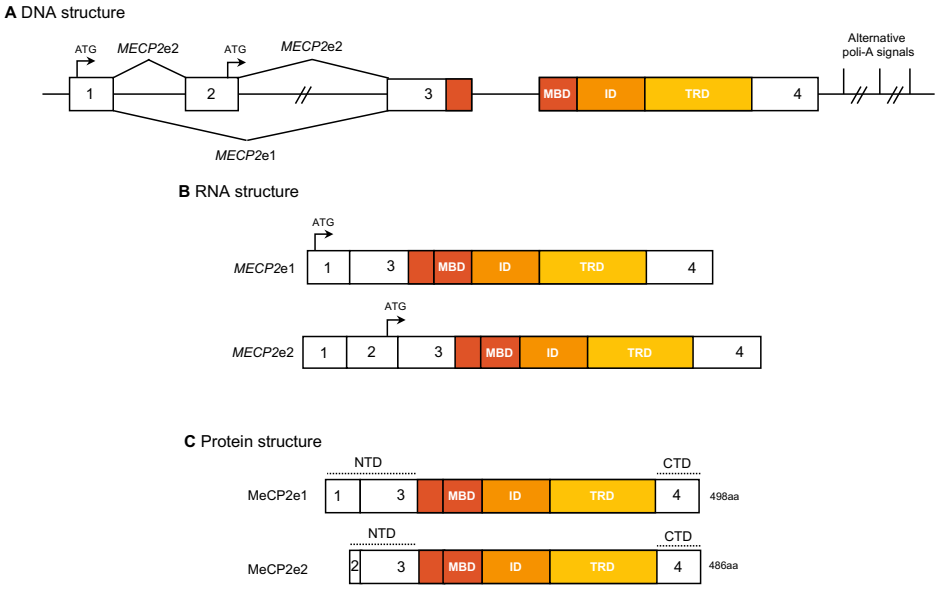


Figure 1. Structure of the *MECP2* gene, transcript isoforms and protein.

The lack of correlation between protein and RNA levels indicates that there may be a post-transcriptional regulation of translation[33]. Differential use of polyadenylation sites in the 3'UTR of *MECP2* produces three different transcripts, which suggests a post-transcriptional control system of MeCP2 levels that may involve miRNA regulation[34].

Expression

MeCP2 is ubiquitously expressed, with the highest expression in brain, lung, and spleen[35]. Within the brain, its highest expression levels are found in the cortex and cerebellum, and it is predominantly expressed in neurons although it is also present in astrocytes, oligodendrocytes and microglia[36]. The mainly neurologic nature of RTT phenotypes underscores the vital role that MeCP2 plays in the brain, but the extra neurological manifestations, such as breathing abnormalities, scoliosis and cardiac defects denote its importance also outside the nervous system[2,37]. MeCP2e1 is the predominant isoform in the brain, while MeCP2e2 is more abundant

in other tissues and cell types, such as fibroblasts[1,31]. Both protein isoforms coexist in the brain, although MeCP2e2 shows a more restrictive pattern and a later onset of expression[41,42].

Brain-specific deletion of *Mecp2* in mouse brain recapitulates RTT features as in the global KO model, including abnormal gait, hindlimb clasping, respiratory abnormalities, and early lethality[38,39]. Region and cell-type-specific KO in the brain correlates with the appearance of different phenotypes, revealing their contribution to RTT: deletion in the brainstem and spinal cord was linked to abnormal heart rate and breathing patterns; in dopaminergic and noradrenergic neurons led to motor incoordination; in amygdala induces anxiety-like behaviour [2]; in forebrain neurons, it triggers abnormal social behaviours and autistic features[2]; in hypothalamic neurons, it produces stress response and hyperagressiveness[2]; and in inhibitory neurons produced a severe RTT phenotype, which highlights the importance of E/I balance in the pathogenesis of RTT[40].

MeCP2 is expressed since the early stages of embryonic development, increasing during the final period, and reaching its peak in mature neurons[33,43] (Figure 2).

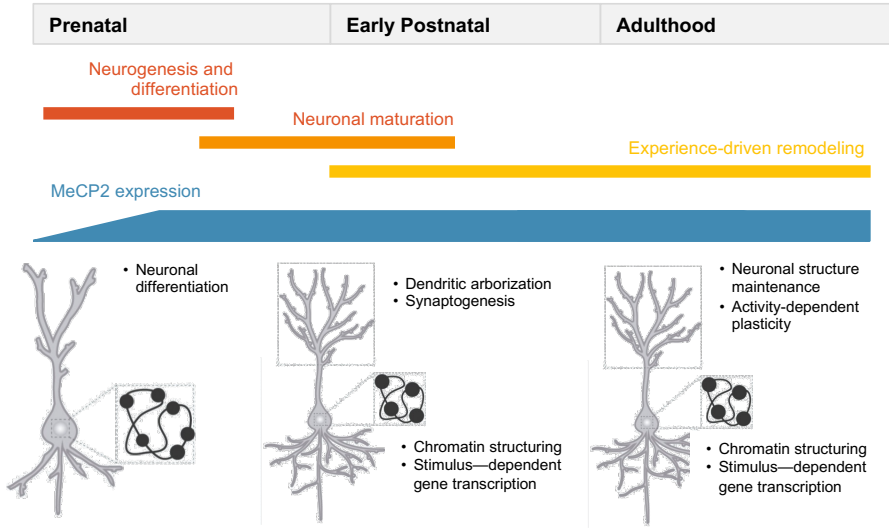


Figure 2. MeCP2 function throughout development. MeCP2 regulates brain development and maintains the function of mature neurons through adulthood. It regulates neuronal differentiation in prenatal stages, neuronal maturation, and experience-driven remodelling in later developmental stages. Modified from Gulmez-Karaca *et al.* (2019)[43].

The increase in expression occurs after neurogenesis and differentiation, and coincides with dendritic growth, branching and spine morphogenesis for synapse formation[33]. The apparent normal early development observed in individuals with RTT is attributed to the delayed expression pattern of MeCP2. High expression levels are maintained throughout adulthood, underscoring an important function not only in the developing brain, but also in the maintenance of neuronal circuits[34,43]. Moreover, several studies in mice models have demonstrated the reversibility of RTT features, even at late stages of disease progression, emphasizing the idea that the function of MeCP2 is essential beyond neurodevelopment [44,45].

Function

The main molecular function of MeCP2 is to recognize and bind specifically to symmetrically methylated cytosine residues in CpG dinucleotides with adjacent A/T-rich motifs through its MBD domain[46]. MeCP2 is known to influence various biological mechanisms, such as transcriptional regulation, chromatin remodelling, alternative splicing, and micro-RNA processing[2,40,47]. Via all these processes, MeCP2 becomes a multifunctional protein that serves as a regulator of gene expression at both transcriptional and post-transcriptional levels, exerting control over numerous genes across the entire genome[48,49].

Moreover, MeCP2 activity is regulated by post-translational modifications such as phosphorylation. The protein has several phosphorylation sites that can be modulated *in vivo* by neuronal activity. Phosphorylation at serine 421 controls the ability of MeCP2 to regulate dendritic arborization and spine morphogenesis, while phosphorylation at serine 80 is essential for its binding to several gene promoters and is crucial for locomotor activity[50–52].

Transcriptional regulation

MeCP2 was initially described as a transcriptional repressor[53,54]. Its MBD domain binds methylated CpG dinucleotides while its TRD recruits repressor complexes containing SIN3A and histone deacetylases (HDACs)[54]. Through MeCP2 binding, HDACs approach methylated areas and proceed to deacetylation to enhance gene silencing.

In expression studies, many genes have been identified as MeCP2 regulatory targets, although opposing effects on their transcription have been noted[36]. The role of MeCP2 in regulating transcription appears to be different depending on its interacting protein partners. When interacting with activator complexes that contain cAMP response element-binding protein (CREB), it can mediate transcriptional activation[55]. Considering its abundance, versatility, and the limited size of its binding site, MeCP2 can play a role as a global transcriptional regulator.

Chromatin remodelling

Transcriptional analyses in RTT models have consistently shown subtle alterations, suggesting that MeCP2 may play a role in fine-tuning gene expression instead of as a classical transcriptional regulator. Recent studies point towards MeCP2 acting as a broad regulator of chromatin structure, rather than being a gene-specific transcription factor. This function would be essential for dampening transcriptional noise and ensuring the proper functioning of transcriptional machinery[43].

Histone H1 is present in most cell types at an approximate stoichiometry of one molecule per nucleosome, but uniquely, in neurons, this is reduced to one molecule every two nucleosomes. Several studies have shown that MeCP2 can compete with histone H1 for binding to methylated chromatin and may function as a substitute linker histone, promoting local chromatin compaction, which could explain the transcription inhibitory role of MeCP2[53,56,57]. Interestingly, MeCP2 deficiency prompts a 2-fold elevation in histone H1 and a higher frequency of spurious transcription of repetitive elements[57].

Moreover, MeCP2 accumulates at chromocenters and plays a role in the organization of pericentric heterochromatin (PCH) by promoting chromocenter clustering, translating DNA methylation signal into transcriptional repression by recruiting silencing complexes and HDACs[53,58]. MeCP2 has been demonstrated to induce chromatin re-organization during postnatal neuronal maturation, possibly acting through the induction of chromocenter clustering[59]. MeCP2 interacts with ATRX, a SWI2/SNF2 DNA helicase/ATPase with key roles in brain development that is mutated in ATRX syndrome, a severe intellectual disability disorder, and recruits it to PCH[48,60]. The interaction of MeCP2 with ATRX, cohesin and CTCF is involved

in the correct regulation of expression of imprinted genes by controlling nucleosome positioning and the formation of silent-chromatin loops[61–63].

Moreover, MeCP2 becomes phosphorylated in response to neuronal activation, and this activity-dependent regulation promotes synapse development. This modification facilitates a genome-wide response of chromatin to neuronal activity during nervous system development[64].

Alternative splicing

Consistent with the finding that MeCP2 can bind to RNA in vitro, it has also been proposed that MeCP2 regulates processes such as alternative splicing and miRNA processing. MeCP2 might therefore also serve to regulate gene expression at a post-transcriptional level[65].

Aberrant alternative splicing patterns have been described in RTT models[66]. Consistently with these findings, MeCP2 has been found to interact with YB1, a Y-box transcription factor that can regulate alternative splicing, and PRPF3, a spliceosome-associated protein[66,67]. Included alternatively spliced exons are enriched in DNA methylation and MeCP2 occupancy, and MeCP2 depletion increases histone acetylation and aberrant exon skipping events, indicating that MeCP2 contributes to exon definition and alternative splicing modulation[65,68].

Micro-RNA processing

Micro RNAs (miRNAs) play a significant role in numerous developmental processes, such as neurogenesis and synaptic plasticity, and have been found dysregulated in several RTT *in vitro* and *in vivo* models[40,65]. MeCP2 can control miRNA expression via two different mechanisms. It can repress the transcription of primary miRNAs by binding to their promoter regions, or it can inhibit miRNA processing by binding the microprocessor protein DiGeorge syndrome critical region 8 (DGCR8) and interfering with the assembly of the DGCR8-DROSHA microprocessor complex[69,70]. These two scenarios lead to increased levels of the target miRNAs in a MeCP2-deficient context. Moreover, some miRNAs can act in a feedback loop, as both MeCP2 regulatory targets and by binding sequences in MeCP2 3'UTR to fine-tune MeCP2 levels[71].

Types of mutations and phenotypic correlation

Loss-of-function mutations in *MECP2* are the most common genetic cause underlying RTT. These mutations are mostly *de novo* and tend to appear on the paternal chromosome[72]. According to the natural history studies, *MECP2* pathogenic variants are identified in 95% of patients with typical RTT and over 70% of patients with atypical RTT[72]. These include nonsense, missense, and frameshift variants, as well as large deletions encompassing several exons or the whole gene. Even though over 500 pathogenic variants have been reported in the specialized database RettBASE[73], there are eight recurrent point mutations that account for almost 50% of all genetically diagnosed cases[72,73] (Figure 3). Remarkably, pathogenic variants disrupting specifically MeCP2e1 have been described in RTT patients, but no mutations affecting exclusively MeCP2e2 have been reported[74].

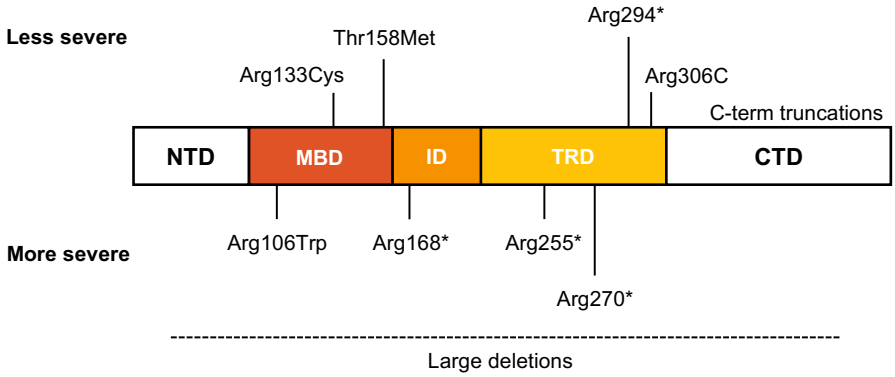


Figure 3. Most recurrent *MECP2* mutations. The 8 most recurrent point mutations in *MECP2* are located in the methyl-binding domain (MBD), the inter-domain (ID) and the transcription repression domain (TRD). Protein truncating variants in the C-terminal domain (CTD), and large deletions are also common. Early-truncating mutations and large deletions tend to produce a more severe phenotype than missense mutations and late-truncating mutations.

No clear correlation could be established between genotype and phenotypic severity. Nevertheless, patients with some missense and late-truncating mutations, as well as C-terminal deletions, tend to present a milder phenotype in relation to ambulation, hand use and language than those with nonsense mutations and large

deletions[75,76] (Figure 3). These are the most common types of mutations among patients with the preserved speech variant[13].

Some of the RTT-causing pathogenic variants have been demonstrated to impair interactions of MeCP2 and its protein partners, thus disrupting many of its functions[48]. MeCP2 truncating mutations disrupt its binding with the spliceosome factor PRPF3 and with the miRNA processing complex DGCR8, while missense mutations apparently affect PCH binding[40,67,77]. Moreover, some of the most prevalent mutations such as p.Arg168* and p.Arg306Cys abolish binding with co-repressor complexes such as SMRT[48].

Not only loss-of-function MeCP2 mutations are pathogenic. Gain-of-function variants consisting of a chromosomal duplication encompassing the MeCP2 gene cause MeCP2 duplication syndrome (MDS), another neurodevelopmental disorder with overlapping features with RTT such as intellectual disability, speech impairment and seizures[78]. This indicates that MeCP2 dosage needs to be tightly regulated to achieve key levels for proper cellular function.

X chromosome inactivation

X chromosome inactivation (XCI) is the process by which mammalian female cells silence one of the two X chromosomes to compensate for gene dosage in females (XX) versus males (XY). This phenomenon is presumably stochastic, and the maternally and the paternally inherited X chromosomes share the same probability of inactivation. XCI takes place in early embryonic stages and is maintained through cell divisions, producing a mosaic expression of X-linked genes in the adult organism. Because XCI happens randomly in cells during early developmental stages, we expect an approximate 1:1 ratio of cells expressing the maternal and paternal X chromosomes[79,80].

Even so, there is a wide margin of variation from the 1:1 ratio in the normal female population. Since most females will carry variants in one or more X-linked genes, some of these variants may confer some cells a selective advantage versus the others. In some cases, this leads to a significant excess of one of the cell populations, a phenomenon called unbalanced XCI or skewing[79]. Alternatively, the

same X chromosome may be preferentially inactivated in most cells via a genetically determined mechanism[80,81].

MeCP2, like most X-linked genes, is subject to XCI. Therefore, it has been suggested that XCI might play a role in RTT phenotypic variability by controlling how many and which cells express the wild-type and the mutant copy of the gene. However, XCI studies in RTT patients have yielded conflicting results[82]. Some cases have been reported where female carriers of *MECP2* disease-causing mutations have extremely skewed XCI patterns, preventing them from developing the disease[83,84]. Nevertheless, most RTT patients exhibit random XCI patterns, suggesting that adaptive skewing (predominance of the cells with the wild-type copy of the gene on the active X chromosome) is uncommon, and suggests that cases of obligate carriers with skewed XCI probably involve a selective effect in another X-linked locus[79]. Cases with less extreme skewing have been reported as atypical RTT patients without the full spectrum of symptoms, and it is generally considered that modest deviations in XCI patterns may affect the clinical development of RTT and could potentially account for phenotypic differences among patients with identical mutations, although the extent of this influence is yet to be determined[85].

1.2.2. Molecular consequences of MeCP2 deficiency

In a cellular context, MeCP2 is involved in the maturation of neuronal function, through processes including synaptogenesis, synaptic plasticity, and establishment of functional circuits[40]. Over 60 published studies have ivestigated the molecular consequences of MeCP2 deficiency and have detected thousands of deregulated genes in a variety of cell-types, tissues, and organisms. The heterogeneity of these experiments complicates conducting a proper meta-analysis with all current data, and their results are hardly overlapping when not contradictory. Nevertheless, some cellular functions and molecular pathways appear repeatedly altered in several studies and have been linked to physiological alterations found in patients with RTT and RTT animal models.

Many different studies have identified possible links implicating MeCP2 in regulating cytoskeletal organisation via the interaction with different proteins, such as CAPG and MEF2C[86–88]. Impaired cytoskeletal organisation has been linked to the

observations of altered neuron structures with decreased dendritic complexity and immature synaptic spine morphology observed in post-mortem brain of RTT patients and RTT mice models, which cause neurons to be abnormally small and densely packed causing a reduction of brain volume, and both actin and tubulin cytoskeleton related proteins have been found to be dysregulated in MeCP2 deficient contexts[89–94].

Abnormal levels of neurotransmitters and neurotransmitter receptors leading to excitatory inhibitory imbalance involving GABAergic, glutamatergic, and dopaminergic neurotransmission has also been identified in MeCP2 deficient contexts[47,95,96]. The transcription factor DLX5, which regulates the differentiation and maturation of GABAergic neurons, has been identified as a direct MeCP2 target, and GABA transporters were found to be downregulated in *Mecp2*-null brains[97,98]. GABAergic dysfunction has been associated to several RTT features, such as ataxia, repetitive behaviours, and deficits in motor coordination[40]. It seems to play a crucial role in RTT pathophysiology since the restoration of MeCP2 in inhibitory neurons of *Mecp2* KO mice partially rescued RTT-like symptoms[99,100].

Mitochondrial dysfunction and inflammatory processes were also repeatedly involved in RTT pathophysiology in several studies[88,101–103]. Several proteins that belong to mitochondrial complexes have been found altered (up and downregulated) in RTT patients[102–104]. Enlargement of mitochondria with altered structure has indeed been identified in *Mecp2* KO mice neurons, as well as increased oxygen consumption rates and reactive oxygen species generation[101].

These altered biological processes are probably contributing to the pathophysiology of RTT and constitute sources for the identification of therapeutic targets that can potentially revert some of the effects of MeCP2 deficiency in diverse cellular contexts.

1.3. Treatments for Rett syndrome

Since there are currently no specific treatments for RTT, clinical management is mainly symptomatic. Frequent drug treatments include antiepileptic drugs for seizures and serotonin reuptake inhibitors for anxiety. Preventive nutritional management has also become common, including ketogenic diet and

supplementation with vitamin D and omega-3 fatty acids to prevent seizures and orthopaedic complications[105].

Several studies have demonstrated that the absence of MeCP2 during development does not cause irreversible damage in mice. The re-activation of MeCP2 in neurons of male (hemizygous) and female (heterozygous) *Mecp2*-null mice has proven to increase survival and improve many RTT features, such as hind-limb clasping, tremor, irregular breathing patterns, general motor coordination, anxiety, and epileptiform activity, rescuing their phenotype to levels comparable to their wild-type littermates[44,106,107]. MeCP2 re-activation also produced a rescue in neuronal morphology defects commonly observed in MeCP2-deficient contexts, increasing soma size and dendritic complexity and length[106]. These studies provide evidence that RTT is not a strictly neurodevelopmental disorder. MeCP2 is mainly expressed in differentiated neurons and is more involved in neuronal maturation and maintenance rather than early cell-fate decisions, fine-tuning gene expression of pathways pivotal for dendritic development and synaptogenesis[34,44,108–110] . The lack of MeCP2, however, does not produce neurodegeneration, but rather freezes neurons in an immature state, where MeCP2 target sites are established and maintained normally and enable the consequences of MeCP2 deficiency to be reversed when re-activating its expression[44].

Considering the evidence of reversibility of the RTT phenotype, several strategies to restore MeCP2 levels are being considered (Figure 4):

- **Site-directed gene repair**. Gene editing technologies such as CRISPR-Cas9 have been used to repair disease-causing mutations in the *MECP2* gene in cell lines, with efficiencies around 20-30%[111]. Although this type of gene therapy would be a curative, single-administration approach, the possibility of off-target effects makes it still far away from the clinical practice. RNA editing would be an alternative approach with no risk of permanent genetic changes. Hippocampal injection of adeno-associated viruses (AAVs) carrying an adenosine deaminase and a *Mecp2* guide RNA corrected the mutation in 50% of transcripts and rescued MeCP2 protein localization, but the potentially immunogenic effects of the enzyme has prevented it from reaching clinical trials[112].

- **Replacement therapy**. Replacement therapy consists in delivering the defective protein to the target cells, using gene constructs, mRNA, or recombinant protein. Replacement therapy for RTT must consider MeCP2 dosage carefully since an excess of protein also causes the neurological symptoms characteristic of MDS. The most successful approach to date has been NGN-401, which consists of an adeno-associated virus (AAV) delivered *MECP2* expression cassette with a modified promoter to limit transcription, a synthetic miRNA, and a 3'UTR with binding sites for the miRNA that regulate *MECP2* expression. This construct enables self-regulated transgene expression at single-cell level through the miRNA feedback loop that eliminates excess transcript[113–115].

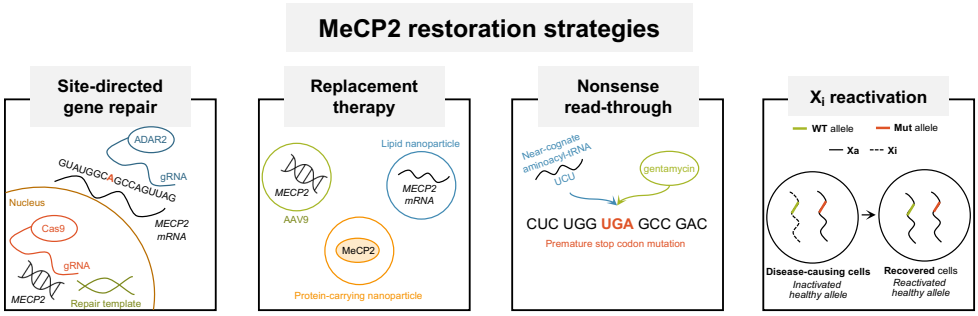


Figure 4. MeCP2 restoration strategies. Several strategies are being considered to restore functional MeCP2 levels in RTT. Some of the most relevant are site-directed gene repair, replacement therapy, nonsense read-through and reactivating the inactive X chromosome. Modified from Grimm, et al. (2022)[116].

- **Nonsense read-through**. Since about one third of disease-causing mutations in *MECP2* are nonsense mutations, therapy with aminoglycosides such as gentamycin could promote the read-through of these mutations producing a complete protein. The ability of these compounds to cross the blood-brain barrier, their potential toxicity and the inconsistent results when tried on *MECP2* mutations have hindered the application of this type of therapies to restore MeCP2 levels[117].
- **Inactive X chromosome reactivation**. Another potential strategy to increase the levels of MeCP2 is to use the WT copy that is already present in every cell in female RTT patients. Pharmacologic strategies have been directed to selectively re-activate the copy of *MECP2* on the inactive X chromosome in

neurons, successfully correcting reduced soma size and branch points *in vitro*[118]. However, care needs to be taken to produce a sufficient and safe dosage not only of MeCP2, but also of all the other proteins that are encoded by the X chromosome[116].

While gene therapy approaches to restore MeCP2 levels are still in pre-clinical stages, pharmacological approaches with trophic factors that promote brain growth and development have reached clinical trials. Insulin-like growth factor 1 (IGF-1) is a compound required for proper brain development that modulates synaptic development and function. When used to treat *Mecp2*-null mice, recombinant IGF-1 (a drug named mecasermin) and an IGF-1 analogue (a drug named trofinetide) demonstrated an increase in the number of glutamatergic synapses, soma size and dendritic complexity. In clinical trials, mecasermin showed no benefits in patients treated in phase 2, but trofinetide demonstrated safety, tolerability and significant improvement in breathing problems, mood abnormalities and repetitive hand movements in phase 3 of the clinical trials and was recently approved by the FDA for the treatment of RTT in the USA[119–121]. Brain-derived neurotrophic factor (BDNF) is a neurotrophin implicated in neuronal development, synaptic transmission, and plasticity. Several studies have identified MeCP2 as a regulator of *BDNF* gene expression, implicating it as a potential therapeutic target for RTT[122]. BDNF overexpression in neurons of *Mecp2*-null mice significantly increased their lifespan, improved locomotor function, and reversed impaired dendritic arborization[123]. Fingolimod, a drug that enhances BDNF expression and is already used to treat multiple sclerosis, entered a clinical trial for RTT, but did not demonstrate any improvements for RTT patients involved in the trial[124].

With the growing evidence of downstream effects of MeCP2 deficiency, another therapeutic strategy is to modulate the cellular processes known to be impaired in models of RTT. Loss of MeCP2 alters the expression levels of many genes, suggesting that it is unlikely that a single molecular pathway can be targeted for therapeutic intervention, but rather many cellular processes are affected, including synaptic transmission and plasticity, protein synthesis, mitochondrial function, and lipid metabolism[4]. Gaining insight into the molecular landscape of MeCP2 deficient

cells will help to identify potential therapeutic targets to tackle the altered molecular functions in patients with RTT.

2. Rett syndrome spectrum disorders

2.1. New genes

After the association of MeCP2 mutations with RTT, the fact that a yet significant number of patients remained genetically undiagnosed, especially patients with atypical RTT, raised the idea that it might be a genetically heterogeneous disorder.

2.1.1. *CDKL5*

In 2004, pathogenic variants in cyclin-dependent kinase-like 5 (*CDKL5,* OMIM*300203), a gene located in Xp22, were identified in patients who had been clinically diagnosed with the early-onset-seizure variant of RTT[1,125,126]. *CDKL5* had been previously associated with X-linked infantile spasms or X-linked West syndrome, and later in patients with encephalopathy with refractory epilepsy, suggesting that the clinical presentation of patients with *CDKL5* mutations may be heterogeneous and only a subset of them, approximately 25%, may fulfil criteria for an atypical RTT diagnosis[127–129].

Despite the significant phenotypic overlap in these cases, recent studies have highlighted the differences between patients with *MECP2* and *CDKL5* mutations, considering *CDKL5* Deficiency Disorder (CDD) a separate and specific clinical entity[127,130]. Distinctive features that characterise CDD include early-onset seizures that evolve into severe encephalopathy with mostly refractory epilepsy, marked hypotonia and very limited developmental progress, while the characteristic RTT pattern of normal early development followed by regression is rarely observed[128–130]. CDD patients also exhibit some characteristic RTT features, such as acquired microcephaly, stereotypies, hand apraxia, sleep disturbances and bruxism, although autonomic dysfunctions such as breathing disturbances and gastrointestinal dysfunction are rare[129]. Unlike *MECP2* mutations, *CDKL5* mutations are relatively common in male patients presenting with epileptic encephalopathy[131].

CDKL5 is a serine/threonine kinase that shares homology with members of the mitogen-activated protein (MAP) kinase and the cyclin-dependent kinase (CDK)

families, and whose expression patterns overlap with that of MeCP2 during neuronal maturation and synaptogenesis[132,133]. CDKL5 can mediate the phosphorylation of MeCP2 via its kinase domain, which suggests that they are part of the same molecular pathway and partially explains the overlapping phenotypes[133]. Moreover, MeCP2 can bind the methylated CpG dinucleotides on the *CDKL5* promoter and modulate its transcription, establishing a feedback loop[51].

No significant correlation has been found in CDD patients between the phenotypic variability and the type and location of the mutations or XCI patterns. However, it is hypothesized that the phenotypic variability may be attributed to the specific molecular consequences on CDKL5 activity caused by each mutation[129]. In particular, *CDKL5* mutations identified in patients with the early-onset seizures variant of RTT are believed to disrupt the phosphorylation of MeCP2, leading to alterations in MeCP2 activity that contribute to generating a subset of RTT features[51].

CDKL5 is required for dendritic arborization, correct neuronal migration and neuronal morphogenesis[134,135]. CDKL5 interacts with a protein complex containing Rac1, a regulator of actin remodelling implicated in neuronal morphogenesis, whose overexpression rescued dendrite growth in CDKL5 deficient neurons. These results suggest that a dysfunction in Rho GTPase signalling may contribute to CDD pathogenesis. Dendritic spine formation is also impaired in RTT neurons[135].

2.1.2. *FOXG1*

In 2008, chromosome rearrangements and interstitial deletions affecting the region 14q12 were identified in patients with the congenital variant of RTT, in which the forkhead box protein G1 (*FOXG1*, OMIM*164874) gene was highlighted as an important candidate[136,137]. Subsequent screening for *FOXG1* intragenic mutations in RTT patients with no molecular diagnosis confirmed that *FOXG1* was frequently the causative gene in patients with the congenital variant[138]. Since it is not an X-liked gene, *FOXG1* mutations are equally identified in male and female patients.

The initial clinical characterization of patients with *FOXG1* mutations was through the identification of mutations in patients with the congenital variant of RTT, but with the widespreading of NGS diagnostic methods new *FOXG1* variants are being

detected in patients with more unspecific clinical phenotypes[139]. While there is a significant phenotypic overlap with patients with *MECP2* mutations and the congenital variant of RTT, it has been suggested that *FOXG1* syndrome may be designated as a new clinical entity based on its distinctive set of features. These include the absence of a period of normal development, a movement disorder including dyskinesias, chorea and dystonia, and brain malformations consisting of a simplified gyral pattern, reduced white matter volume and callosal hypogenesis[140].

FOXG1 encodes a transcriptional repressor from the forkhead (FOX) family with restricted expression to fetal and adult brain and testis[139]. *FOXG1* has a DNA binding fork-head domain and represses transcription of target genes by recruiting two families of corepressors, Groucho and JARID1B, that associate with specific domains of the protein[141,142]. It is localized in the nucleus but not found in MeCP2-positive heterochromatic foci, although it does colocalize with MeCP2 in other nuclear compartments[136]. This suggests FOXG1 is not a TF stably associated with heterochromatin, but it may be involved in some gene expression regulatory activities alongside MeCP2 in differentiating and mature neurons, thus partially explaining the overlapping phenotypes[136].

FOXG1 is expressed in neural structures derived from the telencephalon such as the cerebral cortex, the hippocampus, and the basal ganglia, and is involved in regional patterning during development by promoting the proliferation and differentiation of progenitor cells[143]. It also controls interneuron development and is implicated in neurite growth and migration[144].

Frameshift and nonsense variants in the N-terminal domain and the forkhead domain are predicted to result in a truncated protein with loss of DNA binding and correlate with the most severe phenotypes, while truncating variants affecting the C-terminal domain and missense variants in the forkhead domain may produce a protein with residual function, including preserved binding sites for corepressors and correlate with milder phenotypes[139]. It has been hypothesized that these less impactful variants produce the phenotypes with a greater degree of overlap with RTT, while higher impact variants produce more severe phenotypes[145].

2.1.3. Other genes

After considering *MECP2*, *CDKL5* and *FOXG1*, some patients remain mutation-negative for all these genes and yet fulfil many criteria for a clinical diagnosis of RTT[146]. The generalization of NGS technologies for genetic diagnosis prompted the detection of mutations in many different genes in these cases. Patients who present a combination of distinct RTT features, yet do not fulfil established clinical criteria for either typical or atypical RTT, can be described as RTT-like. RTT-like patients lack formal consensus diagnostic criteria, and therefore represent a genetically heterogeneous group with phenotypes overlapping RTT to different degrees[146].

In the past few years, disease-causing variants in more than 90 genes have been identified in RTT-like patients[147]. Some of these genes had previously been associated with other well-defined, monogenic disorders with considerable phenotypic overlap with RTT, such as Pitt-Hopkins syndrome (*TCF4,* OMIM*602272), Angelman syndrome (*UBE3A*, OMIM*601623), or Cornelia de Lange syndrome (*SMC1A*, OMIM*300040). Others were linked to epileptic encephalopathies, such as *STXBP1* (OMIM*602926), *SCN2A* (OMIM*182390), *KCNQ2* (OMIM*602235) or *GRIN2B* (OMIM*138252), while others were associated with more unspecific phenotypes of intellectual disability with epilepsy, such as *MEF2C* (OMIM*600662) and *SYNGAP1* (OMIM*603384). Finally, novel potentially disease-causing genes for RTT-like phenotypes, have also been identified, such as *JMJD1C* (OMIM*604503) and *GABBR2* (OMIM*607340)[146–158]. Interestingly, many patients with mutations in these genes still fulfil the necessary criteria for the clinical diagnosis of RTT[159].

Considering the genetic diversity yet phenotypic overlap of RTT-spectrum (typical and atypical RTT and RTT-like) disorders, it is likely that some of the implicated genes converge in common molecular pathways and processes that are implicated in the pathophysiology behind these disorders[160].

2.2. Common molecular alterations in RTT-spectrum disorders

The current list of RTT-spectrum genes is especially enriched in genes involved in chromatin remodelling functions, such as histone deacetylases and chromatin-modifying enzymes, as well as genes involved in synaptic function, essential for

GABAergic, glutamatergic and dopaminergic neurotransmission, synaptic vesicle trafficking, ion homeostasis in neurons and circadian entrainment[147,152]. There are also some genes involved in ubiquitination and proteasome degradation processes.

Molecular pathways related to chromatin structuring and synaptic circuits are impaired in patients with RTT, as well as in RTT animal models, and could establish a link between RTT-spectrum genes and suggest a reason why patients with mutations in different genes present overlapping phenotypes[161,162].

The characterisation of the molecular consequences of the deficiency of different RTT-spectrum genes could contribute to the understanding of their functions and interrelationships. This could help in identifying common molecular pathways that may be affected in RTT-spectrum patients, contributing to the shared phenotypes, and could propose common potential therapeutic targets.

3. Omics technologies

Omics technologies are high-throughput technical approaches that allow for non-targeted detection of different types of molecules in biological samples, generating non-biased data for analytic purposes. Omics approaches permit the acquisition and analysis of all available data, which can lead to hypothesis generation in agnostic experiments and allow to detect unpredicted connections and interrelationships in the generated data. Omics sciences can focus onto many different fields, some of the most relevant being the genome (genomics), the transcriptome (transcriptomics) and the proteome (proteomics).

3.1. Omics for the diagnosis of rare diseases

3.1.1. Genomics

Genomics involves the comprehensive study of the DNA, including all its genes and regulatory structures (the genome). Next-Generation Sequencing (NGS) is the technological approach for the simultaneous sequencing of multiple regions in one single experiment. NGS provides a fast and comprehensive analysis of genetically heterogeneous clinical entities, such as RTT-spectrum disorders, and has progressively substituted successive single-gene molecular testing which may be more expensive and inefficient[163]. Instead of limiting the scope of the genetic study

to one single candidate gene, NGS allows to extend or redirect a genetic analysis if needed[163,164]

NGS-based diagnostic tests can vary in the number and type of target regions[165]:

- **Target gene panels**: the smallest NGS designs, which usually cover a set of disease-specific genes (ranging from a few to thousands).
- **Mendeliome or Clinical Exome Sequencing (CES)**: specific type of target gene panel that covers all exons of genes currently associated with monogenic (mendelian) disorders[166] (more than 4,000 genes).
- **Whole Exome Sequencing (WES)**: sequencing of all the exons of all protein-coding genes (more than 20,000 genes).
- **Whole Genome Sequencing (WGS)**: sequencing of all the genes and intergenic regions of the genome, with no targeted capture.

Whole Exome Sequencing (WES)

The American College of Medical Genetics and Genomics (ACMG) currently recommends WES as the gold standard of clinical practice in children with intellectual disability, developmental delay, or multiple congenital anomalies due to its cost-effectiveness[167]. The diagnostic yield of WES in patients with paediatric rare diseases is typically around 28% when performed in a proband-only approach, although it varies depending on the group of genetic disorders being considered[163,168–170]. In a recent systematic review, neurodevelopmental disorders (NDDs) were found to have a 23.7% overall diagnostic yield by NGS (22.6% using gene panels and 27.3% using WES). Among NDD subtypes, patients with intellectual disability showed the highest diagnostic yields (28.2%), while patients with autism spectrum disorder (ASD) showed the lowest (17.1%)[171]. Trio-based analysis, which consists of including data from the proband's parents, can help filtering out rare benign variants, identifying *de novo* variants and phasing the variants in recessive or imprinted genes, which can increase the diagnostic yield up to 50% compared to proband-only analyses[163,172]. Since *de novo* mutations are the most common cause of NDDs such as RTT, the trio-based approach can streamline the genetic diagnosis[173]. Even though the sequencing costs of the experiments are

higher when compared to proband-only analyses, the reduction in costs of segregation studies can compensate for this fact.

In terms of technical capacities, WES is especially sensitive to detecting single nucleotide variants (SNVs) and small insertions or deletions (indels), but copy number variant (CNV) detection is also possible through read depth analysis[163,174]. In fact, the power of CNV detection in WES could be superior to that of low-resolution genomic microarrays[175]. Therefore, WES studies that include CNV analysis usually have a higher diagnostic yield.

Even so, more than 50% of patients remain with no molecular diagnosis after WES analysis. WES can capture protein-coding regions, some UTRs, and intron-exon boundaries[176], which represent only 2% of the human genome, and therefore it is most likely missing disease-causing variants in non-coding regions. Moreover, WES presents non-uniform coverage, particularly in first exons and low-complexity regions, and is limited by the specificity of capture probes, which makes variant detection challenging in some genomic regions[177,178].

Whole Genome Sequencing (WGS)

Short-read WGS data provide many benefits when compared to WES. The library preparation protocol free of targeted capture and PCR amplification steps results in a more uniform coverage and higher quality data across the exome, producing fewer sequencing artifacts than WES, particularly in GC-rich exons[179–181]. Moreover, extending the coverage to noncoding regions of the genome enables the detection of potentially disease-causing deep intronic or regulatory variants, as well as complex structural variants that cannot be resolved simply by depth of coverage analysis with breakpoint base-pair resolution.

Noncoding variants

Despite not being part of the final protein sequence, noncoding variants have demonstrated disease-causing potential through various mechanisms: they may affect transcript splicing, gene transcription, protein translation, RNA processing and stability, or they may shape chromatin structure and interactions (Figure 5).

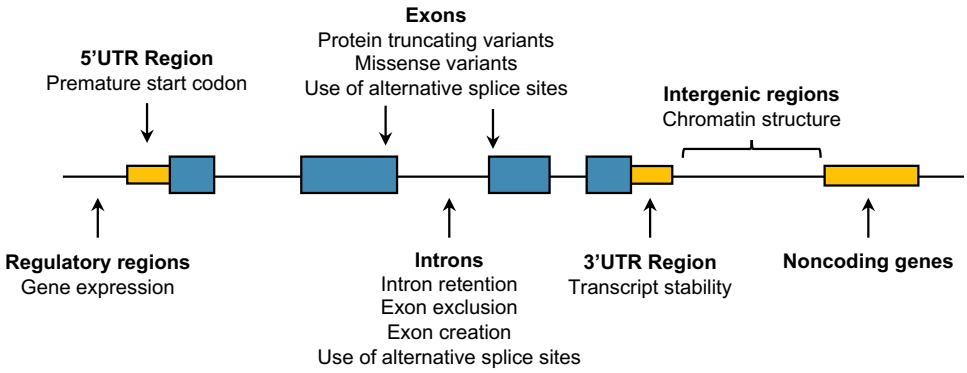


Figure 5. Types of genomic variants and their consequences on gene and protein function.

- **Intronic regions**: Intronic regions can contain variants with splicing altering consequences, which may alter transcript structure. These variants may cause intron retention, exon exclusion or creation, and the use of 3' and 5' alternative splicing sites. The alterations of transcript structure may generate premature stop codons that trigger the degradation of the aberrant transcript[182].
- **Promoters and enhancers**: Promoters and enhancers are regulatory regions which are implicated in the modulation of gene expression. Variants affecting these regions can alter the affinity of certain transcription factors, affecting the expression levels of the regulated genes[183,184].
- **Untranslated regions (UTRs)**: UTRs are sequences that are part of the mature mRNA (they are transcribed), but they are not translated into protein. The 5'UTR region is the nucleotide sequence before the translation initiation codon, and the 3'UTR region is the nucleotide sequence after the termination codon. Variants in the 5'UTR can affect gene functioning creating premature start codons, affecting splicing, and affecting the interaction with ribosomes and transcript elongation. On the other hand, 3'UTRs are implicated in transcript stability, and variants in these regions may affect miRNA interaction, splicing, and poli-A tail addition signals[183].
- **Noncoding RNAs**: noncoding RNAs are those that will not end up generating a protein, such as transference RNAs (tRNAs), ribosomal RNAs (rRNAs), microRNAs (miRNAs), small nuclear RNAs (snRNAs), and long non-coding RNAs (lncRNAs)[183].

- **Intergenic regions**: Despite not containing gene sequences, intergenic regions can contain functional elements important for chromatin structure. One of their functions is to delimit regions known as topologically associated domains (TADs), which are chromatin domains folded in a way that enables the interaction between functional elements contained inside them, such as genes, promoters, and enhancers, to modulate gene expression. TADs are delimited by sequences which can be bound by proteins that structure chromatin, and variants affecting these sequences can cause TAD restructuring in such a way that prevents or enhances ectopic interaction between functional elements and influencing gene expression[183].

Structural variants

The extension and the uniformity of coverage in WGS data also enable the resolution of complex structural variants (SVs), which usually have their breakpoints in noncoding regions. SVs can consist in different combinations of gains, losses, and rearrangements of the DNA sequence, which are usually classified in deletions, duplications, inversions, insertions, translocations, and complex rearrangements[185]. These variants are less common than smaller genetic variants such as SNVs and indels but have a greater potential phenotypic impact since given their size have higher probabilities of altering gene structure or dosage, causing alterations in the functionality or regulation of affected genes[185,186].

SV detection from short-read NGS data can be performed using different methods, including the analysis of read depth, discordant read pairs, and split reads, or a *de novo* assembly approach of the sequencing data[186,187]. Methods beyond read depth analysis enable the detection of balanced SVs which are nearly impossible to capture in WES and can resolve their breakpoints at nucleotide level.

With the progressive decrease in sequencing costs, the current limitation of WGS consists in the interpretation of the vast quantity of data generated, due to the lack of insight into the functional consequences of most of these variants. Even though some noncoding and structural variants may have potential disease-causing consequences, they are usually excluded from WGS analysis or are classified as variants of unknown significance because of the complexity of predicting their

biological impact with no functional evidence. This is why complementary omics technologies such as transcriptomics and proteomics can play a role in the prioritization and interpretation of variants detected by WGS.

3.1.2. Transcriptomics and proteomics

Aberrant RNA phenotypes

Transcriptomics consists in the study of gene expression (i.e., transcripts) in a biological sample. The current gold standard for transcriptomics analysis is RNA sequencing (RNAseq), which is an NGS approach that enables the capture, sequencing, and quantification of multiple transcripts in a single experiment. Transcriptome analysis provides functional information of the impact of genetic variants at a molecular level. RNAseq data can be used to detect different aberrant molecular phenotypes which may indicate gene dysfunction[188,189].

- **Aberrant expression**: Aberrant expression analysis is based on the detection of expression outliers, i.e., samples with expression levels of one or several genes significantly higher or lower than the rest of the population. Low expression may be caused by truncating variants that cause transcript degradation via the activation of the nonsense-mediated decay (NMD) system, variants in regulatory regions affecting gene expression, or SVs causing a rearrangement or loss of genetic material that breaks the reading frame of a gene.
- **Aberrant splicing**: The analysis of exon-exon junctions enables the detection of alterations in splicing patterns, such as exon exclusion, intron retention and the use of alternative splicing sites. The detection of aberrant splicing helps the correct interpretation of variants in non-canonical splicing regions and deep intronic variants.
- **Monoallelic expression**: The analysis of allelic frequency in RNAseq reads allows to determine whether there is a preferential transcription of one of the two alleles. This may indicate that the allele with lower expression may have a loss-of-function variant causing transcript degradation, or a regulatory variant hindering its transcription. In cases where one of the two alleles is significantly more expressed than the other, a variant identified in heterozygosity in DNAseq may still

be causing a disorder with a recessive mode of inheritance, since its expression is hemizygous.

Aberrant protein phenotypes

Mass spectrometry-based proteomics enable the identification and quantification of all proteins present in a sample (the proteome) in a single experiment[190]. Tandem mass tag (TMT) proteomics allow to multiplex several samples in a single experiment, reducing the costs[191]. As with RNAseq aberrantly expressed transcripts, expression outliers in proteomics may help reprioritize variants detected by WGS within those genes and might indicate that those variants affect protein stability or post-translational modifications[192,193]

3.1.3. Multi-omics data integration for diagnosis

Transcriptomics and proteomics evidence enables the reinterpretation, and, in some cases, the reclassification of variants identified in WGS, whose functional consequences are not accurately predicted *in silico*. Transcriptome and proteome analyses provide information about gene activity and are a strong indication of whether a functional protein is being produced (Figure 6).

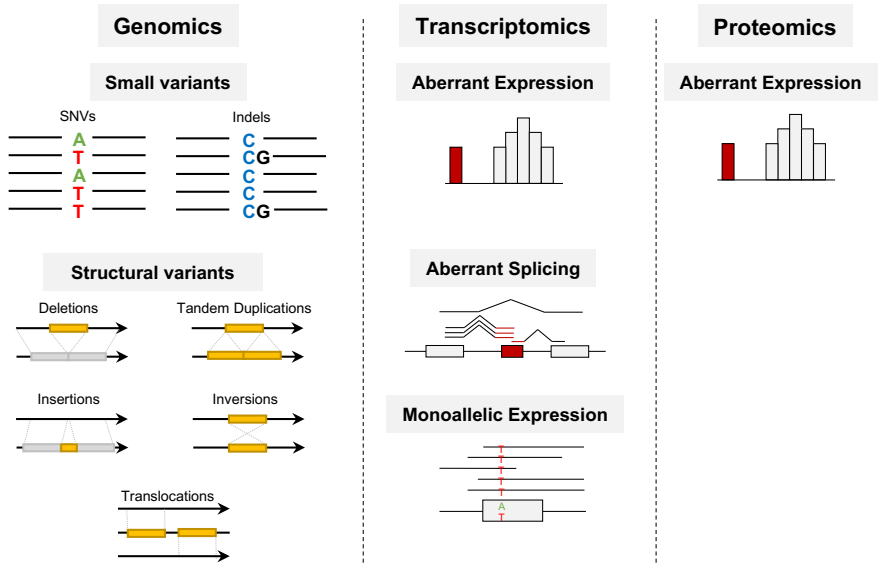


Figure 6. Genetic variation identified in genomics, transcriptomics, and proteomics. Figure modified from Yépez et al. (2021)[188].

Detecting normal gene expression levels does not necessarily imply that the protein is functioning adequately, but detecting alterations in gene expression constitutes strong evidence of gene dysfunction. Using multi-omics data, several studies have unravelled previously unsolved cases, increasing diagnostic yield in 10-36%[194–196].

Since gene expression is tissue and cell-type specific, selecting a relevant tissue to conduct the complementary omics analyses is crucial. In neuromuscular disorders, for example, muscle biopsies have proven to generate more robust data for RNAseq analysis compared with blood and cultured fibroblasts[197]. The most relevant tissue to study RTT-spectrum phenotypes would probably be the brain, which is not an easily accessible sample to obtain. However, recent studies show that cultured fibroblasts reliably express almost 70% of disease-causing genes registered in the OMIM database, and they express 70-75% of known RTT-spectrum-related genes according to GTEx data[193,198]. These data suggest that a minimally invasive procedure, such a skin biopsy, could provide a useful sample for multi-omics analyses in RTT-spectrum patients with no molecular diagnosis.

3.2. Omics for the characterization of molecular alterations in disease

Besides providing data for disease-causing variant interpretation, omics experiments offer a high-throughput, agnostic framework to comprehensively characterise the molecular landscape in patients with rare disorders.

Transcriptomics and proteomics generate a picture of gene expression landscape that can be analysed searching for patterns in biological samples from a disease group. Differential expression analysis is an approach that compares two or several groups of samples and detects significant variations in gene expression levels between groups and may be applied through specific algorithms onto both transcriptomics and proteomics data. RNAseq and TMT-proteomics experiments can capture expression data from thousands of genes in each experiment, which allows for comprehensive profiling, but also generates false-positive hits even when correcting for statistical significance in multiple testing. Therefore, integrating data in a multi-omics analysis can highlight the most robust differential expression hits.

Differential expression analysis can yield from tenths to thousands of hits per analysis. In order to understand some of the biological implications of these results, enrichment analysis attempts to stratify them under biological processes, molecular functions, cellular components or known molecular pathways or gene sets. This highlights whether some set of genes is commonly deregulated, indicating that some particular functions or pathways could be implicated in the disease pathogenesis and suggesting potential therapeutic targets to revert the molecular effects of the malfunctioning of the disease-causing gene.

OBJECTIVES

The main objective of this work is to characterise the molecular landscape of RTT and RTT-spectrum disorders to identify potential biomarkers or therapeutic targets.

The specific objectives are:

1. To determine the degree of implication of X chromosome inactivation in phenotypic heterogeneity in patients with classical RTT with *MECP2* mutations, as a potential modifier of the disease phenotype.
2. To characterize the alterations of GABAergic synapse proteins expression as a consequence of MeCP2 deficiency and as a potential therapeutic target for RTT.
3. To characterize gene expression alterations in fibroblasts from RTT patients, as well as common alterations shared in fibroblasts from RTT-spectrum patients that may constitute bases explaining phenotypic overlap, contributing to the understanding of the pathomechanisms behind these disorders and to the identification of potential biomarkers and therapeutic targets.
4. To design and apply a diagnostic pipeline integrating data from genomics, transcriptomics, and proteomics to maximize the diagnostic efficiency in RTT-spectrum patients.

MATERIAL AND METHODS

1. Samples and data

1.1. Patients

1.1.1. Patient recruitment and informed consent

All studies were approved by the ethical committees of Hospital Sant Joan de Déu, CEIC: Comitè d'Ètica d'Investigació Clínica – Fundació Sant Joan de Déu. Written informed consent from the legal guardians of the patients was obtained in accordance with the corresponding ethical protocols to perform the genetic studies, and tissue samples from patients and controls were obtained according to the Helsinki Declaration of 1964, as revised in 2001.

1.1.2. Patient sample collection

Peripheral blood samples

Samples of blood genomic DNA (gDNA) were obtained from peripheral blood leukocytes. 221 patients with one of the following recurrent mutations in the *MECP2* gene were recruited for X chromosome inactivation (XCI) analysis, of which 174 had available blood samples and 181 had available clinical data to complete the checklists for the calculation of clinical severity scores (CSS) (Table 3).

***MECP2* Variant**	**Total**	**Blood**	**CSS**
c.455C>G | p.Pro152Arg	6	6	5
c.473C>T | p.Thr158Met	36	33	30
c.502C>T | p.Arg168*	38	29	34
c.674C>G | p.Pro225Arg	2	2	1
c.763C>T | p.Arg255*	47	36	39
c.806delG | p.Gly269fs	13	11	7
c.808C>T | p.Arg270*	31	20	29
c.880C>T | p.Arg294*	21	20	13
c.916C>T | p.Arg306Cys	25	15	22
Large deletions	2	2	1
Total	**221**	**174**	**181**

Table 3. Patients recruited for XCI studies. Total number of patients with each MECP2 mutation, number of patients with available blood samples (Blood) and with available clinical data to calculate clinical severity scores (CSS).

Moreover, 80 patients were recruited for genetic testing using several NGS approaches. 67 patients were recruited for clinical exome sequencing (CES), 50 patients for whole exome sequencing (WES), and 10 patients and their parents were recruited for whole genome sequencing (WGS) in trio analysis. Blood DNA samples from parents were also used when available to validate the segregation patterns of candidate variants identified in singleton CES and WES analyses.

Post-mortem brain samples

For XCI analysis, we used samples of brain gDNA obtained post-mortem from several brain regions (frontal, occipital, temporal and parietal cortex; white matter, brain stem, striatum and cerebellum) of two patients with the recurrent *MECP2* c.763C>T mutation. RNA was also obtained from the frontal and occipital cortices of such patients. DNA samples were isolated using the saline extraction kit PUREGENE® DNA Isolation Kit of Gentra Systems®, and brain RNA samples were extracted using TRIzol Reagent from Invitrogen.

For RNAseq, we used post-mortem brain samples from the same two RTT patients and an intra-assay control for RNAseq (i.e., an extra sample that due to its pathology could not be used as a *bona fide* control, but that allowed us to run technical verifications). The patients were 10 and 15 years old at *exitus*. RNA was isolated from two brain regions (frontal and occipital cortex).

Skin biopsies

47 individuals (34 patients and 13 healthy age-matched controls) participated in the differential expression study. Patients were recruited after clinical and genetic confirmation of their pathology as described elsewhere[150]. We studied 22 patients with RTT and mutations in *MECP2* (21 females, 1 male); 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). Clinical severity of patients with RTT and RTT-like phenotypes was measured using the clinical severity score designed by Dr Pineda[21]. Skin biopsies were obtained from the internal part of the upper arm and primary fibroblast cell lines were established using standard procedure.

DNA was isolated from fibroblast cell lines using the DNAeasy Blood & Tissue Kit (Qiagen) according to manufacturer's instructions. XCI was performed in all female samples as described by Allen et al.[199]. XCI was considered skewed with an allele ratio of 80:20 or greater. RNA was isolated from cultured fibroblast pellets using the RNeasy Fibrous Tissue Mini Kit (Qiagen) following 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.

1.1.3. Patient phenotype evaluation and correlation analysis

When clinical data were available, the RTT phenotype was evaluated, and a score was assigned following the scale of evaluation of the RTT phenotype published by Monrós, *et al.*[21]. The linear correlation between the inactivation patterns of the mutated alleles and the CSS of each patient was evaluated using statistical methods that are based on Ordinary Least Squares (OLS) regression models, grouping patients with the same mutation.

1.2. Mouse colony

Cortex samples from one and six-month old *Mecp2*+/- female mice Bird's model[200] (B6.129P2(C)-Mecp2tm1.1Bird/J) were obtained after mouse sacrifice and brain dissection. The proteins from cortex were extracted by tissue homogenization with an ice-cold RIPA buffer with protease inhibitors (complete, mini, EDTA-free protease inhibitor cocktail, Merck), for 30 minutes at 4ºC followed by 15 minutes of centrifuge at 4ºC. The protein samples were quantified by Bradford assay and stored at -80ºC.

1.3. External public data

Publicly available data from the GTEx project were downloaded and used for comparison with the samples used in this study. Gene expression counts of all GTEx samples and all available brain areas (amygdala, basal ganglia, cerebellar hemisphere, cerebellum, cortex, hippocampus, hypothalamus, spinal cord and susbstantia nigra) were used to assess gene expression levels in the brain and compare them with the transcriptomic landscape identified in fibroblast cell lines used in this study. Cortex gene expression counts of five samples (GTEX-12126-1026-SM-5P9JJ, GTEX-15ER7-3126-SM-7KUGH, GTEX-T2IS-3026-SM-32QPM,

GTEX-T5JC-2426-SM-3NMDB, and GTEX-WHSE-3026-SM-3P5ZH) were used for transcriptomic analysis of GABAergic synapse proteins in comparison to our patients.

2. Cell culture

2.1. Primary fibroblast cell culture

Fibroblast lines were grown on plates with Dulbecco's Modified Eagle's Medium (DMEM) high glucose with glutamine, supplemented with 10% heat-inactivated foetal bovine serum (FBS) and 1% penicillin, streptomycin and B amphotericin (Thermo Fisher). Cultures were kept at 37ºC with 5% CO_2 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 la Investigació, which is integrated into the Spanish Biobank Network of ISCIII for the sample and data procurement.

2.2. Neuro2a cell culture

Immortalized Neuro2a cells (also known as N2A cells, a fast-growing mouse neuroblastoma cell line) were grown following standard conditions in DMEM supplemented with 1% glutamine, 10% FBS, and 1% penicillin, streptomycin and B amphotericin (Thermo Fisher).

2.3. Neuronal primary cell culture

For neuronal primary cell cultures, the protocol described in the literature was followed[201]. All of the experimental procedures were carried out according to European Union guidelines (Directive 2010/63/EU) and following protocols that were approved by the Ethics Committee of the Bellvitge Biomedical Research Institute (IDIBELL). Briefly, mouse embryos (embryonic day E18) were obtained from pregnant CD1 females, the cortexes were isolated and maintained in cold Hank's Balanced Salt Solution supplemented with 0.45% glucose (HBSS-Glucose) and digested mildly with trypsin for 17 minutes at 37ºC and dissociated. The cells were washed three times in HBSS and resuspended in a Neurobasal medium

supplemented with 2mM Glutamax (Gibco) before filtering in 70mm mesh filters (BD Falcon). The cells were then plated onto glass coverslips (5 x 10^4 cells/cm^2) coated with 0.1mg/mL poly-L-lysine (Sigma), and 2h after seeding, the plating medium was substituted by a complete growth medium, consisting of a Neurobasal medium supplemented with 2% B27 (Invitrogen) and 2mM Glutamax.

3. Molecular biology

3.1. X chromosome inactivation assays

3.1.1. Enzymatic digestion

Digestion of gDNA samples was performed with one of the methylation-sensitive restriction enzymes *HpaII* or *HinfI* (New England Biolabs Inc.), depending on the presence of the relevant enzyme target sequences near the studied loci. In the androgen receptor gene, c.455C>G, c.473C>T, c.502C>T, c.674C>G, c.763C>T, c.806delG, c.808C>T, c.880C>T, c.916C>T and deletion 2 (NM_004992.3: c.887_10015 + 18460del) loci assays *HpaII* was used, while in the deletion 1 (NM_004992.3: c.27-10677_1192del) locus assay *HinfI* was used. A total volume of 500ng of gDNA was digested with 0.5μL of enzyme in a 25μL reaction volume in CutSmart 1x Buffer (New England Biolabs Inc.). Digestions were incubated at 37ºC for 20 minutes followed by another 20 minutes at 80ºC for enzyme inactivation, as established in the enzyme protocol.

3.1.2. PCR amplification and fragment analysis

A pair of primers with the sequences described in Cutler Allen *et al.* was used to amplify the AR polymorphic locus[199]. Allele-specific primers were designed for each *MECP2* recurrent mutation included in the study. Primer design was carried out following the recommendations in Liu, *et al.* [202]. For the deletion assays, a forward primer was designed inside the deletion locus and another primer immediately after the deletion; they were both amplified with a reverse primer outside the deleted region. All primers used were designed using Primer3 Software v4.1.0[203,204]. One primer of each pair was FAM-labelled at the 5' end. Primers and PCR conditions are shown in Table 4, Table 5 and Figure 7.

Locus	Forward Primer		Reverse Primer	
AR	CCAGAATCTGTTCCAGAGCGTGC		*GCTGTGAAGGTTGCTGTTCCTCAT	
MECP2 c.455C>G (p.Pro152Arg)	**WT**	GCGACACATCCCTGGACTC	*AGCTTCCCAGGACTTTTCTCC	
	Mut	GCGACACATCCCTGGACTG		
MECP2 c.473C>T (p.Thr158Met)	**WT**	GGACCCTAATGATTTTGACTTTAC		
	Mut	GGACCCTAATGATTTTGACTTTAT		
MECP2 c.502C>T (p.Arg168*)	*TCGAAAAGGTAGGCGACACATC		**WT**	CTTAGGTGGTTTCTGCTCCCG
			Mut	CTTAGGTGGTTTCTGCTCCCA
MECP2 c.674C>G (p.ProP225Leu)			**WT**	CCCCTGGCGAAGTTTGAAAGG
			Mut	CCCCTGGCGAAGTTTGAAAGC
MECP2 c.763C>T (p.Arg255*)	**WT**	AAACGCCCCGGCAGGAGGC	*AGTCCTTTCCCGCTCTTCTC	
	Mut	AAACGCCCCGGCAGGAGGT		
MECP2 c.806delG (p.Gly269fs)	**WT**	ATTCCCAAGAAACGGGGCCG		
	Mut	ATTCCCAAGAAACGGGCCGA		
MECP2 c.808C>T (p.Arg270*)	**WT**	CATTCCCAAGAAACGGGACC		
	Mut	CATTCCCAAGAAACGGGACT		
MECP2 c.880C>T (p.Arg294*)	*TCGAAAAGGTAGGCGACACATC		**WT**	GGTCTCCTGCACAGAACG
			Mut	GGTCTCCTGCACAGAACA
MECP2 c.916C>T (p.Arg306Cys)	**WT**	ACCGTACTCCCCATCAAGAGGC	*TCTGAGTGGTGGTGATGGTG	
	Mut	ACCGTACTCCCCATCAAGAGGT		
MECP2 c.27-10677_1192del (**deletion 1**; exons 3-4)	**WT**	ACATTCAGGTTGGCTTGGTC	*CTCTGGGCATCTTCTCCTCTT	
	Mut	AAAAATAAAATGTGCAATTCAGTGTC		
MECP2 c.887_10015+18460del (**deletion 2**; exon 4)	**WT**	TCTGAGTGGTGGTGATGGTG	*CAGGCCATTCCCAAGAAAC	
	Mut	CTCAGCTGGAAGGGAAAATG		

Table 4. Primer sequences for XCI assays. Primer sequences to amplify each locus. Bold yellow formatting indicates the mutated nucleotides and bold red formatting indicates the mismatch introduced for primer specificity. The bold green formatting indicates the nucleotide only present in the WT allele and not in the mutated allele with the G deletion. * Indicates the FAM label at the 5' end of the primer to allow detection in fragment analysis. Mutation nomenclature is referenced to *MECP2* NM_004992.3.

	PCR Mix 1		PCR Mix 2		PCR Mix 3	
	Volume	*Final concentration*	*Volume*	*Final concentration*	*Volume*	*Final concentration*
H_2O	19.3μL	-	18.25μL	-	17.25μL	-
Buffer	2.5μL	1x	2.5μL	1x	2.5μL	1x
$MgCl_2$	*in buffer*	2.5mM	*in buffer*	2.5mM	1μL	1mM
dNTPs	0.5μL	0.2mM	1μL	0.4mM	1μL	0.4mM
Primers	0.5μL + 0.5μL	0,2μM	1μL + 1μL	0.4μM	1μL + 1μL	0.4μM
Taq Pol	0.2μL	0.028U/μL	0.25μL	0.035U/μL	0.25μL	0.035U/μL
Mix	23.5μL		24μL		24μL	
DNA	1.5μL	1.2ng/μL	1μL	0.8ng/μL	1μL	0.8ng/μL
Total	25μL		25μL		25μL	

Table 5. Amplification conditions. Amplification conditions for all assays. PCR Mix 1 was used for the AR, c.455C>G, c.473C>T, c.502C>T, c.763C>T, c.808C>T and c.916C>T assays; PCR Mix 2 was used for the c.674C>G, c.880C>T assays and the deletion assays; PCR Mix 3 was used for the c.806delG assay. The buffer used in PCR Mix 1 and 2 already contains $MgCl_2$, whereas in PCR Mix 3 the $MgCl_2$ is added separately.

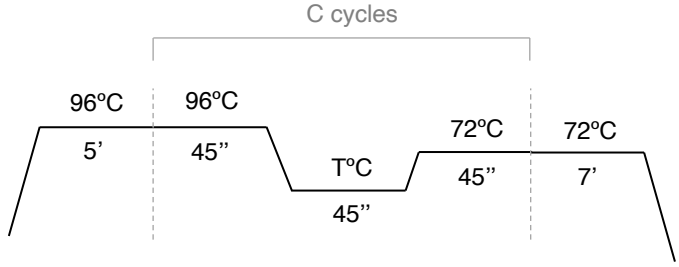


Figure 7. Amplification programs. Annealing temperature (T) varied between primer pairs: AR – 60ºC; c.455C>G – 58ºC; c.473C>T – 61ºC; c.502C>T – 61ºC; c.674C>T – 60ºC; c.763C>T – 60ºC; c.806delG – 61ºC; c.808C>T – 58ºC; c.880C>T – 56ºC; c.916C>T – 60ºC; deletion assays – 59ºC. The number of cycles (C) also varied between the AR assay (25 cycles) and the allele-specific assays (30 cycles).

PCR amplification was performed on gDNA before and after the digestion of each sample. PCR products were analysed on a 3500 Genetic Analyzer (Applied Biosystems) using GeneScan – 500LIZ Size Standard of Applied Biosystems as an internal size standard and Peak Scanner Software v1.0. The X chromosome inactivation ratios were calculated as described elsewhere[199].

3.1.3. RNA analysis

RT-PCR was performed with frontal and occipital cortex RNA of two patients with the c.763C>T mutation, following the recommendations provided with the SuperScript III First-Strand Synthesis SuperMix for qRT-PCR from Invitrogen. Subsequently, Sanger sequencing of the cDNA obtained in the RT-PCR reaction was performed with primers and conditions indicated in Table 6.

Locus	Forward Primer	Reverse Primer
MECP2 c.763C>T (p.Arg255*)	GGACCCTAATGATTTTGACTTTAC	GGTCTCCTGCACAGAACG

Table 6. Primer sequences for Sanger sequencing c.763C>T variant.

qPCR was performed in a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems) with TaqMan Gene Expression Master Mix (Applied Biosystems) and specific TaqMan MGB probes to amplify the mutant and the wild-type alleles. qPCR data were analysed using the comparative Ct method. Primers and probes were designed using Primer3 Software v4.1.0 [203,204] and they are listed in Table 7.

Forward Primer	CCAGGTCATGGTGATCAAACG
Reverse Primer	AGACTCCTTCACGGCTTTCT
***MECP2* c.763C>T *C* Probe (WT)**	CAGCTTTTC**G**CTTCCTG
***MECP2* c.763C>T *T* Probe (Mut)**	CAGCTTTTC**A**CTTCCTG

Table 7. TaqMan probe and primer sequences. Primer and probe sequences for TaqMan qPCR assay. Bold coloured formatting indicates the different nucleotides in the TaqMan probes.

3.2. *MECP2* deficiency experiments

3.2.1. Plasmids and transfection

In certain experiments, we attempted to silence the endogenous *MECP2* expression and re-express the human gene either in the wild type or mutated form. For the *MECP2* silencing, transient transfection with a vector containing shRNA targeting m*MECP2* was performed. Silencing was done with the MISSION shRNA technology (Sigma Aldrich; Clone TRCN0000304464), and the efficiency was checked at the protein level. To ensure the sole silencing of the endogenous gene, and not the re-expressed form, we used a shRNA directed to the 3'UTR region of the gene, absent in the transfected cDNA. We used, as a control for transfection, a vector with the same backbone but no shRNA.

The MeCP2 c.763C>T genetic variant was introduced by site-directed mutagenesis using the QuickChange II XL Kit (Agilent Technologies), in the pEGFP-C1-hMeCP2 (WT) mammalian expression vector (kindly provided by Dr. Landsberger). The mutation was confirmed by Sanger sequencing. Both vectors, together with the p.EGFP-C1 vector (BD Clontech) and a mock vector with the same backbone, were used for the experiments that required plasmid transfections.

These were carried out using Lipofectamine 2000 (Thermofisher) following manufacturer recommendations. For neuronal primary cultures 0,8μg of total DNA was mixed with Lipofectamine 2000 and incubated with cortical neurons (at DIV11). The transient expression was allowed for 96h, and the neurons were fixed at DIV14. For the N2A cells, 4μg of DNA was transfected over 300,000 cells grown in 10cm^2 plates, and the cells were collected four days after transfection for RNA analysis.

3.2.2. qRT-PCR

RNA for RNAseq (from post-mortem human brain samples) and for qRT-PCR (from post-mortem human brain samples and N2A cells) was extracted using RNeasy Fibrous Tissue Mini Kit (Qiagen), following manufacturer's instructions. The total RNA was eluted in 40μL of RNAse-free water and stored at -80ºC. The RNA concentration was measured using the NanoDrop 2000 Spectrophotometer

(ThermoScientific), and integrity was assessed by running the samples through a 1% agarose gel.

qPCRs were carried out following a two-step protocol. First, cDNA was synthesised from a total of 500ng of RNA per reaction, following the recommendations provided with SuperScript III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen). After the RT-PCR reaction, the post-mortem brain cDNA from frontal and occipital cortex samples was pooled. Second, qPCR was performed in a QuantStudio 6 Felx Real Time PCR sustem (AppliedBiosystems) with PowerUp SYBR Green Master Mix (AppliedBiosystems). The data were analysed using a comparative method, correlating the initial template concentration with the cycle threshold (Ct) so as to obtain the relative quantity (RQ) of the RNA. The RQ is defined as $2^{-\Delta\Delta Ct}$, where $\Delta\Delta Ct$ is the ΔCt of the patient cell line minus the ΔCt of the control cell line, and ΔCt is the Ct of the target gene minus the Ct of the endogenous gene (*RPLP0* and *GUSB*).

The probes design was done through the selection of the exonic areas present in all of the transcript variants of each gene, by the selection of common fragments in the UCSC genome browser, based on GRCh38/hg38 version. Primers for N2A-derived qPCR experiments were, in 5'-3' sense, as follows:

Locus	Forward Primer	Reverse Primer
m-*Mecp2*	ACCATCATCACCACCATCAC	GGGCATCTTCTCTTCTTTGC
h-*MECP2*	AGGAGAGACTGGAAGAAAAGT	CTTGAGGGGTTTGTCCTTGA
m-*Gabra1*	ACCAGTTTCGGACCAGTTTC	TACAGCAGAGTGCCATCCTC
m-*Gabrb2*	TCGCTGGTTAAAGAGACGGT	TCTCTCCAGGCTTGCTGAAA
m-*Gabrg2*	TGGGCTACTTCACCATCCAG	GCCATACTCCACCAAAGCAG

Table 8. Primer sequences for gene expression quantification in transfection experiments.

3.2.3. Western blotting and ICC

Western blot analysis of the cortex protein samples from *Mecp2*$^{+/-}$ female mice was performed. The proteins were subjected to SDS-PAGE and transferred to a nitrocellulose membrane using the Pierce Power Station (Thermo Scientific). The membranes were blocked with milk, as follows: PBST buffer 5% for 1h at room temperature. Incubation with primary antibodies was directed against GABA-A1

(Neuromab, 73-136) at a concentration of 1:500, MeCP2 (ab2828; Abcam) at a concentration of 1:1000, and vinculin (ab129002, Abcam) was performed o/n. The secondary antibodies used were horseradish peroxidase-conjugated goat anti-rabbit and goat anti-mouse IgG antibodies (Life Technologies); these were detected using the Enhanced Chemiluminescence System (GE Healthcare).

Immunochemistry experiments were performed as described in the literature[201]. Anti-GABA A1R (Neuromab, 75-136) was used at a concentration of 1:100, and MeCP2 (ab2828; Abcam) at a concentration of 1:250. The conjugated secondary antibodies for the confocal microscopy were used.

Fluorescence was visualized with a Leica TCS-SL spectral confocal microscope (Leica Microsystems) using a Plan-Apochromat 63x/1.4 N.A. immersion oil objective (Leica Microsystems). To excite the different fluorophores, the confocal system was equipped with excitation laser beams at 488 and 546nm. The images were analysed with ImageJ software. For the intensity quantification of the ICCs, pictures from independent primary cultures were used. Regions Of Interest (ROIs) were defined on the green channel (GFP positive neurons) applying a threshold to only select the desired neuron. The same ROI was exported to the red channel pictures, and the mean grey value was measured.

3.2.4. RNAseq data analysis

A frontal and occipital cortex paired-end RNAseq was performed for two RTT patients bearing the *MECP2* c.763C>T, p.Arg255* mutation and the mentioned intra-assay control. Technical replicates (three separate RNA extractions) were run for each of the two brain regions. The RNA samples were sent to the Centre Nacional d'Anàlisi Genòmica (CNAG), where the RNAseq experiment was performed, within the framework of the project FIS PI15/01159 Rett Syndrome (IP: Judith Armstrong, PhD). Both brain areas were sequenced separately, and as no significant differences were observed between them, they were analysed as a sole data pool. To discard the differences between the brain areas, we performed a principal component analysis (PCA). In such an analysis, we compare the variance between all of the samples (patients and areas). At this point, we also added an "intra-assay control" sample. Principal Component 1 (explaining 75% of the variance), clearly

discriminated between the Rett and non-Rett samples, while there was not a principal component separating the brain areas.

Because of the lack of true control data, the data from our RNAseq were compared to various public controls' data. We used data from public controls available on the GTEx (Genotype-Tissue Expression) Portal. Cortex RNAseq data from five controls were used, two of which were female (GTEX-15ER7 and GTEX-T2IS) and three were male (GTEX-12126, GTEX-T5JC, and GTEX-WHSE), all of them with ages ranging 20-29.

As a result of the low performance of the sequencing experiment, the internal control's occipital cortex data was excluded from the analysis. The RNAseq analysis pipeline was run by the Bioinformatics Unit form the Genetics and Molecular Medicine Service at the Hospital Sant Joan de Déu. The FASTQ files passed through a first quality control, after which a trimming was performed, and the adapters were removed. Then, low quality bases were eliminated so only reads longer than 70bp were left to continue the analysis. Here, a second quality control was performed, and the reads were mapped with TopHat2[205]. The counting was performed with HTseq[206] and the R package DESeq2[207] was used for library normalization and differential expression analysis. The frontal and occipital cortex data from the two RTT patients were averaged. In order to compare the data obtained from our RNAseq experiment to public data, an internal normalization over the endogens *RPLP0* and *GUSB* of every patient's and control's data was performed.

A validation of the results was carried out at the Hospital Sant Joan de Déu using RT-qPCR, comparing patients with an internal RNAseq control; that is, a sample that could be used for a later comparison of the results through qPCR and therefore RNAseq technical validation, but could not be used to biologically validate the results, as it was not a healthy brain, as previously described. Thus, following the RNAseq analysis, the targeted gene expression of a subset of 21 genes corresponding to different nodes of the GABAergic pathway and differentially expressed between patients and the internal control was validated by RT-qPCR, showing a strong correlation (20 out of 21 transcripts deregulated), with an overall

coincidence between RQ values (RT-qPCR experiments) and fold-change (RNAseq experiment).

3.3. Next-Generation Sequencing

3.3.1. DNAseq library prep

DNA samples were obtained from peripheral blood leukocytes or primary fibroblast cell cultures depending on availability. WGS libraries were generated following a PCR free protocol (Illumina) and were sequenced using paired-end 150bp reads in a NovaSeq6000. CES and WES libraries were generated using SureSelect (Agilent Technologies) probes following manufacturer's instructions and were sequenced using paired-end 75pb reads in a NextSeq500.

3.3.2. RNAseq library prep

For each sample, 2500ng of RNA were used for library preparation. TruSeq Stranded mRNA kit (Illumna) was used following the manufacturer's protocol. Libraries were quantified in a 4200 TapeStation (Agilent Technologies) and their integrity was checked. Sequencing was performed on an Illumina NextSeq500 sequencer and 75bp 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.

3.3.3. RNAseq results validation

To further confirm the quality of the isolated RNA and to diminish undesirable gene alterations due to cell stress conditions[208], we performed 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, 500ng of total RNA were processed according to the manufacturer's instructions, and double-stranded cDNA was generated in the presence of random hexamers using the SuperScript III First-Strand Synthesis SuperMix for qRT-PCR kit (Thermo Fisher). Primers for the five mitochondrial genes and two additional housekeeping genes (*RPLP0* and *ALAS1*) were designed with Primer3 software. The qRT-PCR was performed with PowerUp SYBR Green Master Mix (Thermo Fisher) in a QuantStudio 6 Flex Real-Time PCR System (Applied

Biosystems). All reactions were conducted in triplicate and the average of each triplicate group was used for the analysis, which was 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.

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.

3.3.4. Proteomics prep

Proteomics was performed at the BayBioMS core facility at the Technical University Munich (TUM), Germany, as described elsewhere[191]. Fibroblast cell pellets containing 0.5 million cells were sent frozen. These cells were thawed and lysed under denaturing conditions in urea containing buffer and quantified using BCA Protein Assay Kit (ThermoScientific). 15μg of protein extract were further reduced, alkylated and the tryptic digest was performed using Trypsin Gold (Promega). Digests were acidified, desalted and tandem mass tag (TMT)-labelling was performed according as described elsewhere[209] using TMT 11-plex labelling reagent (ThermoFisher Scientific). TMT-batches were organised to always contain one reference control common to all batches to allow for data normalization between batches. Each TMT 11-plex peptide mix was fractionated using trimodal mixed-mode chromatography as described elsewhere[210]. Liquid chromatography-mass spectrometry (LC-MS) measurements were conducted on a Fusion Lumos Tribrid mass spectrometer (Thermo Fisher) which was operated in data-dependent acquisition mode and multi-notch MS3 mode. Peptide identification was performed using MaxQuant version 1.6.3.4[211]. And protein groups obtained. Missing values were imputed with the minimal value across the dataset.

4. Data analysis

4.1. XCI ratio analysis

The XCI ratio analysis was calculated as described elsewhere [199].

4.2. DNAseq analysis pipelines

4.2.1. Quality control, pre-processing, and alignment

Quality control of sequencing data was carried out using FastQC v4.11.9[212]. The cutadapt software[213] was used to eliminate the first 10 bases from the 5' end of each read and low-quality (less than 20) bases and high-quality Gs (sequencing artifacts) at the 3' end, and only reads of at least 100bp after pre-processing were kept. Pre-processed reads were aligned to the GRCh37/hg19 reference genome using the Burrows-Wheeler Alignment Tool (BWA-mem v0.7.17)[214]. Depth of coverage was over 20x in 90% of coding regions of the genome and over 30x in 50% of coding regions, indicating sufficient quality to proceed with WGS analysis.

4.2.2. Variant calling of SNVs and indels

Variant calling of SNVs and indels was carried out using three different tools: the Genome Analysis ToolKit (GATK, v4.3.0.0)[215], DeepVariant (v1.4.0)[216] and Octopus (v0.7.4)[217], chosen after considering the sensibility and specificity metrics calculated in a recent benchmarking paper[218]. GATK analysis pipeline was designed according to their best practices manuals[219,220]. The first step consists in a base quality score recalibration (BQSR), in which a new bam file is generated with new quality values, which is used for the variant calling. To perform this recalibration, two files containing known and validated SNVs and indels were used[221]. Called variants were filtered according to the GATK hard filters[222], and genotype calls were refined and filtered using a file of validated genotypes[223]. DeepVariant was used with a WGS model and with default specifications. Octopus was used with a standard filter and default specifications.

After variant calling with any software, biallelic block substitutions were decomposed in their constituent SNVs and indels were normalized (left-aligned). SNVs and indels were separated in two files which were treated independently, and variants not

meeting quality standards for each software were removed. SNVs and indels called by each software were combined, sorted by chromosome and position, and annotated with how many and which detectors had called them.

The SnpEff software was used to annotate variant effects (affected genes and transcripts, impact and predicted effects on the transcript or protein)[224]. SnpSift was used to annotate variant ids from dbSNP, population frequencies from gnomAD v2.1.1[225], functional predictions from REVEL[226], CADD and SpliceAI, variant classification form ClinVar[227], and imprinted genes from GeneImprintDB[228]. Variants affecting more than one transcript were separated in the corresponding rows and a python script was used to annotate phenotypes associated to each gene according to OMIM[198,229] and gnomAD o/e metrics.

4.2.3. Variant calling of SVs

Variant calling of SVs was performed with four different tools, chosen because of their precision and specificity metrics, and their ability to detect different kinds of SVs, according to two recent benchmarking papers[187,230]. We chose tools based on different kinds of methods to identify SVs: discordant read pairs (RP), split reads (SR), read depth (RP) and de novo assembly (AS) (Table 9).

Software	Methods	Type of SVs detected
Lumpy[186]	RP, SR, RD	DEL, DUP, INV, TRA
Manta[231]	RP, SR, AS	DEL, DUP, INV, INS, TRA
Delly[232]	RP, SR	DEL, DUP, INV, INS, TRA
Wham[233]	RP, SR	DEL, DUP, INS, INV

Table 9. SV detection tools. SV detection tools and their detection methods and types of SVs detected.

SV detection was performed using the default specifications of each software, but excluding badly solved regions in the reference genome, which include centromeres and telomeres of chromosomes, as well as low complexity regions described in Li, et al. (2014)[234]. Moreover, we filtered out from the output those variants with both break ends mapping to the same repetitive region.

To merge the sets of SVs detected by each caller we considered that the detection of breakpoints may not be exact, so each detector may identify slightly different breakpoints for the same variant. ACMG guidelines for SV detection establish a minimum of 90% of reciprocal overlap between variants to be considered the same, so we used this threshold to merge structural variant sets[235]. SVs were annotated with population frequencies in DGV, dbVar and gnomAD, and with haploinsufficiency and triplosensitivity curations from ClinGen[236,237].

4.2.4. Detection of Runs of Homozygosity (ROHs)

ROH detection was carried out using the AutoMap software fed with a VCF file containing genotype and allelic depth information, with default parameters (sliding window of 7bp, and a minimum resolution of 1Mb and 25 variants to call an ROH)[238].

4.3. RNAseq analysis pipelines

4.3.1. Quality control, pre-processing, and alignment

The FastQC v4.11.9 software was used for sequencing data quality control[212]. The cutadapt v4.1 software as used to remove low-quality bases (less than 10) from the 3' end and high-quality Gs (sequencing artifacts) from the 5' end of each read)[213]. Only sequences longer than 25bp after pre-processing were kept for downstream analysis.

RNAseq reads were aligned to the human reference genome (GRCh37/hg19) and the reference transcriptome GENCODEv34 using STAR (v2.4.2a) in a strand-specific manner[239]. We obtained approximately 40M of paired-end reads per sample, of which almost 90% belonged to exonic reads. A low proportion of rRNA reads confirmed the efficacy of the poli-A tail selection steps of the library prep protocol. We detected a variety of transcripts corresponding approximately to 18,000 genes. These data ensured the quality of data to proceed with analysis.

4.3.2. Differential expression

Uniquely mapped reads were counted for each gene using the HTSeq package (v2.0.0) with gene models from GENCODE release 29[206]. A final count matrix for analysis was generated by averaging the values of raw counts from different replicates of the same sample. Counts per mullion 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. 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)[207]. We used a Bejamini-Hochberg (BH) corrected p-value of 0.05 to consider significant differences.

4.3.3. Enrichment and upstream regulator analysis

Enrichment analysis was performed using the clusterProfiler (v4.2.2)[240] and ReactomePA (v1.38.0)[241] 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)[242], the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database[243], WikiPathways (WP)[244], and the Reactome Pathway (RP) database[245]. 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 term significantly enriched 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[246]. 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.

4.3.4. Aberrant RNA phenotypes

Aberrant Expression

BAM files obtained from the alignment against hg19/GENCODEv34 were used for aberrant expression events using the tool OUTRIDER, included as a module in the DROP framework with default parameters[188,247]. OUTIRDER computes aligned reads internally using HTseq, and it includes a denoising autoencoder to correct batch effect when combining sets of samples. A publicly available count matrix was used to complete the sample set for aberrant expression analysis. This matrix is built from 127 fibroblast cell lines, sequenced with strand-specific chemistry and about 70M of mapped reads, produced by the Technical University of Munich (TUM)[248,249].

Aberrant Splicing

The tool FRASER2, included as a module in the DROP framework, was used for the detection of aberrant splicing events, with default specifications[250]. We used a publicly available count matrix of split reads and intron-spanning reads from the same samples that we used in aberrant expression analysis with OUTRIDER. FRASER2 also uses the denoising autoencoder to combine exon junction usage counts from different sample sets, and provides metrics of this usage and a p-value to determine which samples present aberrant splicing events.

Monoallelic Expression

The monoallelic expression analysis module from DROP is fed with a VCF of detected variants in DNAseq and a BAM file from RNAseq, from which it will quantify the reads corresponding to each allele in heterozygous positions and will calculate a p-value for statistical significance of a preferential usage of one of the two alleles[188].

4.4. Proteomics analysis pipelines

4.4.1. Differential Expression

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[191]. 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. Therefore, we carried out the differential expression analysis using the limma (v3.50.3) package in R, including the MS batch as a covariate in the model to adjust for this confounding factor[251]. 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).

4.4.2. Aberrant Expression

Aberrant expression analysis was performed using the PROTRIDER software[191]. To detect protein expression outliers while controlling for known and unknown sources of variation, PROTRIDER uses a denoising autoencoder based method, analogous to methods for calling RNA expression outliers and splicing outliers[247,250]. Sizefactor normalised and log-transformed protein intensities were centred protein-wise and used as input to a denoising autoencoder model with three layers (encoder, hidden space, decoder). As protein intensities varied strongly between batches, we included the batch as a covariate in the input of the encoder and in the input of the decoder. During the process of fitting the autoencoder model as well as statistical tests, missing data are masked as unavailable and ignored.

RESULTS AND DISCUSSION

Chapter 1. X chromosome inactivation in patients with Rett syndrome

Despite carrying the same mutation, patients with RTT present a large degree of phenotypic variation. These clinical differences have traditionally been attributed, at least in part, to differences in XCI patterns. To investigate this hypothesis, we studied XCI in blood samples of RTT patients and assessed the severity of their clinical presentation using a clinical severity score (CSS)[21]. To check whether XCI status in blood was an accurate predictor of the XCI landscape in brain, the primarily affected tissue in RTT, we also compared XCI patterns in blood and several brain areas in two RTT patients. We finally assessed the expression levels of each *MECP2* allele in brain tissue and whether these levels correlated with XCI status.

1.1. Allele-specific XCI assay design

In most XCI studies in patients with RTT, the method used to assess XCI patterns was the classic methylation-based assay on the androgen receptor gene (XCI-AR; Figure 8A). This assay consists in digesting the DNA with a methylation-sensitive restriction enzyme that targets the palindromic sequence CCGG and cuts it only when it is not methylated. The androgen receptor gene contains a highly polymorphic CAG repeat close to one of these palindromic sequences, which enables the distinction of the two alleles by fragment analysis after PCR amplification. The PCR products from digested and undigested samples are quantified to assess the XCI ratio. The main limitation of this assay is that the phase of the androgen receptor alleles and the *MECP2* mutation is not determined, so the XCI pattern can only be classified as either skewed or random, but no evidence is provided of whether the preferentially inactivated chromosome is the one carrying the mutant or the wild-type allele. Moreover, given that the separation process in fragment analysis is based on PCR product size, if the patient is homozygous for the CAG repeat allele, the assay results are noninformative. To overcome these shortcomings, we designed an XCI assay specifically targeting several recurrent *MECP2* mutations.

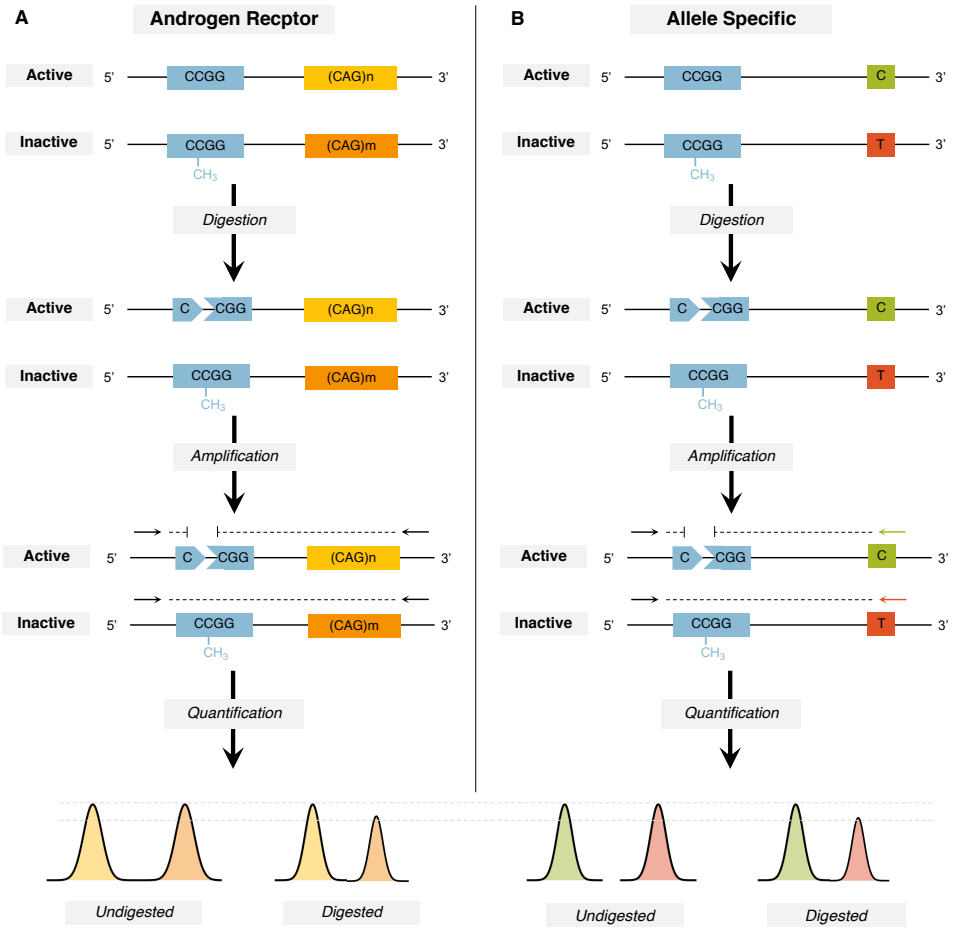


Figure 8. XCI assays outline. A) Androgen receptor assay. DNA samples are digested with a methylation-sensitive restriction enzyme that targets the palindromic sequence CCGG, which will only be cut when not methylated. PCR amplification is therefore only possible in unmethylated sequences (originating from the inactive X chromosome), and the resulting PCR products are quantified by analysing the area under the curve obtained in fragment analysis before and after enzymatic digestion, allowing to calculate XCI ratio. **B) Allele Specific assay.** Restriction enzyme digestion is performed as abovementioned. In this case, specific primers are designed for each *MECP2* mutation and two PCR reactions, one for each allele, are performed separately. The resulting PCR products are also quantified by fragment analysis before and after digestion for XCI ratio calculation.

Given that the *MECP2* gene also contains the palindromic sequence CCGG, we have developed an allele-specific XCI assay on the loci of several recurrent *MECP2* mutations based on the digestion with the same methylation-sensitive enzyme. This assay allows for evaluation of the XCI pattern while considering which is the mutant, and which is the wild-type allele (XCI-AS; Figure 8B). We designed allele-specific PCR primers for each mutation, with the mutated nucleotide in the 3' end and containing a mismatch one or two nucleotides upstream to prevent the amplification of the other allele (Table 4, Material and Methods). The use of allele-specific PCR primers to selectively amplify only the wild-type or the mutant *MECP2* allele allows for specific quantification after fragment analysis.

We compared the results from both assays, the XCI-AR and the XCI-AS (Figure 9).

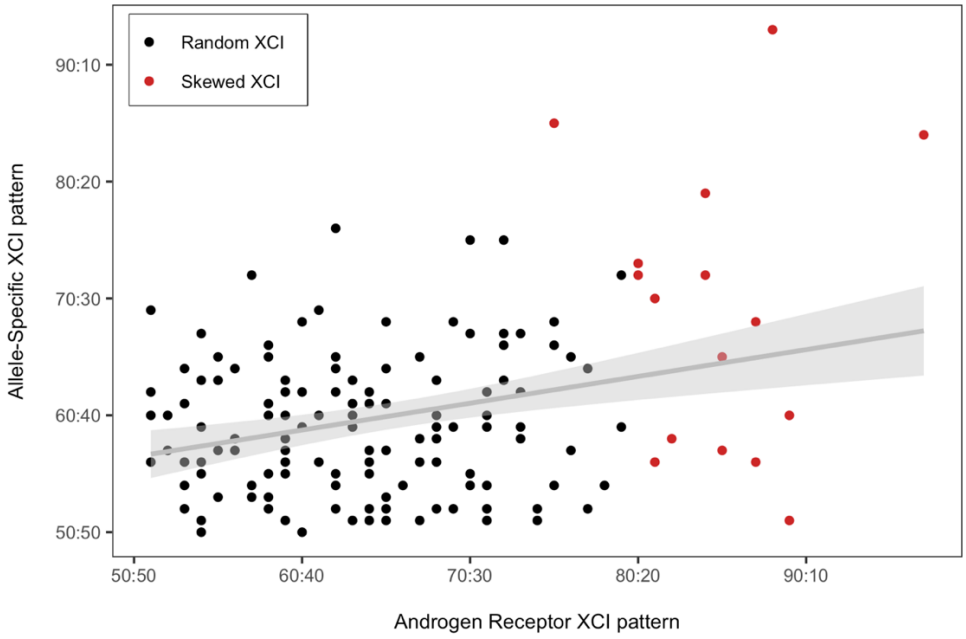


Figure 9. Comparison of XCI patterns with the two assays: androgen receptor and allele-specific. Samples with skewed XCI (80:20 or higher) according to at least one assay were coloured in red. The correlation between the two assays is depicted by the grey line (r = 0.3, p-value < 0.001).

Although the concordance is limited, there is a significant correlation between the results of both assays (r = 0.3, p-value < 0.001). Differences between the XCI

patterns obtained by both assays may be explained because in each of them the methylation status is only analysed at a single locus, and the methylation of a single cytosine residue may not be representative of the inactivation status of the entire X chromosome[252,253]. Different studies have shown that when methylation is assessed in several loci of the X chromosome, the calculated XCI ratios can vary, with different of up to 27%[252,253]. Therefore, the use of several loci for characterising XCI would indicate the true XCI pattern more consistently[252].

The XCI-AS assay allowed us to describe XCI patterns of patients previously classified as noninformative by the classical XCI-AR assay and to identify which *MECP2* allele (WT or mutated) was preferentially inactivated in cases of skewed XCI pattern.

1.2. Allele-specific X chromosome inactivation and skewing in blood samples

We recruited 221 patients to study XCI patterns in blood and to collect clinical data for the calculation of CSS according to the Pineda scale (Table 3, Material and methods). For each patient, we performed the classical XCI-AR assay and the corresponding XCI-AS assay when blood samples were available (174/221 patients), and we calculated the CSS when clinical data were available (181/221 patients). In agreement with literature, we established an 80:20 ratio for considering an XCI pattern as skewed[80,254]. The entire list of XCI results and clinical scores for all patients can be found in Table S1 (Annex 1).

The vast majority (more than 90%) of patients presented a random pattern of XCI (Table 10). Only 17 (9.8%) showed a skewed XCI pattern (80:20 or higher), which is similar to what was found in other studies[255,256]. Different studies have found a considerably higher incidence of skewing, up to 43%, among RTT patients[257]. Some authors claim that most of the patients who meet diagnostic criteria for RTT have a random XCI pattern, while those with skewed XCI patterns may not meet all the criteria and therefore are not included in RTT studies[258].

Mutation	Class	MeCP2 region	Skewed XCI	% Skewed XCI
c.455C>G (p.Pro152Arg)	Missense	MBD	0/6	0%
c.473C>T (p.Thr158Met)	Missense	MBD	0/33	0%
c.502C>T (p.Arg168*)	Nonsense	IDR	5/29	17.2%
c.674C>G (p.Pro225Arg)	Missense	TRD	0/2	0%
c.763C>T (p.Arg255*)	Nonsense	TRD	4/36	11.1%
c.806delG (p.Gly269fs)	Frameshift	TRD-NLS	1/11	9.1%
c.808C>T (p.Arg270*)	Nonsense	TRD-NLS	4/20	20%
c.880C>T (p.Arg294*)	Nonsense	TRD	1/20	5%
c.916C>T (p.Arg306Cys)	Missense	TRD	1/15	6.7%
Large deletions	Deletion	Exons 3-4	1/2	50%
All	**-**	**-**	**17/174**	**9.8%**

Table 10. Skewed XCI. Proportion of patients per mutation with skewed XCI pattern according to at least one of the two techniques used for assessing XCI (XCI-AR and XCI-AS).

Interestingly, the percentage of patients in our cohort with skewed XCI patterns varied across different types of mutations (Table 10). No patients with p.Pro152Arg, p.Thr158Met or p.Pro225Arg mutations showed skewed XCI patterns in either XCI-AR or XCI-AS. Skewed XCI patterns were more common in RTT patients with deletions, nonsense, and frameshift mutations than in those with missense mutations, and mutations that produce a truncated protein tend to result in a more severe phenotype than missense mutations[257]. This relationship between XCI and mutation type could reflect a protective effect related to the severity of the mutation. It is possible that mutations producing a less functional, truncated protein (deletions and nonsense mutations) cause cells to preferentially inactivate the X chromosome harbouring the mutation. It has been shown that skewed XCI can be the consequence of a selective advantage of cells with a particular active X chromosome proliferating faster than cells where the other X chromosome is active[79,81,259]. This type of skewing has been described in up to 50% of familial cases of X-linked mental retardation disorders[260].

Despite the higher tendency to present skewed XCI in patients with nonsense mutations and large deletions, we did not observe a consistent preference for the

inactivation of the wild-type or the mutant allele when analysing cases with skewed XCI (Table 11).

Patient	Variant	Class	XCI-AR	XCI-AS	CSS	Z-score
P47	c.502C>T | p.Arg168*	Nonsense	n.i.	81:19	13	0.01
P60	c.502C>T | p.Arg168*	Nonsense	84:16	28:72	16	0.88
P68	c.502C>T | p.Arg168*	Nonsense	75:25	15:85	NA	NA
P70	c.502C>T | p.Arg168*	Nonsense	85:15	35:65	NA	NA
P74	c.502C>T | p.Arg168*	Nonsense	81:19	55:45	NA	NA
P83	c.763C>T | p.Arg255*	Nonsense	85:15	57:43	NA	NA
P84	c.763C>T | p.Arg255*	Nonsense	87:13	55:45	13	-0.68
P85	c.763C>T | p.Arg255*	Nonsense	80:20	28:72	14	-0.37
P107	c.763C>T | p.Arg255*	Nonsense	87:13	68:32	11	-1.3
P139	c.806delG | p.Gly269fs	Frameshift	82:18	58:42	NA	NA
P143	c.808C>T | p.Arg270*	Nonsense	97:3	16:84	18	0.86
P144	c.808C>T | p.Arg270*	Nonsense	84:16	21:79	NA	NA
P145	c.808C>T | p.Arg270*	Nonsense	81:19	30:70	9	-1.47
P146	c.808C>T | p.Arg270*	Nonsense	80:20	73:27	13	-0.44
P191	c.880C>T | p.Arg294*	Nonsense	89:11	49:51	NA	NA
P195	c.916C>T | p.Arg306Cys	Missense	89:11	60:40	9	-0.76
P220	Deletion	Deletion	88:12	7:93	NA	NA

Table 11. XCI and CSS results for patients with skewed XCI according to at least one of the two assays. The XCI-AR column shows the results of the AR XCI assay (percentage of inactivation of each allele). The XCI-AS column shows the results of the allele-specific XCI assay (percentage of inactivation of WT:Mut alleles; mean of two to three replicates). The CSS is shown when available, next to the Z-score of each CSS value with respect to the mean CSS for patients of our cohort with each variant.

Skewed proliferation could explain some cases, such as several patients with p.Arg168* (P60, P68, P70; Table 11), p.Arg255* (P85; Table 11), and p.Arg270* (P143, P144, P145; Table 11) mutations. The most extreme case would be that of patient P220 (Table 11), who had a large deletion in *MECP2* and showed a skewed XCI pattern (88:12) by the XCI-AR assay. In this patient, the XCI-AS assay confirmed an extremely skewed XCI pattern and that the preferentially inactivated

allele was the mutated allele at a ratio of 7:93 (WT:Mut). However, there were other patients with these same mutations and with skewed XCI according to the XCI-AR assay who showed a preferential inactivation of the WT allele when the XCI-AS assay was performed, such as P146 (Table 11). Patient P47 (Table 11), who was noninformative for the XCI-AR assay, also showed a preferential inactivation for the WT allele when the XCI-AS assay was performed. These last patients do not support the abovementioned hypothesis.

In order to test for significant differences in CSS in patients with skewed XCI patterns with respect to the average CSS of patients with the same mutation, we calculated Z-scores, that represent the number of standard deviations from each CSS value to the population mean. We found no significant differences between CSS in patients with skewed XCI patterns when comparing to the average CSS of patients with the same mutation (|Z-score| < 2), and most patients were included in the central 68% of individuals in a normal distribution (|Z-score| < 1) (Table 11, Figure 10).

We found no consistent increases or decreases in the CSS of RTT patients with a preferential inactivation of the WT and mutated alleles in blood samples (Figure 10) and no substantial correlation between the XCI patterns in blood and the clinical presentation of RTT according to the CSS scale (Figure 11).

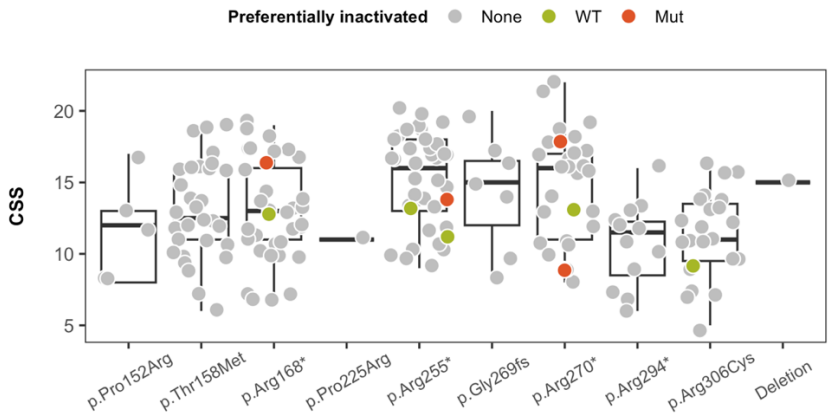


Figure 10. CSS distribution by *MECP2* variant. Boxplots of CSS values for each specific mutation, where each dot represents one patient. Coloured dots correspond to patients with skewed XCI patterns according to at least one of the two XCI assays (green for patients with preferential inactivation of the WT allele and red for patients with preferential inactivation of the mutated allele).

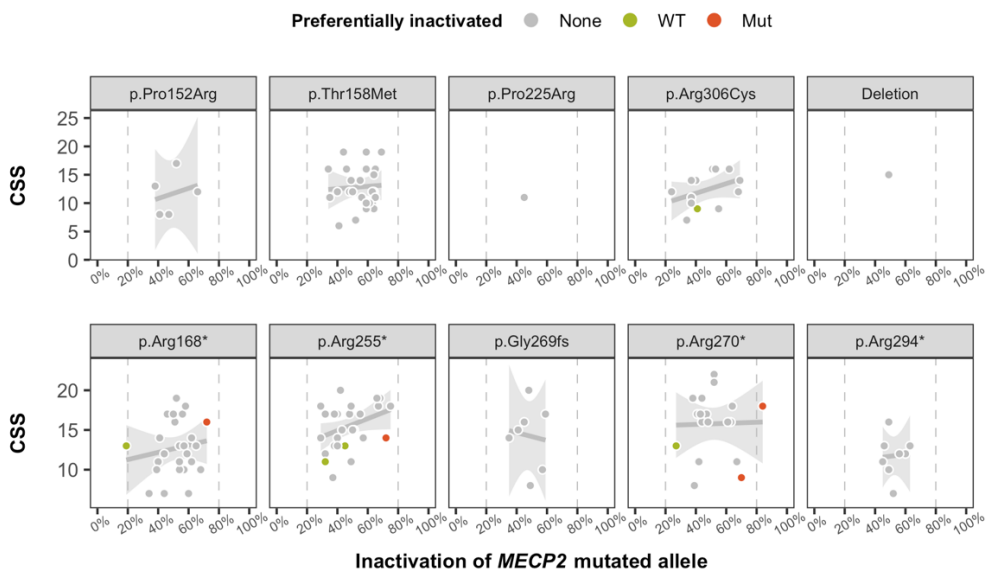


Figure 11. Correlation analysis between inactivation of the *MECP2* mutated allele and CSS. Scatter plots of CSS and degree of inactivation of the mutated allele, where each dot represents one patient and coloured dots correspond to patients with skewed XCI patterns according to at least one of the two XCI assays (green for patients with preferential inactivation of the WT allele and red for patients with preferential inactivation of the mutated allele). The grey lines represent a linear model attempting to correlate the inactivated fraction of the mutated allele and the CSS (not significant).

1.3. Allele-specific X chromosome inactivation and skewing in brain samples

In order to determine if blood could be an accurate predictor of brain XCI status, we assessed XCI patterns in several brain regions of two patients with the c.763C>T (p.Arg255*) mutation and compared them to the results obtained in blood samples (Table 12). Although no samples showed skewed XCI by either assay, there was no clear homogeneity among blood and brain samples, and although small, there were also differences in the XCI patterns between different brain regions of the same patient. Some samples, such as the frontal cortex or the white matter samples of patient P109, showed an XCI pattern closer to the skewing threshold than other regions, such as the cerebellum, of the same patient. In patient P119, most samples were close to the random XCI pattern, but the temporal cortex sample showed an

XCI pattern closer to the skewing threshold. The consequences of having heterogeneous XCI patterns in different brain regions could potentially contribute to the wide spectrum of RTT phenotypes, given that RTT is not characterised by a dysfunction in a specific brain area but rather impacts the entirety of the brain.

Patient P109 (CSS = 20)			***Patient P119 (CSS = 19)***		
Brain region	**XCI-AR**	**XCI-AS**	**Brain region**	**XCI-AR**	**XCI-AS**
Frontal Cortex	65:35	26:74	*Frontal Cortex*	n.i.	48:52
Occipital Cortex	58:42	59:41	*Occipital Cortex*	n.i.	NA
Parietal Cortex	64:36	40:60	*Parietal Cortex*	n.i.	56:44
Temporal Cortex	60:40	32:68	*Temporal Cortex*	n.i.	73:27
White matter	59:41	23:77	*White matter*	n.i.	46:54
Brain stem	59:41	31:69	*Brain stem*	n.i.	38:62
Striatum	61:39	51:49	*Striatum*	n.i.	50:50
Cerebellum	55:45	43:57	*Cerebellum*	n.i.	50:50
Blood	73:27	64:36	*Blood*	n.i.	34:66

Table 12. XCI patterns in blood and several brain areas in patients with the c.736C>T mutation. The XCI-AR column shows the results of the XCI androgen receptor assay (percentage of inactivation of each allele), and the XCI-AS shows the results of the XCI allele-specific assay (percentage of inactivation of each allele, WT:Mut). n.i. = polymorphism noninformative for the assay. NA = data not available.

It has been observed that XCI patterns can vary among different tissues[253,261]. Blood is particularly prone to XCI skewing, because of the proliferation of different clones of lymphocytes under different conditions[253,262]. In fact, blood XCI patterns have shown variations at different time points in different studies[80]. For two of the patients included in the study (P9 and P199; Table S1, Annex 1) we compared two different blood samples from different extractions. Both patients showed some differences in the results of the XCI assays in the two extraction samples. Therefore, blood XCI patterns may not accurately reflect the XCI landscape in other organs and tissues such as the brain.

We observed that XCI patterns in blood and different regions of the brain are not necessarily homogenous. Therefore, if RTT symptoms are caused mainly by the

lack of *MECP2* function in the brain, it is expected that the severity of the phenotype will be more related to the XCI pattern in the brain than to the XCI pattern in the blood, explaining the lack of a direct correlation between the XCI patterns in blood and the clinical presentation of RTT.

RTT symptoms arise from a loss of function of *MECP2* in neurons, and the severity of the male phenotype points towards a dose-dependent mechanism of action of *MECP2*[4,35,255]. It is possible that in females, slight deviations from random 50:50 XCI ratios do not cause sufficient changes in the levels of the mutant *MECP2* in the brain to be translated into a different phenotype. It is possible that the effect is only remarkable in extreme cases, such as some familial cases of RTT, in which a healthy mother with extremely skewed XCI can be a carrier of a pathogenic mutation responsible for causing RTT in her offspring, although she remains asymptomatic [84,258,263,]. Some authors have claimed that these familial cases of RTT are only possible due to the presence of two coincident traits: RTT and the trait for skewed XCI, which would be genetically determined, and would therefore not be caused by a selective proliferation of cells expressing the WT allele [80,84].

1.4. Brain RNA analysis

Finally, we analysed RNA samples from frontal and occipital cortex of the two RTT patients with the c.763C>T (p.Arg255*) mutation. We checked if we could detect an overrepresentation of one of the alleles by Sanger sequencing of cDNA samples obtained by RT-PCR (Figure 12A). In cDNA samples from patient P109, the T allele (mutated allele) was overrepresented, while in samples from patient P119, the C allele (WT allele) was overrepresented. However, both patients presented a severe form of RTT, with CSS of 20 and 19, respectively.

cDNA sequencing analysis seemed to indicate that one allele was more frequently present than the other, although these results were not conclusive since Sanger sequencing is not a quantitative technique. We later confirmed our findings in the frontal cortex samples by RT-qPCR, a proper quantitative technique for quantifying RNA levels (Figure 12B). We confirmed that in samples from patient P109, the mutated allele was overrepresented, while in samples from P119, the WT allele was overrepresented. The XCI assay results showed inactivation patterns that did not

reach the threshold to be classified as skewed in any of the two patients and regions, and the expression levels of each allele did not correlate with XCI patterns (Figure 12B).

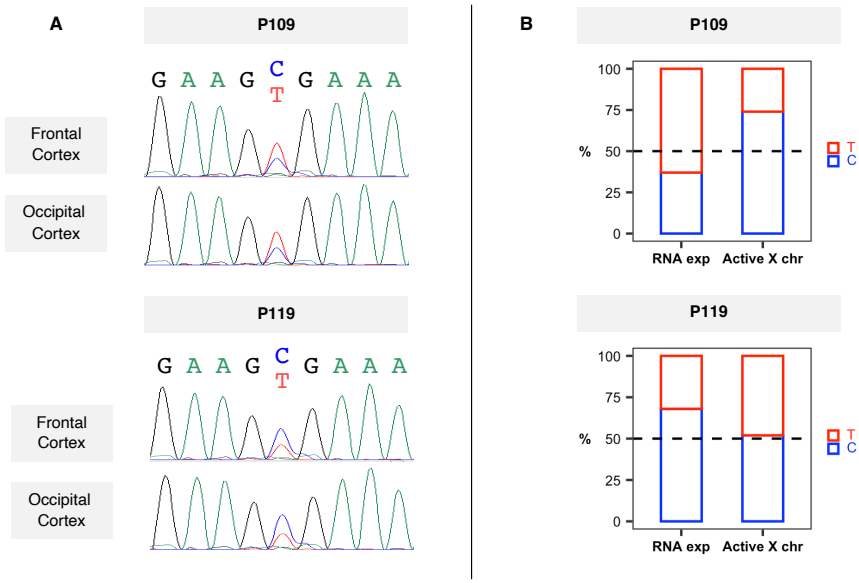


Figure 12. *MECP2* expression in brain samples. A) Sanger sequencing. Chromatograms obtained by Sanger sequencing cDNA obtained from frontal and occipital cortex samples of the two patients. At the mutated locus (c.763C>T) patient P109 shows a higher expression of the mutated allele (T), while patient P119 shows a higher expression of the wild-type allele (C). **B) RT-qPCR and XCI ratio in frontal cortex.** Bar plots showing the quantification of allelic expression by RT-qPCR (RNA exp), and XCI pattern by XCI-AS (Active X chr) in frontal cortex samples of the same two patients.

The differences between XCI patterns measured and the levels of each allele observed in Sanger sequencing and RT-qPCR could be due to RNA degradation in the post-mortem interval. RNA degradation during life through the nonsense-mediated mRNA decay (NMD) pathway would be unlikely since this *MECP2* variant affects the last exon of the gene. However, brain RNA levels of each allele seemed to show discrepancies with the XCI patterns identified in our XCI assays. Some authors have noticed discordances between the XCI pattern according to the XCI-AR assay and the quantification of the AR gene expression [264]. These discrepancies

suggest, first, that the methylation assay may not always be representative of XCI and, second, that gene transcript levels may be regulated by more factors than XCI.

The difference between the XCI pattern and the final RNA levels of each allele suggests that the levels of *MECP2* are not directly determined by the XCI pattern and that there could be mechanisms other than XCI involved in regulating *MECP2* transcript levels. There might be other genes involved in regulating *MECP2* transcription and/or RNA degradation, causing changes in the final levels of functional *MECP2* [47]. Therefore, XCI may not necessarily be determining the severity of the clinical presentation of RTT, which would be more related to the levels of functional *MECP2* in the brain [4,33]. Nevertheless, it is important to keep in mind that we are measuring *MECP2* transcript levels from bulk RNA. Since different cell types have shown diverse transcriptional profiles in several studies, the levels of the *MECP2* transcripts measured do not necessarily reflect these transcripts levels in neurons relevant for RTT pathophysiology [265].

Although one patient showed higher levels of *MECP2* mutant transcript than the other, the CSS of both patients was not dissimilar (20 vs 19). This score similarity supports the hypothesis that slight deviations from a 1:1 ratio of each allele produce little to no change in the RTT phenotype. It is possible that more consistent differences would be noticeable if one allele was more prevalent than the other, such as in asymptomatic carriers with an XCI pattern close to the 100:0 ratio.

In conclusion, our results show that the relationship between XCI and the severity of the RTT phenotype is not straightforward. *MECP2* transcript levels are presumably regulated by factors other than XCI, such as genetic polymorphisms, the expression of other genes, and environmental conditions, whose combination and addition may influence the clinical presentation to a greater extent than solely the XCI pattern in the brain. Therefore, probably only extremely skewed XCI patterns affecting relevant cell types can produce significantly different phenotypes.

Chapter 2. Consequences of MeCP2 deficiency for the expression of GABAergic synapse proteins

Although *MECP2* mutations are known to be the cause of RTT, the pathomechanism downstream MeCP2 deficiency leading to RTT remains elusive. Although the entirety of the brain is affected by MeCP2 deficiency, GABAergic neurons have demonstrated to play a key part in RTT pathophysiology, since mice with a conditional *Mecp2* KO in GABA-releasing neurons have proved to recapitulate most of the RTT phenotypes[266]. Moreover, restoring *Mecp2* expression specifically in GABAergic neurons of *Mecp2*-null mice rescued many RTT features, suggesting GABAergic modulation as a potential therapeutic option in RTT[99]. Based on these previous reports supporting the contribution of GABAergic dysfunction in RTT progression and the alteration of the expression of GABA receptors, we hypothesized that MeCP2 disturbance might directly affect the density these receptors, rather than such an altered expression being a secondary effect of an overall GABAergic dysfunction[99,266].

Fast GABAergic neurotransmission is mediated by GABA ionotropic receptors ($GABA_A$). These are ligand-gated chloride (Cl^-) channels consisting of five subunits of eight subfamilies[267]. Mechanistically, the activation of $GABA_A$ receptors allows for a selective Cl^- influx, triggering a hyperpolarization of the postsynaptic neuron that reduces its likelihood of starting an action potential[268]. Among the multiple stoichiometric combinations of the heteromeric $GABA_A$ receptors, α1-β2-γ2 is the major molecular combination, with the α1 subunit being present in over 60% of $GABA_A$ heteromers and being widely expressed in brain areas.

2.1. Expression of GABAergic synapse proteins in cellular models

To test whether MeCP2 expression could directly affect the levels of these proteins, we studied the expression pattern of GABAergic ionotropic receptors upon *Mecp2* knockdown in cellular models. First, we used Neuro2a (N2A) mouse neuroblastoma cells transfected with a shRNA-anti-3'UTR *Mecp2* plasmid (shRNA) to knock-down endogenous mouse *Mecp2* (m*Mecp2*) expression. On this knocked-down

background, we assessed the effects of co-transfecting these cells either with a plasmid expressing the WT human *MECP2* (WT-h*MECP2*) or with a plasmid expressing a human *MECP2* harbouring the pathogenic c.763C>T variant (Mut-h*MECP2*).

First, we evaluated the efficiency and specificity of the *Mecp2* expression interference. For this, total RNA was isolated from the cells under the four scenarios analysed: untransfected, transfected only with shRNA, and co-transfected either with shRNA and WT-h*MECP2* or Mut-h*MECP2*. The shRNA was designed to specifically target m*Mecp2*, while the h*MECP2* was predicted to be insensitive to its knock-down effects. The system was proven to be reliable, as the endogenous m*Mecp2* levels were drastically reduced upon transfection of N2A cells with the shRNA vector, while exogenous expression of the human isoform, however, was not affected by the shRNA co-expression (Figure 13A).

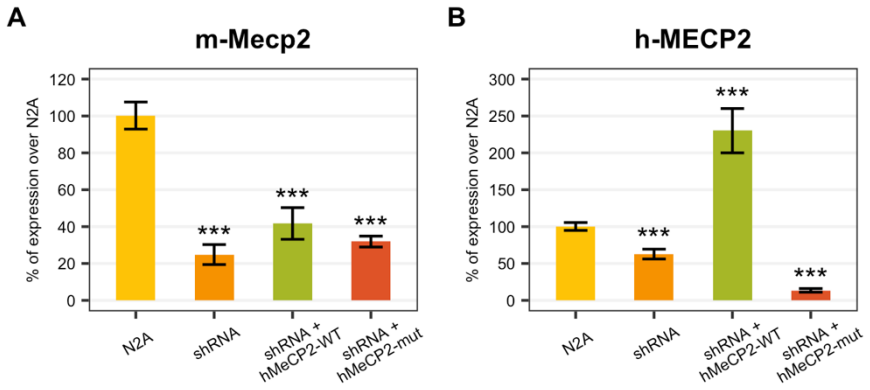


Figure 13. Validation of the inhibition/re-expression system. Bar graph representing the relative expression of **(A)** mouse *Mecp2* (m-Mecp2) and **(B)** human *MECP2* (h-MECP2), measured by RT-qPCR under the four transfection scenarios (non-transfected cells (N2A), transfected with the shRNA-anti-(3'UTR)Mecp2 (shRNA), and co-transfected with the shRNA-anti-(3'UTR)Mecp2 and either WT or c.763C>T mutated *MECP2* carrying plasmids. Error bars represent the standard deviation of the average values. Statistical significance was calculated through an unpaired two-tailed Student's t-test (*** p-value < 0.001).

The amplification with specific h*MECP2* primers revealed an increased expression only upon transfection with the WT-h*MECP2* vector. While the shRNA has

demonstrated specificity for the m*Mecp2* transcript, the exclusivity of the h*MECP2* detection cannot be ensured, as endogenous m*Mecp2* is amplified in the untransfected N2A cells using the h*MECP2* primers (Figure 13B). Given that the human and mouse MeCP2 are identical in almost 90% of their sequence, we assume we are detecting undesired amplification due to cross-binding between the mouse and human forms. While the forward primer binds to a sequence similar in 90% of both forms (19 out of 20 bases), the reverse primer was designed to only bind the human transcript but is still producing some degree of undesired amplification.

Under basal conditions, the transcript encoding for the major $GABA_A$ receptors subunit (m*Gabra1*) was not detectable in N2A cells. Remarkably, the target was amplified upon transfection with the plasmid carrying the WT form of h*MECP2* (Figure 14A), supporting a positive relationship between MeCP2 and *Gabra1* expression.

Noteworthy, the $GABA_A$ receptor subunit expression was undetectable upon transfection with the plasmid containing Mut-h*MECP2*. The expression of the remaining principal $GABA_A$ subunits, *Gabrb2* and *Gabrg2* was also increased in the presence of WT-h*MECP2* but not Mut-h*MECP2* (Figure 14B and C). These results point towards a role of MeCP2 as a positive regulator of the most common $GABA_A$ subunits.

We then assessed the effects of MeCP2 deficiency on another gene reported to be affected in RTT patients, *Slc12a5*[269]. *Slc12a5* is the gene encoding the KCC2 protein, a neuron-specific potassium-chloride symporter responsible for the maintenance of intracellular chloride concentrations. Interestingly, *Slc12a5* expression did not appear to be affected by MeCP2 inhibition (Figure 14D), suggesting that both alterations affecting the same synapse can occur through different mechanisms.

In order to evaluate the identified MeCP2 positive regulatory effect on the $GABA_A$ receptor expression in dendritic processes, primary cultures of WT murine cortical neurons were established. Following shRNA-mediated MeCP2 silencing, the expression of m*Gabra1* was assessed.

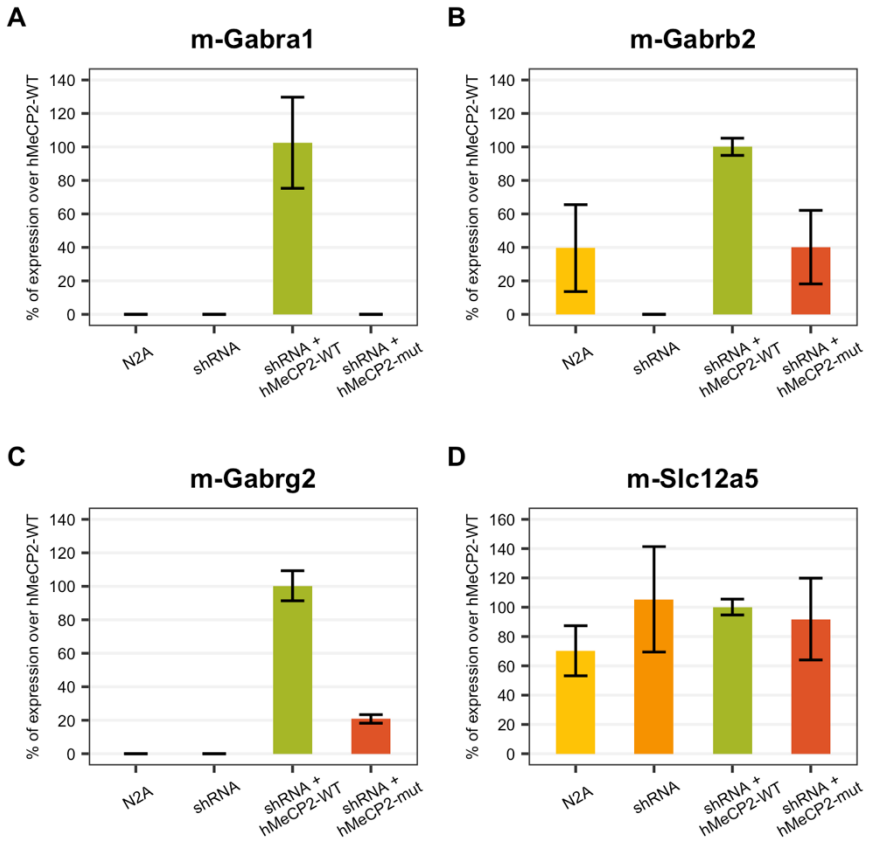


Figure 14. In vitro analysis of MeCP2 altered activity over GABAergic synapse proteins. Bar graph representing the relative expression of **(A)** mouse *Gabra1* (m-Gabra1), **(B)** mouse *Gabrbr2* (m-Gabrb2), **(C)** mouse *Gabrg2* (m-Gabrg2) and **(D)** mouse *Slc12a5* (m-Slc12a5), measured by RT-qPCR under the four transfection scenarios (non-transfected cells (N2A), transfected with the shRNA-anti-(3'UTR)Mecp2 (shRNA), and co-transfected with the shRNA-anti-(3'UTR)Mecp2 and either WT or c.763C>T mutated *MECP2* carrying plasmids. Error bars represent the standard deviation of the average values.

Primary cortical neurons were transiently co-transfected at day in vitro 7 (DIV7) with either the shRNA or a mock plasmid, together with a pcDNA-EGFP vector in a 1:7 ratio, allowing for the identification of shRNA-transfected neurons. An immunofluorescence analysis was performed in the early mature primary neuronal cultures (DIV11), showing a complete lack of *Mecp2* detection in shRNA-(GFP-positive) transfected neurons (Figure 15A), and thus validating the inhibition system.

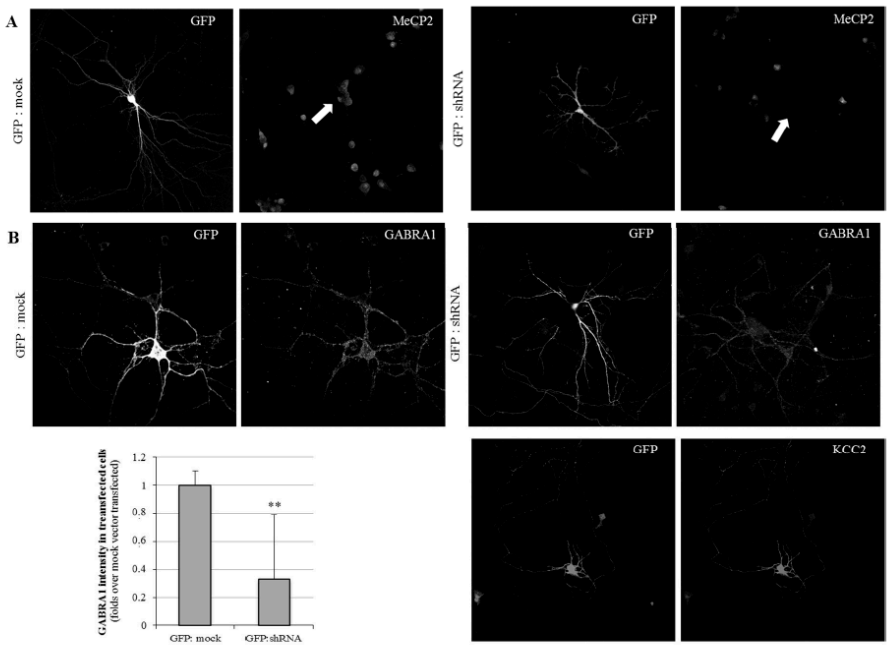


Figure 15. Immunofluorescence analysis of MeCP2, GABRA1, and KCC2 expression in cortical primary neuronal cultures. Cells were transfected with either GFP and mock DNA, or GFP and shRNA-anti-(3'UTR)Mecp2. Images show neurons at DIV11. Images were taken at 63x with constant time of exposure. Transfected neurons were labelled with anti-GFP and anti-Mecp2 **(A)**. For the GABRA1 and KCC2 immunostaining **(B)**, different neurons are shown. The quantification of the mean GABRA1 immunosignal in mock and shRNA transfection conditions is shown in the bar graph. N=10 different neurons, from two independent cultures. ** p-value < 0.01.

When we next evaluated the $GABA_A$-A1R expression on the silenced neurons, we observed a severe decrease in the detection in the GFP-positive neurons, compared with the cultures transfected with GFP and mock vector (Figure 15B), considered as the control conditions. The mean detection on the GABRA1 marking was 37% compared with the control conditions. These results are complementary to the N2A cells observations, as both support a positive and straightforward relationship between MeCP2 and $GABA_A$-A1 receptor expression.

2.2. Expression of GABAergic synapse proteins in animal models at different developmental time points

As RTT is a neurodevelopmental disorder, we aimed to evaluate these gene expression changes in different evolutionary stages. Thus, we evaluated the GABAergic synapse proteins expression in the cortex of mice at pre-symptomatic and symptomatic stages (one and six months, respectively). We used female *Mecp2*$^{+/-}$ heterozygous mice from the Bird strain, a RTT mice model that recapitulates RTT-like abnormalities [270]. Western blot analysis revealed a significant decrease of $GABA_A$-A1 at a young, pre-symptomatic stage (one month old), while no significant differences were detected between RTT and control mice at the symptomatic stage (six months old) (Figure 16A, B and C).

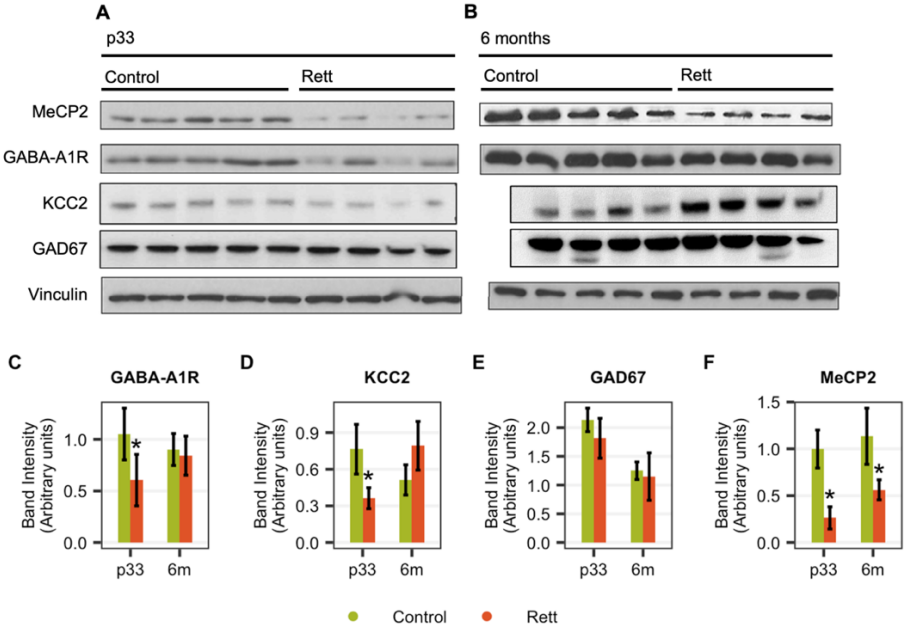


Figure 16. Developmental expression analysis of GABAergic proteins in the *Mecp2*-/+ mouse brain cortex. Representative western blot analysis of MeCP2, GABA-A1R, KCC2 and GAD67 expression in the cortex of female RTT and control mice. Vinculin was used as a loading control. **(A)** Expression in p33 pre-symptomatic mice (control and pre-symptomatic RTT mice). **(B)** Expression in six-months old mice (control and symptomatic RTT mice). Both western blots **(A, B)** are cropped stripes of two different membranes each, and were

incubated with each antibody separately. Each lane is a different animal. **(C, D, E, F)** Quantification of protein expression detected by western blot. Relative protein western blot quantification. Data shown as mean ± standard deviation. The asterisk (*) represents statistical significance ($p < 0.05$) in the unpaired two-tailed t-test comparing the means of RTT vs control mice in each developmental stage.

To contextualise this difference between the prodromic stages and considering whether it was a $GABA_A$-A1 receptor subunit specific variation or a generalized GABAergic synapse downregulation, we analysed the expression of two other proteins in the same conditions: the abovementioned potassium-chloride cotransporter KCC2 and GAD67, the enzyme that catalyses the decarboxylation of glutamate to GABA and CO_2, which is widely used as a GABAergic marker. A decreased expression of KCC2 was observed only in the pre-symptomatic stage, suggesting a generalized decrease of the GABAergic function without a decrease in GABAergic neuronal population (Figure 16A, B, D and E).

As a control, the MeCP2 expression was assessed in the same samples and appeared to be lower in all the RTT samples, regardless of the disease stage (Figure 16A, B and F).

These findings are in line with what we observed in the *in vitro* models, and suggest an important component of development in RTT, showing that there are gene expression changes throughout development and that MeCP2 may act differently at different time points of the disease.

2.3. Transcriptomic profile of GABAergic pathway in post-mortem brain of RTT patients

In order to study the potential translation of our previous findings onto RTT patients, we then profiled the expression of the genes implicated in the GABAergic pathway in brain samples of two RTT patients with the c.763C>T *MECP2* mutation. We used a sample of a patient with a different neurological disease as an internal RNAseq technical control and publicly available data to assess biologically differential expression. The internal RNAseq control was a patient with a neurological pathology with a non-genetic basis, which was not related to RTT and was age-matched to our RTT patients (who were 10 and 15 years old at exitus). We used this sample to

validate RNAseq results, making it possible to confirm by RT-qPCR the accuracy of the expression data obtained.

We analysed gene expression of all genes in the GABAergic pathway according to KEGG Pathway Database (n = 108, map04727). We found that 27% of genes were differentially expressed when compared to the internal control, among which we found the abovementioned subunits of the $GABA_A$ receptors: *GABRA1*, *GABRB2* and *GABRG2*. From these results, we chose a subset of genes (n = 22) to validate the RNAseq results by RT-qPCR (Figure 17).

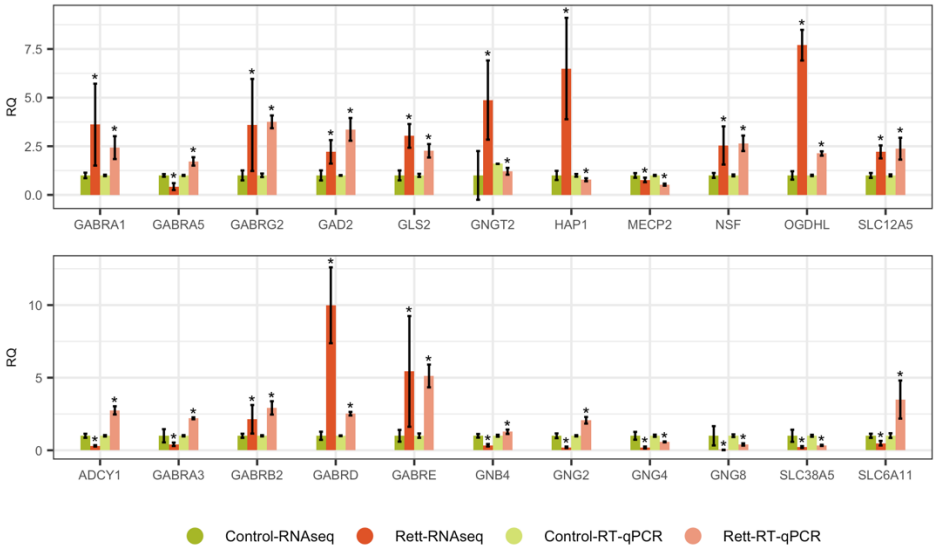


Figure 17. RNAseq vs RT-qPCR results for a subset of genes involved in the GABAergic synapse and *MECP2*. Comparison of RNAseq vs RT-qPCR results for each gene, as relative quantification (RQ) of RTT samples vs our internal RNAseq technical control. *RPLP0* and *GUSB* were used as endogens for data normalization in RT-qPCR. Triplicates of each sample and target were run, and results were averaged for both endogens (standard deviation shown by error bars). Statistical significance was calculated using an unpaired two-tailed Student's t-test (* $p < 0.05$).

We chose genes that were differentially expressed (p-adjusted < 0.05) with a remarkable effect size (|Fold Change| > 2) between RTT patients and the internal control in RNAseq. This validation subset included the abovementioned *GABRA1*, *GABRB2* and *GABRG2*. RT-qPCR data confirmed the expression changes obtained

in RNAseq for most part of the genes analysed including *GABRA1*, *GABRB2* and *GABRG2*, all of which showed overexpression in RTT patients compared to the internal control (Figure 17).

Upon the validation of the RNAseq experiment, and given that we only had one control brain for differential expression analysis, the raw data from RTT patients were compared with RNAseq data from control individuals available at GTEx Portal[271]. We compared our data with data derived from five controls (two females and three males) aged 20-29, being the closest age-matched group to our patients. Gene expression was normalized with the housekeeping gene *RPLP0* to allow for the comparison between internal and public data.

The transcripts encoding for the major $GABA_A$ receptor subunits and the potassium-chloride transporter KCC2 were slightly increased in RTT samples compared with controls (Figure 18). As expected from western blot studies, GAD67 levels remained unaltered compared with controls.

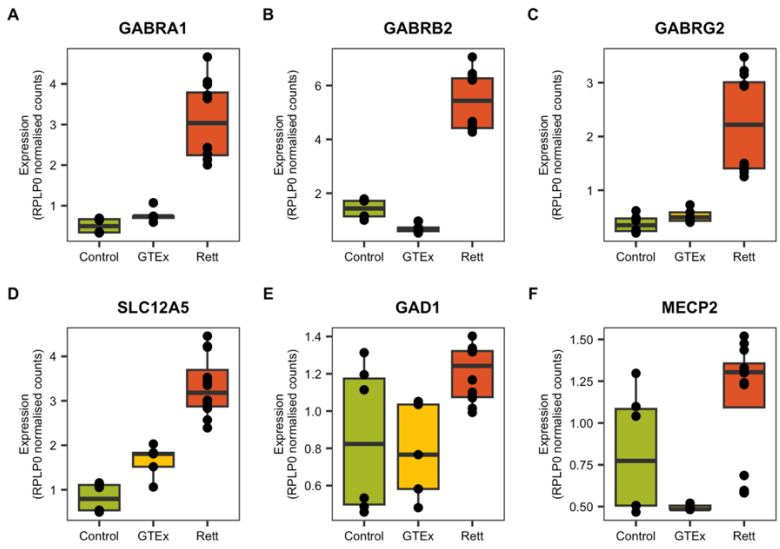


Figure 18. Expression of GABAergic synapse genes in human brain. Boxplots of RNAseq normalised counts of **(A)** *GABRA1*, **(B)** *GABRB2*, **(C)** *GABRG2*, **(D)** *SLC12A5*, **(E)** *GAD1* (GAD67), and **(F)** *MECP2*, using a housekeeping gene (*RPLP0*) as a reference in cortex samples of our intra-assay control (Control), publicly available controls from GTEx (GTEX) and patients with RTT (Rett).

Many studies have reported that the excitatory/inhibitory (E/I) balance and GABAergic inhibition are largely disturbed in RTT, and this malfunctioning affecting several brain regions could give rise to a number of neurological symptoms, such as motor dysfunction, abnormalities in breathing patterns and seizures, all common manifestations of RTT[40,272].

Previous studies have explored the effects of *MECP2* deficiency on GABAergic synapses [97,273,274]. An increased MeCP2 expression in GABAergic neurons has been reported, and a reduced GABA release was reported upon MeCP2 knocking out in forebrain GABAergic neurons [275]. This positive relationship has been observed in other neuron types and brain areas, such as CA3 hippocampal neurons or brainstems[273]. Moreover, the restoration of MeCP2 expression exclusively in GABAergic neurons was demonstrated to be sufficient to rescue some disease features in a mouse model of RTT[276]. One study even suggested a time-dependent alteration in GABAergic neurotransmission in *Mecp2*-deficient male mice[274].

In the present study, we have evaluated how alterations in GABAergic synapse proteins were occurring during development, setting the focus on the pre-symptomatic stages of RTT. We have observed a direct relationship between the MeCP2 altered expression and GABAergic receptors disruption that is strongly dependent on the prodromic stage of the disease, which will set the focus on the time frame as a key factor when evaluating therapeutic options.

Previous studies pointing out a relationship between GABA receptors and MeCP2 had been done in the context of whole brain models analysis, so it remained unclear if the potential decrease in GABA receptors was strictly related to MeCP2 expression or was a secondary effect to global dysfunction. The results we obtained in *in vitro* models where MeCP2 expression was silenced and posteriorly re-expressed strongly suggest a direct relationship between MeCP2 and the expression of post-synaptic GABA ionotropic receptors, rather than this being a secondary effect of an overall altered homeostasis. Moreover, the fact that the Mut-h*MECP2* vector was not able to induce the same expression changes observed with the WT vector suggests that this is one of the consequences of *MECP2* malfunctioning due to RTT-causing mutations and that may contribute to the disease pathomechanism.

Opposite to that, we found that the KCC2 expression was unaltered in the presented scenario. KCC2 is a potassium-chloride channel, essential for GABAergic correct functioning, and described to be downregulated in RTT [269]. The fact that its expression was not significantly altered by MeCP2 inhibition suggests that such an altered regulation is more related to the pathophysiology of the disorder rather than to the straightforward MeCP2 mutations. In fact, it has been well described how the KCC2 expression and function is, indeed, regulated by the GABA function itself [277,278].

To evaluate different prodromic stages of the disease, we used female RTT mice, with a null *Mecp2* allele and a WT allele, as they better recapitulate the disorder pathophysiology and reproduce the cases of patients with RTT, who are mostly female[279]. We focused our analysis on the following three proteins: $GABA_A$-A1R, KCC2 and GAD67. The first two proteins were selected because of the previous observations, and because they have been proven to be not only crucial in RTT development, but also potential actionable targets of the disease [280,281]. Our results showed a markedly reduced expression of $GABA_A$-A1R and KCC2 in pre-symptomatic mice, while the GAD67 population (used as a marker for GABAergic neurons) remained unaltered. These results suggest a reduced GABAergic activity without the affectation of the GABAergic general population, which are aligned with the previously described results. We did not observe any reliable difference in these proteins' expression in fully symptomatic mice, enhancing the importance of the time frame when addressing RTT. The variability in KCC2 expression, and its activation through phosphorylation, is a field that further studies should explore, especially under the recent scenario in which KCC2 is being addressed as a therapeutic target [280]. Preliminary results have pointed towards an over-phosphorylation of KCC2 in symptomatic mice, drawing a scenario in which KCC2 will be under expressed in early pre-symptomatic stages and inactivated in symptomatic phases – again, shaping different therapeutic strategies on different prodromic stages. KCC2 activation and membrane diffusion has been related with GABAergic activity itself, increasing its therapeutic interest[277,278].

Completing the observed results, we analysed the aforementioned targets in two RTT patients' necropsied brains. Backing up our previous description, the results

showed an even enhanced expression of $GABA_A$-A1R and KCC2, without alterations on GAD67. The overall analysis of the GABAergic pathway showed a slightly increased expression of almost all the implicated genes. These results confirm the previous findings, pointing towards the importance of pre-symptomatic damage. These results agree with the dataset reported by Renthal et al. [282]. An increasing body of evidence pointing towards the importance of early intervention has been reported in the last few years, raising awareness of subtle signs displayed by RTT patients in early stages of development that will probably be accompanied by significant molecular alterations[283].

Despite the limitations of the RNAseq experiment in post-mortem brain samples of RTT patients, this constitutes a novel approach with the potential to unravel many molecular alterations involved in RTT pathophysiology. Given the rareness of RTT, post-mortem brain samples of RTT patients ware extremely uncommon, and this type of transcriptomic studies is yet infrequent. To date, only four transcriptome-level studies in RTT have used patients post-mortem brain samples; three of them were carried out with microarray technology and only one employed RNAseq [88,284]. There is a huge variation in the samples used in these studies (different brain regions, mutation types and age groups), complicating the task to conduct a proper meta-analysis of their results. Our decision to compare our samples with publicly available controls was motivated by the lack of true healthy biological control samples. Results found with this approach will need to go through further validation and replication studies when more samples become available, but they can serve as starting point in the search for candidate genes to become potential therapeutic targets for RTT. To overcome the dependence on healthy controls, one study employed single-cell RNAseq technology, where neurons expressing the mutated *MECP2* allele were compared to those expressing the wild type copy, disregarding the need of external controls [282]. Even though cells expressing the wild type MeCP2 may be influenced by surrounding cells expressing the mutated MeCP2 allele, they serve as an expression control with the advantage of sharing the same genetic background as mutated cells and thus eliminating variability in expression because of that fact. Furthermore, these types of studies can elucidate cell-type-specific effects of

MeCP2 dysfunction, which would allow to better profile the pathomechanisms behind neuronal dysfunction in the RTT brain.

Some reports have provided neurophysiological evidence of temporal alterations in GABAergic dysfunction during development in RTT [285]. MeCP2 is involved in neuronal maturation, through influencing synaptogenesis and establishment of functional circuits [40]. However, additional evidence indicates that MeCP2 also has an important function in the stabilization and maintenance of mature neural networks, since its conditional KO in adulthood in a mouse model recapitulates most RTT features[286]. Also, in a mouse model in which *Mecp2* was silenced by insertion of a lox-stop cassette, a conditional reactivation of *Mecp2* in adulthood rescued the RTT phenotype, suggesting that the lack of *Mecp2* during development does not permanently damage neurons [44]. Increasing evidence suggests that MeCP2 has diverse and distinct roles in all stages of brain development and maintenance, and it could carry out its different functions through distinct gene expression changes throughout development[44]. MeCP2 has spatiotemporally diverse effects in different brain regions, stages of development and cell types, and the importance of each effect for cellular dysfunction might depend on the type and state of the cell [4,40]. As recent studies reveal that neurological features in RTT mouse models are reversible, even at late stages of disease progression, a treatment applied to RTT patients could reasonably improve their condition [87].

Diverse downstream effects of MeCP2 at different time points could explain the different levels of GABA ionotropic receptors expression in our experiments. MeCP2 may be involved in positively regulating their expression in early stages of development, while it might not carry on with this function later, when its role is more involved in neural circuit maintenance than in development.

Most $GABA_A$ receptor coding genes are clustered in four chromosomal regions in chromosomes 4, 5, 15 and 19[287]. These subunits comprising the pentameric GABA receptor formation have a coordinated expression, and as revealed by human brain transcriptome analysis, this produces a subject and region-specific expression signature of $GABA_A$ receptor subunits[288,289]. Further studies should be made to elucidate whether MeCP2 acts as a transcriptional regulator of these clusters, the mechanisms through which this regulation takes place, and the time during

development. The differences in expression levels of GABA receptor subunits at different time points of disease progression reinforces the idea of time-dependence in the MeCP2 control of the GABAergic cluster during the specific developmental stages.

Inhibitory dysfunction is a major feature of RTT, so its treatment seems a feasible approach to improving RTT patients' condition[40]. Currently, there is an antiepileptic drug called Vigabatrin which increases GABA levels that is already FDA-approved for use in epilepsy syndromes and clinical trials for RTT are underway[290]. As MeCP2 may have very diverse targets, it is unlikely that restoring the function of a single pathway can revert all RTT symptoms. However, given the implication of the GABAergic function in the RTT phenotype, targeting the GABAergic pathway is still a good approach as *Mecp2* re-expression only in inhibitory neurons in *Mecp2*-null mice rescued motor coordination deficits[40].

Our results suggest a potential implication of MeCP2 as a positive regulator of GABA receptors expression in early stages of RTT. The lack of the most common subunit of the $GABA_A$ receptor in a *MECP2* deficient environment is one of the potential causes of the alteration of GABAergic neurotransmission in RTT. Therefore, early GABAergic modulation in RTT may represent a promising therapeutic strategy. According to our findings, the time window of intervention would be critical, which raises awareness of the importance of the stage of disease when trying to identify potential therapeutic targets for RTT

Chapter 3. Identification of molecular signatures and pathways involved in RTT-spectrum disorders using fibroblast cell lines and a multi-omics approach

3.1. Validation of fibroblasts as a surrogate tissue to study RTT

One of the drawbacks in studying the downstream molecular effects of MeCP2 dysfunction is the lack of accessibility to patient samples of the primarily affected tissue, the brain. The most accessible surrogate tissues to characterize molecular alterations are blood and skin fibroblasts. Fibroblasts have demonstrated greater consistency in gene expression studies and express more OMIM and neurologically relevant genes compared with whole blood[188,291]. Therefore, we considered using primary fibroblast cell cultures obtained from skin biopsies to characterize the molecular landscape in a MeCP2-deficient, patient-derived context, employing a multi-omics approach integrating transcriptomic and proteomic data. To this end, we recruited 22 patients with RTT and *MECP2* mutations, 12 patients with RTT-like phenotypes and mutations in different genes, and 13 healthy controls to generate primary fibroblasts cell cultures, which we subsequently studied by high-throughput transcriptomics (RNAseq) and proteomics (TMT-mass spectrometry).

3.1.1. Cell culture quality control

An initial exploratory inspection of RNAseq data in which we interrogated the primary sources of variation via principal components analysis (PCA) revealed a set of samples, including patients and controls, far away from clustering with the rest (Figure 19A). Intriguingly, some of the outlier samples were in fact technical duplicates from samples that clustered within the main data cloud. The separation of the outlier samples was mainly driven by PC1, encompassing approximately 25% of variance, and strongly determined by the influence of mitochondrial respiratory chain genes (Figure 19B). We hypothesised that those cell lines could have been subject to higher oxidative stress conditions, therefore producing a stress response and overexpressing mitochondrial respiratory chain genes[208].

Following these findings, we re-cultured the outlier samples and performed RT-qPCR of five of these mitochondrial respiratory chain genes (*MT-CO1*, *MT-CO2*, *MT-*

CYB, *MT-ND4* and *MT-ATP6*) in the outlier samples before and after re-culture, and we observed a normalisation of the expression levels of the mitochondrial respiratory chain genes (Figure 19C).

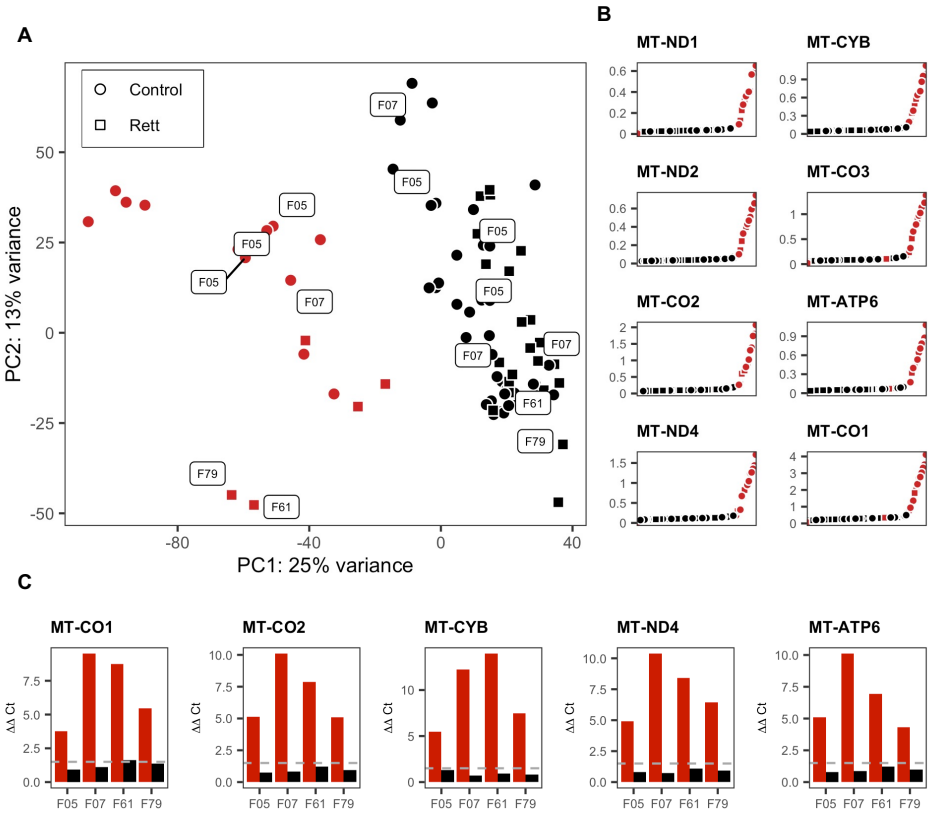


Figure 19. Principal Components Analysis (PCA). A) PCA was performed using the top 10000 variable genes (close to the total number of expressed genes). Samples separated towards the negative values of the x axis (coloured in red) overexpressed many mitochondrial respiratory chain genes with respect to the samples in the more uniform black cluster towards the positive end, and this separation explained approximately 25% of the variance between samples. Some of the genes contributing the most to the separation of samples along PC1 were mitochondrial respiratory chain genes. **B)** Representative scatter plots of some of the mitochondrial respiratory chain genes primarily driving the separation along the x axis of the PCA (PC1). The y axis shows millions of normalized counts. Each dot corresponds to one sample, and they are represented from left to right according to their ranking in expression levels of each gene. **C)** Representative results of RT-qPCR in outlier samples before (red) and after (black) re-culturing.

These results proved that the aberrant expression of mitochondrial respiratory chain genes was related to cell culture conditions and not to the inherent biology of the cell lines. Therefore, all samples displaying an overexpression exceeding 1.5-fold of the control values in two or more of the tested genes were excluded from analysis and subsequently re-cultured and tested by RT-qPCR before re-sequencing to avoid inaccurate results due to cell culture conditions.

3.1.2. Transcriptomic profiles in primary fibroblast cell cultures

To understand how many of the molecular alterations that we identify could be extrapolated to neural tissues, we examined the similarity between the transcriptomic profiles obtained from primary fibroblast cell cultures and those from several brain areas. We used publicly available data from the 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. 41% of brain-expressed transcripts are expressed in fibroblasts (TPM>0.5), and more than 99% of transcripts expressed in fibroblasts correspond to genes with some degree of expression in the nervous system (Figure 20). This indicates that the vast majority of the data that we are analysing may be extrapolated to biological processes occurring in the brain and may therefore reflect neurological phenotypes.

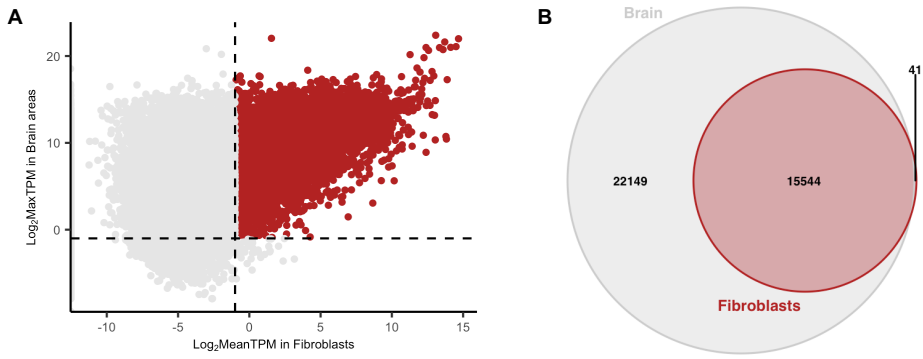


Figure 20. A) Transcript expression levels in fibroblasts versus brain**. B)** Number genes captured in RNAseq fibroblasts and all brain areas in GTEx.

3.1.3. Correlation of MeCP2 expression and phenotypic severity

We next evaluated *MECP2* gene expression in RTT versus control fibroblasts. Whereas there were no differences in expression levels in RNAseq data, we observed a decrease in MeCP2 protein levels in RTT patients compared with control fibroblasts (Figure 21A and B). Moreover, this decrease in MeCP2 protein abundance was significantly correlated with an increase in phenotypic severity indicated by higher Pineda CSS values[21] (Figure 21C).

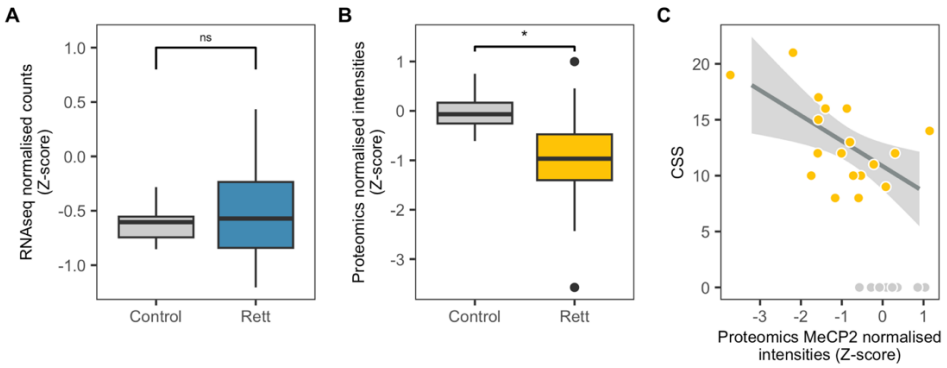


Figure 21. A) *MECP2* RNA expression levels measured by RNAseq data for RTT and control individuals. **B)** MeCP2 protein expression levels measured by proteomics for RTT and control individuals. **C)** Pearson's correlation between the CSS of RTT patients and MeCP2 protein levels (r = -0.56, adjusted R^2 = 0.28, p-value = 0.0122).

All these data suggest that primary fibroblast cell cultures are a suitable surrogate tissue that can capture to a significant extent the consequences of MeCP2 deficiency in a patient-derived context, although away from the primarily affected organ, and we therefore used the obtained data to profile the molecular landscape in patients with RTT.

3.1.4. RNAseq validation

Given that RNAseq is a high-throughput technique, we validated some of the findings to confirm the reliability of the obtained results. To this aim, we chose 21 genes that were called as differentially expressed in RTT samples versus controls in RNAseq and performed RT-qPCR (Figure 22). All the analysed genes displayed

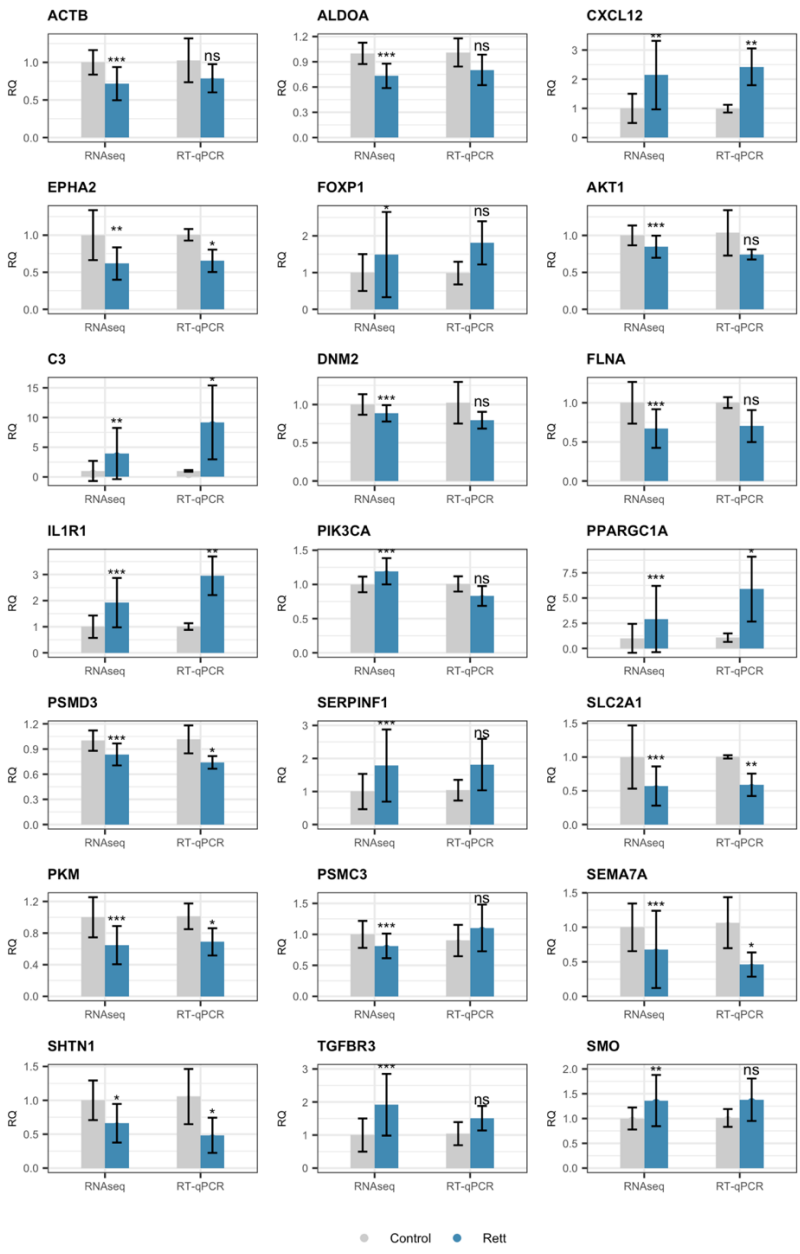


Figure 22. RNAseq results validation by RT-qPCR. 21 genes that were prioritized as significant in the differential expression analysis of the RNAseq data of RTT vs control samples were tested by RT-qPCR. The y axis displays the relative quantification (RQ) of each gene in RTT samples with respect to controls, and significance is denoted with asterisks (*** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05; ns not significant).

an expression change following the same direction (up or downregulated) in RT-qPCR results compared to RNAseq, and 10 (48%) were also deemed significant (p-value < 0.05). These results extensively validate the high-throughput findings and we therefore carried on with the downstream analysis of RNAseq data with no further validation.

3.2. Characterization of molecular alterations in RTT fibroblasts

High-throughput analyses of RNA and protein profiles are efficient approaches to extract a great amount of molecular data from biological samples. However, the obtained data tend to be noisy and complicate the identification of robust gene expression changes. Integrative multi-omics has the potential to distinguish the most reliable targets from these analyses and can contribute to the understanding of the pathomechanisms downstream MeCP2 deficiency and the identification of potential biomarkers and therapeutic targets for RTT.

3.2.1. Differential expression 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 23A). We subsequently used these DEGs as input for upstream regulator analysis with ChIP-X Enrichment Analysis 3 software (ChEA3)[246]. 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-value < 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. 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.

Proteomics differential expression analysis revealed 224 DEPs, 123 upregulated and 101 downregulated (Figure 23B). 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 because mass spectrometry identified roughly half (5918) of the number of genes mapped in the RNAseq experiment (12448). Almost 97% of the proteins detected via mass spectrometry were also identified in RNAseq.

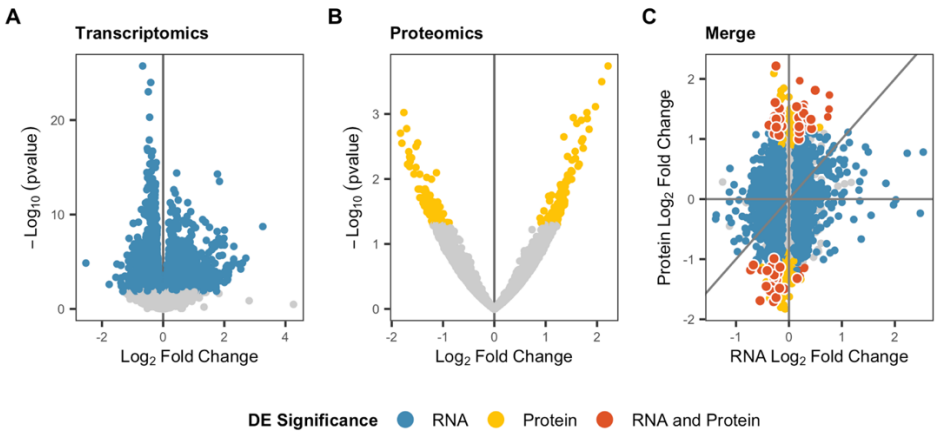


Figure 23. Multi-omics analysis results. A and B) Volcano plots of RNAseq and proteomics DE analysis, blue and yellow dots, represent genes with significant DE in RNAseq and proteomics, respectively. **C)** Fold Change comparison in RNAseq versus proteomics. Red dots represent genes with significant DE both in RNAseq and proteomics, and highlighted red dots correspond to SRF and/or CREB1 potential regulatory targets.

Although the correlation between transcriptome and proteome differential expression findings was not high (Pearson's correlation coefficient = 0.09, p-value = 1.8e-11), we found 75 genes deregulated at both the RNA and protein levels in patients with RTT, many of which were potential targets of SRF and CREB1 (Figure 23C, Table S2, Annex 1). Even though the overlap between DEGs and DEPs was not significantly higher than expected by chance (Fisher's exact test p-value = 0.1397, OR= 1.18), 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 13).

Biological process	Direction	Gene	Potential TFs
Cytoskeletal processes	*Upregulated*	*AFAP1*	SRF
		FMNL2	CREB1
		FNBP1L	CREB1
		KIF3A	-
		MARCKSL1	-
		PLS3	SRF
	Downregulated	*ARMC9*	SRF
		ARHGEF1	-
		CDC42EP1	-
		IQGAP3	SRF
		PLXNB2	CREB1, SRF
RNA processing	*Upregulated*	*EIF4G3*	-
		NUDT12	-
	Downregulated	*SART1*	CREB1, SRF
		DDX31	CREB1, SRF
		DDX54	SRF
		MYBBP1A	SRF
Vesicular activity	*Upregulated*	*NCALD*	-
		PREPL	CREB1
	Downregulated	*TMED1*	SRF
		ZFPL1	CREB1, SRF
Metabolism	*Upregulated*	*CTBS*	-
		HS2ST1	-
	Downregulated	*AGPAT3*	-
		DCAKD	CREB1
		AACS	CREB1
		ORMDL2	-
		PCK2	-
		PI4KB	-
		UAP1L1	CREB1
		COMT	CREB1

Table 13. Genes with concordant differential expression in transcriptomics and proteomics, that are involved in the main biological processes identified via enrichment analysis.

3.2.2. Enrichment analysis: altered pathways in patients with RTT

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 these are especially interesting when trying to elucidate the pathomechanisms underlying RTT. The most remarkable pathways that repeatedly appeared significantly enriched with DEGs and DEPs were

cytoskeletal processes, vesicular activity, rRNA processing and mRNA splicing (Table 13, Figure 24). 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 data.

Cytoskeletal processes

Cytoskeletal actin-filament-based processes play a crucial role in neuronal development, and their dysregulation is associated with cognitive disorders like RTT [86,91,292,293]. We discovered a significant enrichment of cytoskeleton related DEGs and DEPs in RTT patients compared with healthy controls, some of which play important roles in neuronal morphology (Figure 24A).

Our study found that *ARMC9* (OMIM*617612), a gene involved in cilium assembly, signalling 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[294]. 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 which controls dendritic spine morphogenesis and excitatory synapse formation[295]. Moreover, Roux et al. 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[296]. Our study also showed a downregulation of protein levels of Ca^{2+}/calmodulin-activated Ser-Thr kinase (*CASK*, OMIM*300172), a scaffolding protein that is involved in synaptic transmembrane protein anchoring in the brain[297]. CASK dysfunction could be a promising route towards understanding some of the pathomechanisms behind RTT since it has been linked to neurodevelopmental disorders with overlapping phenotypes with RTT [298].

Vesicular activity

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 (Figure 24B). We found

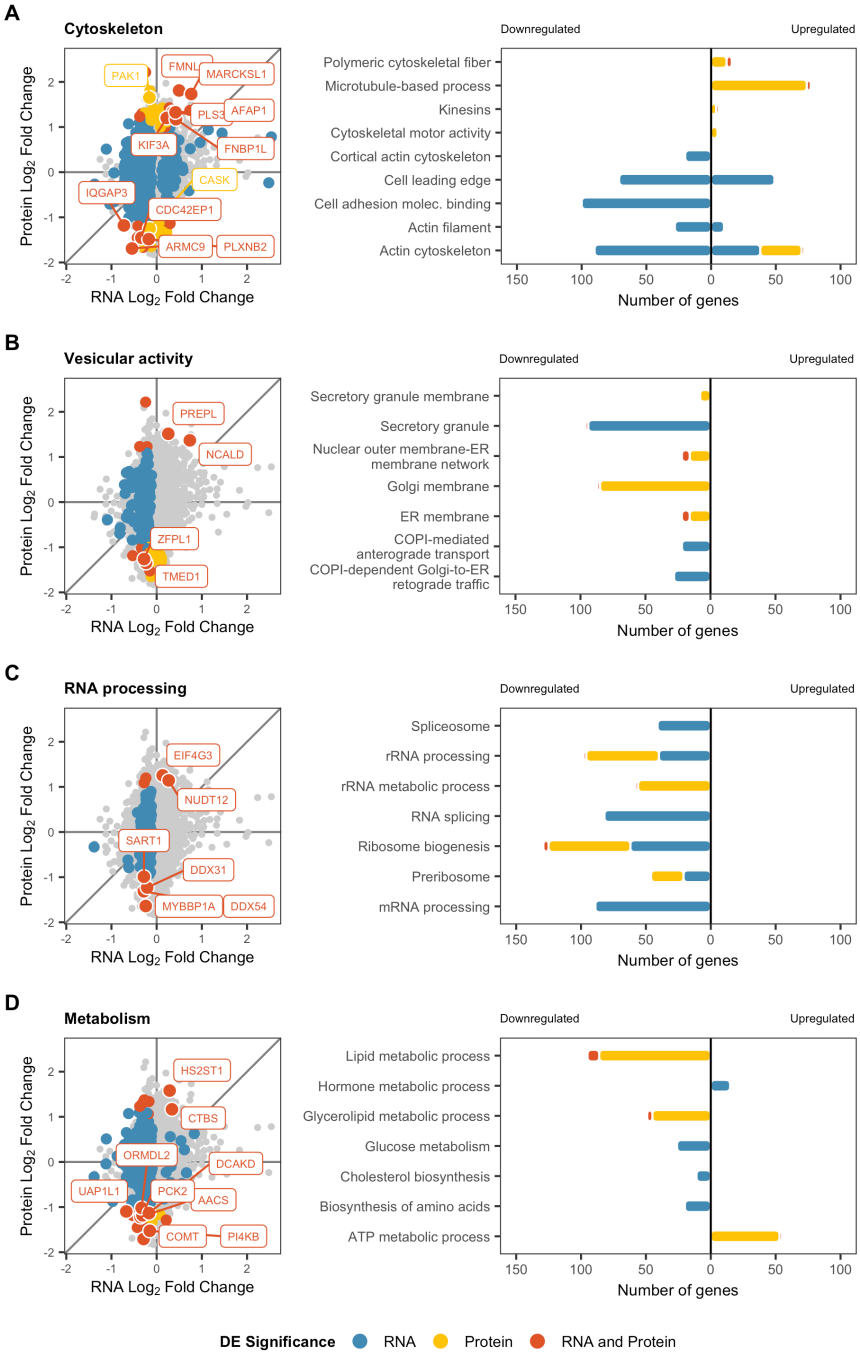


Figure 24. Enrichment analysis results. Left panels represent Fold Changes in RNAseq (x axis) versus proteomics (y axis), and interesting genes are labelled by GeneSymbol. Right panels display the number of genes involved in significantly enriched terms).

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[299].

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[300]. 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[301].

RNA processing

In agreement with literature, we identified a downregulation of genes and proteins involved in rRNA processing and ribosome biogenesis in patients with RTT[302–304] (Figure 24C). This could affect general protein translation in affected cells, possibly due to a reduction in mTORC1 activity[302,303]. The downregulation of three proteins that interact with MeCP2, *MYBBP1A*, *DDX31*, and *DDX54*, could explain alterations in rRNA processing and mRNA splicing[305]. 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, but a recent publication questions its role as a global regulator of splicing[67,162,306,307]. Additional studies are needed to clarify the role of MeCP2 in splicing since many genes involved in mRNA splicing are repeatedly dysregulated in different transcriptomics experiments.

Metabolism

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 (Figure 24D)[308]. 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[309].

TFs potentially driving gene expression changes

CREB1 (OMIM*123810), which is a known MeCP2 interactor, regulates transcription in processes of relevance for neuronal survival and memory consolidation, among others[310,311]. In astrocytes, it regulates genes related to mitochondrial function, vesicle dynamics, and the cytoskeleton[312]. 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)[313]. Besides, SRF regulates neuronal morphology and activity-dependent transcription, 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 a RTT-spectrum phenotype[314,315].

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. Discordance in transcriptomics and proteomics data was similar in other recent studies and may be explained by cellular compensatory processes and further mechanisms of protein dosage regulation[316].

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 RNA processing, which are altered in patients with

RTT. TF analysis supports CREB1 and SRF transcriptional regulation as potential therapeutic targets.

3.3. Characterisation of molecular alterations in RTT-spectrum fibroblasts

3.3.1. Differential expression and upstream regulator analysis

The RTT-like cohort of patients was recruited considering their phenotypic 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 BH-adjusted p-value threshold of < 0.1 (instead of the 0.05 threshold used with RTT data) for transcriptomics to be able to call DEGs despite the data heterogeneity. We interpreted these results in comparison with those obtained in typical RTT patients, searching for shared molecular alterations that could constitute common grounds in the pathogenesis of overlapping disorders of diverse genetic nature.

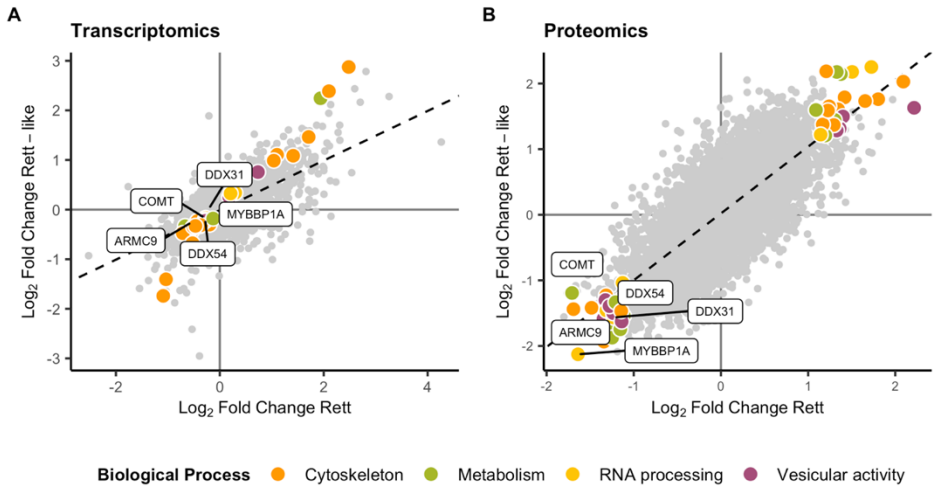


Figure 25. Common DE findings between the analysis of RTT and RTT-like patients versus healthy controls. Change in expression relative to healthy controls in RTT (x axis) versus RTT-like (y axis) in **A)** transcriptomics and **B)** proteomics. Highlighted genes presented a consistently significant DE in both RTT and RTT-like cohorts and are related to the main

biological processes that were identified in enrichment analysis (and are coloured accordingly). Dashed lines depict the correlation in DE results from both cohorts (**A)** r = 0.69, adj-R^2 = 0.47, p < 0.001; **B)** r = 0.75, adj-R^2 = 0.56, p < 0.001).

DE analysis of transcriptomics data revealed 63 genes consistently altered in patients with RTT and RTT-like phenotypes (25 upregulated and 38 downregulated) (Figure 25A). 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 25B). Interestingly, transcriptomic and proteomics 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), indicating common gene expression changes (Figure 25A and B). No gene was altered both in transcriptomics and proteomics reaching significance, but 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 14).

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

Table 14. 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.

In order to study the similarity of the transcriptional profiles observed in patients with RTT and patients with RTT-spectrum phenotypes with mutations in different genes, we correlated gene expression changes (fold changes) versus healthy controls obtained in RNAseq (Figure 26). Patients with mutations in *CDKL5* (n=4), *FOXG1* (n=2) and *NR2F1* (n=1) presented transcriptomic profiles with significant

correlations with those of patients with *MECP2* mutations (Figure 26A, B and C), whereas the fibroblast transcriptome of the patient with a *GRIN2B* mutation showed no correlation (Figure 26D).

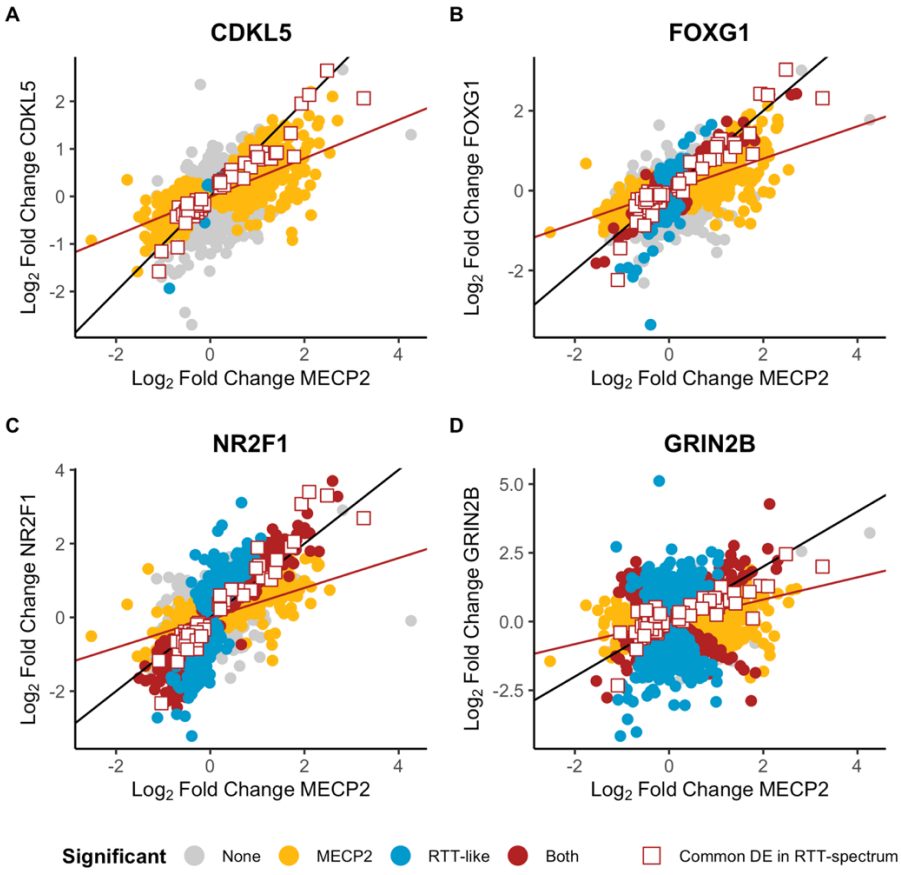


Figure 26. Correlation of gene expression levels quantified in transcriptomics analysis. Each panel depicts the correlation in gene expression levels (Log_2 Fold Change) in patients with mutations in *MECP2* (x axis) and patients with mutations in a RTT-like gene (y axis): **A)** *CDKL5*; **B)** *FOXG1*; **C)** *NR2F1*; **D)** *GRIN2B*. Yellow dots represent genes exclusively DE in patients with *MECP2* mutations, blue dots represent genes exclusively DE in patients with mutations in each RTT-like gene, and red dots represent genes DE in both cohorts of patients. Squares correspond to genes commonly DE in all RTT-like patients. The black trend line corresponds to the formula y = x. The red trend line depicts the linear regression model of the data in each panel (**A)** r = 0.56 | Adj-R^2= 0.32 | p-value = 0; **B)** r = 0.44 | Adj-R^2 = 0.19 | p-value = 0; **C)** r = 0.7 | Adj-R^2 = 0.49 | p-value = 0; **D)** r = 0 | Adj-R^2 = 0 | p-value = 0.6459).

The *GRIN2B* gene encodes a glutamate receptor and has an expression profile limited to the nervous system, and therefore it is reasonable that its dysfunction does not cause extra-neurologic gene expression changes compatible with those caused by MeCP2 deficiency.

3.3.2. Enrichment analysis: altered pathways in patients with RTT-spectrum phenotypes

Enrichment analysis of common DEGs and DEPs revealed terms related to cytoskeletal organisation, RNA processing, vesicular activity, and metabolism, which constitute molecular alterations shared in patients with typical RTT and RTT-like phenotypes and could explain phenotypic overlap to some extent (Figure 27A and B).

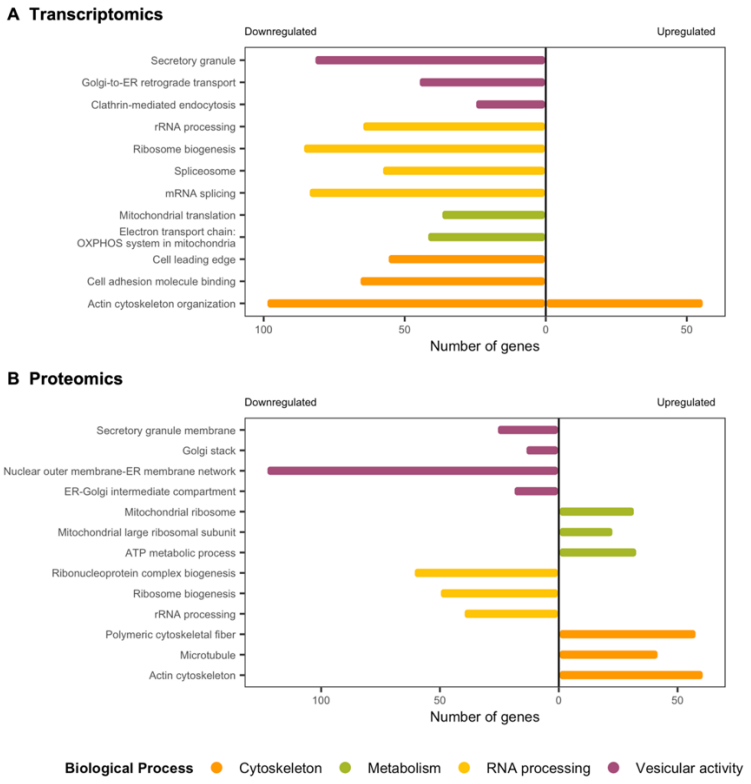


Figure 27. Common Enrichment Analysis findings between the analysis of RTT and RTT-like patients versus healthy controls. Gene Set Enrichment Analysis (GSEA) results for the shared 63 DEGs **(A)** and 81 DEPs **(B)**, coloured by biological process.

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[317,318]. 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 is mirrored in patients with RTT-like phenotypes, constituting a link between RTT-spectrum disorders and overlapping the phenotype caused by loss-of-function variants in this gene (Joubert syndrome).

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[95,303]. *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 autism spectrum disorder (ASD)[319].

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.

Chapter 4. Rett syndrome spectrum disorders: molecular diagnosis and characterization of molecular landscape

4.1. Molecular diagnosis of Rett syndrome spectrum patients

In the last few years, the use of massively parallel sequencing in patients with a clinical diagnosis of RTT has uncovered mutations in many different genes. Some of these genes were novel, while others were already associated with other neurodevelopmental disorders with overlapping phenotypes with RTT. These phenotypes are sometimes termed RTT-spectrum disorders and constitute a vague group of overlapping clinical entities caused by mutations in different genes. Biological interactions between these genes may explain, at least in part, the overlapping phenotypes.

4.1.1. Rett syndrome spectrum candidate gene list

Given the genetic heterogeneity of RTT-spectrum disorders, our first approach was to restrict the initial analysis to a limited list of genes related to RTT-spectrum phenotypes. To this aim, we curated a candidate gene list using phenotypic data from the Online Mendelian Inheritance in Men (OMIM) and the Human Phenotype Ontology (HPO) databases.

We quantified the similarity of the clinical phenotype associated with each gene with current clinical RTT diagnostic criteria using a scoring system[11] (Table 15). We assigned the highest scoring to main criteria and developmental regression, since they are necessary to consider a diagnosis of RTT, and the lowest scoring to supportive criteria and microcephaly. We assigned a medium scoring to major descriptions of neurodevelopmental disorders and epileptic encephalopathies (developmental delay, intellectual disability, and seizures), to cover unspecific overlapping phenotypes.

Main criteria (*score = 3*)	Supportive criteria (*score = 1*)	Major description (*score = 2*)
1. Stereotypic hand movements 2. Loss of purposeful hand skills 3. Loss of spoken language 4. Gait abnormalities 5. Developmental regression	1. Breathing disturbances 2. Bruxism 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 12. Microcephaly	1. Developmental delay 2. Intellectual disability 3. Seizures

Table 15. Scoring of phenotypic descriptions according to RTT diagnostic criteria.

Using this scoring system, a patient fulfilling criteria for typical RTT should have a minimum score of 15 (developmental regression and 4 main criteria), and a patient fulfilling criteria for atypical RTT should have a minimum score of 14 (developmental regression, 2 main criteria and 5 supportive criteria). Using phenotypic data from the OMIM and HPO databases, we scored all genes previously related to neurodevelopmental disorders, epileptic encephalopathies and implicated in glutamatergic and GABAergic neurotransmission. The highest scoring gene, with a score of 32, was *MECP2*, for which OMIM and HPO databases contained all mentioned criteria except for intense eye communication (Table 16). *CDKL5* and *FOXG1* were also contained within the top 10 scoring genes, with scores of 30 and 26, respectively (Table 16). After scoring all genes, we specified a cut-off score of 6 to include a gene in the curated RTT-spectrum gene set, which would correspond to 2 main criteria or all 3 major descriptions, considering this threshold would retain genes with unspecific overlapping phenotypes and leave out genes not associated to any known clinical entity with phenotypic characterization or with spurious similarities with a RTT phenotype. A score of 6 was also the minimal score assigned

Gene	Phenotype	Developmental regression	Hand skills impairment	Language impairment	Gait abnormalities	Stereotypic movements	Microcephaly	Breathing disturbances	Bruxism	Sleep disturbances	Abnormal muscle tone	Vasomotor disturbances	Growth retardation	Cold hands and feet	Laughing and screaming spells	Diminished pain sensitivity	Intense eye communication	Developmental delay	Intellectual disability	Seizures	Total
MECP2	Rett syndrome	3	3	3	3	3	1	1	1	1	1	1	1	1	1	1	0	2	2	2	**32**
CDKL5	Developmental and epileptic encephalopathy 2	3	3	3	3	3	1	1	1	1	1	0	1	0	1	1	0	2	2	2	**30**
GABBR2	Developmental and epileptic encephalopathy 59	3	3	3	3	3	1	1	1	1	1	0	1	0	1	1	0	2	2	2	**30**
NTNG2	NDD with behavioural abnormalities, absent speech, and hypotonia	3	3	3	3	3	1	0	1	1	1	0	0	1	1	0	0	2	2	2	**28**
STXBP1	Developmental and epileptic encephalopathy 4	3	3	3	3	3	1	1	0	1	1	0	1	0	0	1	0	2	2	2	**28**
FOXG1	Rett syndrome, congenital variant	3	0	3	3	3	1	1	1	1	1	0	1	0	1	0	0	2	2	2	**26**
HTT	Lopes-Maciel-Rodan syndrome	3	0	3	3	3	1	1	1	1	1	0	0	1	0	0	0	2	2	2	**25**
SMC1A	Cornelia de Lange syndrome 2	3	0	3	3	3	1	0	0	1	1	0	1	0	0	1	0	2	2	2	**24**
ACTL6B	Developmental and epileptic encephalopathy 76	3	0	3	3	3	1	0	0	1	1	0	1	0	0	0	0	2	2	2	**23**
GRIN1	Developmental and epileptic encephalopathy 101	3	0	3	3	3	1	0	0	1	1	0	1	0	0	0	0	2	2	2	**23**

Table 16. Top scoring genes according to phenotypic similarity with RTT diagnostic criteria.

to genes that have published variants in HGMD linked to a RTT-spectrum phenotype. With this threshold, we obtained a list of 1245 candidate RTT-spectrum genes to use as a first-tier filter to evaluate NGS findings (Table S3, Annex 1).

4.1.2. Exome, genome, and multi-omics for diagnosis: diagnostic yield

After an initial search for mutations in the genes *MECP2*, *CDKL5* and *FOXG1* by direct sequencing and MLPA analysis, we followed a sequential workflow of NGS approaches of increasing coverage to look for alternative genetic causes in negative cases (Figure 28). The first NGS approach was mendeliome, or clinical exome, sequencing (CES), which consists of a targeted panel that covers the coding regions of all genes associated with human disease. Next, for cases with no pathogenic, likely pathogenic or candidate variants identified in CES first analysis or re-analysis, we performed whole exome sequencing (WES), which consists of a targeted assay of exonic regions of all genes in the human genome.

CES and WES analyses were performed as a singleton test, and therefore the identified candidate variants were segregated in the family trio by Sanger sequencing, MLPA or qPCR when parental samples were available to assess their inheritance pattern.

Some cases that remained negative after WES first analysis and re-analysis were analysed by trio whole genome sequencing (WGS). To overcome the difficulties in variant interpretation from WGS, we combined it with transcriptomic and proteomic profiling in fibroblast cell cultures searching for aberrant splicing and aberrant and monoallelic expression events that could prioritise candidate genes and variants.

The bioinformatics pipeline for variant calling consisted of aligning the sequencing reads to the GRCh37/hg19 reference genome after initial quality control and pre-processing, followed by variant detection using several variant callers. GATK, DeepVariant and Octopus were used for SNV and indel variant calling in all data, while CNVs were called by read depth analysis using ExomeDepth on exonic data (CES and WES) and four variant callers (Lumpy, Manta, Delly and Wham) considering read depth, split reads and discordant read pairs in genomic data. In the

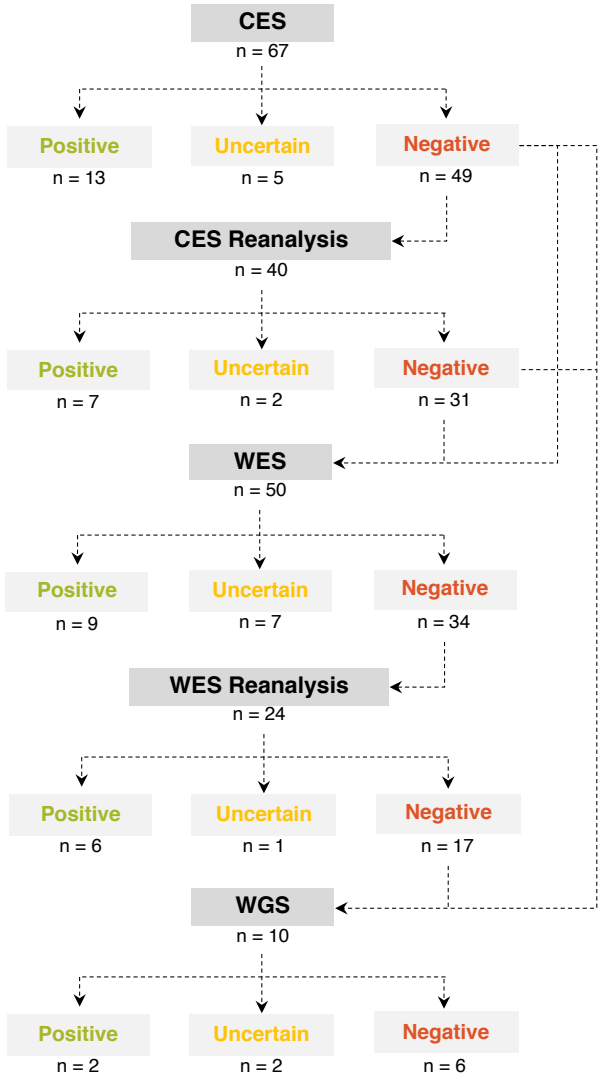


Figure 28. Sequential NGS workflow for negative cases in initial analysis. Initially, 67 patients were analysed by CES, obtaining 13 positive, 5 uncertain and 49 negative results. 40 of the negative results were re-analysed at least after the first analysis, obtaining another 7 positive and 2 uncertain results. Two of the negative CES cases were directly passed on to WGS analysis and multi-omics analysis due to the fibroblast availability. The analysis of 37 other negative cases was widened to WES, together with 13 new patients, obtaining 9 positive, 7 uncertain and 34 negative results. 24 of the negative cases were re-analysed at least one year after the first analysis, obtaining 6 new diagnoses and one candidate result. Finally, negative cases with both parental blood samples and proband's fibroblast availability were analysed by trio WGS and multi-omics profiling of the fibroblasts.

data analysis of each assay, SNVs and indels were initially filtered according to a stringent population frequency threshold (< 1%), a high or moderate impact on the protein and having been called by at least two of the three variant callers as a quality filter. As a first-tier analysis, we examined the genes in the curated RTT-spectrum gene list, and when negative, we examined all rare, high-quality variants with high or moderate impact on the protein in an agnostic manner (with no candidate gene list). In trio WGS analysis, we also filtered and analysed *de novo* variants, plus all variants (coding and noncoding) in genes with aberrant expression phenotypes in fibroblasts. Structural variant analysis was initially limited to variants with low internal and external population frequencies and containing coding sequences, and later extended to any *de novo* variation (in *trio* analysis).

We considered that an NGS test result was positive when pathogenic or likely pathogenic variants, matching the patient's phenotype and observed mode of inheritance, were identified; uncertain when variants of uncertain significance, matching the patient's phenotype and observed mode of inheritance, or pathogenic or likely pathogenic variants, matching the patient's phenotype but not the observed mode of inheritance, were identified; and negative when no diagnostic or candidate variants were identified in the analysis. Positive and candidate results of all analysis are summarised in Table 17.

CES - Positive								
Gene	**OMIM***	**MOI**	**OMIM#**	**Phenotype**	**Variant**	**Inh**	**ACMG**	**Criteria**
SLC6A1	137165	AD	616421	Myoclonic-atonic epilepsy	DEL: chr3:11058811-11078744	NA	**Pathogenic**	2A(+1.00)
SYNGAP1	603384	AD	612621	Intellectual developmental disorder, autosomal dominant 5	NM_006772.2: c.1167_1168delAG / p.Gly391fs	NA	**Pathogenic**	PVS1-vstr, PS2-str, PM2-sup
CSTB	601145	AR	254800	Epilepsy, progressive myoclonic 1A (Unverricht and Lundborg)	NM_000100.3: c.67-1G>C / p.?	Mat	**Pathogenic**	PVS1-str, PS4-str, PM2-sup, PM3-vstr
					NM_000100.3: c.67-1G>C / p.?	Pat	**Pathogenic**	PVS1-str, PS4-str, PM2-sup, PM3-vstr
136 genes	.	.	.	.	DUP: chr20:42743391-49552886	NA	**Pathogenic**	2K(+0.30), 3C(+0.90)
TCF4	602272	AD	610954	Pitt-Hopkins syndrome	DEL: chr18:53017529-53254395	*De novo*	**Pathogenic**	2C(+0.90), 4L(+0.10)
CHAMP1	616327	AD	616579	Neurodevelopmental disorder with hypotonia, impaired language, and dysmorphic features	NM_001164144.1: c.959dupC / p.Arg321fs	*De novo*	**Pathogenic**	PVS1-vstr, PM2-sup, PM6-mod
TRAPPC9	611966	AR	613192	Intellectual developmental disorder, autosomal recessive 13	NM_031466.7: c.2994-2A>T / p.?	Mat	**Likely Pathogenic**	PVS1-vstr, PM2-sup, PM3-sup
					NM_031466.7: c.2994-2A>T / p.?	Pat	**Likely Pathogenic**	PVS1-vstr, PM2-sup, PM3-sup
297 genes	.	.	.	.	DUP: chr14:90871003-106322353	*De novo*	**Pathogenic**	3C(+0.90), 4L(+0.15)
12 genes	.	.	.	.	DEL: chr1:243419424-245027648	NA	**Pathogenic**	2A(+1.00), 4L(+0.15)

88 genes	.	.	.	.	DEL: chr9:130216796-133376418 (*STXBP1*)	NA	**Pathogenic**	2A(+1.00), 3C(+0.90), 4L(+0.15)
TCF4	602272	AD	610954	Pitt-Hopkins syndrome	NM_001243228.2: c.1751G>A / p.Arg584His	NA	**Pathogenic**	PS4-str, PS3-sup, PM6-str, PM1-mod, PP2-sup, PM2-sup, PM5-mod, PP3-mod
36 gens	.	.	.	.	DEL: chr22:50297395-51169828 (*SHANK3*)	NA	**Pathogenic**	2A(+1.00), 3B(+0.45), 4L(+0.15)
SETBP1	611060	AD	616078	Intellectual developmental disorder, autosomal dominant 29	NM_015559.2: c.2572G>A / p.Glu858Lys	*De novo*	**Likely Pathogenic**	PM2-sup, PM1-mod, PM6-str
			269150	Schinzel-Giedion midface retraction syndrome				
CES - Uncertain								
Gene	**OMIM***	**MOI**	**OMIM#**	**Phenotype**	**Variant**	**Inh**	**ACMG**	**Criteria**
HUWE1	300697	XL	309590	Intellectual developmental disorder, X-linked syndromic, Turner type	NM_031407.7: c.12648A>G / p.Thr4216=	Not Mat	**VUS**	PM2-sup, PP3-mod
ZBTB18	608433	AD	612337	Intellectual developmental disorder, autosomal dominant 22	NM_205768.2: c.510delA / p.Gly171fs	NA	**VUS**	PVS1-str, PM2-sup
SCN2A	182390	AD	613721	Developmental and epileptic encephalopathy 11	NM_001040142.1: c.379G>T / p.Val127Leu	*De novo*	**VUS**	PM2-sup, PP2-sup, PM6-mod

UNC80	612636	AR	616801	Hypotonia, infantile, with psychomotor retardation and characteristic facies 2	NM_032504.1: c.2045G>T / p.Cys682Phe	Mat	VUS	PM2-sup, PP3-sup, PP2-sup
					NM_032504.1: c.8407G>C / p.Val2803Leu	Pat	VUS	PM2-sup, PP2-sup
WDFY3	617485	AD	617520	?Microcephaly 18, primary, autosomal dominant	NM_014991.4: c.7906G>A / p.Gly2636Arg	*De novo*	VUS	PM2-sup, PP2-sup, PM6-mod
CES Re-analysis - Positive								
Gene	**OMIM***	**MOI**	**OMIM#**	**Phenotype**	**Variant**	**Inh**	**ACMG**	**Criteria**
DNM1	602377	AD	616346	Developmental and epileptic encephalopathy 31A	NM_001288739.1: c.1612A>G / p.Lys538Glu	*De novo*	Likely Pathogenic	PM2-sup, PP2-sup, PP3-sup, PM6-mod
10 genes	.	.	.	.	DEL: chrX:146993433-148713298	NA	Pathogenic	2A(+1.00),4L(+0.10)
MAGEL2	605283	AD	615547	Schaaf-Yang syndrome	NM_019066.4: c.2874G>A / p.Trp958*	*De novo*	Pathogenic	PVS1-vstr, PM2-sup, PM6-str
HIVEP2	143054	AD	616977	Intellectual developmental disorder, autosomal dominant 43	NM_006734.3: c.2842_2852delGAGCACTCCTC / p.Glu948fs	*De novo*	Pathogenic	PVS1-vstr, PM2-sup, PM6-mod
DYRK1A	600855	AD	614104	Intellectual developmental disorder, autosomal dominant 7	NM_001396.3: c.557_560delCCAT / p.Ala186fs	NA	Likely Pathogenic	PVS1-vstr, PM2-sup
ARID1B	614556	AD	135900	Coffin-Siris syndrome 1	NM_001374828.1: c.2458C>T / p.Arg820*	*De novo*	Pathogenic	PVS1-vstr, PM2-sup, PM6-mod
SHANK3	606230	AD	606232	Phelan-McDermid syndrome	NM_001372044.2: c.5093_5111del / p.Ser1698CysfsTer41	*De novo*	Pathogenic	PVS1-vstr, PM2-sup, PM6-mod

CES Re-analysis - Uncertain								
Gene	**OMIM***	**MOI**	**OMIM#**	**Phenotype**	**Variant**	**Inh**	**ACMG**	**Criteria**
PDHA1	300502	XLD	312170	Pyruvate dehydrogenase E1-alpha deficiency	NM_001173454.1: c.541G>C / p.Gly181Arg	Mat	VUS	PM2-sup, PM1-mod, PP3-sup
PGAP1	611655	AR	615802	Neurodevelopmental disorder with dysmorphic features, spasticity, and brain abnormalities	NM_024989.3: c.2709T>G / p.Tyr903*	Mat	VUS	PVS1-mod, PM2-sup, PM3-sup
					NM_024989.3: c.2709T>G / p.Tyr903*	Pat	VUS	PVS1-mod, PM2-sup, PM3-sup
WES - Positive								
Gene	**OMIM***	**MOI**	**OMIM#**	**Phenotype**	**Variant**	**Inh**	**ACMG**	**Criteria**
DDX3X	300160	XL	300958	Intellectual developmental disorder, X-linked syndromic, Snijders Blok type	NM_001193416.3: c .543+5G>T / p.?	*De novo*	Likely Pathogenic	PM2-sup, PM6-str, PP3-mod
GABBR2	607340	AD	617904	Developmental and epileptic encephalopathy 59	NM_005458.8: c.1699G>A / p.Ala567Thr	*De novo*	Pathogenic	PS3-sup, PS4-str, PM2-sup, PM6-mod, PP2-sup, PP3-sup
			617903	Neurodevelopmental disorder with poor language and loss of hand skills				
MECP2	300005	XLD	312750	Rett syndrome	DEL: chrX:153295564-153296520	NA	Pathogenic	2D(+0.90),4L(+0.15)
SYNGAP1	603384	AD	612621	Intellectual developmental disorder, autosomal dominant 5	NM_006772.3: c.3583-6G>A / p.?	*De novo*	Likely Pathogenic	PS2-str, PM2-sup, PP3-mod

GATAD2B	614998	AD	615074	GAND syndrome	NM_020699.4: c.1283C>T / p.Thr428Ile	NA	**Likely Pathogenic**	PM2-sup, PP3-str, PP2-sup
SHANK3	606230	AD	606232	Phelan-McDermid syndrome	NM_001372044.2: c.3264_3274delCCCGGCCAAGG (p.Lys1091Alafs*276)	*De novo*	**Likely Pathogenic**	PVS1-vstr, PM2-sup
RAB11B	604198	AD	617807	Neurodevelopmental disorder with ataxic gait, absent speech, and decreased cortical white matter	NM_004218.4: c.202G>A (p.Ala68Thr)	*De novo*	**Pathogenic**	PS3-sup, PS4-str, PM2-sup, PM6-mod, PP3-mod
UBTF	600673	AD	617672	Neurodegeneration, childhood-onset, with brain atrophy	NM_014233.4: c.628G>A / p.Glu210Lys	NA	**Pathogenic**	PS2-str, PS3-sup, PS4-str, PM2-sup, PP2-sup, PP3-sup
SYNGAP1	603384	AD	612621	Intellectual developmental disorder, autosomal dominant 5	NM_006772.3: c.1716G>A / p.Trp572*	*De novo*	**Pathogenic**	PVS1-vstr, PM2-mod, PM6-mod
WES - Uncertain								
Gene	**OMIM***	**MOI**	**OMIM#**	**Phenotype**	**Variant**	**Inh**	**ACMG**	**Criteria**
MED12	300188	XLD	301068	Hardikar syndrome	NM_005120.3: c.3844G>A / p.Val1282Met	*De novo*	**VUS**	PM2-sup, PP3-sup, PP2-sup
		XLR	309520	Lujan-Fryns syndrome				
		XLR	300895	Ohdo syndrome, X-linked				
		XLR	305450	Opitz-Kaveggia syndrome				
UNC13A	609894	.	.	.	NM_001080421.3: c.2422G>A / p.Gly808Ser	NA	**VUS**	PM2-sup, PP2-sup
DNMT3A	602769	AD	618724	Heyn-Sproul-Jackson syndrome	NM_022552.5: c.820G>A / p.Ala274Thr	NA	**VUS**	PM2-sup, PP2-sup
			615879	Tatton-Brown-Rahman syndrome				

KDM6B	611577	AD	618505	Neurodevelopmental disorder with coarse facies and mild distal skeletal abnormalities	NM_001348716.2: c.2013T>G / p.Phe671Leu	NA	VUS	PM2-sup
SCN1A	182389	AD	619317	Developmental and epileptic encephalopathy 6B, non-Dravet	NM_006920.6: c.2556+2_2556+3insC / p.?	NA	VUS	PVS1-str, PM2-sup
			607208	Dravet syndrome				
JMJD1C	604503	.	.	.	NM_0327776.3: c.488C>T / p.Pro163Leu	*De novo*	VUS	PS3-sup, PM2-sup, PM6-mod
CACNA1E	601013	AD	618285	Developmental and epileptic encephalopathy 69	NM_001205293.3: c.5077C>T / p.Arg1693Cys	NA	VUS	PM2-sup, PP3-mod, PP2-sup
WES Re-analysis - Positive								
Gene	**OMIM***	**MOI**	**OMIM#**	**Phenotype**	**Variant**	**Inh**	**ACMG**	**Criteria**
NR2F1	132890	AD	615722	Bosch-Boonstra-Schaaf optic atrophy syndrome	NM_005654.5: c.421T>G / p.Cys141Gly	*De novo*	Likely Pathogenic	PM2-sup, PM1-mod, PM6-mod, PP2-sup, PP3-mod
HNRNPH2	300610	XLD	300986	Intellectual developmental disorder, X-linked syndromic, Bain type	NM_001257293.2.5: c.617G>A /p.Arg206Gln	*De novo*	Likely Pathogenic	PM2-sup, PM5-mod, PP2-sup, PM6-mod
IQSEC2	300522	XLD	309530	Intellectual developmental disorder, X-linked 1	NM_001111125.3: c.1638G>A / p.Trp546*	*De novo*	Likely Pathogenic	PVS1-vstr, PM2-sup, PS2-str
PHIP	612870	AD	617991	Chung-Jansen syndrome	NM_017934.5: c.225G>C / p.Leu75Phe	*De novo*	Likely Pathogenic	PM2-sup, PP2-sup,PS2-str

PURA	600473	AD	616158	Neurodevelopmental disorder with neonatal respiratory insufficiency, hypotonia, and feeding difficulties	NM_005859.5: c.810dupC / p.Phe271fs	*De novo*	**Pathogenic**	PVS1-vstr, PM2-sup, PS2-str
DNMT3A	602769	AD	618724	Heyn-Sproul-Jackson syndrome	NM_022552.4: c.2095G>A / p.Gly699Ser	*De novo*	**Pathogenic**	PM2-sup, PP3-str, PP2-sup, PS2-str
			615879	Tatton-Brown-Rahman syndrome				
WES Re-analysis - Uncertain								
Gene	**OMIM***	**MOI**	**OMIM#**	**Phenotype**	**Variant**	**Inh**	**ACMG**	**Criteria**
ZNF142	604083	AR	618425	Neurodevelopmental disorder with impaired speech and hyperkinetic movements	NM_001105537.2: c.3887G>C / p.Ser1296Thr	Mat	**VUS**	PM2-sup
					NM_001105537.2: c.3406C>G / p.His1136Asp	Pat	**VUS**	PM2-sup, BP4
WGS - Positive								
Gene	**OMIM***	**MOI**	**OMIM#**	**Phenotype**	**Variant**	**Inh**	**ACMG**	**Criteria**
TCEAL1	300237	XLD	301094	Hijazi-Reis syndrome	DEL chrX:102589677-103221990	*De novo*	**Pathogenic**	2A(+1.00),4L(+0.15)
SCN8A	600702	AD	614306	Cognitive impairment with or without cerebellar ataxia	NM_001330260.2: c.671T>A / p.Val224Asp	*De novo*	**Pathogenic**	PM2-sup, PP3-str, PP2-sup, PS2-str
			614558	Developmental and epileptic encephalopathy 13				

WGS - Uncertain								
Gene	**OMIM***	**MOI**	**OMIM#**	**Phenotype**	**Variant**	**Inh**	**ACMG**	**Criteria**
ASXL3	615115	AD	615485	Bainbridge-Ropers syndrome	NM_030632.3:c.2435T>C / p.Ile812Thr	*De novo*	**VUS**	PM2-sup, PS2-str
QARS1	603727	AR	615760	Microcephaly, progressive, seizures, and cerebral and cerebellar atrophy	NM_005051.3: c.1614+3A>G / p.?	Mat	**Likely Pathogenic**	PM2-sup, PP3-mod, PS3-str
					NM_005051.3: c.869A>G / p.Tyr290Cys	Pat	**VUS**	PM2-sup, PP3-sup, PM3-mod

Table 17. Diagnostic and candidate NGS results by CES, WES and WGS.

We analysed 67 patients by CES, with a diagnostic yield of 19% (13/67) after the first analysis. Candidate variants of unknown significance were identified in another 7% of patients (5/67) (Figure 29). A re-analysis of the genomic data at least one year after the initial analysis increased the diagnostic yield to 30% (20/67) and identified candidate variants in 2 more patients. The main reasons for the reporting of new variants in re-analysis were that they were either in genes not analysed in the previous analysis (5 cases), or re-classified during the time between analyses (1 case). Therefore, CES produced a definitive or a candidate result for 40% of patients (27/67).

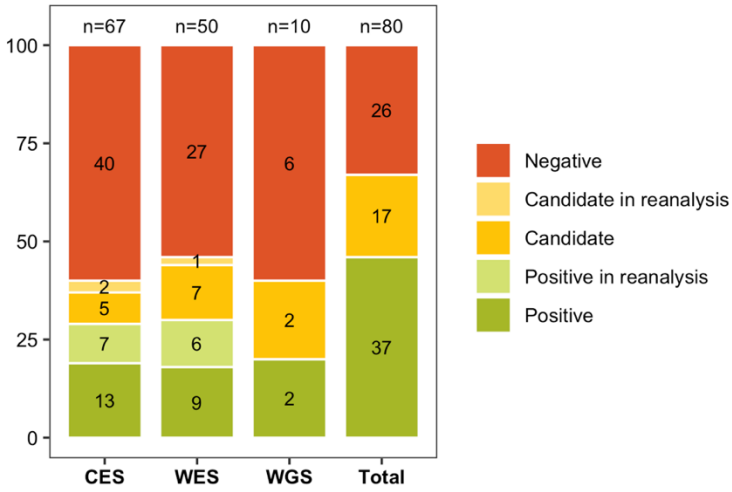


Figure 29. Results of CES, WES, WGS and in total. 67 patients were analysed by CES, obtaining 20 positive results (13 in first analysis and 7 in re-analysis) and 7 candidate variants (5 in first analysis and 2 in re-analysis). 50 patients were analysed by WES, obtaining 15 positive results (9 in first analysis and 6 in re-analysis) and 8 candidate variants (7 in first analysis and 1 in re-analysis). Trio WGS analysis combined with multi-omics profiling in 10 patients produced 2 positive results and 2 candidate variants. Globally, we obtained a positive diagnostic result in 37/80 cases (46%), and a candidate variant in 17/80 cases (21%), leaving 32% (26/80) of patients with no informative results from genetic testing.

Next, we analysed 50 patients by WES (37 with a negative CES result and 13 with WES as a first-tier diagnostic test) with a diagnostic yield of 18% (9/50). We identified candidate variants of unknown significance in another 7 patients (14%). Most variants reported in WES analysis after a negative CES study were not

covered by the previous assay (6 cases) or not detected or filtered out in the previous analysis due to issues in the uniformity of coverage (6 cases), including a deletion in the *MECP2* gene. Some of the variants were also re-classified during the time between analysis (2 cases). Negative WES cases were re-analysed at least one year after the initial analysis, obtaining 6 new diagnoses and 1 candidate result, and increasing diagnostic yield to 30%. These new variants were reported because of new gene-disease relationships being published and/or patient phenotype re-evaluation. In total, WES produced a definitive or a candidate result for 46% of patients (Figure 29).

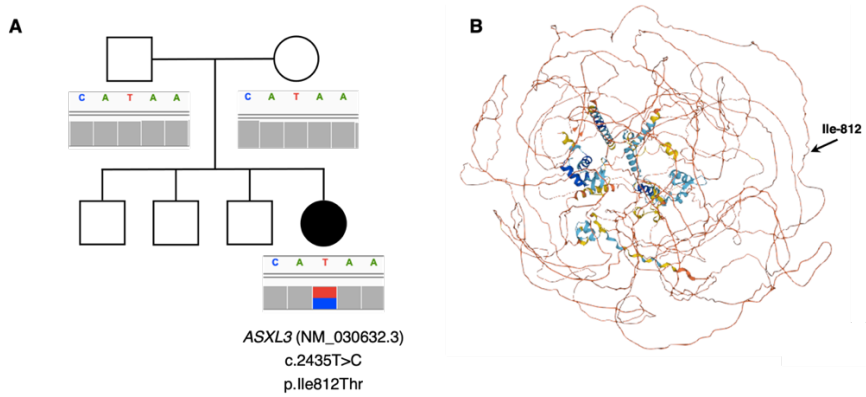
Finally, we analysed 8 families with 10 girls clinically diagnosed with RTT by trio-WGS. We combined WGS data with transcriptomic and proteomic data from fibroblast cell cultures of the probands to help prioritise candidate genes and variants using evidence from aberrant expression events. This approach yielded 2 new diagnoses and 2 candidate results, producing a definitive or candidate result for 40% of studied patients (Figure 29).

Trio WGS and multi-omics

Case 1: Re-classification of a variant

The index patient in this family was a 42-year-old woman with three healthy brothers. After normal pregnancy and neonatal period, she started evidencing a psychomotor delay at age 5 months and lost contact with the environment at age 6 months. She achieved independent sitting at age 16 months, and she used her hands to grasp objects at age 20 months, but lost both skills at age 2 years, and never achieved independent sitting. She babbled at the age of 1, but never acquired purposeful language. She started to present stereotypical hand movements including hand washing and hand flapping at age 5, and she presented apnoeas, kyphoscoliosis and peripheral vasomotor disturbances at age 6. At age 42, she shows acquired microcephaly with a cephalic perimeter of 51cm (-3.5 sd), she keeps intense eye communication, and she has refractory epilepsy presenting with several crisis per day. She also presents facial dysmorphism with broad eyebrows, micrognathia, broad lips, tubular nose, hypertelorism, wide nasal bridge and pointed palate.

Directed genetic testing for *MECP2*, CDKL5 and *FOXG1*, as well as CES, yielded no candidate variants for molecular diagnosis, so we pursued trio WGS and fibroblasts multi-omics. In WGS data filtered using the RTT-spectrum curated candidate gene list, we identified a *de novo* missense variant in the gene *ASXL3* (NM_030632.3): c.2435T>C (p.Ile812Thr) (Figure 30A). It is absent from gnomAD, but it is predicted to be benign by the missense effect prediction algorithm REVEL. This variant had already been detected in CES analysis but was discarded because it was classified as a cold VUS with the ACMG criteria PM2-sup, BP4-sup.



C

	G	A	A	C	C	C	A	T	C	A	T	A	A	T	G	A	G	T	T	C	T
Proband	E				P			I			T			M			S			S	
Human	E				P			I			I			M			S			S	
Chimp	E				P			I			I			M			S			S	
Gorilla	E				P			I			I			M			S			S	
Organgutan	E				P			I			V			M			S			S	
Rhesus	E				P			I			I			M			S			S	
Babboon	E				P			I			I			M			S			S	
Mouse	E				P			I			N			M			S			-	
Rat	E				P			I			N			M			S			-	
Rabbit	E				P			I			K			M			G			P	
Pig	E				P			I			K			M			G			T	

Figure 30. *ASXL3* candidate variant. A) Pedigree of the proband with the identified *de novo* missense variant in the *ASXL3* gene. **B)** AlphaFold-based simulation of the ASXL3 protein, with a vast majority of sequence with intrinsically disordered structure and indicating the position of the amino acid with the missense substitution in our proband. **C)** Amino acid sequence of the studied regions in several vertebrate species. The isoleucine in position 812 is conserved in human, chimp, gorilla, rhesus and baboon, while orangutan presents a valine, another hydrophobic amino acid. Other more distant species present different amino acids in this position.

De novo mutations in the *ASXL3* gene (OMIM*615115) cause Bainbridge-Ropers syndrome (BRPS, OMIM#615485), in an autosomal dominant manner. The clinical presentation of BRPS patients significantly overlaps that of patients with RTT, including microcephaly, delayed psychomotor development, profound intellectual disability, poor or absent speech, inability to walk, seizures, autistic features, and stereotypic hand flapping. Moreover, they present with dysmorphic features significantly matching the facies of our patient.

The majority of published BRPS causing mutations are truncating variants, while some pathogenic missense variants have been described, mostly in ASD patients lacking the characteristic BRPS phenotype. Nevertheless, given the phenotypic resemblance of BRPS with our patient and the fact that the mutation was identified *de novo*, we pursued further *in silico* analyses of the variant. We fed the primary amino acid sequence to Phyre2, an AlphaFold-based software, to predict ASXL3 wild-type protein structure, since it was not contained in the pre-computed AlphaFold Protein Structure Database. The results predicted an 85% of the protein sequence to be intrinsically disordered, with no specific structure (Figure 30B). The isoleucine at position 812 is placed within the extensive intrinsically disordered region of the ASXL3 protein but is relatively conserved in most primates (Figure 30C). Despite the lack of an accurate prediction of the wild-type protein, we can hypothesise that the substitution of a hydrophobic amino acid as isoleucine by a polar amino acid such as threonine may impact protein function. Unfortunately, this protein was not captured in fibroblast proteomics profiling, so we could not ascertain whether protein expression levels were altered in the patient.

With the information from trio sequencing and the extended *in silico* analysis, we were able to re-classify the identified variant as a hot VUS with the ACMG criteria PM2-sup and PS2-str. Although further functional studies are needed to ascertain the pathogenicity of this variant, the evidence so far places it as a strong candidate.

Case 2: Detection of a mosaic variant

The proband was a 22 years-old woman with a history of congenital encephalopathy and drug-resistant epilepsy. At age 7 months started having seizures and hand stereotypies, and had never achieved purposeful use of hands or language. She had severe mental retardation, scoliosis and osteoporosis.

Trio WGS in this family revealed a *de novo* missense variant (NM_001330260.2: c.671T>A - p.Val224Asp) in the *SCN8A* (OMIM*600702) gene (Figure 31). Mutations in *SCN8A* can cause various phenotypic presentations, ranging from familial benign infantile seizures (OMIM#617080) to developmental and epileptic encephalopathy (OMIM#614558) or cognitive impairment with or without cerebellar ataxia (OMIM#614306), all in an autosomal dominant mode of inheritance. The most severe phenotypes associated with mutations in *SCN8A* significantly overlap that of RTT, including the characteristic psychomotor regression.

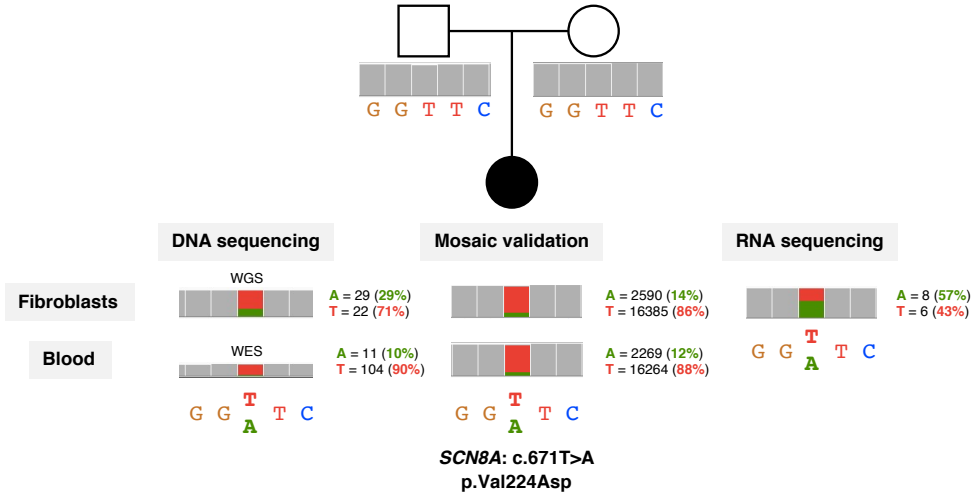


Figure 31. Detection of a mosaic variant in *SCN8A* by trio-WGS. The variant was initially detected in fibroblasts 30x WGS despite having been missed in blood WES with a coverage higher than 100x in this region. We conducted a very high coverage NGS sequencing experiment (~20 000x) targeting the mutated region and quantified a mosaic of 14% in fibroblasts and 12% in blood of the *SCN8A*: c.671T>A (p.Val224Asp) variant. Fibroblasts RNA sequencing confirmed the expression of the identified mosaic variant in approximately half of the covered transcripts.

Given that it is an exonic variant, we wondered why it had been missed by CES and WES in the previous analyses. WGS data had a coverage of 31x on this nucleotide position, and the variant was detected in 29% of reads (9/31), while WES data had a coverage of 115x on this position and the variant was detected in 10% of reads (11/115) (Figure 31). Therefore, the variant had been previously filtered out due to low variant allele frequency (VAF) in the CES and WES analyses, which could indicate a mosaic variant or a sequencing artifact. Transcriptomics data indicated that 57% of produced transcripts contained the variant, diminishing the chance that this should be a sequencing artifact (Figure 31).

We then carried out a mosaic validation experiment by amplifying the target region by long-range PCR and very high coverage sequencing of the PCR product using NGS. To this end, we used DNA isolated from peripheral blood leukocytes of both parents and the proband, plus DNA isolated from the proband's fibroblasts to quantify the mosaic in more than one tissue. The *SCN8A* variant was not identified in either of the parents but was detected as a mosaic in the proband: 12% of the sequencing reads in blood DNA and 14% in fibroblast DNA harboured the variant, indicating that approximately 25% of the cells would be carrying the variant in heterozygosity (Figure 31). It is possible that WES library preparation masked the mosaic variant because of a preferential amplification of the reference allele, while the PCR-free library preparation workflow for WGS retained DNA copies containing the variant more successfully.

Case 3: CNV prioritization by aberrant gene expression

The proband in this family was a 22-year-old girl with a clinical history compatible with RTT. At age 18 months she evidenced a psychomotor delay accompanied with a loss of social interaction and purposeful hand use, which was subsequently replaced with hand stereotypies. She started babbling, but in time lost purposeful language. She presents acquired microcephaly, hyperventilation crises, sleep disturbances and behavioral manifestations, and she has never achieved standing or walking independently. She does not present seizures, but she has shown white matter alterations in brain imaging, and she has some unspecific dysmorphic facial features.

Transcriptomic profiling of the proband's fibroblast cell lines revealed an aberrant expression of 4 genes located consecutively on the X chromosome: *TCEAL1*, *TCEAL3*, *TCEAL4* and *TCEAL9* (Figure 32A). *TCEAL3* and *TCEAL4* also presented low protein expression levels in fibroblasts, while *TCEAL1* and *TCEAL9* were inconsistently detected through proteomics batches (Figure 32B). These expression data directed us to the genomic region containing these genes (chrX:102609819-102887415). Trio WGS revealed a 755kb *de novo* deletion affecting this region (chrX:102589677-103221990) that was confirmed by CGHa.

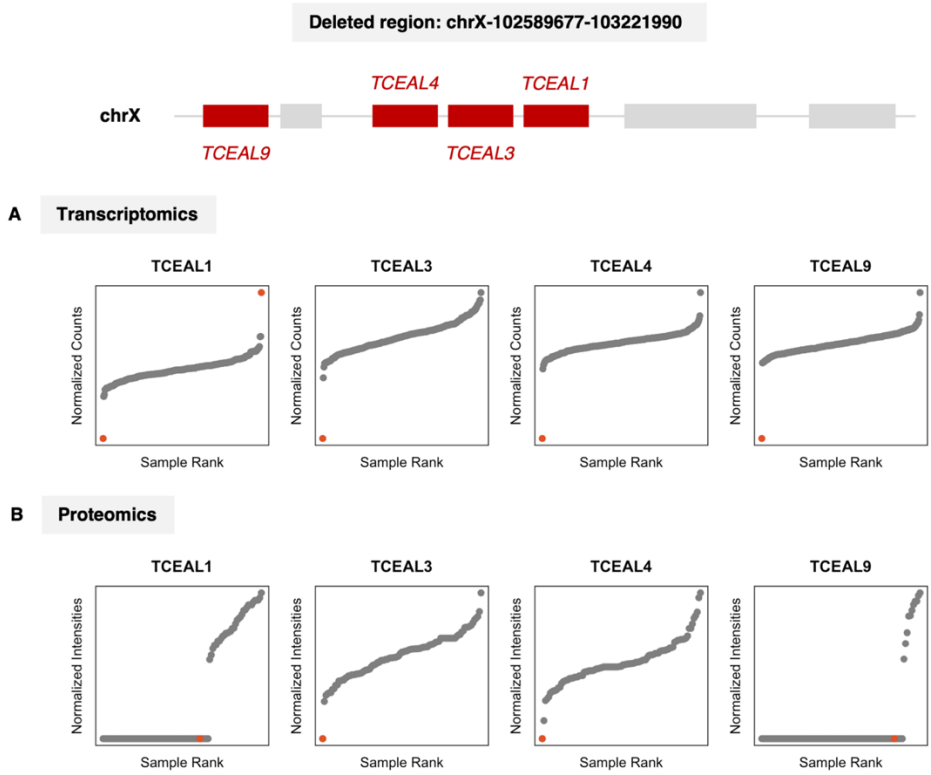


Figure 32. Identification of an interstitial deletion in chromosome X using aberrant expression events. Expression of the genes *TCEAL1*, *TCEAL3*, *TCEAL4* and *TCEAL9* in **A)** transcriptomics and **B)** proteomics profiling in the proband's fibroblasts. *TCEAL3* and *TCEAL4* were detected as aberrantly expressed in both assays, while *TCEAL1* and *TCEAL9* only were in transcriptomics due to inconsistent peptide detection through proteomics batches. These aberrant expression events led to the identification of a deletion affecting these four genes and as well as others in the chrX-102589677-103221990 region.

Mutations in *TCEAL1* were recently described as the cause of Hijazi-Reis syndrome, an X-liked dominant clinical entity that overlaps the phenotype of RTT, presenting global developmental delay, hypotonia, gait abnormalities, impaired intellectual development, poor or absent speech, seizures, and autistic and repetitive behaviours[320]. The first report of Hijazi-Reis syndrome patients was published in December 2022, and implicates the *TCEAL1* as the driver gene causing the phenotype in patients with Xq22.2 deletions. This phenotype consistently matches our patient, including the myelination defects observed on brain imaging.

Case 4: Variant interpretation through aberrant gene expression

The patient was a 28-year-old girl, the first child of healthy, non-consanguineous parents. After a normal early development, at age 9 months concerns were raised due to the lack of independent sitting, generalized hypotonia and psychomotor delay. At age 2, she started using purposeful monosyllabic words, and at age 3 she achieved independent ambulation. She showed a myelinization delay in the telencephalon and microcephaly, with a cephalic perimeter of -2sd according to her age, and she started presenting refractory seizures. At age 28, she presents a light cerebellum atrophy and significant behavioural and agitation disorders accompanied by epileptic seizures.

She underwent genetic testing searching for mutations in *MECP2*, *CDKL5* and *FOXG1*, as well as CES, but no candidate variants were identified. Transcriptomic and proteomic profiling in fibroblast cell lines revealed several aberrant expression events in the *QARS1* (OMIM*603727) gene. *QARS1* was flagged as aberrantly expressed, since both transcript and protein expression were significantly lower than in any other sample (Figure 33A and B). Aberrant splicing analysis showed an aberrant isoform containing retention of 16pb in intron 17, that could produce transcript degradation by nonsense-mediated decay (NMD) (Figure 33C). Finally, monoallelic expression analysis detected an allelic imbalance in chr3:49138795 with 93% of transcripts containing the variant NM_005051.3:c.869A>G (Figure 33D).

Mutations in *QARS1* produce progressive microcephaly with seizures and cerebral and cerebellar atrophy, with an autosomal recessive inheritance pattern (OMIM#615760). This clinical entity closely resembles the phenotypic description of

our patients, including microcephaly, seizures, and cerebellar atrophy. Therefore, and given aberrant expression data, the *QARS1* gene was prioritised in WGS analysis. We identified two compound heterozygous variants in the *QARS1* gene: a paternally inherited missense variant (NM_005051.3: c.869A>G - p.Tyr290Cys), and a maternally inherited intronic variant (NM_005051.3: c.1614+3A>G) (Figure 33E).

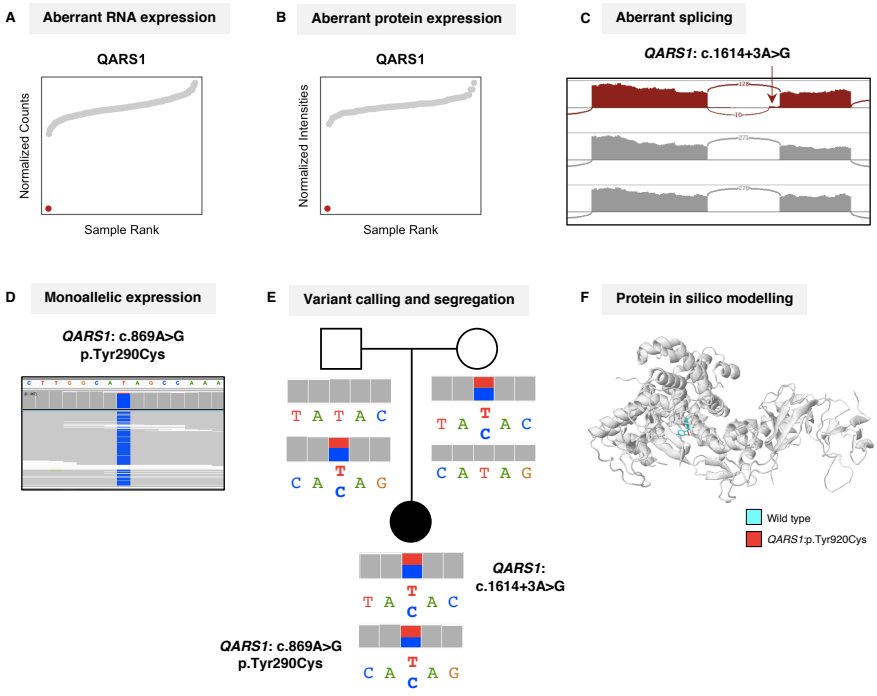


Figure 33. Identification of two compound heterozygous variants in the *QARS1* gene. **A)** Aberrant expression analysis of transcriptomic data revealed an abnormally decreased amount of the *QARS1* transcript. **B)** Aberrant expression analysis of proteomic data also revealed an abnormally decreased amount of the QARS1 protein. **C)** Aberrant splicing analysis revealed a 16bp intronic retention in the exon-intron junction of exon 17 and intron 17, encompassing the c.1614+3A>G variant in the *QARS1* gene. **D)** Monoallelic expression analysis revealed an allelic imbalance strongly favouring variant c.869A>G (p.Tyr260Cys) in the *QARS1* gene, suggesting that the other allele could be less transcribed or degraded. **E)** The two variants were found to be in compound heterozygosity in the proband: the splicing variant was inherited from the mother and the missense variant was inherited from the father. **F)** Missense3D analysis of the impact of the missense variant on the protein structure predicted a structural damage on the protein, changing the size of a cavity generated on the protein surface.

Aberrant splicing data showed that the c.1614+3A>G variant disrupts the intronic sequence of a donor splice site, producing a donor site loss and a 16pb intronic retention in intron 17, producing a frameshift. The fact that only 7% of sequenced transcripts contain the intronic retention, could implicate that this isoform is likely to cause aberrant transcript degradation by NMD or that the splicing sequence alteration does not produce the intron retention in all transcripts. Nevertheless, the allelic imbalance observed in monoallelic expression analysis reveals that the major allele being expressed is the one containing the paternally inherited missense variant, and therefore the transcript containing the maternally inherited intronic retention is likely to be degraded. The tyrosine in position 290 is highly conserved in vertebrates, and the substitution of this aromatic, polar amino acid with a cysteine (non-polar) is predicted to impact protein structure (Figure 33F).

None of these two variants were previously published in the literature. The c.1614+3A>G variant has a very low frequency in the gnomAD database, and according to the recommendations for using the expression evidence derived from transcriptomic data can be classified as likely pathogenic with the PS3-str, PM2-sup and PP3-mod criteria. The p.Tyr290Cys missense variant is absent from gnomAD and can be classified as a hot VUS with the PM2-sup, PM3-mod and PP3-sup criteria, taking into account that it is found *in trans* with a likely pathogenic variant.

4.1.3. Rett syndrome spectrum genes

We identified disease-causing or candidate variants in 54 of the 80 participants (67.5%). 6 patients had large CNVs encompassing several genes, and the remaining had SNVs and indels distributed in 44 different genes (Table 18). The only genes with reported variants in more than one patient were *SYNGAP1* (3 patients), *DNMT3A* (2 patients), *SHANK3* (2 patients) and *TCF4* (2 patients).

Gene	N	Gene	N	Gene	N	Gene	N
SYNGAP1	3	*DYRK1A*	1	*MED12*	1	*SETBP1*	1
DNMT3A	2	*GABBR2*	1	*NR2F1*	1	*SLC2A1*	1
SHANK3	2	*GATAD2B*	1	*PDHA1*	1	*SLC6A1*	1
TCF4	2	*HIVEP2*	1	*PGAP1*	1	*TCEAL1*	1

ARID1B	1	*HNRNPH2*	1	*PHIP*	1	*TRAPPC9*	1
ASXL3	1	*HUWE1*	1	*PURA*	1	*UBTF*	1
CACNA1E	1	*IQSEC2*	1	*QARS1*	1	*UNC13A*	1
CHAMP1	1	*JMJD1C*	1	*RAB11B*	1	*UNC80*	1
CSTB	1	*KDM6B*	1	*SCN1A*	1	*WDFY3*	1
DDX3X	1	*MAGEL2*	1	*SCN2A*	1	*ZBTB18*	1
DNM1	1	*MECP2*	1	*SCN8A*	1	*ZNF142*	1

Table 18. Number of patients with reported variants in each gene.

SYNGAP1-related disorder and Pitt-Hopkins syndrome (caused by mutations in *TCF4*) have been found to be two of the most recurrent monogenic disorders to phenotypically overlap with RTT and fulfil all main diagnostic criteria for typical RTT except for developmental regression in Pitt-Hopkins syndrome[159]. According to phenotypic databases, Phelan-McDermid syndrome (caused by mutations in *SHANK3*) and Heyn-Sproul-Jackson syndrome (caused by mutations in *DNMT3A*) share at least two main criteria and several supportive criteria, so patients with mutations in these genes may fit the description of atypical RTT. The close resemblance in the clinical characteristics of these patients poses a challenge in the clinical diagnosis before a molecular confirmation is reached.

Nearly all the genes in which we reported disease-causing, or candidate variants were contained in our curated RTT-spectrum gene set. The only exceptions were *TCEAL1*, with no reported clinical phenotype until December 2022; *UNC13A* and *JMJD1C*, with no reported clinical phenotype to date but with significant functions in neuronal trafficking[157,321,322]; and *QARS1*, *CSTB* and *PHIP*, with scores lower than 6 in our scoring system according to the phenotypic information found in databases at the time of curating the gene set. Mutations in *CTSB* cause progressive myoclonic epilepsy and according to databases these patients would not fulfil clinical criteria for RTT diagnosis. However, re-scoring of both *QARS1* and *PHIP* according to current phenotypic data revealed a higher phenotypic overlap with RTT and would pass the inclusion threshold, so they have prospectively been included in the curated RTT-spectrum gene set together with *TCEAL1*, increasing the number of included genes from 1245 to 1248.

We wondered whether multi-omics profiling in fibroblasts would be an informative test, since not using the primarily affected tissue would hinder the detection of potentially important genes with limited expression patterns. 1012 and 700 of these 1248 genes (81% and 56%) were detected in our transcriptomics and proteomics experiments in fibroblasts, respectively, and 691 (55%) were detected in both (Figure 34). This data confirmed that despite not being able to capture de biological consequences of variants in all candidate genes, a significant fraction of these could be assayed using these high-throughput experiments.

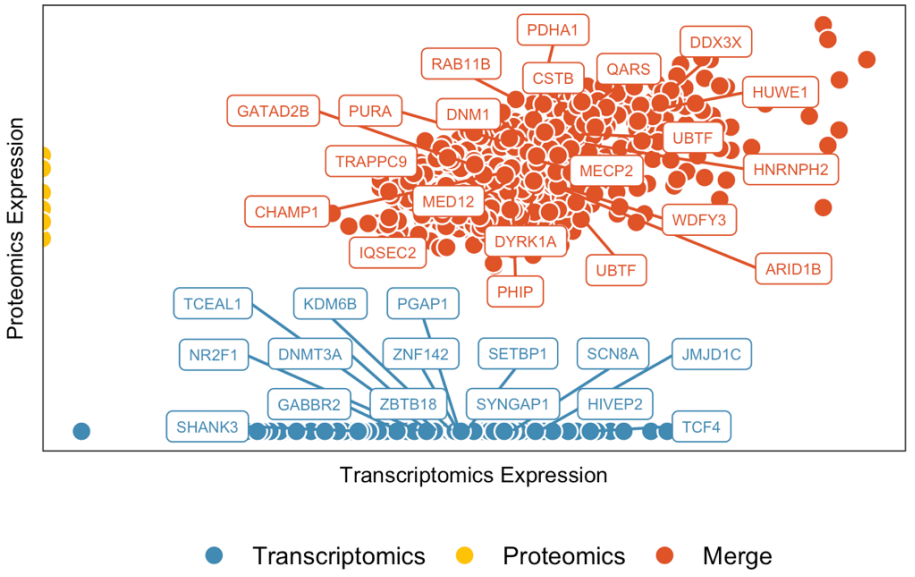


Figure 34. Detection of transcript and protein expression levels in fibroblasts. The x and y axes represent log-normalized expression in transcriptomics and proteomics, respectively, for the genes included in the curated RTT-spectrum dataset and other genes with reported variants in this study that were detected in fibroblasts expression profiling.

This work used a sequential pipeline of genomic analyses of increasing coverage for the diagnosis of RTT-spectrum patients with no candidate variants identified in *MECP2*, *CDKL5* and *FOXG1*.

CES and WES produced a diagnostic or candidate result in 40% and 46% of patients, respectively, of which 30% (20/67 and 15/50, respectively) obtained a definitive result. These outcomes align with the diagnostic yield typically reported in

the literature for singleton exome sequencing in patients with NDDs[163,168–171,323]. Studies specifically targeting RTT-spectrum patients showed varying diagnostic yields, ranging from 20% to almost 70%[146,149–153,324–326]. This wide range of variation can be explained by the heterogeneity in study designs. Studies with *trio* sequencing and with exome as a first-tier diagnostic test tend to have higher diagnostic yields than singleton analyses and exome analysis after a negative result from another genetic test (like CHGa, CES or targeted analysis of *MECP2*, *CDKL5* and *FOXG1*). Therefore, our results are in line with the expected yield from singleton exome studies performed after a prior negative genetic test.

Periodic re-analysis of genomic data has proven to be a cost-efficient approach to increase the diagnostic yield without increasing sequencing efforts. Recent literature estimates that 30% of cases solved by WGS after negative WES could have been solved by reanalysing WES data[327]. Several studies have reported increases in diagnostic yield of 10-15% after re-analysis within approximately one year of the first analysis, in line with the results we obtained in this study[328–331].

After the sequential NGS analysis pipeline, 67.5% (54/80) of patients had a definitive or candidate result, of which 46% (37/50) obtained a final molecular diagnosis. WGS and multi-omics profiling in primary fibroblast cell cultures modestly increased the diagnostic yield by solving 2 cases and identifying strong candidate variants in another 2. Transcriptomic and proteomic profiling effectively helped to prioritise and interpret WGS data to detect disease-causing variants in two individuals, providing direct functional evidence of gene malfunctioning. It is remarkable that these findings were obtained by analysing expression profiles of primary fibroblast cell cultures, proving that these samples are a suitable proxy for studying some NDDs.

Notwithstanding, *trio* analysis in WGS has proven to be valuable in itself, since it allowed to prioritise a candidate *de novo* variant that had been previously filtered out because of a benign *in silico* prediction without inheritance data. A closer inspection of the variant prioritised by inheritance data revealed that the protein was intrinsically disordered and had no solid structure prediction, casting serious doubts on the *in silico* prediction of benignity. This finding must be taken into account when analysing singleton data and when using *in silico* prediction algorithms.

Given the phenotypic overlap and genetic heterogeneity of RTT-spectrum disorders, comprehensive genomic tests may the most cost-effective diagnostic approach, being more feasible than carrying out several targeted analyses sequentially. WGS data have the potential to be re-analysed as many times as needed, changing the phenotypic orientation according to the clinical evolution of the patient and being able to accommodate new knowledge. Considering the rate at which new disease-causing genes are identified in NDDs, the future of their molecular diagnosis will probably involve more comprehensive genomic tests.

Chapter 5. Discussion: Diagnosis, prognosis, and treatments

5.1. Diagnosis of RTT-spectrum disorders: the molecular roots

Almost 60 years after the first description of RTT, its diagnosis still relies mainly on clinical features. Current consensus criteria contemplate pathogenic *MECP2* mutations as a genetic confirmation of RTT when clinical criteria are fulfilled, but some *MECP2* mutations can cause phenotypes other than RTT and therefore are not sufficient to establish a definitive diagnosis by themselves[11]. Moreover, these clinical criteria contemplate regression as an essential feature to consider in RTT diagnosis, but with the generalisation of NGS technologies, a disease-causing variant may be detected in young children before the hallmark features of RTT become apparent[4]. With increasing knowledge about the pathomechanisms downstream MeCP2 deficiency, diagnostic criteria of RTT and MeCP2-related disorders will probably need to be updated, and molecular findings will have a pivotal role in determining the final diagnosis.

Considering the phenotypic overlap between RTT and other neurodevelopmental disorders, a molecular diagnosis has become an essential piece of evidence to distinguish some of these clinical entities. A molecular diagnosis not only enables accurate clinical management and genetic counselling, but with the arrival of precision medicine, it will also become crucial to access clinical trials and targeted therapies which will become gene and/or mutation specific. Comprehensive genetic tests that cover more potentially disease-causing genes and variants, such as WGS combined with multi-omics to assess gene expression, are a valuable tool to streamline the process of detecting the molecular roots of neurodevelopmental disorders and can potentially shorten the diagnostic odyssey and facilitate clinical management.

5.2. Factors influencing the RTT phenotype

Phenotypic variability in RTT is evident, with patients presenting diverse clinical phenotypes ranging from severe infantile encephalopathies to intellectual disability and autism. Besides some exceptions, such as late-truncating mutations and C-

terminal deletions that seem to produce a less severe phenotype than other *MECP2* mutations, the efforts for establishing a genotype-phenotype correlation have not reached robust conclusions that can help in disease prognosis[75,76].

With phenotypical differences between RTT patients with the same mutation and *MECP2* being an X-linked gene, X chromosome inactivation (XCI) has long been thought to be a contributing factor to the clinical variability[332]. However, based on the results presented in this work, we could find no clear correlation between XCI in blood and clinical severity. We found heterogeneous XCI patterns in blood and several brain regions, suggesting that blood is not an accurate proxy for determining XCI in the most relevant tissue for RTT pathogenesis. Moreover, we observed that XCI patterns in brain do not directly correlate with the expression levels of each *MECP2* allele, suggesting that slight deviations from a random XCI pattern may not be sufficient to cause relevant variations in the clinical phenotype. We did observe, however, a correlation between MeCP2 protein levels in fibroblasts and disease severity, indicating that MeCP2 dosage is indeed related to the clinical presentation. Taken together, these data suggest that MeCP2 protein dosage is modulated by factors other than *MECP2* gene expression, and therefore these factors may impact the disease severity. Moreover, other factors such as genetic modifiers (like polymorphisms in the *BDNF* gene) and environmental factors (like early therapeutic enrichment), are likely playing their roles and influencing the final phenotype[4,333,334].

There are currently no biomarkers (besides *MECP2* mutations) that can help in establishing a diagnosis or prognosis in RTT patients, or that can help in assessing the efficacy of preclinical trials. Understanding the precise downstream effects of *MECP2* dysfunction in a cellular context is essential to establish reliable biomarkers for RTT that can inform prognosis and constitute potential therapeutic targets.

5.3. Counteracting MeCP2 deficiency from the molecular bases

Traditional treatments for RTT are generalist and tackle the symptoms but not the roots of the disorder. To design specific treatments, it is essential to understand the pathophysiology of the disorder and the specific consequences of MeCP2 deficiency in a cellular context.

Many studies have investigated the molecular landscape downstream MeCP2 deficiency. These studies have used a variety of models, including *Mecp2*-null mice models, modified cell lines, and different kinds of samples directly derived from patients with RTT. Gene expression deregulation due to MeCP2 deficiency has shown model, tissue, and cell type-specific effects, complicating the options of conducting a proper meta-analysis of the existing data and producing a set of minimally overlapping putative MeCP2 targets. Moreover, most of the studies conducted in mice have used hemizygous *Mecp2*-null male mice, that are not the most accurate representation of the majority of patients with RTT, who are heterozygous females, complicating the extrapolation of the obtained results[4].

Despite the heterogeneous results of gene expression studies to date, some cellular functions and molecular pathways have been repeatedly found dysregulated in different disease models, making them robust candidates for therapeutic intervention. Many expression studies have found altered expression of genes involved in mitochondrial function, associated with increased oxidative stress and ROS production[102–104,335]. The dysregulation of mitochondrial respiratory chain genes would explain morphological and functional abnormalities in the mitochondria of RTT patients and mouse models and can constitute strong therapeutic targets through modulation of mitochondrial function[101]. Genes implicated in lipid metabolism, in particular cholesterol metabolism, have also been found to be dysregulated in RTT models[336,337]. Cholesterol metabolism is important for neurite outgrowth, synaptogenesis, and myelination, and therefore statin treatment was considered for treating RTT-like symptoms in mice models, but results were not conclusive[337–339].

Aberrant expression of genes implicated in dendrite development and neurotransmission has been reported by many studies, highlighting the central function of MeCP2 in synapsis formation and maintenance[47,340,341]. According to these studies, MeCP2 deficiency causes a general deregulation of genes encoding post-synaptic membrane proteins, such as ion channels, synaptic scaffolding proteins and neurotransmitter receptors[49]. In line with these results, our study of GABAergic synapsis proteins in MeCP2-deficient contexts demonstrates that MeCP2 regulates the expression of genes crucial for synaptic function and highlights

the importance of the developmental stage when considering therapeutic intervention in RTT. The results of our multi-omics expression study in fibroblasts also emphasise the importance of cytoskeleton regulation in the pathogenesis of RTT, which could be related to the dendritic arborization abnormalities observed in MeCP2 deficient neurons. *Mecp2* reactivation in *Mecp2*-null mouse models has shown a rescue in behavioural phenotypes together with significant improvements in dendritic morphology, making the cytoskeleton regulatory proteins interesting therapeutic targets. Some studies have suggested that alterations in the NF-κB signalling pathway could be causing the dendritic arborisation defects observed in MeCP2 deficient neurons[341]. Upstream regulator analysis of the DEGs in RTT fibroblasts in our study pointed to SRF as a potential TF governing these expression changes. SRF is a TF that regulates the expression of activity-regulated cytoskeleton-associated proteins modulating neuronal morphology and activity-dependent transcription[313]. Some of the genes regulated by SRF are immediate early genes (IEGs), which are rapidly transcribed genes in response to stimuli like neuronal activity and are involved in synaptogenesis and synaptic plasticity[342]. Interestingly, a recent study also identified a dysregulation of IEGs in *Mecp2*-null mice neurons and confirmed that MeCP2 was directly binding some of these genes. They also detected aberrant expression of IEGs in peripheral blood and post-mortem brain samples of RTT patients, verifying that these alterations occur in diverse cellular contexts and validating the results we obtained on fibroblast cell cultures[343]. Taken together, these results present the regulation of IEGs transcription as a promising therapeutic target for rescuing the dendritic arborization and synaptogenesis defects observed in RTT neurons.

It is noteworthy that, despite the cell type-specific effects of MeCP2, some of the molecular findings identified in fibroblasts can be translated to other more relevant cell types, such as neurons. More extensive validation studies in affected tissues are needed to confirm how much of the transcriptional landscape profiled in fibroblasts can be extrapolated to a neuronal context, and neuron-specific gene expression changes need to be deeply characterised to cover any alterations that could not be detected in fibroblast cell lines. Moreover, further studies will be needed to link MeCP2 function with the observed gene expression alterations and

distinguish primary consequences of MeCP2 deficiency from secondary alterations prompted by an altered cellular homeostasis. ChIP-seq could indicate whether MeCP2 is involved in gene expression regulation as a traditional TF by binding gene promoters, and Hi-C analysis of the chromatin structure may explain how MeCP2 is regulating chromatin compaction to modulate the expression of certain genes.

Finally, potential connections between MeCP2 and other RTT-spectrum genes can contribute to the understanding of the roles that this multifunctional protein is playing in neuronal cells. Finding interactions and common molecular pathways between MeCP2 and these genes can explain phenotypic overlap and encourage the discovery of new therapeutic targets that may be useful for different neurodevelopmental disorders sharing molecular alterations. Our differential expression studies indicated a common dysregulation in genes under the transcriptional regulation of SRF, indicating that modulating its function on dendritic arborization could be a potential therapeutic target for several neurodevelopmental disorders sharing this phenotype. In summary, profiling the molecular landscape of neurodevelopmental disorders can pave the way to finding successful therapeutic targets to improve the quality of life of affected children.

CONCLUSIONS

This work has characterised the molecular landscape in RTT and RTT-spectrum disorders from several perspectives and with complementary techniques, from diagnosis to the identification of potential therapeutic targets. Specifically, we can conclude the following:

- XCI patterns in blood from RTT patients are rarely skewed, and do not directly correlate with the severity of the phenotype.
- XCI patterns in blood and different brain areas are heterogeneous, and therefore blood is not an accurate proxy for determining the XCI patterns in disease-relevant brain areas.
- Given the lack of correlation between allele-specific XCI patterns and the relative levels of WT and mutated *MECP2* transcripts, *MECP2* transcript abundance must be influenced by factors other than XCI.
- Our results support the implication of the GABAergic pathway in the pathophysiology of RTT, identifying a role of MeCP2 as a positive regulator of the expression of GABA ionotropic receptors.
- Different experimental approaches have revealed temporally dynamic results regarding this positive regulatory effect, highlighting the importance of the developmental stage and the progression of disease in the state of the neural circuits in RTT and qualifying GABA ionotropic receptors as a potential therapeutic target in early stages of the disorder.
- The identification of gene expression changes in neurologically relevant genes in fibroblasts demonstrates that this is a suitable surrogate tissue to study the molecular alterations occurring downstream *MECP2* deficiency.
- MeCP2 protein levels in fibroblasts significantly correlate with the severity of the phenotype in RTT patients, with lower protein levels associated to more severe phenotypes.
- Cytoskeletal regulation, vesicular activity, RNA processing and metabolic pathways are significantly enriched in differentially expressed genes (DEGs) dysregulated in RTT fibroblasts, implicating these cellular functions in RTT pathophysiology. Specifically, SRF-regulated cytoskeletal remodelling is potentially implicated in the dysregulation of cytoskeleton related genes and

could be related with the dendritic arborization abnormalities observed in RTT neurons, constituting a potential therapeutic target.

- We identified significant correlation in transcriptomic landscape of some RTT-spectrum patients versus typical RTT patients, indicating common molecular footprints in genetically heterogeneous disorders. Interestingly, SRF regulatory targets were also enriched in the common DEGs, indicating a potential connection between the overlapping disorders.
- A diagnostic pipeline integrating data from WGS with multi-omics profiling in fibroblasts was useful for prioritising and interpreting variants in RTT-spectrum patients, increasing diagnostic yield.

REFERENCES

1. Weaving, L. S., Ellaway, C. J., Gécz, J. & Christodoulou, J. Rett syndrome: Clinical review and genetic update. *Journal of Medical Genetics* vol. 42 1–7 Preprint at https://doi.org/10.1136/jmg.2004.027730 (2005).
2. Liyanage, V. R. B. & Rastegar, M. Rett syndrome and MeCP2. *Neuromolecular Med* **16**, 231–264 (2014).
3. Rett, A. Uber ein eigartiges hirnatrophisches Syndrom bei Hyperammoniamie in Kindesalter. *Wien Med Wochenschr* **116**, 723–738 (1966).
4. Leonard, H., Cobb, S. & Downs, J. Clinical and biological progress over 50 years in Rett syndrome. *Nat Rev Neurol* **13**, 37–51 (2016).
5. 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).
6. Hagberg, B. Clinical manifestations and stages of Rett syndrome. *Ment Retard Dev Disabil Res Rev* **8**, 61–65 (2002).
7. Einspieler, C. & Marschik, P. B. Regression in Rett syndrome: Developmental pathways to its onset. *Neurosci Biobehav Rev* **98**, 320–332 (2019).
8. Marschik, P. B. *et al.* Early development in Rett syndrome–the benefits and difficulties of a birth cohort approach. *Dev Neurorehabil* **21**, 68–72 (2018).
9. Chahrour, M. & Zoghbi, H. Y. The Story of Rett Syndrome : From Clinic to Neurobiology. *Neuron* **56**, 422–437 (2007).
10. Hagberg, B., Hanefeld, F., Percy, A. & Skjedal, O. An update on clinically applicable diagnostic criteria in Rett syndrome. *European Journal of Paediatric Neurology* **6**, 293–297 (2002).
11. Neul, J. L. *et al.* Rett syndrome: Revised diagnostic criteria and nomenclature. *Ann Neurol* **68**, 944–950 (2010).
12. Trevathan, E. & Adams, M. J. The Epidemiology and Public Health Significance of Rett Syndrome. *J Child Neurol* **3**, S17–S20 (1988).
13. Renieri, A. *et al.* Diagnostic criteria for the Zappella variant of Rett syndrome (the preserved speech variant). *Brain Dev* **31**, 208–216 (2009).
14. Zappella, M. The rett girls with preserved speech. *Brain Dev* **14**, 98–101 (1992).
15. Goutières, F. & Aicardi, J. Atypical forms of Rett syndrome. *Am J Med Genet* **24**, 183–194 (1986).
16. Hagberg, B. A. & Skjeldal, O. H. *Rett Variants: A Suggested Model for Inclusion Criteria*. (1994).
17. Hanefeld, F. The clinical pattern of the rett syndrome. *Brain Dev* **7**, 320–325 (1985).
18. Rolando, S. Rett syndrome: Report of eight cases. *Brain Dev* **7**, 290–296 (1985).
19. Erlandson, A. & Hagberg, B. MECP2 Abnormality Phenotypes: Clinicopathologic Area With Broad Variability. *J Child Neurol* **20**, 727–732 (2005).
20. Percy, A. K. Rett syndrome: clinical correlates of the newly discovered gene. *Brain Dev* **23**, S202–S205 (2001).
21. Monrós, E. *et al.* Rett syndrome in Spain: mutation analysis and clinical correlations. *Brain Dev* **23 Suppl 1**, S251–S253 (2001).
22. Kerr, A. M. *et al.* Guidelines for reporting clinical features in cases with MECP2 mutations. *Brain Dev* **23**, 208–211 (2001).
23. Coker, S. B. & Melnyk, A. R. Rett Syndrome and Mitochondrial Enzyme Deficiencies. *J Child Neurol* **6**, 164–166 (1991).
24. Zoghbi, H. Genetic Aspects of Rett Syndrome. *J Child Neurol* **3**, S76–S78 (1988).

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

26. Ellison, K. A. *et al. Examination of X Chromosome Markers in Rett Syndrome: Exclusion Mapping with a Novel Variation on Multilocus Linkage Analysis*. *Am. J. IHum. Genet* vol. 50 (1992).

27. Schanen, N. C. *et al.* A New Rett Syndrome Family Consistent with X-Linked Inheritance Expands the X Chromosome Exclusion Map. *The American Journal of Human Genetics* **61**, 634–641 (1997).

28. Schanen, C. & Francke, U. A Severely Affected Male Born into a Rett Syndrome Kindred Supports X-Linked Inheritance and Allows Extension of the Exclusion Map. *The American Journal of Human Genetics* **63**, 267–269 (1998).

29. 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).

30. Ariani, F. *et al.* Real-time quantitative PCR as a routine method for screening large rearrangements in Rett syndrome: Report of one case of MECP2 deletion and one case of MECP2 duplication. *Hum Mutat* **24**, 172–177 (2004).

31. 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).

32. 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. *Journal of Biological Chemistry* **282**, 15057–15064 (2007).

33. 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*. *Human Molecular Genetics* vol. 11 (2002).

34. Kishi, N. & Macklis, J. D. Dissecting MECP2 Function in the Central Nervous System. *J Child Neurol* **20**, 753–759 (2005).

35. Shahbazian, M. D. & Zoghbi, H. Y. Rett syndrome and MeCP2: Linking epigenetics and neuronal function. *Am J Hum Genet* **71**, 1259–1272 (2002).

36. Zachariah, R. M. & Rastegar, M. Linking Epigenetics to Human Disease and Rett Syndrome: The Emerging Novel and Challenging Concepts in MeCP2 Research. *Neural Plast* **2012**, 1–10 (2012).

37. 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* ddw269 (2016) doi:10.1093/hmg/ddw269.

38. 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).

39. 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).

40. Ip, J. P. K., Mellios, N. & Sur, M. Rett syndrome: Insights into genetic, molecular and circuit mechanisms. *Nat Rev Neurosci* **19**, 368–382 (2018).

41. 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).

42. Dragich, J. M., Kim, Y.-H., Arnold, A. P. & Schanen, C. Differential Distribution of the Mecp2 Splice Variants in the Postnatal Mouse Brain. *Journal of Comparative Neurology* **501**, 526–542 (2007).

43. 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**, (2019).

44. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of Neurological Defects in a Mouse Model of Rett Syndrome. *Science (1979)* **315**, 1143–1147 (2007).

45. Tropea, D. *et al.* Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice. *Proc Natl Acad Sci U S A* **106**, 2029–2034 (2009).

46. Meehan, R. R., Lewis, J. D. & Bird, A. P. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res* **20**, 5085–5092 (1992).

47. Ehrhart, F. *et al.* Rett syndrome - Biological pathways leading from MECP2 to disorder phenotypes. *Orphanet Journal of Rare Diseases* vol. 11 Preprint at https://doi.org/10.1186/s13023-016-0545-5 (2016).

48. Ragione, F. Della, Vacca, M., Fioriniello, S., Pepe, G. & D'Esposito, M. MECP2, a multi-talented modulator of chromatin architecture. *Brief Funct Genomics* **15**, 420–431 (2016).

49. Marano, D., Fioriniello, S., Esposito, M. D. & Ragione, F. Della. Transcriptomic and Epigenomic Landscape in Rett Syndrome. *Biomolecules* **11**, (2021).

50. Zhou, Z. *et al.* Brain-Specific Phosphorylation of MeCP2 Regulates Activity-Dependent Bdnf Transcription, Dendritic Growth, and Spine Maturation. *Neuron* **52**, 255–269 (2006).

51. Carouge, D., Host, L., Aunis, D., Zwiller, J. & Anglard, P. CDKL5 is a brain MeCP2 target gene regulated by DNA methylation. *Neurobiol Dis* **38**, 414–424 (2010).

52. Tao, J. *et al.* Phosphorylation of MeCP2 at serine 80 regulates its chromatin association and neurological function. *Proc Natl Acad Sci U S A* **106**, 4882–4887 (2009).

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

54. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, (1998).

55. Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science (1979)* **320**, 1224–1229 (2008).

56. Ghosh, R. P., Horowitz-Scherer, R. A., Nikitina, T., Shlyakhtenko, L. S. & Woodcock, C. L. MeCP2 Binds Cooperatively to Its Substrate and Competes with Histone H1 for Chromatin Binding Sites. *Mol Cell Biol* **30**, 4656–4670 (2010).

57. Skene, P. J. *et al.* Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. *Mol Cell* **37**, 457–468 (2010).

58. Brero, A. *et al.* Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *Journal of Cell Biology* **169**, 733–743 (2005).

59. Singleton, M. K. *et al.* MeCP2 is required for global heterochromatic and nucleolar changes during activity-dependent neuronal maturation. *Neurobiol Dis* **43**, 190–200 (2011).

60. Nan, X. *et al.* Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proceedings of the National Academy of Sciences* **104**, 2709–2714 (2007).

61. Kernohan, K. D., Vernimmen, D., Gloor, G. B. & Bérubé, N. G. Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. *Nucleic Acids Res* **42**, 8356–8368 (2014).

62. Kernohan, K. D. *et al.* ATRX Partners with Cohesin and MeCP2 and Contributes to Developmental Silencing of Imprinted Genes in the Brain. *Dev Cell* **18**, 191–202 (2010).

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

64. Cohen, S. *et al.* Genome-Wide Activity-Dependent MeCP2 Phosphorylation Regulates Nervous System Development and Function. *Neuron* **72**, 72–85 (2011).

65. Lyst, M. J. & Bird, A. Rett syndrome: A complex disorder with simple roots. *Nat Rev Genet* **16**, 261–274 (2015).

66. 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).

67. Long, S. W., Ooi, J. Y. Y., Yau, P. M. & Jones, P. L. A brain-derived MeCP2 complex supports a role for MeCP2 in RNA processing. *Biosci Rep* **31**, 333–343 (2011).

68. Maunakea, A. K., Chepelev, I., Cui, K. & Zhao, K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res* **23**, 1256–1269 (2013).

69. Szulwach, K. E. *et al.* Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *Journal of Cell Biology* **189**, 127–141 (2010).

70. Cheng, T. L. *et al.* MeCP2 Suppresses Nuclear MicroRNA Processing and Dendritic Growth by Regulating the DGCR8/Drosha Complex. *Dev Cell* **28**, 547–560 (2014).

71. Klein, M. E. *et al.* Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. *Nat Neurosci* **10**, 1513–1514 (2007).

72. Percy, A. K. *et al.* Rett syndrome diagnostic criteria: Lessons from the Natural History Study. *Ann Neurol* **68**, 951–955 (2010).

73. Krishnaraj, R., Ho, G. & Christodoulou, J. RettBASE: Rett syndrome database update. *Hum Mutat* **38**, 922–931 (2017).

74. Gianakopoulos, P. J. *et al.* Mutations in MECP2 exon 1 in classical rett patients disrupt MECP2_e1 transcription, but not transcription of MECP2_e2. *American Journal of Medical Genetics, Part B* **159 B**, 210–216 (2012).

75. Bebbington, A. *et al.* Updating the profile of C-terminal MECP2 deletions in Rett syndrome. *J Med Genet* **47**, 242–248 (2010).

76. Neul, J. L. *et al.* Specific mutations in Methyl-CpG-Binding Protein 2 confer different severity in Rett syndrome. *Neurology* **70**, 1313–1321 (2008).

77. Agarwal, N. *et al.* MeCP2 Rett mutations affect large scale chromatin organization. *Hum Mol Genet* **20**, 4187–4195 (2011).

78. Van Esch, H. MECP2 duplication syndrome. *Mol Syndromol* **2**, 128–136 (2012).

79. Gartler, S. M. & Goldman, M. A. *Biology of the X chromosome*. *Curr Opin Pediatr* vol. 13 http://journals.lww.com/co-pediatrics (2001).

80. Vacca, M., Della Ragione, F., Scalabrì, F. & D'Esposito, M. X inactivation and reactivation in X-linked diseases. *Seminars in Cell and Developmental Biology* Preprint at https://doi.org/10.1016/j.semcdb.2016.03.009 (2016).

81. Clerc, P. & Avner, P. Random X-chromosome inactivation: skewing lessons for mice and men. *Current Opinion in Genetics and Development* vol. 16 246–253 Preprint at https://doi.org/10.1016/j.gde.2006.04.001 (2006).

82. Schluth, C. *et al.* Phenotype in X chromosome rearrangements: pitfalls of X inactivation study. *Pathologie Biologie* **55**, 29–36 (2007).

83. Hoffbuhr, K. *et al.* MeCP2 mutations in children with and without the phenotype of Rett syndrome. *Neurology* **56**, 1486–1495 (2001).

84. Sirianni, N., Naidu, S. B., Pereira, J. L., Pillotto, R. F. & Hoffman, E. P. Rett Syndrome: Confirmation of X-Linked Dominant Inheritance, and Localization of the Gene to Xq28. *Am. J. Hum. Genet* **63**, 1552–1558 (1998).

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

86. Miller, R. A. *et al.* Beyond pathway analysis: Identification of active subnetworks in rett syndrome. *Front Genet* **10**, (2019).

87. 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).

88. Shovlin, S. & Tropea, D. Transcriptome level analysis in Rett syndrome using human samples from different tissues. *Orphanet Journal of Rare Diseases* vol. 13 Preprint at https://doi.org/10.1186/s13023-018-0857-8 (2018).

89. Jugloff, D. G. M., Jung, B. P., Purushotham, D., Logan, R. & Eubanks, J. H. Increased dendritic complexity and axonal length in cultured mouse cortical neurons overexpressing methyl-CpG-binding protein MeCP2. *Neurobiol Dis* **19**, 18–27 (2005).

90. Abuhatzira, L., Shemer, R. & Razin, A. MeCP2 involvement in the regulation of neuronal α-tubulin production. *Hum Mol Genet* **18**, 1415–1423 (2009).

91. 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).

92. Armstrong, D., Dunn, J. K., Antalffy, B. & Trivedi, R. Selective Dendritic Alterations in the Cortex of Rett Syndrome. *J Neuropathol Exp Neurol* **54**, 195–201 (1995).

93. Kaufmann, W. E. Dendritic Cytoskeletal Protein Expression in Mental Retardation: An Immunohistochemical Study of the Neocortex in Rett Syndrome. *Cerebral Cortex* **10**, 992–1004 (2000).

94. Francke, U. Mechanisms of Disease: neurogenetics of MeCP2 deficiency. *Nat Clin Pract Neurol* **2**, 212–221 (2006).

95. Colantuoni, C. *et al.* Gene expression profiling in postmortem Rett Syndrome brain: Differential gene expression and patient classification. *Neurobiol Dis* **8**, 847–865 (2001).

96. Livide, G. *et al.* GluD1 is a common altered player in neuronal differentiation from both MECP2-mutated and CDKL5-mutated iPS cells. *European Journal of Human Genetics* **23**, 195–201 (2015).

97. Kang, S. K., Kim, S. T., Johnston, M. V. & Kadam, S. D. Temporal- and Location-Specific Alterations of the GABA Recycling System in Mecp2 KO Mouse Brains. *J Cent Nerv Syst Dis* **6**, JCNSD.S14012 (2014).

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

99. Ure, K. *et al.* Restoration of Mecp2 expression in GABAergic neurons is sufficient to rescue multiple disease features in a mouse model of Rett syndrome. *Elife* **5**, 1–21 (2016).

100. Cobb, S., Guy, J. & Bird, A. Reversibility of functional deficits in experimental models of Rett syndrome. *Biochem Soc Trans* **38**, 498–506 (2010).

101. 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**, (2019).

102. 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).

103. Kriaucionis, S. *et al.* Gene Expression Analysis Exposes Mitochondrial Abnormalities in a Mouse Model of Rett Syndrome. *Mol Cell Biol* **26**, 5033–5042 (2006).

104. Gibson, J. H. *et al.* Downstream targets of methyl CpG binding protein 2 and their abnormal expression in the frontal cortex of the human Rett syndrome brain. *BMC Neurosci* **11**, (2010).

105. 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).

106. Robinson, L. *et al.* Morphological and functional reversal of phenotypes in a mouse model of Rett syndrome. *Brain* **135**, 2699–2710 (2012).

107. Lang, M. *et al.* Rescue of behavioral and EEG deficits in male and female Mecp2-deficient mice by delayed Mecp2 gene reactivation. *Hum Mol Genet* **23**, 303–318 (2014).

108. Smrt, R. D. *et al.* Mecp2 deficiency leads to delayed maturation and altered gene expression in hippocampal neurons. *Neurobiol Dis* **27**, 77–89 (2007).

109. Johnston, M. V, Jeon, O.-H., Pevsner, J., Blue, M. E. & Naidu, S. *Neurobiology of Rett syndrome: a genetic disorder of synapse development*. www.elsevier.com/locate/braindev.

110. Ronnett, G. V. *et al.* Olfactory biopsies demonstrate a defect in neuronal development in Rett's syndrome. *Ann Neurol* **54**, 206–218 (2003).

111. Huong Le, T. T. *et al.* Efficient and precise CRISPR/Cas9-mediated MECP2 modifications in human-induced pluripotent stem cells. *Front Genet* **10**, (2019).

112. Sinnamon, J. R. *et al.* In Vivo Repair of a Protein Underlying a Neurological Disorder by Programmable RNA Editing. *Cell Rep* **32**, 107878 (2020).

113. Garg, S. K. *et al.* Systemic delivery of MeCP2 rescues behavioral and cellular deficits in female mouse models of Rett syndrome. *Journal of Neuroscience* **33**, 13612–13620 (2013).

114. Sinnett, S. E. *et al.* Improved MECP2 Gene Therapy Extends the Survival of MeCP2-Null Mice without Apparent Toxicity after Intracisternal Delivery. *Mol Ther Methods Clin Dev* **5**, 106–115 (2017).

115. Collins, B. E., Merritt, J. K., Erickson, K. R. & Neul, J. L. Safety and efficacy of genetic MECP2 supplementation in the R294X mouse model of Rett syndrome. *Genes Brain Behav* **21**, (2022).

116. Grimm, N. B. & Lee, J. T. Selective Xi reactivation and alternative methods to restore MECP2 function in Rett syndrome. *Trends in Genetics* vol. 38 920–943 Preprint at https://doi.org/10.1016/j.tig.2022.01.007 (2022).

117. Brendel, C. *et al.* Readthrough of nonsense mutations in Rett syndrome: evaluation of novel aminoglycosides and generation of a new mouse model. *J Mol Med* **89**, 389–398 (2011).

118. 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).

119. 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).

120. Glaze, D. G. *et al.* Double-blind, randomized, placebo-controlled study of trofinetide in pediatric Rett syndrome. *Neurology* **92**, E1912–E1925 (2019).

121. Neul, J. L. *et al.* Design and outcome measures of LAVENDER, a phase 3 study of trofinetide for Rett syndrome. *Contemp Clin Trials* **114**, (2022).

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

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

124. Naegelin, Y. *et al.* Fingolimod in children with Rett syndrome: the FINGORETT study. *Orphanet J Rare Dis* **16**, 19 (2021).

125. Tao, J. *et al.* Mutations in the X-linked cyclin-dependent kinase-like 5 (CDKL5/STK9) gene are associated with severe neurodevelopmental retardation. *Am J Hum Genet* **75**, 1149–1154 (2004).

126. Scala, E. *et al.* CDKL5/STK9 is mutated in Rett syndrome variant with infantile spasms. *J Med Genet* **42**, 103–107 (2005).

127. Kalscheuer, V. M. *et al.* Disruption of the serine/threonine kinase 9 gene causes severe X-linked infantile spasms and mental retardation. *Am J Hum Genet* **72**, 1401–1411 (2003).

128. Archer, H. L. *et al.* CDKL5 mutations cause infantile spasms, early onset seizures, and severe mental retardation in female patients. *J Med Genet* **43**, 729–734 (2006).

129. Bahi-Buisson, N. *et al.* Key clinical features to identify girls with CDKL5 mutations. *Brain* **131**, 2647–2661 (2008).

130. Jakimiec, M., Paprocka, J. & Śmigiel, R. CDKL5 deficiency disorder—a complex epileptic encephalopathy. *Brain Sci* **10**, (2020).

131. Roche Martínez, A. *et al.* CDKL5 in different atypical rett syndrome variants: Description of the first eight patients from Spain. *J Pediatr Epilepsy* **1**, 27–35 (2012).

132. Montini, E. *et al. Identification and Characterization of a Novel Serine- Threonine Kinase Gene from the Xp22 Region*. http://www.sanger.ac.uk/HGP/blast_server.shtml. (1998).

133. Mari, F. *et al.* CDKL5 belongs to the same molecular pathway of MeCP2 and it is responsible for the early-onset seizure variant of Rett syndrome. *Hum Mol Genet* **14**, 1935–1946 (2005).

134. Chen, Q. *et al.* CDKL5, a protein associated with Rett syndrome, regulates neuronal morphogenesis via Rac1 signaling. *Journal of Neuroscience* **30**, 12777–12786 (2010).

135. Kadam, S. D., Sullivan, B. J., Goyal, A., Blue, M. E. & Smith-Hicks, C. Rett syndrome and CDKL5 deficiency disorder: From bench to clinic. *Int J Mol Sci* **20**, (2019).

136. Ariani, F. *et al.* FOXG1 Is Responsible for the Congenital Variant of Rett Syndrome. *Am J Hum Genet* **83**, 89–93 (2008).

137. Papa, F. T. *et al.* A 3 Mb deletion in 14q12 causes severe mental retardation, mild facial dysmorphisms and rett-like features. *Am J Med Genet A* **146**, 1994–1998 (2008).

138. Roche-Martínez, A., Gerotina, E., Armstrong-Morón, J., Sans-Capdevila, Ó. & Pineda, M. FOXG1, un nuevo gen responsable de la forma congénita del síndrome de Rett. *Rev Neurol* **52**, 597–602 (2011).

139. Mitter, D. *et al.* FOXG1 syndrome: Genotype-phenotype association in 83 patients with FOXG1 variants. *Genetics in Medicine* **20**, 98–108 (2018).

140. 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).

141. Yao, J., Lai, E. & Stifani, S. The Winged-Helix Protein Brain Factor 1 Interacts with Groucho and Hes Proteins To Repress Transcription. *Mol Cell Biol* **21**, 1962–1972 (2001).

142. Tan, K. *et al.* Human PLU-1 has transcriptional repression properties and interacts with the developmental transcription factors BF-1 and PAX9. *Journal of Biological Chemistry* **278**, 20507–20513 (2003).

143. Hanashima, C., Shen, L., Li, S. C. & Lai, E. *Brain Factor-1 Controls the Proliferation and Differentiation of Neocortical Progenitor Cells through Independent Mechanisms*. (2002).

144. Yang, Y. *et al.* Impaired interneuron development after Foxg1 disruption. *Cerebral Cortex* **27**, 793–808 (2017).

145. Philippe, C. *et al.* Phenotypic variability in Rett syndrome associated with FOXG1 mutations in females. *J Med Genet* **47**, 59–65 (2010).

146. Lucariello, M. *et al.* Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype. *Hum Genet* **135**, 1343–1354 (2016).

147. Vidal, S. *et al.* Genetic landscape of rett syndrome spectrum: Improvements and challenges. *International Journal of Molecular Sciences* vol. 20 Preprint at https://doi.org/10.3390/ijms20163925 (2019).

148. Schönewolf-Greulich, B. *et al.* Clinician's guide to genes associated with Rett-like phenotypes-Investigation of a Danish cohort and review of the literature. *Clin Genet* 1–10 (2018) doi:10.1111/cge.13153.

149. Iwama, K. *et al.* Genetic landscape of Rett syndrome-like phenotypes revealed by whole exome sequencing. *J Med Genet* 1–12 (2019) doi:10.1136/jmedgenet-2018-105775.

150. Vidal, S. *et al.* The utility of Next Generation Sequencing for molecular diagnostics in Rett syndrome. *Sci Rep* **7**, 1–11 (2017).

151. Lopes, F. *et al.* Identification of novel genetic causes of Rett syndrome-like phenotypes. *J Med Genet* **53**, 190–199 (2016).

152. Sajan, S. A. *et al.* Enrichment of mutations in chromatin regulators in people with Rett syndrome lacking mutations in MECP2. *Genetics in Medicine* **19**, 13–19 (2017).

153. Yoo, Y. *et al.* GABBR2 mutations determine phenotype in rett syndrome and epileptic encephalopathy. *Ann Neurol* **82**, 466–478 (2017).

154. Srivastava, S. *et al.* Monogenic disorders that mimic the phenotype of Rett syndrome. *Neurogenetics* **19**, 41–47 (2018).

155. Jiang, Y. H. *et al.* Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am J Hum Genet* **93**, 249–263 (2013).

156. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).

157. Sáez, M. A. *et al.* Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. *Genetics in Medicine* **18**, 378–385 (2016).

158. Vuillaume, M., Xue, L. & Blesson, S. A Novel Mutation in the Transmembrane 6 Domain of GABBR2 Leads to a Rett-like Phenotype Acknowledgment This work was supported by grants from the French Potential Conflicts of Interest of Neurology. *Ann Neurol* **83**, 437–439 (2018).

159. Vidal, S. *et al.* The most recurrent monogenic disorders that overlap with the phenotype of Rett syndrome. *European Journal of Paediatric Neurology* **23**, 609–620 (2019).

160. Frankel, E. *et al.* Genetic and Protein Network Underlying the Convergence of Rett-Syndrome-like (RTT-L) Phenotype in Neurodevelopmental Disorders. *Cells* **12**, (2023).

161. 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. *Human Mutation* vol. 40 2184–2196 Preprint at https://doi.org/10.1002/humu.23887 (2019).

162. Ehrhart, F. *et al.* Integrated analysis of human transcriptome data for Rett syndrome finds a network of involved genes. *World Journal of Biological Psychiatry* **21**, 712–725 (2020).

163. Wright, C. F., FitzPatrick, D. R. & Firth, H. V. Paediatric genomics: Diagnosing rare disease in children. *Nature Reviews Genetics* vol. 19 253–268 Preprint at https://doi.org/10.1038/nrg.2017.116 (2018).

164. Mardis, E. R. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133–141 (2008).

165. Marwaha, S., Knowles, J. W. & Ashley, E. A. A guide for the diagnosis of rare and undiagnosed disease: beyond the exome. *Genome Med* **14**, 1–22 (2022).

166. Saudi Mendeliome Group. Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. *Genome Biol* **16**, (2015).

167. Manickam, K. *et al.* Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine* (2021) doi:10.1038/s41436-021-01242-6.

168. Yang, Y. *et al.* Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. *New England Journal of Medicine* **369**, 1502–1511 (2013).

169. Retterer, K. *et al.* Clinical application of whole-exome sequencing across clinical indications. *Genetics in Medicine* **18**, 696–704 (2016).

170. Farwell, K. D. *et al.* Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: Results from 500 unselected families with undiagnosed genetic conditions. *Genetics in Medicine* **17**, 578–586 (2015).

171. Stefanski, A. *et al.* Clinical sequencing yield in epilepsy, autism spectrum disorder, and intellectual disability: A systematic review and meta-analysis. *Epilepsia* **62**, 143–151 (2021).

172. Clark, M. M. *et al.* Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med* **3**, 1–10 (2018).

173. DDD Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2018).

174. Cheng, A. Y., Teo, Y. Y. & Ong, R. T. H. Assessing single nucleotide variant detection and genotype calling on whole-genome sequenced individuals. *Bioinformatics* **30**, 1707–1713 (2014).

175. de Ligt, J. *et al.* Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum Mutat* **34**, 1439–1448 (2013).

176. Guo, Y. *et al.* Exome sequencing generates high quality data in non-target regions. *BMC Genomics* **13**, (2012).

177. Wang, Q., Shashikant, C. S., Jensen, M., Altman, N. S. & Girirajan, S. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Sci Rep* **7**, (2017).

178. Meienberg, J., Bruggmann, R., Oexle, K. & Matyas, G. Clinical sequencing: is WGS the better WES? *Hum Genet* **135**, 359–362 (2016).

179. Nisar, H. *et al.* Whole-genome sequencing as a first-tier diagnostic framework for rare genetic diseases. *Experimental Biology and Medicine* vol. 246 2610–2617 Preprint at https://doi.org/10.1177/15353702211040046 (2021).

180. Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* **112**, 5473–5478 (2015).

181. Soden, S. E. *et al.* Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci Transl Med* **6**, 1–14 (2014).

182. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nature Reviews Genetics* vol. 17 19–32 Preprint at https://doi.org/10.1038/nrg.2015.3 (2016).

183. Ellingford, J. M. *et al.* Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *medRxiv* 2021.12.28.21267792 (2021) doi:10.1186/s13073-022-01073-3.

184. Hitomi, Y. & Tokunaga, K. Significance of functional disease-causal/susceptible variants identified by whole-genome analyses for the understanding of human diseases. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences* vol. 93 657–676 Preprint at https://doi.org/10.2183/pjab.93.042 (2017).

185. Mahmoud, M. *et al.* Structural variant calling: The long and the short of it. *Genome Biology* vol. 20 Preprint at https://doi.org/10.1186/s13059-019-1828-7 (2019).

186. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: A probabilistic framework for structural variant discovery. *Genome Biol* **15**, 1–19 (2014).

187. Cameron, D. L., Di Stefano, L. & Papenfuss, A. T. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun* **10**, 1–11 (2019).

188. Yépez, V. A. *et al.* Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc* **16**, 1276–1296 (2021).

189. Ferraro, N. M. *et al.* Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science (1979)* **369**, (2020).

190. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).

191. Kopajtich, R. *et al.* Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders. *medRxiv* (2021).

192. Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nature Reviews Genetics* vol. 19 299–310 Preprint at https://doi.org/10.1038/nrg.2018.4 (2018).

193. 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. *Journal of Inherited Metabolic Disease* vol. 43 25–35 Preprint at https://doi.org/10.1002/jimd.12130 (2019).

194. Kremer, L. S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* **8**, (2017).

195. Gonorazky, H. *et al.* RNAseq analysis for the diagnosis of muscular dystrophy. *Ann Clin Transl Neurol* **3**, 55–60 (2016).

196. Cummings, B. B. *et al.* Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med* **9**, (2017).

197. 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).

198. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. *Nucleic Acids Res* **43**, D789–D798 (2015).

199. Cutler Allen, R., Zoghbi, H. Y., Annemarie Moseley, I. B., Rosenblatt, H. M. & Belmont, J. W. Methylation of Hpall and Hhal Sites Near the Polymorphic CAG Repeat in the Human Androgen-Receptor Gene Correlates with X Chromosome Inactivation. *Am. J. Hum. Genet* **51**, 1229–1239 (1992).

200. 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).

201. Soto, D. *et al.* L-Serine dietary supplementation is associated with clinical improvement of loss-of-function GRIN2B-related pediatric encephalopathy. *Sci Signal* **12**, 1–16 (2019).

202. Liu, J. *et al.* An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods* **8**, (2012).

203. Untergasser, A. *et al.* Primer3-new capabilities and interfaces. *Nucleic Acids Res* **40**, 1–12 (2012).

204. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291 (2007).

205. Kim, D. *et al.* TopHat2 : accurate alignment of transcriptomes in the presence of insertions , deletions and gene fusions. *Genome Biol* **14**, 1–13 (2013).

206. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

207. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).

208. Al-Kafaji, G., Sabry, M. A. & 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* **13**, 1774–1780 (2016).

209. Zecha, J. *et al.* TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Molecular & Cellular Proteomics* **18**, 1468–1478 (2019).

210. Yu, P. *et al.* Trimodal Mixed Mode Chromatography That Enables Efficient Offline Two-Dimensional Peptide Fractionation for Proteome Analysis. *Anal Chem* **89**, 8884–8891 (2017).

211. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301–2319 (2016).

212. Babraham Bioinformatics. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

213. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* **17**, 10–12 (2011).

214. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

215. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).

216. Poplin, R. *et al.* A universal snp and small-indel variant caller using deep neural networks. *Nat Biotechnol* **36**, 983 (2018).

217. Cooke, D. P., Wedge, D. C. & Lunter, G. A unified haplotype-based method for accurate and comprehensive variant calling. *Nat Biotechnol* **39**, 885–892 (2021).

218. Barbitoff, Y. A., Abasov, R., Tvorogova, V. E., Glotov, A. S. & Predeus, A. V. Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery. *BMC Genomics* **23**, (2022).

219. GATK. Germline short variant discovery (SNPs + Indels) - Best Practices. (2022).

220. GATK. Data pre-processing for variant discovery - Best Practices. (2022).

221. Broad Institute. GATK Resource Bundle - hg19. (2019).

222. Broad Institute. Hard-filtering germline short variants. https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants (2022).

223. Sherry, S. T. *et al.* DbSNP: The NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308–311 (2001).

224. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118 ; iso-2; iso-3. *Landes Bioscience* **6**, 80–92 (2012).

225. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

226. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* **32**, 894–899 (2011).

227. Landrum, M. J. *et al.* ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062–D1067 (2018).

228. Cingolani, P. *et al.* Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front Genet* **3**, 1–9 (2012).

229. McKusick-Nathans Institute of Genetic Medicine, J. H. U. Online Mendelian Inheritance in Man, OMIM®. (2022).

230. Kosugi, S. *et al.* Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol* **20**, 8–11 (2019).

231. Chen, X. *et al.* Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).

232. Rausch, T. *et al.* DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, 333–339 (2012).

233. Kronenberg, Z. N. *et al.* Wham: Identifying Structural Variants of Biological Consequence. *PLoS Comput Biol* **11**, 1–19 (2015).

234. Li, H. & Wren, J. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* vol. 30 2843–2851 Preprint at https://doi.org/10.1093/bioinformatics/btu356 (2014).

235. Raca, G. *et al.* Points to consider in the detection of germline structural variants using next-generation sequencing: A statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine* (2022) doi:10.1016/j.gim.2022.09.017.

236. Lappalainen, I. *et al.* DbVar and DGVa: Public archives for genomic structural variation. *Nucleic Acids Res* **41**, 936–941 (2013).

237. Rehm, H. L. *et al.* ClinGen — The Clinical Genome Resource. *N Engl J Med* **372**, 2235–2242 (2015).

238. Quinodoz, M. *et al.* AutoMap is a high performance homozygosity mapping tool using next-generation sequencing data. *Nat Commun* **12**, 1–7 (2021).

239. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

240. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).

241. Yu, G. & He, Q. Y. ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* **12**, 477–479 (2016).

242. The Gene Ontology Consortium *et al.* The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res* **49**, D325–D334 (2021).

243. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**, 27–30 (2000).

244. Martens, M. *et al.* WikiPathways: Connecting communities. *Nucleic Acids Res* **49**, D613–D621 (2021).

245. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res* **50**, D687–D692 (2022).

246. Keenan, A. B. *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res* **47**, W212–W224 (2019).

247. Brechtmann, F. *et al.* OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data. *Am J Hum Genet* **103**, 907–917 (2018).

248. Gagneur, V. Y. M. G. R. K. T. M. H. P. J. Gene expression and splicing counts from the Yepez, Gusic et al study - fibroblast, hg19, strand-specific, low seq depth (1.0) [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.7436579 (2022).

249. Yépez, V. A., Murdock, D. R. & Lee, B. Gene expression counts from fibroblast, strand-specific, BCM UDN. *Zenodo* https://doi.org/10.5281/zenodo.3963474 (2020).

250. Mertes, C. *et al.* Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat Commun* **12**, (2021).

251. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

252. Bertelsen, B., Tümer, Z. & Ravn, K. Three new loci for determining X chromosome inactivation patterns. *Journal of Molecular Diagnostics* **13**, 537–540 (2011).

253. De Hoon, B., Monkhorst, K., Riegman, P., Laven, J. S. E. & Gribnau, J. Buccal swab as a reliable predictor for X inactivation ratio in inaccessible tissues. *J Med Genet* **52**, 784–790 (2015).

254. Ørstavik, K. H. X chromosome inactivation in clinical practice. *Human Genetics* vol. 126 363–373 Preprint at https://doi.org/10.1007/s00439-009-0670-5 (2009).

255. Amir, R. E. *et al. Influence of Mutation Type and X Chromosome Inactivation on Rett Syndrome Phenotypes*. *Ann Neurol* vol. 47 (2000).

256. Van Den Veyver, I. B. & Zoghbi, H. Y. Mutations in the gene encoding methyl-CpG-binding protein 2 cause Rett syndrome. *Brain Dev* **2001**, S147–S151 (2001).

257. Weaving, L. S. *et al.* Effects of MECP2 mutation type, location and X-inactivation in modulating Rett syndrome phenotype. *Am J Med Genet* **118A**, 103–114 (2003).

258. 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).

259. Peeters, S. B., Yang, C. & Brown, C. J. Have humans lost control: The elusive X-controlling element. *Seminars in Cell and Developmental Biology* vol. 56 71–77 Preprint at https://doi.org/10.1016/j.semcdb.2016.01.044 (2016).

260. Plenge, R. M., Stevenson, R. A., Lubs, H. A., Schwartz, C. E. & Willard, H. F. *Skewed X-Chromosome Inactivation Is a Common Feature of X-Linked Mental Retardation Disorders*. *Am. J. Hum. Genet* vol. 71 (2002).

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

262. Gale, R. E., Wheadon, H., Boulos, P. & Linch, D. C. Tissue Specificity of X-Chromosome Inactivation Patterns. *Blood* **83**, 2899–2905 (1994).

263. Zhang, Q. *et al.* Familial cases and male cases with MECP2 mutations. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics* **174**, 451–457 (2017).

264. Swierczek, S. I. *et al.* Methylation of AR locus does not always reflect X chromosome inactivation state. *Blood* **29**, e100–e109 (2012).

265. Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science (1979)* **352**, 1586–1590 (2016).

266. Chao, H. T. *et al.* Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468**, 263–269 (2010).

267. Tretter, V. & Moss, S. J. GABAA receptor dynamics and constructing GABAergic synapses. *Front Mol Neurosci* **1**, (2008).

268. Çiçek, S. S. & Serhat Sezai Çiçek. Structure-Dependent Activity of Natural GABA(A) Receptor Modulators. *Molecules* **23**, (2018).

269. Duarte, S. T. *et al.* Abnormal Expression of Cerebrospinal Fluid Cation Chloride Cotransporters in Patients with Rett Syndrome. *PLoS One* **8**, (2013).

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

271. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (1979)* **369**, 1318–1330 (2020).

272. Calfa, G., Hablitz, J. J. & Pozzo-Miller, L. Network hyperexcitability in hippocampal slices from Mecp2 mutant mice revealed by voltage-sensitive dye imaging. *J Neurophysiol* **105**, 1768–1784 (2011).

273. Calfa, G., Li, W., Rutherford, J. M. & Pozzo-Miller, L. Excitation/inhibition imbalance and impaired synaptic inhibition in hippocampal area CA3 of Mecp2 knockout mice. *Hippocampus* **25**, 159–168 (2015).

274. El-Khoury, R. *et al.* GABA and glutamate pathways are spatially and developmentally affected in the brain of Mecp2-deficient mice. *PLoS One* **9**, (2014).

275. Zhang, W., Peterson, M., Beyer, B., Frankel, W. N. & Zhang, Z. W. Loss of MeCP2 from forebrain excitatory neurons leads to cortical hyperexcitation and seizures. *Journal of Neuroscience* **34**, 2754–2763 (2014).

276. Meng, X. *et al.* Manipulations of MeCP2 in glutamatergic neurons highlight their contributions to Rett and other neurological disorders. *Elife* **5**, (2016).

277. Côme, E., Heubl, M., Schwartz, E. J., Poncer, J. C. & Lévi, S. Reciprocal regulation of KCC2 trafficking and synaptic activity. *Frontiers in Cellular Neuroscience* vol. 13 Preprint at https://doi.org/10.3389/fncel.2019.00048 (2019).

278. Heubl, M. *et al.* GABAA receptor dependent synaptic inhibition rapidly tunes KCC2 activity via the Cl--sensitive WNK1 kinase. *Nat Commun* **8**, (2017).

279. Ricceri, L., De Filippis, B. & Laviola, G. Mouse models of Rett syndrome: From behavioural phenotyping to preclinical evaluation of new therapeutic approaches. *Behavioural Pharmacology* vol. 19 501–517 Preprint at https://doi.org/10.1097/FBP.0b013e32830c3645 (2008).

280. 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).

281. Moore, Y. E. *et al.* Developmental Regulation of KCC2 Phosphorylation Has Long-Term Impacts on Cognitive Function. *Front Mol Neurosci* **12**, (2019).

282. Renthal, W. *et al.* Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. *Nat Neurosci* **21**, 1670–1679 (2018).

283. 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).

284. Lin, P. *et al.* Transcriptome analysis of human brain tissue identifies reduced expression of complement complex C1Q Genes in Rett syndrome. *BMC Genomics* **17**, (2016).

285. Lozovaya, N. *et al.* Early alterations in a mouse model of Rett syndrome: the GABA developmental shift is abolished at birth. *Sci Rep* **9**, (2019).

286. McGraw, C. M., Samaco, R. C. & Zoghbi, H. Y. Adult neural function requires MeCP2. *Science* vol. 333 186 Preprint at https://doi.org/10.1126/science.1206593 (2011).

287. Simon, J., Wakimoto, H., Fujita, N., Lalande, M. & Barnard, E. A. Analysis of the set of GABAA receptor genes in the human genome. *Journal of Biological Chemistry* **279**, 41422–41435 (2004).

288. Enoch, M. A., Baghal, B., Yuan, Q. & Goldman, D. A Factor Analysis of Global GABAergic Gene Expression in Human Brain Identifies Specificity in Response to Chronic Alcohol and Cocaine Exposure. *PLoS One* **8**, (2013).

289. Sequeira, A., Shen, K., Gottlieb, A. & Limon, A. Human brain transcriptome analysis finds region- and subject-specific expression signatures of GABAAR subunits. *Commun Biol* **2**, (2019).

290. Pozzo-Miller, L., Pati, S. & Percy, A. K. Rett Syndrome: Reaching for Clinical Trials. *Neurotherapeutics* vol. 12 631–640 Preprint at https://doi.org/10.1007/s13311-015-0353-y (2015).

291. Murdock, D. R. *et al.* Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *Journal of Clinical Investigation* **131**, (2021).

292. Pecorelli, A. *et al.* Proteomic analysis of 4-hydroxynonenal and nitrotyrosine modified proteins in RTT fibroblasts. *International Journal of Biochemistry and Cell Biology* **81**, 236–245 (2016).

293. Varderidou-Minasian, S. *et al.* Quantitative proteomic analysis of Rett iPSC-derived neuronal progenitors. *Mol Autism* **11**, 1–15 (2020).

294. Latour, B. L. *et al.* Dysfunction of the ciliary ARMC9/TOGARAM1 protein module causes Joubert syndrome. *Journal of Clinical Investigation* **140**, 4423–4439 (2020).

295. 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. *Journal of Neuroscience* **25**, 3379–3388 (2005).

296. Roux, J. C. *et al.* Modification of Mecp2 dosage alters axonal transport through the Huntingtin/Hap1 pathway. *Neurobiol Dis* **45**, 786–795 (2012).

297. Setou, M., Nakagawa, T., Seog, D. H. & Hirokawa, N. Kinesin superfamily motor protein KIF17 and mLin-10 in NMDA receptor- containing vesicle transport. *Science (1979)* **288**, 1796–1802 (2000).

298. Patel, P. A. *et al.* Complete loss of the X-linked gene CASK causes severe cerebellar degeneration. *J Med Genet* jmedgenet-2021-108115 (2022) doi:10.1136/jmedgenet-2021-108115.

299. 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).

300. Riessland, M. *et al.* Neurocalcin Delta Suppression Protects against Spinal Muscular Atrophy in Humans and across Species by Restoring Impaired Endocytosis. *Am J Hum Genet* **100**, 297–315 (2017).

301. Chiu, C. F. *et al.* ZFPL1, a novel ring finger protein required for cis-Golgi integrity and efficient ER-to-Golgi transport. *EMBO Journal* **27**, 934–947 (2008).

302. 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).

303. 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).

304. 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.e11 (2020).

305. Huttlin, E. L. *et al.* Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).

306. Cheng, T. L. *et al.* Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain. *Sci Rep* **7**, 1–12 (2017).

307. 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**, (2020).

308. Chen, J. *et al.* Orientation and cellular distribution of membrane-bound catechol-O-methyltransferase in cortical neurons: Implications for drug development. *Journal of Biological Chemistry* **286**, 34752–34760 (2011).

309. 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).

310. Chen, Y., Shin, B. C., Thamotharan, S. & Devaskar, S. U. Creb1-Mecp2-mCpG complex transactivates postnatal murine neuronal glucose transporter isoform 3 expression. *Endocrinology* **154**, 1598–1611 (2013).

311. Kaldun, J. C. & Sprecher, S. G. Initiated by CREB: Resolving Gene Regulatory Programs in Learning and Memory. *BioEssays* **41**, (2019).

312. Pardo, L. *et al.* CREB Regulates Distinct Adaptive Transcriptional Programs in Astrocytes and Neurons. *Sci Rep* **7**, (2017).

313. Tabuchi, A. & Ihara, D. Regulation of Dendritic Synaptic Morphology and Transcription by the SRF Cofactor MKL/MRTF. *Front Mol Neurosci* **14**, 1–8 (2021).

314. Knöll, B. & Nordheim, A. Functional versatility of transcription factors in the nervous system: the SRF paradigm. *Trends Neurosci* **32**, 432–442 (2009).

315. Shiota, J. *et al.* Developmental expression of the SRF co-activator MAL in brain: Role in regulating dendritic morphology. *J Neurochem* **98**, 1778–1788 (2006).

316. Zlatic, S. A. *et al.* Systemic Proteome Phenotypes Reveal Defective Metabolic Flexibility in Mecp2 Mutants. doi:10.1093/hmg/ddad154/7274644.

317. Xu, X., Miller, E. C. & Pozzo-Miller, L. Dendritic spine dysgenesis in Rett syndrome. *Front Neuroanat* **8**, 1–8 (2014).

318. Phillips, M. & Pozzo-Miller, L. Dendritic spine dysgenesis in autism related disorders. *Neurosci Lett* **601**, 30–40 (2015).

319. Shah, S. & Richter, J. D. Do Fragile X Syndrome and Other Intellectual Disorders Converge at Aberrant Pre-mRNA Splicing? *Front Psychiatry* **12**, 1–7 (2021).

320. Hijazi, H. *et al.* TCEAL1 loss-of-function results in an X-linked dominant neurodevelopmental syndrome and drives the neurological disease trait in Xq22.2 deletions. *Am J Hum Genet* **109**, 2270–2282 (2022).

321. Engel, A. G., Selcen, D., Shen, X. M., Milone, M. & Harper, C. M. Loss of MUNC13-1 function causes microcephaly, cortical hyperexcitability, and fatal myasthenia. *Neurol Genet* **2**, (2016).

322. Lipstein, N. *et al.* Synaptic UNC13A protein variant causes increased neurotransmission and dyskinetic movement disorder. *Journal of Clinical Investigation* **127**, 1005–1018 (2017).

323. Fitzgerald, T. W. *et al.* Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223–228 (2015).

324. Olson, H. E. *et al.* Mutations in epilepsy and intellectual disability genes in patients with features of Rett syndrome. *Am J Med Genet A* **167**, 2017–2025 (2015).

325. Henriksen, M. W. *et al.* Genetic and clinical variations in a Norwegian sample diagnosed with Rett syndrome. *Brain Dev* **42**, 484–495 (2020).

326. Allou, L. *et al.* Rett-like phenotypes: expanding the genetic heterogeneity to the KCNA2 gene and first familial case of CDKL5-related disease. *Clin Genet* **91**, 431–440 (2017).

327. Alfares, A. *et al.* Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genetics in Medicine* **20**, 1328–1333 (2018).

328. Li, J. *et al.* Reanalysis of whole exome sequencing data in patients with epilepsy and intellectual disability/mental retardation. *Gene* **700**, 168–175 (2019).

329. Jalkh, N. *et al.* The added value of WES reanalysis in the field of genetic diagnosis: Lessons learned from 200 exomes in the Lebanese population. *BMC Med Genomics* **12**, (2019).

330. Al-Nabhani, M. *et al.* Reanalysis of exome sequencing data of intellectual disability samples: Yields and benefits. *Clin Genet* **94**, 495–501 (2018).

331. Ewans, L. J. *et al.* Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. *Genetics in Medicine* **20**, 1564–1574 (2018).

332. Archer, H. *et al.* Correlation between clinical severity in patients with Rett syndrome with a p.R168X or p.T158M MECP2 mutation, and the direction and degree of skewing of X-chromosome inactivation. *J Med Genet* **44**, 148–152 (2007).

333. Ben Zeev, B. *et al.* The common BDNF polymorphism may be a modifier of disease severity in Rett syndrome. *Neurology* **72**, 1242–1247 (2009).

334. Kondo, M. *et al.* Environmental enrichment ameliorates a motor coordination deficit in a mouse model of Rett syndrome - Mecp2 gene dosage effects and BDNF expression. *European Journal of Neuroscience* **27**, 3342–3350 (2008).

335. De Felice, C. *et al.* Effects of ω-3 polyunsaturated fatty acids on plasma proteome in rett syndrome. *Mediators Inflamm* **2013**, (2013).

336. 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).

337. Buchovecky, C. M. *et al.* A suppressor screen in Mecp2 mutant mice implicates cholesterol metabolism in Rett syndrome. *Nat Genet* **45**, 1013–1020 (2013).

338. Lopez, A. M., Chuang, J. C., Posey, K. S. & Turley, S. D. Suppression of brain cholesterol synthesis in male Mecp2-deficient mice is age dependent and not accompanied by a concurrent change in the rate of fatty acid synthesis. *Brain Res* **1654**, 77–84 (2017).

339. 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? *Frontiers in Neuroscience* vol. 13 Preprint at https://doi.org/10.3389/fnins.2019.00680 (2019).

340. Johnson, B. S. *et al.* Biotin tagging of MeCP2 in mice reveals contextual insights into the Rett syndrome transcriptome. *Nat Med* **23**, 1203–1214 (2017).

341. Kishi, N. *et al.* Reduction of aberrant NF-κB signalling ameliorates Rett syndrome phenotypes in Mecp2-null mice. *Nat Commun* **7**, 1–13 (2016).

342. Thomas, A. X. & Brooks-Kayal, A. R. Excitation–Inhibition Epilepsies. in *Neural Circuit Development and Function in the Brain* 709–730 (Elsevier, 2013). doi:10.1016/B978-0-12-397267-5.00042-X.

343. Petazzi, P. *et al.* Global Impairment of Immediate-Early Genes Expression in Rett Syndrome Models and Patients Linked to Myelination Defects. *Int J Mol Sci* **24**, (2023).

ANNEX 1

Supplementary Tables

Results of XCI assays

Table S1. Results of XCI assays in blood for all patients included in the XCI studies. The XCI-AR column shows the results of the AR XCI assay (percentage of inactivation of each allele. The XCI-AS WT and Mut columns show the results of the allele-specific XCI assay (percentage of inactivation of each allele, mean of two replicates, or three replicates in the case of the deletions). The CSS column shows the average score and its standard deviation (SD) in brackets for the patients of our cohort with each mutation. n.i. = polymorphism noninformative for the assay. NA = data not available.

Patient	XCI-AR	XCI-AS WT:Mut	CSS (Mean ± SD)
c.455C>G (p.Pro152Arg)			11.6 ± 3.8
P1	n.i.	53.5:46.5	8
P2	60:40	62:38	13
P3	61:39	59.5:40.5	8
P4	74:26	48.5:51.5	17
P5	68:32	59.5:40.5	NA
P6	72:28	34:66	12
c.473C>T (p.Thr158Met)			13.2 ± 3.5
P7	72:28	37: 63	10
P8	54:46	66.5:33.5	16
P9a	58:42	60.5:39.5	12
P9b	56:44	36:64	
P10	54:46	37.5: 62.5	12
P11	55:45	52.5:47.5	12
P12	55:45	65:35	11
P13	62:38	35.5:64.5	16
P14	65:35	36:64	9
P15	61:39	44:56	11
P16	53:47	54:46	16
P17	56:44	41.5:58	18
P18	n.i.	41.5:58.5	16
P19	67:33	35.5:64.5	11
P20	n.i.	57:43	NA
P21	62:38	36.5: 63.5	15
P22	79:21	41:59	19
P23	71:29	46:54	NA
P24	59:41	63:37	14
P25	51:49	31.5:68.5	19
P26	63:37	41.5:58.5	9
P27	NA	NA	13
P28	NA	NA	17
P29	73:27	59:41	6
P30	71:29	51:49	14
P31	62:38	48.5:51.5	NA
P32	63:37	39.5:60.5	10
P33	n.i.	48:52	7
P34	68:32	56:44	19
P35	n.i.	NA	16
P36	71:29	41:59	10
P37	71:29	50.5:49.5	12
P38	55:45	37.5:62.5	NA
P39	77:23	NA	NA
P40	54:46	45:55	14

P41	64:36	52:48	NA
P42	NA	NA	13
c.502C>T (p.Arg168*)			13.1 ± 3.4
P43	66:34	46:54	13
P44	64:36	44:56	17
P45	65:35	60.5: 39.5	11
P46	69:31	32:68	10
P47	n.i.	81.5:18.5	13
P48	70:30	66.5:33.5	7
P49	51:49	56:44	12
P50	68:32	48:52	19
P51	NA	NA	19
P52	64:36	49.5:50.5	16
P53	63:37	40:60	7
P54	NA	NA	17
P55	59:41	38.5:61.5	14
P56	56:44	43:57	13
P57	n.i.	61:39	10
P58	64:36	38.5:61.5	11
P59	n.i.	59:41	14
P60	84:16	28:72	16
P61	58:44	35:65	13
P62	NA	NA	12
P63	70:30	54:46	17
P64	NA	NA	17
P65	65:35	56.5:43.5	7
P66	76:24	43.5:56.5	10
P67	58:42	45:55	NA
P68	75:25	15.5:84.5	NA
P69	n.i.	46:54	10
P70	85:15	35:65	NA
P71	60:40	41.5:58.5	12
P72	n.i.	46.5:53.5	11
P73	59:41	42:58	18
P74	81:19	55.5:44.5	NA
P75	54:46	50:50	17
P76	NA	NA	14
P77	NA	NA	7
P78	NA	NA	12
P79	NA	NA	11
P80	NA	NA	13
c.674C>G (p.P225R)			
P81	57:43	54:46	NA
P82	67:33	55.5:44.5	11
c.763C>T (p.R255X)			15.2 ± 3.2
P83	85:15	57:43	NA
P84	87:13	55.5:44.5	13
P85	80:20	28:72	14
P86	78,5:21,5	71.5:28.5	14
P87	76,5:23,5	63.5:36.5	9
P88	78:22	53.5:46.5	17
P89	75:25	25:75	18
P90	75:25	63:37	17
P91	NA	NA	10
P92	73:27	33:67	18
P93	71:29	62:38	13
P94	65:35	32:68	19
P95	65:35	52:48	17
P96	65:35	51:49	18
P97	64:36	43:57	14

P98	64:36	52:48	NA
P99	63:37	51:49	11
P100	60:40	68:32	12
P101	59:41	60:40	14
P102	57:43	71.5:28.5	18
P103	55:45	57:43	15
P104	51:49	60:40	17
P105	n.i.	45:55	17
P106	62:38	62:38	17
P107	87:13	68:32	11
P108	75:25	68:32	17
P109	73:27	58:42	20
P110	68:32	60:40	13
P111	65:35	49:51	15
P112	64:36	45:55	NA
P113	64:36	56:44	NA
P114	60:40	50:50	15
P115	58:42	53:47	NA
P116	NA	NA	18
P117	52:48	40:60	NA
P118	51:49	62:38	NA
P119	n.i.	34:66	19
P120	58:42	48:52	NA
P121	NA	NA	20
P122	NA	NA	17
P123	NA	NA	10
P124	NA	NA	19
P125	NA	NA	10
P126	NA	NA	16
P127	NA	NA	10
P128	NA	NA	19
P129	NA	NA	12
c.806delG (p.G269fs)			14.3 ± 4.1
P130	77:23	51.5:48.5	8
P131	76:24	36.5:64.5	NA
P132	76:24	65:35	14
P133	n.i.	52.5:47.5	20
P134	n.i.	55.5:44.5	16
P135	59:41	43.5:56.5	10
P136	68:32	59:41	15
P137	54:46	41:59	17
P138	77:23	52:48	NA
P139	82:18	58:42	NA
P140	68:32	57.5:42.5	NA
P141	NA	NA	NA
P142	NA	NA	NA
c.808C>T (p.R270X)			14.7 ± 3.8
P143	97:3	16:84	18
P144	84:16	21:79	NA
P145	81:19	30:70	9
P146	80:20	73:27	13
P147	73:27	62:38	19
P148	71,5:28,5	33:67	11
P149	72,5:27,5	58:42	11
P150	69:31	59:41	17
P151	69:31	48:52	22
P152	65:35	53:47	17
P153	64:36	61:39	8
P154	68:32	37:63	16
P155	NA	NA	17

P156	54:46	56:44	19
P157	53:47	39:61	16
P158	53:47	56:44	16
P159	53:47	52:48	16
P160	53:47	36:64	18
P161	52:48	57:43	17
P162	62:38	54.5:45.5	NA
P163	58:42	48:52	21
P164	NA	NA	13
P165	NA	NA	9
P166	NA	NA	12
P167	NA	NA	16
P168	NA	NA	11
P169	NA	NA	14
P170	NA	NA	10
P171	NA	NA	9
P172	NA	NA	18
P173	NA	NA	13
c.880C>T (p.R294X)			10.5 ± 3
P174	71:29	48:52	7
P175	59:41	55:45	11
P176	58:42	40:60	12
P177	67:33	51:49	16
P178	59:41	37.5:62.5	13
P179	57:43	53.5:46.5	NA
P180	75:25	54:46	13
P181	56:44	NA	7
P182	55:45	65:35	NA
P183	54:46	51:49	10
P184	55:45	52.5:47.5	NA
P185	62:38	NA	NA
P186	53:47	NA	12
P187	56:44	NA	NA
P188	53:47	44:56	12
P189	62:38	46:54	NA
P190	59:41	NA	6
P191	89:11	49:51	NA
P192	67:33	42:58	8
P193	75:25	34:66	NA
P194	NA	NA	9
c.916C>T (p.R306C)			11.2 ± 3.1
P195	89:11	59.5:40.5	9
P196a	70,5:29,5	60:40	14
P196b	45:55	63:37	NA
P197	70:30	45:55	9
P198	63:37	63:37	11
P199	62:38	76:24	12
P200	59:41	49:51	16
P201	59:41	56:44	NA
P202	59:41	63:37	10
P203	58:42	66:34	7
P204	56,5:43,5	47:53	16
P205	61:39	31:69	14
P206	60:40	32:68	12
P207	64:36	38:62	16
P208	70:30	25:75	NA
P209	n.i.	38:62	NA
P210	NA	NA	12
P211	NA	NA	10
P212	NA	NA	7

P213	NA	NA	7
P214	NA	NA	11
P215	NA	NA	11
P216	NA	NA	13
P217	NA	NA	11
P218	NA	NA	5
P219	NA	NA	13
Deletions in *MECP2*			
P220	88:12	7:93	NA
P221	74:26	51:49	15

Differential expression results from multi-omics in RTT patients

Table S2. Differential expression results that were significant in both RNAseq and proteomics for RTT patients vs healthy controls. Fold changes in red and green indicate downregulated and upregulated genes, respectively. Genes in bold are dysregulated in the same direction according to both omics.

	Transcriptomics		Proteomics	
GeneSymbol	p-value	Fold Change	p-value	Fold Change
AACS	**0.00190**	**0.78034**	**0.02774**	**0.42969**
ABCB8	0.00004	0.83470	0.00140	3.04580
ABCC4	**0.02839**	**0.79643**	**0.03440**	**0.43369**
AFAP1	**0.01271**	**1.33460**	**0.02121**	**2.50779**
AGPAT3	**0.00632**	**0.87789**	**0.02318**	**0.45249**
ALS2	0.00271	0.84393	0.00018	4.64430
ARHGEF1	**0.00000**	**0.82919**	**0.01857**	**0.40413**
ARIH1	**0.00000**	**1.14203**	**0.03828**	**2.30512**
ARMC9	**0.00488**	**0.68509**	**0.00596**	**0.30966**
BCL2L12	**0.00421**	**0.79970**	**0.03968**	**0.46033**
CDC42EP1	**0.00153**	**0.79094**	**0.01430**	**0.36564**
CEP170	**0.00445**	**1.21412**	**0.00870**	**2.68824**
COMT	**0.00013**	**0.81785**	**0.00168**	**0.30611**
COX6A1	0.00252	0.87908	0.03695	2.09044
CTBS	**0.00000**	**1.26403**	**0.03620**	**2.24421**
CYB5R2	**0.00090**	**0.69221**	**0.03486**	**0.43918**
DAP	0.03727	0.88191	0.00512	2.85268
DCAF11	0.00637	0.88989	0.00451	2.68662
DCAKD	**0.04477**	**0.88988**	**0.03798**	**0.45593**
DDX31	**0.02383**	**0.86442**	**0.02545**	**0.42627**
DDX54	**0.00000**	**0.82157**	**0.02877**	**0.40241**
EIF4G3	**0.02717**	**1.09643**	**0.03128**	**2.38663**
EIF4G3	**0.04955**	**1.14686**	**0.04626**	**2.16950**
FMNL2	**0.00063**	**1.41061**	**0.00118**	**3.50793**
FNBP1L	**0.00063**	**1.34463**	**0.04159**	**2.25483**
FZD2	**0.00726**	**0.75179**	**0.03924**	**0.43789**
GATD1	0.00136	0.84617	0.01575	2.51809
GLT8D1	0.00381	1.13971	0.03807	0.44019
GUK1	0.04725	0.88866	0.01923	2.52828
HS2ST1	**0.00110**	**1.22008**	**0.00484**	**2.99088**
HSD17B11	0.00972	1.22272	0.04767	0.44934
IDH3G	0.01098	0.80231	0.02987	2.45767
IL17RA	**0.03899**	**0.81173**	**0.00380**	**0.31503**
IQGAP3	**0.00726**	**0.60519**	**0.04348**	**0.44122**
KCTD12	**0.00079**	**1.69126**	**0.00746**	**2.82156**
KIF3A	**0.00002**	**1.16621**	**0.04336**	**2.30031**
KLHL42	**0.01761**	**1.13597**	**0.00225**	**2.55805**
MAP3K20	**0.00353**	**1.24994**	**0.01306**	**2.60508**

MARCKSL1	**0.00507**	**1.69896**	**0.00336**	**3.32128**
MARF1	**0.00659**	**1.10269**	**0.03373**	**2.24537**
MBD1	**0.04282**	**0.93814**	**0.00279**	**0.35514**
MRPL14	0.00000	0.84758	0.02761	2.29228
MRPL28	0.00003	0.82526	0.03987	2.12834
MYBBP1A	**0.00413**	**0.84546**	**0.00536**	**0.32097**
NCALD	**0.00078**	**1.66300**	**0.02347**	**2.58122**
NME4	0.00167	0.82660	0.01274	2.59560
NSMCE3	**0.04018**	**1.10552**	**0.00114**	**2.90562**
NUDT12	**0.04719**	**1.20290**	**0.03456**	**2.20910**
ORMDL2	**0.00001**	**0.79345**	**0.04381**	**0.49329**
OSTF1	0.00547	0.86229	0.04349	2.34214
PCK2	**0.04360**	**0.80472**	**0.04091**	**0.44068**
PI4KB	**0.01014**	**0.89993**	**0.00316**	**0.34733**
PLS3	**0.03108**	**1.28261**	**0.02321**	**2.51230**
PLSCR1	0.00961	1.22911	0.04044	0.45410
PLXNB2	**0.00876**	**0.88532**	**0.01432**	**0.35749**
PPP3CB	**0.00561**	**1.13964**	**0.04754**	**2.00101**
PREPL	**0.00013**	**1.19472**	**0.00772**	**2.85419**
RASSF8	**0.00491**	**1.38058**	**0.04491**	**2.17047**
RPAP1	0.00002	0.79047	0.03807	2.23591
SART1	**0.00009**	**0.82235**	**0.03427**	**0.50295**
SLC12A2	0.02857	1.15858	0.01980	0.40967
SLC7A1	**0.00000**	**0.74237**	**0.01939**	**0.36554**
SMAD5	**0.01179**	**1.15190**	**0.00077**	**3.91699**
SZRD1	0.01555	0.91056	0.03532	2.31565
TMED1	**0.00386**	**0.84918**	**0.02485**	**0.39295**
TMEM214	**0.00005**	**0.84070**	**0.03301**	**0.47075**
TMSB10	0.00031	0.79652	0.03327	2.45351
TRPM4	0.00136	0.78201	0.00662	0.34667
TUBB6	0.00027	0.77073	0.02394	2.34675
UAP1L1	**0.00053**	**0.62762**	**0.04591**	**0.46741**
UBA3	**0.00142**	**1.14409**	**0.03199**	**2.39929**
VPS28	0.00000	0.82388	0.02066	2.51074
WDR54	0.01762	0.83184	0.04354	2.34349
ZFPL1	**0.00001**	**0.82443**	**0.03571**	**0.41699**
ZFR	0.00177	1.11364	0.02279	0.40003

RTT-spectrum disorders candidate gene list

Table S3. RTT-spectrum candidate gene list. Genes that scored >=6 in semantic similarity of their phenotypes with respect to RTT diagnostic criteria.

AAAS	*AP1S2*	*BBS10*	*CCDC88C*	*CLP1*
AARS1	*AP2M1*	*BBS12*	*CCND2*	*CLPB*
AASS	*AP3B2*	*BBS2*	*CCNK*	*CLTC*
ABCB7	*AP4B1*	*BBS4*	*CD96*	*CNKSR2*
ABCC8	*AP4E1*	*BBS5*	*CDC42*	*CNOT1*
ABCD1	*AP4M1*	*BBS7*	*CDC45*	*CNOT3*
ABHD5	*AP4S1*	*BBS9*	*CDC6*	*CNPY3*
ACADM	*APC2*	*BCAP31*	*CDH15*	*CNTNAP1*
ACADS	*APP*	*BCL11A*	*CDH2*	*CNTNAP2*
ACAN	*ARCN1*	*BCL11B*	*CDK10*	*COA8*
ACAT1	*ARFGEF2*	*BCOR*	*CDK13*	*COASY*
ACO2	*ARG1*	*BCS1L*	*CDK5RAP2*	*COG1*
ACOX1	*ARHGAP31*	*BICD2*	*CDK8*	*COG4*
ACP5	*ARHGEF9*	*BIN1*	*CDKL5*	*COG5*
ACSL4	*ARID1A*	*BMP2*	*CDKN1C*	*COG6*
ACTA1	*ARID1B*	*BMP4*	*CDON*	*COG7*
ACTB	*ARID2*	*BOLA3*	*CDT1*	*COG8*
ACTG1	*ARL13B*	*BPTF*	*CENPJ*	*COL13A1*
ACTL6B	*ARL3*	*BRAF*	*CEP104*	*COL2A1*
ACY1	*ARL6*	*BRAT1*	*CEP135*	*COL4A1*
ADAR	*ARMC9*	*BRCA1*	*CEP152*	*COL4A2*
ADGRG1	*ARSA*	*BRCA2*	*CEP290*	*COL6A1*
ADK	*ARX*	*BRPF1*	*CEP41*	*COL6A3*
ADNP	*ASAH1*	*BRWD3*	*CEP57*	*COLEC10*
ADPRS	*ASCL1*	*BSND*	*CEP63*	*COQ2*
ADSL	*ASH1L*	*BTD*	*CERT1*	*COQ4*
AFF2	*ASL*	*BUB1B*	*CFL2*	*COQ5*
AFF3	*ASPA*	*C12orf57*	*CHAMP1*	*COQ8A*
AFF4	*ASPM*	*C2CD3*	*CHD2*	*COQ9*
AFG3L2	*ASS1*	*CA8*	*CHD3*	*COX15*
AGA	*ASXL1*	*CACNA1A*	*CHD4*	*COX7B*
37469	*ASXL2*	*CACNA1B*	*CHD7*	*CPLANE1*
AHDC1	*ASXL3*	*CACNA1C*	*CHD8*	*CPS1*
AHI1	*ATAD3A*	*CACNA1D*	*CHKB*	*CPT1A*
AIFM1	*ATM*	*CACNA1E*	*CHMP1A*	*CRADD*
AIMP1	*ATN1*	*CACNA1G*	*CHRNA4*	*CRB1*
AIPL1	*ATP1A1*	*CACNG2*	*CHRNA7*	*CRBN*
ALDH18A1	*ATP1A2*	*CAD*	*CHRNB2*	*CREBBP*
ALDH3A2	*ATP1A3*	*CAMK2A*	*CHRND*	*CRPPA*
ALDH5A1	*ATP5F1D*	*CAMK2B*	*CHRNG*	*CRX*
ALDH7A1	*ATP6AP2*	*CAMTA1*	*CHST14*	*CSF1R*
ALG1	*ATP6V0A2*	*CARS1*	*CHST3*	*CSNK2A1*
ALG11	*ATP6V1A*	*CARS2*	*CHSY1*	*CSPP1*
ALG13	*ATP6V1B2*	*CASK*	*CIC*	*CTCF*
ALG2	*ATP6V1E1*	*CAV1*	*CISD2*	*CTNNA2*
ALG3	*ATP7A*	*CBL*	*CIT*	*CTNNB1*
ALG9	*ATP8A2*	*CBS*	*CKAP2L*	*CTSD*
ALPL	*ATR*	*CC2D1A*	*CLCN4*	*CUL3*
ALS2	*ATRX*	*CC2D2A*	*CLCNKA*	*CUL4B*
ALX4	*AUH*	*CCBE1*	*CLCNKB*	*CUX2*
AMER1	*AUTS2*	*CCDC22*	*CLIC2*	*CWC27*
AMPD2	*B3GALNT2*	*CCDC32*	*CLN3*	*CYB5R3*
ANKRD11	*B3GALT6*	*CCDC47*	*CLN5*	*CYP27A1*
ANTXR1	*B9D1*	*CCDC88A*	*CLN8*	*CYP2U1*

D2HGDH
DAG1
DARS1
DARS2
DCC
DCHS1
DCX
DDB2
DDC
DDHD2
DDOST
DDX11
DDX3X
DDX6
DEAF1
DEGS1
DENND5A
DEPDC5
DHCR24
DHCR7
DHDDS
DHPS
DHTKD1
DHX16
DHX30
DHX37
DIP2B
DIS3L2
DKC1
DLAT
DLD
DLG3
DLG4
DLL1
DMD
DMPK
DMXL2
DNA2
DNAJC12
DNM1
DNMT3A
DOCK3
DOCK6
DOCK7
DOCK8
DPAGT1
DPF2
DPM1
DPM3
DPYD
DSTYK
DYM
DYNC1H1
DYNC1I2
DYRK1A
EBF3
EBP
EED
EEF1A2
EFNB1
EFTUD2
EGR2
EHMT1
EIF2AK3
EIF2B1
EIF2B2
EIF2B3
EIF2B4
EIF2B5
EIF2S3
ELN
ELOVL4
ELP2
EMC1
ENPP1
ENTPD1
EP300
EPB41L1
EPG5
EPM2A
EPRS1
ERCC1
ERCC2
ERCC3
ERCC4
ERCC5
ERCC6
ERCC8
ERF
ERLIN2
ERMARD
ESCO2
ETHE1
EXOSC3
EXT2
EXTL3
EYA1
EZH2
FAM126A
FAM149B1
FANCA
FANCB
FANCC
FANCD2
FANCE
FANCF
FANCG
FANCI
FANCL
FANCM
FAR1
FARS2
FAT4
FBLN1
FBN1
FBP1
FBXL4
FBXO11
FBXW11
FCSK
FEZF1
FGF12
FGF13
FGF8
FGFR1
FGFR2
FGFR3
FH
FIG4
FKBP14
FKRP
FKTN
FLNA
FLVCR1
FMN2
FMR1
FOLR1
FOXE1
FOXG1
FOXP1
FOXRED1
FRMPD4
FRRS1L
FTCD
FTSJ1
FUCA1
FUT8
GABBR2
GABRA1
GABRA2
GABRA5
GABRB1
GABRB2
GABRB3
GABRD
GABRG2
GAD1
GALC
GALE
GALK1
GALT
GAMT
GATA4
GATA6
GATAD2B
GATM
GBA
GBA2
GBE1
GCDH
GCH1
GDF6
GDI1
GFER
GJA1
GJC2
GK
GLB1
GLE1
GLI2
GLI3
GLRA1
GLRB
GLS
GLUD1
GLUL
GM2A
GMNN
GMPPA
GMPPB
GNAO1
GNAQ
GNAS
GNB1
GNB5
GNE
GNPTAB
GORAB
GOSR2
GOT2
GPAA1
GPC3
GPC4
GPHN
GPSM2
GRIA2
GRIA3
GRIA4
GRIK2
GRIN1
GRIN2A
GRIN2B
GRIN2D
GRM1
GRM7
GRN
GTF2E2
GTF2H5
GTPBP3
H1-4
HACE1
HADH
HADHA
HAX1
HCCS
HCFC1
HCN1
HDAC4
HDAC8
HECW2
HERC1
HERC2
HESX1
HGSNAT
HIBCH
HIVEP2
HLCS
HMGB3
HMGCL
HNF1B
HNRNPH2
HNRNPK
HNRNPU
HPRT1
HRAS
HSD17B10
HSD17B4
HSPD1
HSPG2
HTRA2
HTT
HUWE1
HYLS1
IARS1
IDH2
IDS
IDUA
IER3IP1
IFIH1
IFT140
IGF1
IGF1R
IGF2
IKBKG
IL10
IL1RAPL1
INPP5E
INPP5K
IQSEC2
IRF2BPL
ITGA7
ITPR1
KANSL1
KARS1
KAT6A
KAT6B
KBTBD13
KCNA1
KCNA2
KCNA4
KCNB1
KCNC3
KCNH1
KCNJ1
KCNJ10
KCNJ11
KCNJ13
KCNJ2
KCNJ5
KCNJ6
KCNK4
KCNMA1
KCNN3
KCNQ2
KCNQ5
KCNT1
KCNT2
KCTD7
KDM1A
KDM3B
KDM5B
KDM5C
KDM6A
KDM6B
KIAA0586

KIAA1109
KIDINS220
KIF11
KIF14
KIF1A
KIF4A
KIF5C
KIF7
KIFBP
KLHL15
KMT2A
KMT2B
KMT2C
KMT2D
KMT2E
KMT5B
KPTN
KRAS
L1CAM
L2HGDH
LAGE3
LAMA1
LAMA2
LAMB1
LAMB2
LAMP2
LARGE1
LARP7
LARS2
LAS1L
LBR
LEMD3
LHX3
LHX4
LIAS
LIG4
LINGO1
LINS1
LIPT1
LIPT2
LMBRD1
LMNA
LMX1B
LNPK
LONP1
LRAT
LRP2
LRP4
LRP5
LRPPRC
LYRM7
LYST
LZTR1
MACF1
MAF
MAFB
MAGEL2
MAN1B1
MAN2B1
MANBA
MAP2K1
MAP2K2
MAPK1
MAPK8IP3
MAPRE2
MAPT
MAST1
MBD5
MBOAT7
MCCC1
MCCC2
MCOLN1
MCPH1
MDH2
MECP2
MED12
MED13
MED13L
MED23
MEF2C
MEIS2
METTL5
MFSD2A
MFSD8
MGAT2
MGP
MICU1
MKKS
MKS1
MLC1
MLYCD
MMACHC
MMADHC
MMP13
MMUT
MN1
MOCS1
MOGS
MORC2
MPDU1
MPLKIP
MPV17
MRE11
MRPS2
MRPS34
MSL3
MTHFR
MTM1
MTMR14
MTOR
MTR
MTRFR
MTRR
MYH3
MYO5A
MYPN
MYRF
MYT1L
NAA10
NAA15
NACC1
NADK2
NAGA
NAGS
NALCN
NANS
NAXD
NBEA
NCAPD2
NCAPD3
NCAPG2
NDE1
NDP
NDST1
NDUFA1
NDUFA10
NDUFA11
NDUFA2
NDUFA6
NDUFA9
NDUFAF2
NDUFAF6
NDUFB11
NDUFB8
NDUFS1
NDUFS3
NDUFS4
NDUFS7
NDUFS8
NDUFV1
NEB
NECAP1
NECTIN1
NEDD4L
NEFL
NEK1
NEU1
NEXMIF
NF1
NFIA
NFIX
NGLY1
NHLRC1
NHP2
NIPBL
NKX2-1
NKX2-5
NKX6-2
NLGN3
NLGN4X
NMNAT1
NODAL
NONO
NOP10
NOTCH3
NOVA2
NPC1
NPC2
NPHP1
NR2F1
NRAS
NRXN1
NSD1
NSDHL
NSF
NSUN2
NTNG2
NTRK1
NTRK2
NUBPL
NUP107
NUP133
NUP214
NUP62
NUS1
OCRL
OFD1
OPA1
OPHN1
ORC1
ORC4
ORC6
OSGEP
OTC
OTUD6B
OTX2
PACS1
PACS2
PAFAH1B1
PAH
PAK1
PAK3
PALB2
PANK2
PARN
PAX6
PAX8
PC
PCCA
PCCB
PCDH19
PCGF2
PCNT
PCYT1A
PDE10A
PDE2A
PDE4D
PDGFRB
PDHA1
PDHX
PET100
PEX1
PEX10
PEX11B
PEX12
PEX13
PEX14
PEX16
PEX19
PEX2
PEX26
PEX3
PEX5
PEX6
PEX7
PGAP1
PGAP2
PGAP3
PGK1
PGM3
PHACTR1
PHC1
PHF21A
PHF6
PHF8
PHGDH
PHOX2B
PIEZO2
PIGA
PIGB
PIGG
PIGL
PIGN
PIGO
PIGQ
PIGS
PIGT
PIGU
PIGV
PIGW
PIGY
PIK3CA
PIK3R2
PLA2G6
PLAA
PLCB1
PLK4
PLOD1
PLP1
PLPBP
PMM2
PMPCB
PMS2
PNKP
PNPLA2
PNPO
POC1A
POGZ
POLA1
POLG
POLR2A
POLR3A
POMGNT1
POMGNT2
POMT1
POMT2
POU1F1
POU3F3
PPM1D
PPP1CB
PPP1R15B
PPP2CA
PPP2R1A
PPP2R5D
PPP3CA

PPT1
PQBP1
PREPL
PRICKLE1
PRKAR1A
PRKD1
PRMT7
PRODH
PROP1
PRPS1
PRR12
PRRT2
PRSS12
PRUNE1
PSAP
PSAT1
PSEN1
PSEN2
PSMD12
PSPH
PTCH1
PTCHD1
PTDSS1
PTEN
PTPN11
PTPN23
PTS
PUF60
PURA
PUS7
PYCR1
PYCR2
QDPR
QRICH1
RAB11B
RAB18
RAB39B
RAB3GAP1
RAB3GAP2
RAC1
RAC3
RAD21
RAD51
RAD51C
RAF1
RAI1
RANBP2
RARS2
RBM10
RBM8A
RBPJ
RECQL4
RELN
RERE
RET
RFT1
RHOBTB2
RLIM
RMND1
RNASEH2A
RNASEH2B
RNASEH2C
RNASET2
RNF113A
RNF13
RNF168
RNU4ATAC
ROGDI
RORA
RPE65
RPGRIP1
RPGRIP1L
RPS19
RPS6KA3
RRAS2
RRM2B
RSPRY1
RTEL1
RTTN
RUBCN
RXYLT1
RYR1
SACS
SALL4
SAMD9L
SAMHD1
SATB2
SC5D
SCARB2
SCN1A
SCN1B
SCN2A
SCN3A
SCN4A
SCN8A
SCO2
SCYL1
SDCCAG8
SDHA
SDHAF1
SEC23B
SELENOI
SEMA6B
SEPSECS
SET
SETBP1
SETD1A
SETD1B
SETD2
SETD5
SHANK3
SHH
SHOC2
SIK1
SIL1
SIN3A
SIX3
SKI
SLC12A1
SLC12A2
SLC12A5
SLC12A6
SLC13A5
SLC16A2
SLC17A5
SLC18A2
SLC19A3
SLC1A2
SLC1A3
SLC25A15
SLC25A20
SLC25A22
SLC25A4
SLC2A1
SLC2A10
SLC33A1
SLC35A1
SLC35A2
SLC35C1
SLC39A13
SLC39A8
SLC45A1
SLC46A1
SLC5A5
SLC5A7
SLC6A1
SLC6A17
SLC6A3
SLC6A5
SLC6A8
SLC9A6
SLX4
SMAD4
SMARCA2
SMARCA4
SMARCAL1
SMARCB1
SMARCC2
SMARCD1
SMARCE1
SMC1A
SMC3
SMG9
SMO
SMPD1
SMPD4
SMS
SNAP25
SNAP29
SNIP1
SNORD118
SNRPN
SNX14
SOBP
SON
SOS1
SOX10
SOX11
SOX2
SOX3
SOX4
SOX5
SPATA5
SPEG
SPEN
SPG11
SPR
SPTAN1
SPTBN2
SRCAP
SRD5A3
SRPX2
ST3GAL3
ST3GAL5
STAG1
STAG2
STAMBP
STIL
STRADA
STS
STT3A
STT3B
STX1B
STXBP1
SUCLA2
SUCLG1
SUFU
SUMF1
SURF1
SYNE1
SYNGAP1
SYNJ1
SYP
SYT1
SZT2
TACO1
TACR3
TAF1
TAF13
TAF2
TANGO2
TARS1
TAT
TBC1D20
TBC1D23
TBC1D24
TBCD
TBCE
TBCK
TBL1XR1
TBR1
TBX1
TBX4
TCF20
TCF4
TCTN1
TCTN2
TCTN3
TECPR2
TELO2
TERT
TFAP2A
TGFB1
TGFBR2
TGIF1
TH
THOC2
THOC6
TINF2
TK2
TKT
TLK2
TMCO1
TMEM165
TMEM216
TMEM237
TMEM67
TMEM70
TMEM94
TMTC3
TOE1
TP53
TP53RK
TPM2
TPP1
TPRKB
TRAF7
TRAIP
TRAPPC11
TRAPPC12
TRAPPC9
TREX1
TRIM32
TRIO
TRIP12
TRIP13
TRIP4
TRIT1
TRMT1
TRPS1
TRPV4
TRPV6
TRRAP
TSEN15
TSEN2
TSEN34
TSEN54
TSHB
TSHR
TSPAN7
TTC8
TTI2
TTN
TUBA1A
TUBB
TUBB2B
TUBB3
TUBB4A
TUBGCP4
TUBGCP6
TUSC3
TWIST1
TWIST2
UBA5
UBE2A

UBE2T	*USP27X*	*WAC*	*WNT1*	*ZC4H2*
UBE3A	*USP7*	*WASF1*	*WWOX*	*ZDHHC9*
UBE3B	*USP8*	*WDFY3*	*XPA*	*ZEB2*
UBE4A	*USP9X*	*WDR11*	*XPC*	*ZIC1*
UBR7	*VAC14*	*WDR26*	*XRCC4*	*ZIC2*
UBTF	*VAMP1*	*WDR37*	*XYLT1*	*ZMIZ1*
UFC1	*VAMP2*	*WDR4*	*YWHAG*	*ZMYND11*
UFM1	*VDR*	*WDR45*	*YY1*	*ZNF142*
UGT1A1	*VLDLR*	*WDR45B*	*ZBTB11*	*ZNF462*
UNC80	*VPS13A*	*WDR62*	*ZBTB18*	*ZNF599*
UPF3B	*VPS13B*	*WDR73*	*ZBTB20*	*ZNF711*
UQCRQ	*VPS53*	*WDR81*	*ZC3H14*	*ZSWIM6*

ANNEX 2

Publications derived from the thesis

SCIENTIFIC REPORTS
nature research

OPEN

X chromosome inactivation does not necessarily determine the severity of the phenotype in Rett syndrome patients

Received: 7 March 2019
Accepted: 5 August 2019
Published online: 19 August 2019

Clara Xiol[1], Silvia Vidal[1], Ainhoa Pascual-Alonso[1], Laura Blasco[1], Núria Brandi[2], Paola Pacheco[1], Edgar Gerotina[1], Mar O'Callaghan[5], Mercè Pineda[3], Judith Armstrong[1,3,4] & Rett Working Group*

Rett syndrome (RTT) is a severe neurological disorder usually caused by mutations in the *MECP2* gene. Since the *MECP2* gene is located on the X chromosome, X chromosome inactivation (XCI) could play a role in the wide range of phenotypic variation of RTT patients; however, classical methylation-based protocols to evaluate XCI could not determine whether the preferentially inactivated X chromosome carried the mutant or the wild-type allele. Therefore, we developed an allele-specific methylation-based assay to evaluate methylation at the loci of several recurrent *MECP2* mutations. We analyzed the XCI patterns in the blood of 174 RTT patients, but we did not find a clear correlation between XCI and the clinical presentation. We also compared XCI in blood and brain cortex samples of two patients and found differences between XCI patterns in these tissues. However, RTT mainly being a neurological disease complicates the establishment of a correlation between the XCI in blood and the clinical presentation of the patients. Furthermore, we analyzed *MECP2* transcript levels and found differences from the expected levels according to XCI. Many factors other than XCI could affect the RTT phenotype, which in combination could influence the clinical presentation of RTT patients to a greater extent than slight variations in the XCI pattern.

Rett syndrome (RTT, OMIM #312750) is a severe neurodevelopmental disorder characterized by a period of normal development until 6-18 months of age followed by a regression of neurological traits. RTT features include compromised brain functions, severe mental retardation, epilepsy, regression of purposeful hand use and language, breathing disturbances, gait apraxia and repetitive stereotyped hand movements[1–3]. RTT has an incidence of 1:10,000–20,000 live births and affects mainly young females[4], being the second most common cause of severe mental retardation in females after Down syndrome.

The association of RTT with mutations in methyl-CpG binding protein 2 (*MECP2*; Xq28; OMIM *300005) gene was recognized in 1999[2]. Since then, more than 800 different mutations in *MECP2* have been identified in more than 95% of patients with classic RTT[5,6]. There are also some atypical RTT variants, such as the early onset seizure variant and the congenital variant, which have been associated with mutations in cyclin-dependent kinase-like 5 (*CDKL5*; Xp22; OMIM *300203) and forkhead box protein G1 (*FOXG1*; 14q12; OMIM *164874), respectively[7,8]. However, the vast majority of RTT patients have a *de novo* mutation in *MECP2*, and there are 8 mutation hotpots with recurrent mutations (p.Thr158Met, p.Arg255*, p.Arg168*, p.Arg306Cys, p.Arg294*, p.Arg270*, p.Arg133Cys and p.Arg106Trp), which are responsible for over 60% of all RTT cases[9,10].

Increasing experience has shown that RTT patients present a large degree of phenotypic variation[2]. Patients with truncating mutations in *MECP2* tend to show a more severe phenotype than those with missense mutations[4], and there are also phenotypical presentation differences between patients with the same mutation[11–13].

[1]Molecular and Genetics Medicine Section, Hospital Sant Joan de Déu, Barcelona, Spain. [2]Facultat de Medicina, Universitat de Barcelona, Barcelona, Spain. [3]Institut de Recerca Pediàtrica, Hospital Sant Joan de Déu, Barcelona, Spain. [4]CIBER-ER (Biomedical Network Research Center for Rare Diseases), Instituto de Salud Carlos III, Madrid, Spain. [5]Neurology Service, Hospital Sant Joan de Déu, Barcelona, Spain. *A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to J.A. (email: jarmstrong@sjdhospitalbarcelona.org)

Mutation	Type of mutation	MeCP2 region	Number of patients with skewed XCI	% of patients with skewed XCI
c.455C > G (p.P152R)	Missense	MBD	0/6	0%
c.473C > T (p.T158M)	Missense	MBD	0/33	0%
c.502C > T (p.R168X)	Nonsense	IDR	5/29	17.2%
c.674C > G (p.P255R)	Missense	TRD	0/2	0%
c.763C > T (p.R255X)	Nonsense	TRD	4/36	11.1%
c.806delG (p.G269fs)	Frameshift	TRD-NLS	1/11	9.1%
c.808C > T (p.R270X)	Nonsense	TRD-NLS	4/20	20%
c.880C > T (p.R294X)	Nonsense	TRD	1/20	5%
c.916C > T (p.R306C)	Missense	TRD	1/15	6.7%
Large deletions	Deletion	Exons 3-4	1/2	50%
All	—	—	17/174	9.8%

Table 1. Proportion of patients per mutation with a skewed XCI pattern according to at least one of the two techniques used for assessing XCI (XCI-AR and XCI-AS).

These clinical differences have been attributed, at least in part, to X chromosome inactivation (XCI). Through the XCI process, mammalian female cells inactivate one of the two X chromosomes to compensate for gene dosage. XCI is a stochastic process that takes place in the initial stages of the embryogenesis, causing a mosaic expression of X-linked genes in the adult organism[3,14,15]. Since *MECP2* is located on the X chromosome, the severity of RTT could be theoretically regulated by XCI, showing a more severe phenotype as more cells express the mutated *MECP2*[14].

Some cases of healthy carriers of RTT-causing mutations with highly skewed XCI patterns have been documented[14,16,17], as have cases of RTT patients with milder symptoms who also presented a skewed XCI pattern[13,17,18]. However, in most XCI studies in RTT, the phase of the two X chromosomes was not determined, so the XCI pattern could only be classified as either skewed or random. Therefore, no evidence of whether the preferentially inactivated chromosome was the mutant or the wild-type (WT) could be obtained.

We have developed an allele-specific methylation-based assay to evaluate methylation on the loci of several recurrent *MECP2* mutations, allowing for evaluation of the XCI pattern while taking into account which is the mutant and which is the wild-type allele. We compared the results from the classical androgen receptor assay for evaluating X chromosome inactivation (XCI-AR) with the allele-specific X chromosome inactivation (XCI-AS) assay we developed. We also compared all XCI results with a score of clinical severity of the clinical presentation of RTT to determine if we could correlate the XCI pattern with milder or more severe forms of RTT. Our cohort included 221 RTT patients with several recurrent mutations and two deletions in *MECP2*, for whom we could evaluate XCI patterns in blood. Moreover, we also assessed XCI in brain samples of two patients and compared the XCI status to blood to determine if it could be used as an accurate predictor. Finally, we measured *MECP2* RNA levels in brain samples to determine whether they correlated with the XCI pattern detected.

Results

Allele-specific X chromosome inactivation and XCI skewing in blood samples. For each patient, we performed an XCI-AR and the corresponding XCI-AS when blood samples were available (174/221 patients), and we also calculated the global score of the clinical presentation when clinical data were available (181/221 patients). The reference values for considering an XCI pattern as skewed in the literature are usually established at an 80:20 ratio[14,19], so we also used that threshold to allow the comparison of our results with previous studies. The entire list of XCI results and clinical scores for all patients can be found in Supplementary Table S1.

The overall tendency of our cohort was to have random XCI. However, 9.8% of our patients showed a skewed XCI pattern (80:20 or higher; Table 1), which is similar to what was found in other studies[13,20]. No patients with p. R152R, p.T158M or p. P225R mutations showed skewed XCI patterns in either XCI-AR or XCI-AS.

When we applied the 80:20 skewing threshold, 17 out of 174 patients presented a skewed XCI pattern according to at least one of the two XCI assays performed (Table 2). We compared these patients' clinical severity scores with the average clinical score of RTT patients with the same mutation. We found that, when the clinical score was available, in the majority of cases this value was included in the interval of $\mu \pm \sigma$ (central 68% of individuals in a normal distribution) of the patients with the same mutation.

There were only two patients who had a clinical score lower than the interval μ-σ for their mutation (P107 and P145, Table 2, in bold). In the case of patient P107, the preferentially inactivated allele was the WT allele, while in the case of patient P145 the mutant allele was inactivated. The results from patient P145 seem to be consistent with the theory that when the chromosome that harbors the *MECP2* mutation is preferentially inactivated, the clinical presentation of RTT may be milder.

Allele-specific X chromosome inactivation and XCI skewing in brain samples. We also performed XCI-AR and XCI-AS assays in samples of several brain regions of two patients with the c.763C > T mutation (Table 3). The XCI-AS assay was useful for assessing the XCI pattern in both patients, but especially in patient P119, since in this case, the polymorphism in the AR locus was noninformative for the XCI-AR assay.

Patient Number	XCI-AR	XCI-AS		Global Score
		WT	Mut	
Patients with c.502C > T (p.Arg168*) mutation		$\overline{X} = 13.12$ (SD = 3.361)		
P47	n.i.	81.5	18.5	13
P60	84:16	28	72	16
P68	75:25	15.5	84.5	NA
P70	85:15	35	65	NA
P74	81:19	55.5	44.5	NA
Patients with c.763C > T (p.Arg255*) mutation		$\overline{X} = 15.21$ (SD = 3.213)		
P83	85:15	57	43	NA
P84	87:13	55.5	44.5	13
P85	80:20	28	72	14
P107	87:13	68	32	11
Patients with c.806delG (p.Gly269fs) mutation		$\overline{X} = 14.29$ (SD = 4.112)		
P139	82:18	58	42	NA
Patients with c.808C > T (p.Arg270*) mutation		$\overline{X} = 14.69$ (SD = 3.846)		
P143	97:3	16	84	18
P144	84:16	21	79	NA
P145	81:19	30	70	9
P146	80:20	73	27	13
Patients with c.880C > T (p.Arg255*) mutation		$\overline{X} = 10.46$ (SD = 2.993)		
P191	89:11	49	51	NA
Patients with c.916C > T (p.Arg306Cys) mutation		$\overline{X} = 11.18$ (SD = 3.065)		
P195	89:11	59.5	40.5	9
Patients with deletions in *MECP2*				
P220	88:12	6.73	93.27	NA

Table 2. Data of patients with skewed XCI according to at least one of the two assays. The XCI-AR column shows the results of the AR XCI assay (percentage of inactivation of each allele). The XCI-AS WT and Mut columns show the results of the allele-specific XCI assay (percentage of inactivation of each allele, mean of two replicates n = 2 or three replicates n = 3 in the cases of the deletions). The Global Score column shows the average ($\overline{X}$) score and its standard deviation (SD) in brackets for the patients of our cohort with each mutation. Bold formatting indicates patients with a clinical score lower than the interval μ-σ for the average clinical score of their mutation. n.i. = polymorphism noninformative for the assay. NA = clinical data not available.

Although no samples showed skewed XCI by either assay, there was no clear homogeneity among blood and brain samples. Some samples, such as the frontal cortex or the white matter sample of patient P109, showed an XCI pattern closer to the skewing threshold than other regions, such as the cerebellum, of the same patient. In patient P119, the vast majority of samples were close to the random XCI pattern, but the temporal cortex sample showed an XCI pattern closer to the skewing threshold.

Brain RNA analysis. Finally, we analyzed frontal and occipital cortex RNA samples. We performed RT-PCR to obtain cDNA samples so that we could perform Sanger sequencing to check if we could detect the presence of one allele over the other (Fig. 1).

In cDNA samples from patient P109, the T allele (mutated allele) was overrepresented, while in samples from patient P119, the C allele (WT allele) was overrepresented. However, both patients presented a severe form of RTT, with clinical scores of 20 and 19, respectively.

The cDNA analysis was not conclusive since Sanger sequencing is not the best technique for quantifying the RNA of each allele. However, the sequencing analysis seemed to indicate that one allele was more frequently present than the other, although the XCI assay results showed inactivation patterns that did not reach the threshold for classifying the XCI pattern as skewed in any of the two patients and regions.

We later confirmed our findings in the frontal cortex samples by qRT-PCR, a more suitable technique for quantifying RNA levels (Fig. 2a,b). We found that in samples from patient P109, the mutated allele was overexpressed, while in samples from patients P119, the WT allele was overexpressed.

Discussion

The XCI-AS assay allowed us to describe the XCI patterns of patients previously classified as noninformative by the classical XCI-AR assay and to identify which *MECP2* allele (mutated or WT) was preferentially inactivated in cases of skewed XCI pattern.

Differences between the XCI patterns obtained by both techniques can be explained because in each technique, the methylation status is only analyzed at a single locus, and the methylation of a single cytosine residue may not be representative of the inactivation status of the entire X chromosome[21,22]. Different studies have shown

Sample	XCI-AR	XCI-AS	
		WT	Mut
Patient 109 (Clinical score = 20)			
Frontal Cortex	65:35	26	74
Occipital Cortex	58:42	59	41
Parietal Cortex	64:36	40	60
Temporal Cortex	60:40	32	68
White matter	59:41	23	77
Brain stem	59:41	31	69
Striatum	61:39	51	49
Cerebellum	55:45	43	57
Blood	73:27	64	36
Patient 119 (Clinical score = 19)			
Frontal Cortex	n.i.	48	52
Occipital Cortex	n.i.	NA	NA
Parietal Cortex	n.i.	56	44
Temporal Cortex	n.i.	73	27
White matter	n.i.	46	54
Brain stem	n.i.	38	62
Striatum	n.i.	50	50
Cerebellum	n.i.	50	50
Blood	n.i.	34	66

Table 3. Data of patients P109 and P119 with the c.763C > T mutation. The XCI-AR column shows the results of the AR XCI assay (percentage of inactivation of each allele). The XCI-AS WT and Mut columns show the results of the allele-specific XCI assay (percentage of inactivation of each allele). n.i. = polymorphism noninformative for the assay. NA = data not available.

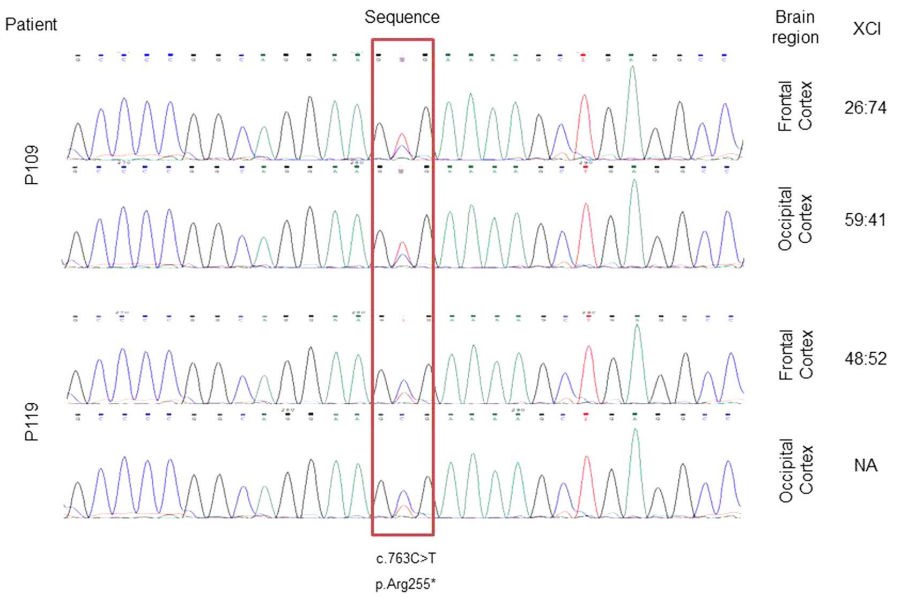
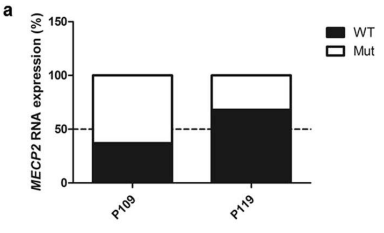


Figure 1. Brain RNA Sanger Sequencing. cDNA analysis of brain samples. Electropherograms obtained from Sanger sequencing of frontal and occipital cortex cDNA samples. Blue peaks correspond to the C allele (WT), while red peaks correspond to the T allele (mutated), and the red box highlights the locus of the c.763C > T mutation in heterozygosis. Inactivation ratios are shown as inactivation WT:inactivation Mut.

that when methylation in several loci of the X chromosome is assessed, different ratios of XCI can be obtained, with up to 27% of variation[21,22]. Therefore, the use of several loci for characterizing XCI would indicate the true XCI pattern more consistently[21].



b

Patient	% Active		% RNA expression	
	WT	Mut	WT	Mut
P109	74	26	37	63
P119	52	48	68	32

Figure 2. Brain RNA qRT-PCR analysis and comparison with XCI-AS assay results. (**a**) cDNA analysis of brain samples. The results obtained by qRT-PCR of frontal cortex RNA samples (% of expression of each allele). The discontinuous line indicates 50% of the expression of each allele (each allele is equally present in the sample). (**b**) Comparison of XCI and qRT-PCR data from patients P109 and P119 with the c.763C > T mutation. Data are shown as % of activation of each X chromosome (% Active) and % RNA expression measured by qRT-PCR.

Gathering data from both XCI assays performed with samples of 174 patients, we found that 9.8% of patients had skewed XCI patterns (80:20 XCI ratio or higher). Other studies have found either similar results[13,20] or a considerably higher incidence of skewing, up to 43%, among RTT patients[23]. Some authors claim that most of the patients who meet the diagnostic criteria for RTT have a random XCI pattern, while those with skewed XCI patterns may not meet all the criteria and therefore are not included in some RTT studies[18].

However, the percentage of patients in our cohort with skewed XCI patterns varied among different types of mutations. Mutations that produce a truncated protein result in a more severe phenotype than missense mutations[23], and skewed XCI patterns were more common in RTT patients with deletions and nonsense mutations than in those with missense mutations. This could be due to a protective effect related to the severity of the mutation. It is possible that mutations producing a less functional, truncated protein (deletions and nonsense mutations) cause cells to preferentially inactivate the X chromosome harboring the mutation. It has been shown that skewed XCI can be caused by a selective advantage of cells with a particular active X chromosome proliferating faster than cells where the other X chromosome is active[15,24,25]. This type of skewing has been described in up to 50% of familial cases of X-linked mental retardation disorders[26].

This skewed proliferation could be the case for patient P220 (Table 2), who had a large deletion in *MECP2* and showed a skewed XCI pattern (88:12) by the XCI-AR assay. In this patient, the XCI-AS assay confirmed an extremely skewed XCI pattern and that the preferentially inactivated allele was the mutated allele at a ratio of 93:7. We also found this tendency in several patients with p.Arg168* (P60, P68, P70; Table 2) p.Arg255* (P85; Table 2) and p.Arg270* (P143, P144, P145; Table 2) mutations. However, there were other patients with these same mutations with skewed XCI according to the XCI-AR assay who showed a preferential inactivation of the WT allele when the XCI-AS assay was performed, such as P146 (Table 2). Patient P47 (Table 2), who was noninformative for the XCI-AR assay, also showed a preferential inactivation of the WT allele at a ratio of 81:19 when the XCI-AS assay was performed. These last patients do not support the abovementioned hypothesis.

We found no substantial correlation between the XCI patterns in blood and the clinical presentation of RTT following the scale of evaluation of the RTT phenotype by Monrós, *et al.*[27] (data not shown). We did not observe consistent increases or decreases in the clinical score of RTT patients with a preferential inactivation of the WT or mutated alleles in blood samples.

It has been published that XCI patterns can vary among different tissues[22,28]. Indeed, we compared the XCI patterns of blood and brain samples of the same patient, and they did not show homogeneous XCI patterns. Although they were small, there was also a slight difference in the XCI patterns between different brain regions of the same patient.

Moreover, it has been shown that blood is especially prone to XCI skewing[29] because of the proliferation of different clones of lymphocytes under different conditions[22,29]. In fact, blood XCI patterns have shown variations at different time points in different studies[14]. For two of the patients included in the study (P9 and P199; Table S5), we compared two different blood samples from two different extractions. Both patients showed some differences in the results of the XCI assays in the two extraction samples.

The lack of a direct correlation between the XCI patterns in blood and the clinical presentation of RTT could be explained by different reasons. First, we observed that the XCI patterns in blood and different regions of the brain are not necessarily homogeneous. Therefore, if RTT symptoms are caused mainly by the lack of *MECP2* function in the brain, it is expected that the severity of the phenotype will be more related to the XCI pattern in the brain than to the XCI pattern in the blood.

Moreover, there are many other factors that can influence the presentation of the RTT phenotype, such as other polymorphisms and genetic variants, the expression levels of other genes and environmental conditions[4].

It is likely that the combination and addition of these additional factors can influence the phenotype to a greater extent than only the XCI pattern in the brain.

RTT symptoms arise from either a partial or a complete loss of function of *MECP2* in neurons[13,30]. RTT affects mainly females, partly because a complete loss of function of *MECP2* in males is so damaging that it can cause death in the first months of life or even before birth. The severity of the male phenotype points towards a dose-dependent mechanism of action of *MECP2*, where the expression of the mutant *MECP2* in a high proportion of cells causes the RTT phenotype[13,31]. It is possible that in females, slight deviations from random 50:50 XCI ratios do not cause sufficient changes in the levels of the mutant *MECP2* in the brain to be translated into a different phenotype.

However, it is possible that in more extreme cases, the effect is more remarkable. This could be similar to the case of female carriers of the *MECP2* duplication who show an extremely skewed XCI pattern with the mutant chromosome inactivated in most of their cells. In these cases, where a greater number of cells have inactivated the mutant chromosome, the effects of the XCI pattern are more important and cause the carrier of the *MECP2* duplication not to present the *MECP2* duplication syndrome. The same phenomenon could occur with pathogenic mutations in *MECP2*. If there is an extremely skewed XCI pattern in the brain, where a greater number of cells express the WT copy of *MECP2*, a threshold of *MECP2* function could be reached, and the RTT phenotype would therefore not be expressed. In some familial cases of RTT, it has been observed that a healthy mother with extremely skewed XCI can be a carrier of a pathogenic mutation responsible for causing RTT in her offspring[16–18], although she remains asymptomatic. Some authors have claimed that these familial cases of RTT are only possible due to the presence of two coincident traits: RTT and the trait for skewed XCI, which would be genetically determined[14,16].

The differences between the XCI patterns measured and the levels of each allele observed in Sanger sequencing and qRT-PCR could be due to RNA degradation, both in the postmortem interval and during life due to the nonsense-mediated mRNA decay (NMD) pathway, which could degrade mutant mRNA because of its potential to be translated into a truncated protein.

However, brain RNA levels of each allele seemed to show discrepancies with the XCI patterns identified in our XCI assays. Some authors have noticed discordances between the XCI pattern according to the XCI-AR assay and the quantification of the AR gene expression[32]. These discrepancies suggest, first, that the methylation assay may not always be representative of XCI and, second, that gene transcript levels may be regulated by more factors than XCI.

The difference between the XCI pattern and the final RNA levels of each allele suggests that the levels of *MECP2* are not directly determined by the XCI pattern and that there could be mechanisms other than XCI involved in regulating *MECP2* transcript levels. Consistent with what we have discussed, there might be other genes involved in regulating *MECP2* transcription and/or RNA degradation, causing changes in the final levels of functional *MECP2*[10]. Therefore, XCI may not necessarily be determining the severity of the clinical presentation of RTT, which would be more related to the levels of functional *MECP2* in the brain[30,31].

Nevertheless, it is important to keep in mind that we are measuring *MECP2* transcript levels from brain bulk RNA. Since different neuronal types have showed diverse transcriptional profiles in several studies[33], the levels of the *MECP2* transcripts we measured do not necessarily reflect these transcript's levels in neurons relevant for RTT pathophysiology.

Although one patient showed higher levels of the *MECP2* mutant transcript than the other, the clinical severity scores of both patients were not dissimilar (20 vs 19). This score similarity supports the hypothesis that slight deviations from a 1:1 ratio of each allele produce little to no change in the RTT phenotype. It is possible that more consistent differences would be noticeable if one allele was more prevalent than the other, such as in asymptomatic carriers with an XCI pattern close to the 100:0 ratio.

In conclusion, our results show that the relationship between XCI and the severity of the RTT phenotype is not straightforward. Factors other than XCI can influence *MECP2* transcript levels, and presumably many additional factors, such as genetic polymorphisms and the expression of other genes, may influence the final clinical presentation of RTT. Therefore, probably only extremely skewed XCI patterns affecting neurons can be correlated with milder forms of RTT or asymptomatic carriers.

Materials and Methods

Sample material. The study cohort consisted of 221 RTT patients with one of the 9 following recurrent mutations in the *MECP2* gene: c.455C > G-p.Pro152Arg (6 patients), c.473C > T-p.Thr158Met (36 patients), c.502C > T-p.Arg168* (38 patients), c.674C > G-p.Pro225Arg (2 patients), c.763C > T-p.Arg255* (47 patients), c.806delG-p.G269fs (13 patients), c.808C > T-p.Arg270* (31 patients), c.880C > T-p.Arg294* (21 patients) and c.916C > T-p.Arg306Cys (25 patients); and 2 patients with a large deletion in *MECP2*.

Samples of blood genomic DNA (gDNA) were obtained from peripheral blood leukocytes. Samples of brain gDNA were obtained postmortem from several brain regions (frontal, occipital, temporal and parietal cortex; white matter, brain stem, striatum and cerebellum) of two patients with c.763C > T mutation. RNA was also obtained from the frontal and occipital cortices of such patients. DNA samples were isolated using the saline extraction kit PUREGENE® DNA Isolation Kit of Gentra Systems®, and brain RNA samples were extracted using TRIzol™ Reagent from Invitrogen™.

Ethical approval and informed consent. 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-101-15). Written informed consent from the legal guardians of the patients was obtained in accordance with the corresponding ethical protocols to perform the genetic studies, and tissue samples from patients and controls were obtained according to the Helsinki Declaration of 1964, as revised in 2001[34].

HpaII and HinfI digestion. Digestion of gDNA samples was performed with one of the methylation-sensitive restriction enzymes *HpaII* or *HinfI* (New England BioLabs® Inc.), depending on the presence of the relevant enzyme target sequences near the studied loci. In the AR, c.455C > G, c.473C > T, c.502C > T, c.674C > G, c.763C > T, c.806delG, c.808C > T, c.880C > T, c.916C > T and deletion 2 (NM_004992.3: c.887_10015 + 18460del) loci assays *HpaII* was used, while in the deletion 1 (NM_004992.3: c.27-10677_1192del) locus assay *HinfI* was used. A total volume of 500 ng of gDNA was digested with 0.5 μL of enzyme in a 25 μL reaction volume in CutSmart 1x Buffer (New England Biolabs® Inc.). Digestions were incubated at 37 °C for 20 minutes followed by another 20 minutes at 80 °C for enzyme inactivation, as established in the enzyme protocol.

PCR amplification and fragment analysis. A pair of primers with the sequences described in Allen, *et al.*[35] was used to amplify the AR polymorphic locus. Allele-specific primers were designed for each *MECP2* recurrent mutation included in the study. Primer design was carried out following the recommendations in Liu, *et al.*[36]. For the deletion assays, a forward primer was designed inside the deletion locus and another primer immediately after the deletion; they were both amplified with a reverse primer outside the deleted region. All primers used were designed using Primer3web version 4.1.0[37,38], and they are shown together with PCR conditions for each pair in Supplementary Tables S2, S3 and S4. One primer of each pair was FAM-labeled at the 5′ end.

PCR amplification was performed using the resulting DNA after the digestion and nondigestion of each sample. PCR products were analyzed on a 3500 Genetic Analyzer (Applied Biosystems®) using GeneScan™ – 500 LIZ® Size Standard of Applied Biosystems® as an internal size standard and Peak Scanner Software v1.0. The X chromosome inactivation ratios were calculated as described elsewhere[35].

Brain RNA analysis. RT-PCR was performed with frontal and occipital cortex RNA of two patients with the c.763C > T mutation, following the recommendations provided with the SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR from Invitrogen™. Subsequently, Sanger sequencing of the cDNA obtained in the RT-PCR reaction was performed. qPCR was performed in a QuantStudio™ 6 Flex Real-Time PCR System (Applied Biosystems™) with TaqMan™ Gene Expression Master Mix (Applied Biosystems™) and specific TaqMan™ MGB probes to amplify the mutant and the wild-type alleles. qPCR data were analyzed using the comparative Ct method. Primers and probes were designed using Primer3web version 4.1.036,37, and they are listed in Supplementary Table S5.

Patient phenotype evaluation and correlation analysis. When clinical data were available (181/221 patients), the RTT phenotype was evaluated, and a score was assigned following the scale of evaluation of the RTT phenotype published by Monrós, *et al.*[27].

The linear correlation between the inactivation patterns of the WT allele and the global score of each patient was evaluated using statistical methods that are based on Ordinary Least Squares (OLS) regression models, grouping patients with the same mutation.

Data Availability

All data from this article is available in the Supplementary Data.

References

1. Neul, J. L. *et al.* Rett syndrome: Revised diagnostic criteria and nomenclature. *Ann. Neurol.* **68**, 944–950 (2010).
2. Weaving, L. S., Ellaway, C. J., Gécz, J. & Christodoulou, J. Rett syndrome: clinical review and genetic update. *J Med Genet* **42**, 1–7 (2005).
3. Ip, J. P. K., Mellios, N. & Sur, M. Rett syndrome: Insights into genetic, molecular and circuit mechanisms. *Nat. Rev. Neurosci.* **19**, 368–382 (2018).
4. Liyanage, V. R. B. & Rastegar, M. Rett syndrome and MeCP2. *NeuroMolecular Med.* **16**, 231–264 (2014).
5. Landucci, E. *et al.* iPSC-derived neurons profiling reveals GABAergic circuit disruption and acetylated α-tubulin defect which improves after iHDAC6 treatment in Rett syndrome. *Exp. Cell Res.* **368**, 225–235 (2018).
6. Vidal, S. *et al.* The utility of Next Generation Sequencing for molecular diagnostics in Rett syndrome. *Sci. Rep.* **7**, 1–11 (2017).
7. Weaving, L. S. *et al.* Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. *Am. J. Hum. Genet.* **75**, 1079–93 (2004).
8. Mencarelli, M. A. *et al.* Novel FOXG1 mutations associated with the congenital variant of Rett syndrome. *J. Med. Genet.* **47**, 49–53 (2010).
9. Percy, A. K. *et al.* Rett syndrome diagnostic criteria: Lessons from the Natural History Study. *Ann. Neurol.* **68**, 951–955 (2010).
10. Ehrhart, F. *et al.* Rett syndrome - Biological pathways leading from MECP2 to disorder phenotypes. *Orphanet J. Rare Dis.* **11**, 1–13 (2016).
11. Gonzales, M. L. & LaSalle, J. M. The role of MeCP2 in brain development and neurodevelopmental disorders. *Curr. Psychiatry Rep.* **12**, 127–134 (2010).
12. Hoffbuhr, K. C., Moses, L. M., Jerdonek, M. A., Naidu, S. & Hoffman, E. P. Associations between MeCP2 mutations, X-chromosome inactivation, and phenotype. *Ment. Retard. Dev. Disabil. Res. Rev.* **8**, 99–105 (2002).
13. Amir, R. E. *et al.* Influence of mutation type and X chromosome inactivation on Rett syndrome phenotypes. *Ann. Neurol.* **47**, 670–679 (2000).
14. Vacca, M., Della Ragione, F., Scalabrì, F. & D'Esposito, M. X inactivation and reactivation in X-linked diseases. *Semin. Cell Dev. Biol.* **56**, 78–87 (2016).
15. Gartler, S. M. & Goldman, M. A. X-Chromosome Inactivation. *Encycl. Life Sci.* 1–6, https://doi.org/10.1038/npg.els.0004172 (2001).
16. Sirianni, N., Naidu, S., Pereira, J., Pillotto, R. F. & Hoffman, E. P. Rett Syndrome: Confirmation of X-Linked Dominant Inheritance, and Localization of the Gene to Xq28. *Am. J. Hum. Genet* **63**, 1552–1558 (1998).
17. Zhang, Q. *et al.* Familial cases and male cases with MECP2 mutations. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **174**, 451–457 (2017).
18. 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).
19. Ørstavik, K. H. X chromosome inactivation in clinical practice. *Hum. Genet.* **126**, 363–373 (2009).

20. Van Den Veyver, I. B. & Zoghbi, H. Y. Mutations in the gene encoding methyl-CpG-binding protein 2 cause Rett syndrome. *Brain Dev.* **23**, 147–151 (2001).
21. Bertelsen, B., Tümer, Z. & Ravn, K. Three new loci for determining X chromosome inactivation patterns. *J. Mol. Diagnostics* **13**, 537–540 (2011).
22. De Hoon, B., Monkhorst, K., Riegman, P., Laven, J. S. E. & Gribnau, J. Buccal swab as a reliable predictor for X inactivation ratio in inaccessible tissues. *J. Med. Genet.* **52**, 784–790 (2015).
23. Weaving, L. S. *et al.* Effects ofMECP2 mutation type, location and X-inactivation in modulating Rett syndrome phenotype. *Am. J. Med. Genet.* **118A**, 103–114 (2003).
24. Clerc, P. & Avner, P. Random X-chromosome inactivation: skewing lessons for mice and men. *Curr. Opin. Genet. Dev.* **16**, 246–253 (2006).
25. Peeters, S. B., Yang, C. & Brown, C. J. Have humans lost control: The elusive X-controlling element. *Semin. Cell Dev. Biol.* **56**, 71–77 (2016).
26. Plenge, R. M., Stevenson, R. A., Lubs, H. A., Schwartz, C. E. & Willard, H. F. Report Skewed X-Chromosome Inactivation Is a Common Feature of X-Linked Mental Retardation Disorders. *Am. J. Hum. Genet* **71**, 168–173 (2002).
27. Monrós, E. *et al.* Rett syndrome in Spain: mutation analysis and clinical correlations. *Brain Dev.* **23**(Suppl 1), S251–S253 (2001).
28. Gibson, J. H., Williamson, S. L., Arbuckle, S. & Christodoulou, J. X chromosome inactivation patterns in brain in Rett syndrome: Implications for the disease phenotype. *Brain Dev.* **27**, 266–270 (2005).
29. Gale, R. E., Wheadon, H., Boulos, P. & Linch, D. C. Tissue specificity of X-chromosome inactivation patterns. *Blood* **83**, 2899–2905 (1994).
30. Leonard, H., Cobb, S. & Downs, J. Clinical and biological progress over 50 years in Rett syndrome. *Nat. Rev. Neurol.* **13**, 37–51 (2016).
31. Shahbazian, M. D., Sun, Y. & Zoghbi, H. Y. Balanced X chromosome inactivation patterns in the Rett syndrome brain. *Am. J. Med. Genet.* **111**, 164–168 (2002).
32. Swierczek, S. I. *et al.* Methylation of AR locus does not always reflect X chromosome inactivation state. *Blood* **119**, e100–e109 (2012).
33. Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-1590 (2016).
34. Carlson, R. V., Boyd, K. M. & Webb, D. J. The revision of the Declaration of Helsinki: Past, present and future. Vol. 57, British Journal of Clinical Pharmacology. p. 695–713 (2004).
35. Cutler Allen, R., Zoghbi, H. Y., Annemarie Moseley, I. B., Rosenblatt, H. M. & Belmont, J. W. Methylation of Hpall and Hhal Sites Near the Polymorphic CAG Repeat in the Human Androgen-Receptor Gene Correlates with X Chromosome Inactivation. *Am. J. Hum. Genet* **51**, 1229–1239 (1992).
36. Liu, J. *et al.* An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods* **8**, 1 (2012).
37. Untergasser, A. *et al.* Primer3-new capabilities and interfaces. *Nucleic Acids Res.* **40**, 1–12 (2012).
38. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291 (2007).

Acknowledgements

We thank all patients and their families who contributed to this study. The work was supported by grants from the Spanish Ministry of Health (Instituto de Salud Carlos III/FEDER, PI15/01159); Crowdfunding program PRECIPITA, from the Spanish Ministry of Health (Fundación Española para la Ciencia y la Tecnología); the Catalan Association for Rett Syndrome; Fondobiorett and Mi Princesa Rett.

Author Contributions

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

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-48385-w.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Consortia
Rett Working Group

Francisco Javier Aguirre[49], Montserrat Aleu[63], Xènia Alonso[5], Mercè Alsius[50], Maria Inmaculada Amorós[69], Guillermo Antiñolo[59], Lourdes Aquino[43], Carmen Arellano[12], Gema Arriola[23], Rosa Arteaga[54], Neus Baena[35], Montserrat Barcos[56], Nuria Belzunces[7], Susana Boronat[32], Tomás Camacho[29], Jaume Campistol[5], Miguel del Campo[32], Andrea Campo[59], Ramon Cancho[17], Ramon Candau[59], Ignacio Canós[53], María del Carmen Carrascosa[20], Francisco Carratalá-Marco[51], Jovaní Casano[42], Pedro Castro[46], Ana Cobo[73], Jaime Colomer[5], David Conejo[10], Maria José Corrales[19], Rocío Cortés[40], Gabriel Cruz[24], Gábor Csányi[45], María Teresa de Santos[13], María de Toledo[57], Miguel Del Campo[32], Mireia Del Toro[32], Rosario Domingo[38], Anna Duat[34], Rosario Duque[55], Ana María Esparza[42], Rosa Fernández[22], Maria Carme Fons[5], Ana Fontalba[54], Enrique Galán[37], Pia Gallano[45], María José Gamundi[12], Pedro Luis García[26], María del Mar García[74], María García-Barcina[71], María Jesús Garcia-Catalan[5], Ángels García-Cazorla[5], Sixto García-Miñaur[36], Juan Jose Garcia-Peñas[34], María Teresa García-Silva[52], Rosa Gassio[5], Esther Geán[5], Belén Gil[22], Sarenur Gökben[62], Luis Gonzalez[34], Veronica Gonzalez[5], Julieta Gonzalez[5], Gloria González[7], Encarna Guillén[38], Miriam Guitart[35], Montserrat Guitet[16], Juan Manuel Gutierrez[67], Eva Gutiérrez[13], Jose Luís Herranz[54], Gemma Iglesias[25], Iva Karacic[9], Carlos H. Lahoz[14], José Ignacio Lao[28], Pablo Lapunzina[36], María Jesús Lautre-Ecenarro[41], María Dolores Lluch[60], Laura López[34], Asunción López-Ariztegui[72], Alfons Macaya[32], Rosario Marín[39], Charles M. Lourenço Marquez[66], Elena Martín[52], Beatriz Martínez[22], Eduardo Martínez-Salcedo[21], María José Mas[47], Gonzalo Mateo[23], Pilar Mendez[37], Amparo Morant Jimenez[33], Sira Moreno[27], Fernando Mulas[64], Juan Narbona[65], Andrés Nascimento[5], Manuel Nieto[59], Tania Fabiola Nunes[5], Núria Núñez[32], María Obón[50], Ignacio Onsurbe[20], Carlos Ignacio Ortez[5], Emilio Orts[19], Francisco Martinez[53], Rafael Parrilla[61], Samuel Ignacio Pascual[36], Ana Patiño[44], Maria Pérez-Poyato[5], Belén Pérez-Dueñas[5], Pilar Póo[5], Eliodoro Puche[38], Feliciano Ramos[68], Miquel Raspall[32], Ana Roche[5], Susana Roldan[58], Jordi Rosell[48], Cesar Ruiz[70], María Luz Ruiz-Falcó[34], Maria Eugenia Russi[5], Jordi Samarra[18], Victoria San Antonio[41], Ivan Sanchez[5], Xavier Sanmartin[5], Ana Sans[5], Alfredo Santacana[11], Sabine Scholl-Bürgi[30], Nuria Serrano[6], Mercedes Serrano[5], Pilar Martin-Tamayo[20], Adrián Tendero[36], Jaime Torrents[31], Diego Tortosa[38], Emma Triviño[8], Ledia Troncoso[40], Eulàlia Turón[45], Pilar Vázquez[46], Carlos Vázquez[11], Ramón Velázquez[36], Clara Ventura[32], Alfonso Verdú[26], Anna Vernet[5], M. Tomás Vila[15] & Cristina Villar[5]

[6]Althaia, Manresa, Spain. [7]Balagué Center, Barcelona, Spain. [8]Catlab, Barcelona, Spain. [9]Clinical Hospital Center Zagreb, Zagreb, Croatia. [10]Complejo asistencial, Burgos, Spain. [11]Complejo Hospitalario Universitario Insular, Las Palmas de Gran Canaria, Spain. [12]Consorci Sanitari, Terrassa, Spain. [13]Hospital de Fuenlabrada, Madrid, Spain. [14]Hospital Central Asturias, Asturias, Spain. [15]Hospital Francesc De Borja, Valencia, Spain. [16]Hospital General de Granollers, Barcelona, Spain. [17]Hospital Universitario Río Hortega, Valladolid, Spain. [18]Hospital General de Vic, Barcelona, Spain. [19]Hospital General Mancha Centro, Ciudad Real, Spain. [20]Hospital General Universitario de Albacete, Albacete, Spain. [21]Hospital General Universitario de Alicante, Alicante, Spain. [22]Hospital Universitario de Getafe, Getafe, Spain. [23]Hospital Universitario de Guadalajara, Guadalajara, Spain. [24]Hospital Universitario de Valme, Sevilla, Spain. [25]Hospital Virgen de la Luz, Cuenca, Spain. [26]Hospital Virgen de la Salud, Toledo, Spain. [27]Hospital Virgen del Camino, Pamplona, Spain. [28]Laboratorio Echevarne, Barcelona, Spain. [29]Lema & Bandin Laboratorios, Vigo, Spain. [30]Medizinische Universität Innsbruck, Innsbruck, Austria. [31]Reference Laboratory, Barcelona, Spain. [32]Hospital de la Vall d'Hebrón, Barcelona, Spain. [33]Centro privado, Valencia, Spain. [34]Hospital Infantil Universitario Niño Jesús, Madrid, Spain. [35]Hospital de Sabadell, Barcelona, Spain. [36]Hospital La Paz, Madrid, Spain. [37]Hospital Materno-Infantil de Badajoz, Badajoz, Spain. [38]Hospital Infantil de La Arrixaca, Murcia, Spain. [39]Hospital Universitario Puerta del Mar, Cádiz, Spain. [40]Hospital San Borja Arriaran, Santiago, Chile. [41]Hospital Clínico San Carlos, Madrid, Spain. [42]Hospital Universitario General de Castellón, Castellón, Spain. [43]Hospital de Mataró, Mataró, Spain. [44]Hospital de Navarra, Pamplona, Spain. [45]Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. [46]Hospital Gregorio Marañón, Madrid, Spain. [47]Hospital Joan XXIII, Tarragona, Spain. [48]Hospital Son Dureta, Palma de Mallorca, Palma, Spain. [49]Hospital Torrecardenas, Almería, Spain. [50]Hospital Universitari Dr. Josep Trueta, Girona, Spain. [51]Hospital Universitari San Juan, Alicante, Spain. [52]Hospital Universitario 12 de Octubre, Madrid, Spain. [53]Hospital Universitario Doctor Peset, Valencia, Spain. [54]Hospital Universitario Marqués de Valdecilla, Santander, Spain. [55]Hospital Universitario Nuestra Señora de la Candelaria, Santa Cruz de Tenerife, Spain. [56]Hospital Universitario Reina Sofía, Córdoba, Spain. [57]Hospital Universitario Severo Ochoa, Madrid, Spain. [58]Hospital Universitario Virgen de las Nieves, Granada, Spain. [59]Hospital Universitario Virgen del Rocio, Sevilla, Spain. [60]Hospital Universitario Virgen Macarena, Sevilla, Spain. [61]Complejo Hospitalario de Jaén, Jaén, Spain. [62]Ege Ünŭversŭtesŭ Tip Fakültesŭ Pedŭatrŭ AD, Ŭzmŭr, Turkey. [63]Consorcio Hospital General Universitario de Valencia, Valencia, Spain. [64]Instituto Valenciano de Neurociencias, Valencia, Spain. [65]Clínica Universitaria de Pamplona, Pamplona, Spain. [66]Medical Genetics Service, Clinics Hospital of Ribeirão Preto, University of São Paulo, São Paulo, Brazil. [67]Hospital Clínico Universitario de Valladolid, Valladolid, Spain. [68]Hospital Clínico Universitario Lozano Blesa, Zaragoza, Spain. [69]Hospital Punta Europa, Cádiz, Spain. [70]Hospital Costa del Sol, Málaga, Spain. [71]Hospital de Basurto, Bilbao, Spain. [72]Hospital de Cruces, Bilbao, Spain. [73]Hospital de Donostia, San Sebastián, Spain. [74]Hospital Cormarcal de Figueres, Girona, Spain.

International Journal of
Molecular Sciences

MDPI

Article

Comprehensive Analysis of GABA$_A$-A1R Developmental Alterations in Rett Syndrome: Setting the Focus for Therapeutic Targets in the Time Frame of the Disease

Alfonso Oyarzabal [1,*], Clara Xiol [2,†], Alba Aina Castells [2,†], Cristina Grau [1], Mar O'Callaghan [3], Guerau Fernández [2], Soledad Alcántara [4], Mercè Pineda [3], Judith Armstrong [2], Xavier Altafaj [5,‡], and Angels García-Cazorla [1,‡]

[1] Synaptic Metabolism Lab, Neurology Department, Institut Pediàtric de Recerca, Hospital Sant Joan de Déu and CIBERER, 08950 Barcelona, Spain; cgrau@fsjd.org (C.G.); agarcia@sjdhospitalbarcelona.org (A.G.-C.)
[2] Genetics Department, Institut Pediàtric de Recerca Hospital Sant Joan de Déu, 08950 Barcelona, Spain; cxiol@sjdhospitalbarcelona.org (C.X.); acastells@fsjd.org (A.A.C.); gfernandezi@sjdhospitalbarcelona.org (G.F.); jarmstrong@sjdhospitalbarcelona.org (J.A.)
[3] Neurology Department, Institut Pediàtric de Recerca Hospital Sant Joan de Déu, and CIBERER, 08950 Barcelona, Spain; mocallaghan@sjdhospitalbarcelona.org (M.O.); pineda@sjdhospitalbarcelona.org (M.P.)
[4] Neural Development Lab, Departament de Patologia i Terapèutica Experimental, Institut de Neurociènces, Universitat de Barcelona, IDIBELL, 08950 Barcelona, Spain; salcantara@ub.edu
[5] Bellvitge Biomedical Research Institute, Neuropharmacology and Pain Unit, University of Barcelona, 08950 Barcelona, Spain; xaltafaj@idibell.cat
* Correspondence: aldeoyarzabal@fsjd.org
† These authors contributed equally to this work.
‡ These authors contributed equally to this work.

Received: 30 November 2019; Accepted: 10 January 2020; Published: 14 January 2020

Abstract: Rett syndrome, a serious neurodevelopmental disorder, has been associated with an altered expression of different synaptic-related proteins and aberrant glutamatergic and γ-aminobutyric acid (GABA)ergic neurotransmission. Despite its severity, it lacks a therapeutic option. Through this work we aimed to define the relationship between MeCP2 and GABAA.-A1 receptor expression, emphasizing the time dependence of such relationship. For this, we analyzed the expression of the ionotropic receptor subunit in different MeCP2 gene-dosage and developmental conditions, in cells lines, and in primary cultured neurons, as well as in different developmental stages of a Rett mouse model. Further, RNAseq and systems biology analysis was performed from post-mortem brain biopsies of Rett patients. We observed that the modulation of the MeCP2 expression in cellular models (both Neuro2a (N2A) cells and primary neuronal cultures) revealed a MeCP2 positive effect on the GABAA.-A1 receptor subunit expression, which did not occur in other proteins such as KCC2 (Potassium-chloride channel, member 5). In the Mecp2+/− mouse brain, both the KCC2 and GABA subunits expression were developmentally regulated, with a decreased expression during the pre-symptomatic stage, while the expression was variable in the adult symptomatic mice. Finally, the expression of the gamma-aminobutyric acid (GABA) receptor-related synaptic proteins from the postmortem brain biopsies of two Rett patients was evaluated, specifically revealing the GABA A1R subunit overexpression. The identification of the molecular changes along with the Rett syndrome prodromic stages strongly endorses the importance of time frame when addressing this disease, supporting the need for a neurotransmission-targeted early therapeutic intervention.

Keywords: Rett syndrome; GABA; autism; GABA-A1R; KCC2; RNAseq

1. Introduction

Rett syndrome (RTT; OMIM #312750) is a severe neurodevelopmental disorder characterized by a regression in the neurological development between 6 and 18 months following a normal early development. Patients experience seizures, autistic features, apnea/hyperpnoea and a loss of all the acquired capabilities, speech and non-verbal communication capacity, stereotypes, loss of purposeful use of hands, and organic dysfunctions [1,2].

The complexity of Rett syndrome derives from the MeCP2 protein function, coded by *MECP2* (Xq28; MIM* 300005), as most reported cases are associated with its defective activity. MeCP2 is a nuclear protein that acts as an epigenetic regulator, controlling the expression of numerous genes (either as transcription activators or repressors) involved in several biological processes [3]. Whilst it is a ubiquitous protein, MeCP2 is most highly expressed in the brain [2,4], most precisely in post mitotic neurons [5,6], and its deficiency results in a global neurodevelopment disturbance [7]. Neurochemically, Rett syndrome has been associated with an aberrant expression of neurotransmitters, neuromodulators, transporters, and receptors [8–11]. Collectively, these alterations may underlie an unbalanced excitatory/inhibitory neurotransmission together with a disturbed synaptic development associated with Rett syndrome [12,13]. In particular, an unbalanced excitatory/inhibitory neurotransmission stands out, with a specific γ-aminobutyric acid (GABA)ergic malfunction. GABA (γ-aminobutyric acid) is the major inhibitory neurotransmitter in the brain [14]. The fast inhibitory actions of GABA are mediated by the GABA(A) receptors, which are ligand-gated chloride (Cl-) channels consisting on assemblies of five different subunits from eight possible subfamilies [15], the 2α1 + 2β2 + 1γ2 conformation being the most prevalent, accounting for 43% of the total $GABA_A$ receptors [16], present in most brain areas. The selective transport of Cl^- when the $GABA_A$ receptors are activated hyperpolarizes the neuron, reducing its likelihood of starting an action potential [17].

GABAergic synapses dysfunction has been associated with several Rett features. This altered performance, nevertheless, seems to be region and developmental-stage dependent. In fact, studies in MeCP2-/y mice brain slices show reduced miniature excitatory postsynaptic currents in the somatosensory cortical neurons, together with unaltered miniature inhibitory postsynaptic currents, which result in an overall reduced excitation. Opposite to that, there is a reduced conductance but increased excitatory/inhibitory ratio in the CA1 and CA3 areas of the hippocampus and V1 pyramidal neurons in in vivo visually evoked responses. Many players appear to be participating in this GABAergic neurotransmission alteration, from GABA receptors [18–20] to the chloride channels NKCC1 and KCC2 [13,21], responsible for the excitatory to inhibitory switch of GABAergic synapses during development, and for which the expression has been found to be altered in Rett patients' cerebrospinal fluid (CSF).

In agreement with this, the restoration of correct GABAergic neurotransmission partially rescued Rett-like phenotypic abnormalities in mouse models [22], supporting the GABAergic pathway pivotal role in Rett's pathophysiology, and opening a window for the treatment-expectancy of the disease [23].

Regarding the need for a treatment for the disease, and proving its mentioned potential reversibility, neurotransmission modulation appears to be an attractive therapeutic approach. However, enlightened by very last reports, the question of "when" adds on to the "how" for treating the disease. Given the severity of Rett syndrome and the lack of therapeutic options, there is an urge for the definition of the molecular alterations during development that set the bases for the window travel to address novel therapeutic targets, as intended throughout this work.

Throughout this work, we aimed to define if there was a direct relationship between GABAergic synapses known to alter elements, and MeCP2. Our results show a direct relationship between MeCP2 and GABA ionotropic receptors' expression, which is not extensible to other GABAergic proteins such as KCC2, altered in the context of MeCP2 dysfunction, but not appearing to be directly regulated by its activity. More important than this, our work points attention towards the importance of the time frame when addressing Rett syndrome, as changes appear to be time-dependent, with greater importance for the pre-symptomatic stages.

2. Results

2.1. MeCP2 Levels Are Associated with GABA$_A$ Receptors' Expression in Cellular Models

Fast GABAergic neurotransmission is mediated by GABA ionotropic receptors (GABA$_A$). These are ligand-gated chloride (Cl^-) channels consisting on five subunits of eight subfamilies [15]. Mechanistically, GABA$_A$ receptors' activation allows for a selective Cl^- influx, triggering a hyperpolarization of the postsynaptic neuron that reduces its likelihood of starting an action potential [17]. Among the multiple stoichiometric combinations of the heteromeric GABA$_A$ receptors, α1-β2-γ2 is the major molecular combination, with the α1 subunit being present in over 60% of GABA$_A$ heteromers, being widely expressed in brain areas.

Based on previous reports supporting the contribution of GABAergic dysfunction in Rett syndrome progression and the alteration of the GABA$_A$-A1 expression, we hypothesized that MeCP2 disturbance might directly affect the density of GABA$_A$ receptors, rather than such an altered expression being a secondary effect of an overall GABAergic dysfunction. To this end, we studied its expression pattern over two days of Mecp2 knockdown in cellular models.

On one side, we interfered mMeCP2 (mouse *MECP2*) expression in Neuro2a (N2A) mouse neuroblastoma cells by the transient transfection of a shRNA-anti-3'UTR Mecp2 plasmid (from now on, mentioned solely as shRNA). Prior to the GABA$_A$-A1R assessment, the efficiency and specificity of the *MECP2* expression interference was evaluated. For this, the total RNA was extracted from the cells under the four scenarios analyzed, that is, un-transfected N2A cells, cells transfected with the shRNA vector or co-transfected with the shRNA vector, and the *hMECP2* gene in either its wild type (wt) or mutated versions. Interestingly, the short hairpin sequence was designed to specifically target mouse Mecp2, while the human MECP2 (hMeCP2) was predicted to be insensitive to shRNA-anti-(3'UTR)Mecp2. The system was proven to be reliable, as the endogenous *MECP2* levels were drastically reduced upon the transfection of N2A cells with the shRNA vector, while the exogenous expression of the human isoform, however, was not affected by the shRNA co-expression. As shown in Figure 1A, the amplification with specific *hMECP2* primers revealed an increased expression only upon transfection with the *hMECP2* wt vector.

While the specificity of the shRNA interference has been shown, the exclusivity of the hMeCP2 cannot be ensured, as endogenous mMeCP2 is amplified in the un-transfected N2A cells using hMeCP2 primers (Figure 1A, right panel). We believe we are detecting undesired amplification due to cross-binding between the mouse and human forms. Human and mouse MeCP2 are identical in 87.41% of their RNA sequence. While the forward primer binds to a sequence similar in 95% of both forms (19 out of 20 bases), the reverse primer was designed to only bind to the human transcript, which, however, seems not to have been properly achieved.

Under basal conditions, the transcript encoding for the major GABA$_A$ receptors subunit (*Gabra1*) was not detectable in N2A cells. Remarkably, the target was amplified upon transfection with the plasmid carrying the *wt* form of *hMECP2* (Figure 1B), supporting a relationship between MeCP2 expression and the gene over-expression. Noteworthy, the GABA$_A$ receptor subunit expression was undetectable upon transfection with the plasmid containing *hMECP2*_c.763C>T (Figure 1B). The same happened with the remaining principal GABA$_A$ subunits *Gabrb2* and *Gabrg2* (data not shown). These results point towards a direct regulation of GABA$_A$-A1R expression by MeCP2 during, at least, a certain time frame of development.

Interestingly, with the KCC2 expression, which has been reported to be affected in Rett syndrome models, patients did not appear to be affected by MeCP2 inhibition, as can be shown in Figure 1B, suggesting that both alterations affecting the same synapse occur through different mechanisms.

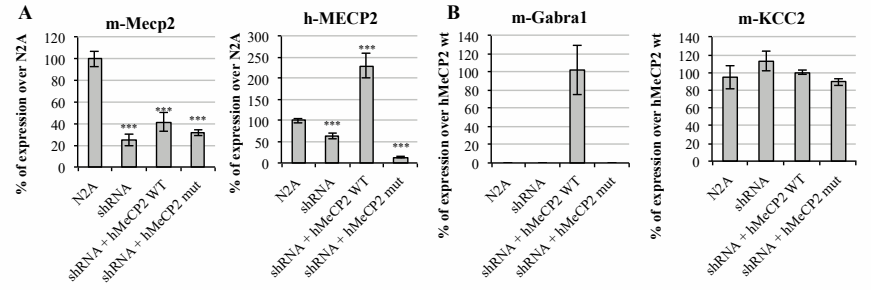


Figure 1. In vitro analysis of MeCP2 altered activity over γ-aminobutyric acid (GABA) ionotropic receptors' expression. Bar graph representing the relative expression of (**A**) mouse MECP2 (mMeCP2), human MECP2 (hMeCP2) and (**B**) mouse GABRA1, and mouse KCC2, measured by qRT-PCR under four different transfection scenarios, namely: non-transfected cells (N2A), transfected with the shRNA-anti-(3'UTR)Mecp2 (shRNA) and co-transfected with the shRNA-anti-(3'UTR)Mecp2 and *wt*, or c.763C>T mutated MeCP2 carrying plasmids. Error bars represent the standard deviation of the average values. Statistical significance was calculated through a Student's t-test (*** p-value < 0.001).

In order to evaluate the identified MeCP2 positive regulatory effect on the $GABA_A$ receptor expression in the dendritic processes, the primary cultures of murine cortical neurons were established. Following shRNA-mediated MeCP2 silencing, the expression of the *GABRA1* subunit was assessed.

The primary cortical neurons were transiently co-transfected at day *in vitro* 7 (DIV7) with either a shRNA-anti-(3'UTR)Mecp2 or a mock plasmid, together with a pcDNA-EGFP vector in a 1:7 ratio, allowing for the identification of shRNA-transfected neurons. An immunofluorescence analysis was performed in the early mature primary neuronal cultures (DIV11), showing a complete lack of Mecp2 detection in shRNA-(GFP-positive) transfected neurons (Figure 2A), and thus validating the inhibition system.

When we next evaluated the $GABA_A$-A1R expression on the silenced neurons, we observed a severe decrease in the detection in the GFP-positive neurons, compared with the cultures transfected with GFP and mock vector (Figure 2B), considered as the control conditions. The mean detection on the GABRA1 marking was 37% compared with the control conditions. These results are complementary to the N2A cells observations, as both support a positive and straightforward relationship between MeCP2 and $GABA_A$-A1 receptor expression.

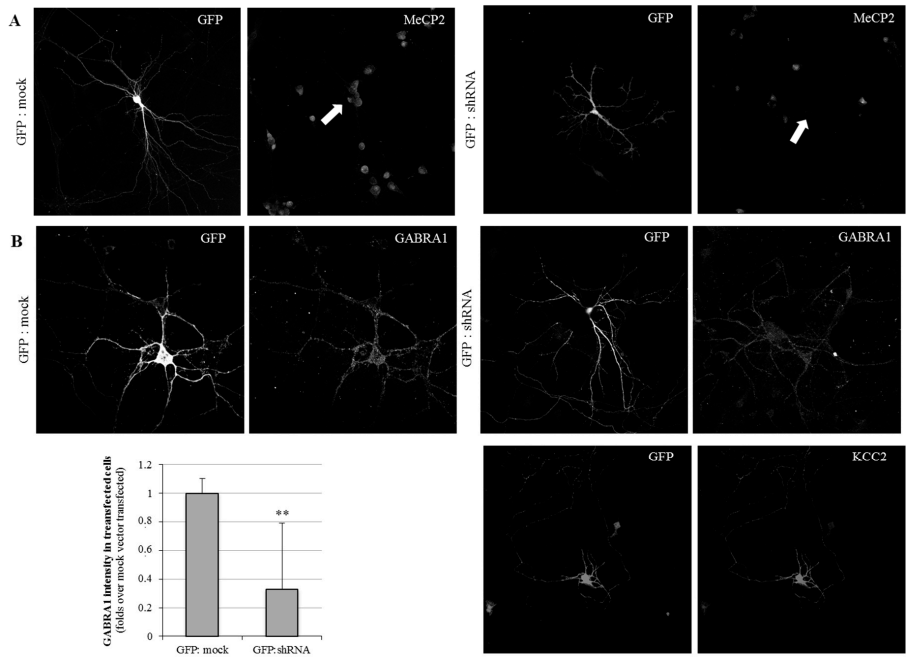


Figure 2. Immunofluorescence analysis of MeCP2, GABRA1, and KCC2 expression in cortical primary neuronal cultures. Cells were transfected with either GFP and mock DNA, or GFP and shRNA-anti-(3'UTR)Mecp2. Images show neurons at DIV11**.** Images were taken at 63x with constant time of exposure. Transfected neurons were labeled with anti-GFP and anti-Mecp2 (**A**). For the GABRA1 and KCC2 immunostaining (**B**), different neurons are shown. The quantification of the mean GABRA1 immunosignal in mock and shRNA transfection conditions is shown in the bar graph. n = 10 different neurons, from two independent cultures. ** refer to *p*-value < 0.01.

2.2. Neurodevelopmental Changes of GABA$_A$-A1R and KCC2 in a Rett Syndrome Mouse Model Point Towards the Importance of Pre-Symptomatic Versus Symptomatic Manifestations

As Rett syndrome is a neurodevelopmental disorder, we aimed to appraise the studied changes in different evolutionary stages.

Thus, we evaluated the GABA$_A$ A1 subunit expression and KCC2 in the cortex of mice at pre-sympromatic and symptomatic stages (one and six months, respectively). As aforementioned, we selected these two proteins for their proven implication in Rett syndrome pathophysiology. We used female Rett mice models from the Bird strain, which recapitulate Rett-like abnormalities [24]. Western blot analysis revealed a significant decrease of the ionotropic GABA receptor subunit at a young stage (one month-old), while no significant differences were detected between the genotypes (Figure 2) at the symptomatic stage (six months old).

To contextualize this difference between the prodromic stages and considering whether it was a GABA$_A$ A1 receptor subunit specific variation or a generalized GABAergic synapse downregulation, we analyzed two other proteins expression in the same conditions, namely: KCC2 and GAD67. The first one, KCC2, is a neuron-specific chloride potassium symporter responsible for the maintenance of intracellular chloride concentrations. On the other side, GAD67 or glutamate decarboxylase is the enzyme that catalyzes the decarboxylation of glutamate to GABA and CO^2, and is widely used as a GABAergic marker. As can be seen in Figure 2B, a decreased expression of both markers was

observed in the pre-symptomatic stage, suggesting a generalized decrease of the GABAergic function.

Parallel to this, and as a control, the MeCP2 expression was assessed in the same samples and appeared to be lower in all of the Rett samples (Figure 3).

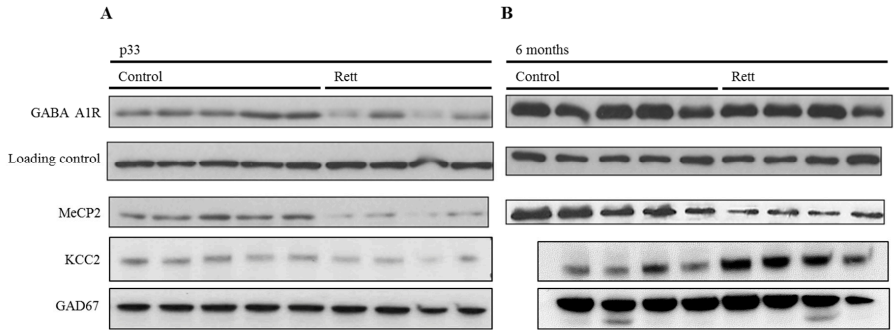


Figure 3. Developmental expression analysis of GABAergic proteins in the MeCP2$^{-/+}$ mouse brain cortex. Representative Western blot analysis of the GABA$_A$-A1R, KCC2, GAD67, and MeCP2 E1 expression in the adult cortex of female Rett and control mice. (**A**) Expression in p33-pre-symptomatic mice (control vs. pre-symptomatic Rett mice). Vinculin was used as a loading control. (**A**) Expression in six-months old mice (control vs. symptomatic Rett mice). Both Western blots (**A,B**) are cropped stripes of two different membranes each, and incubated with each antibody separately (target protein and loading control respectively). Each lane is a different animal. Representative blots shown.

2.3. Transcriptomic Profile of the GABAergic Pathway in Post-Mortem Brain of Rett Syndrome Patients Shows a Generalized GABAergic Pathway Upregulation

The above-mentioned results allowed for the definition of the MeCP2-related expression of GABA$_A$-A1 R, highlighting the difference between pre-symptomatic and phenotypic stages. Furthermore, we evaluated the expression of the 108 genes shaping the GABAergic synaptic pathway, as defined in the Kegg pathways (Kegg pathway: map04727).

Upon the validation of the RNAseq experiment, as described in Materials and Methods, and because of the lack of healthy control samples, the raw data from the Rett patients' necropsies were compared with the RNAseq data of the control individuals, available at GTEx Portal [25]. We crossed our data with the data derived from five controls (two females and three males), all aged 20 to 29, being the closest age-matched group to our patients (10 to 15 years old at exitus). All of the data were normalized with housekeeping genes *RPLP0* and *GUSB*.

The transcripts encoding for the major GABA$_A$ receptor subunits, GABRA1, and the chloride transporter KCC2, were slightly increased in the patients' samples compared with the controls (Figure 4). Similar as expected from the Western blot studies, the GAD67 levels remained unaltered compared with the controls.

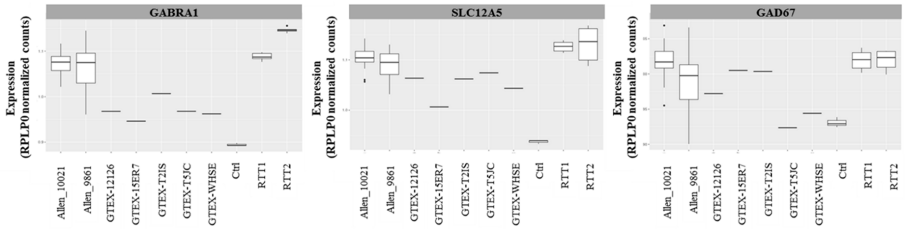


Figure 4. GABAergic pathways genes expression assessment in brain patients by RNAseq. Comparison of the GABA ionotropic receptors subunits *GABRA1,* and KCC2 and GAD67 genes between the patients (RTT 1 and RTT 2), the intra-assay control (Ctrl), and publicly available controls, showing a tendency to increase in the Rett patients. The results were normalized by *RPLP0.*

3. Discussion

The present work arises from the need to deepen into the definition of Rett's syndrome pathophysiology in order to define new therapeutic strategies. To such end, we have explored the GABAergic neurotransmission system in different evolutionary stages of the disease, setting the focus on the main GABA ionotropic receptor, $GABA_A$-A1R. We have observed a direct relationship between the MeCP2 altered expression and GABAergic receptors disruption, which is strongly dependent on the prodromic stage of the disease, angling the focus towards the time frame, which will be a key factor when looking for therapeutic options.

The triangle conformed by Rett syndrome, MeCP2, and GABAergic synapses has been previously explored by other groups [9,12,26]. An increased MeCP2 expression in GABAergic neurons has been reported, and a reduced GABA release was reported upon MeCP2 knocking out in forebrain GABAergic neurons [27]. Such a relationship has been shown to be extensive to other neuron types and brain areas, such as CA3 hippocampal neurons or brainstems [12]. Moreover, in 2016, Dr. Zoghbi's team showed how the restoration of the MeCP2 expression exclusively in GABAergic neurons was sufficient to rescue some disease features in a mouse model of Rett syndrome [28]. Even a time-dependent alteration was suggested in male mice by the authors of [26]. Through our work, we have followed such a line of thought, wondering how the alterations were occurring during development, setting the focus on the pre-symptomatic stages.

We first wondered if there was a direct relationship between MeCP2 and the postsynaptic $GABA_A$-A1R expression. Previous studies pointing out a relationship between GABA receptors and MeCP2 had been done in the context of whole brain models' analysis, so it remained unclear if the potential decrease in GABA receptors was strictly related to MeCP2 expression or was a secondary effect to a global dysfunction. To address that question, we performed two complementary experiments. On first term, we overexpressed MeCP2 in a cell system that almost did not express GABA ionotropic receptors (or its expression was mostly below our detection sensitive), namely N2A cells. To increase the assay restricted conditions, we silenced any potential endogenous MeCP2 expression, and overexpressed the human form in either its wild-type or mutated forms. As expected, the GABA ionotropic receptors were only detected upon transfection with the MeCP2 *wt* form, not occurring when transfected with the mutated version. These results were complemented with the peer experiment in primary cortical neurons; this time, we imaged a decrease in $GABA_A$-A1R expression four days after silencing the MeCP2 expression. Our results strongly suggest a direct relationship between MeCP2 and post-synaptic GABA ionotropic receptors' expression, rather than this being a secondary effect of an overall altered homeostasis. Opposite to that, we found that the KCC2 expression was unaltered in the presented scenario. KCC2 is a chloride channel, essential for GABAergic correct functioning, and described to be down regulated in Rett syndrome [13]. The fact that its expression was not significantly altered by MeCP2 inhibition suggests that such an altered regulation is more related to the syndrome pathophysiology rather

than to the straightforward MeCP2 mutations. In fact, it has been well described how the KCC2 expression and function is, indeed, regulated by the GABA function itself [29,30].

It is largely known that most $GABA_A$ receptor coding genes are clustered in four chromosomal regions in chromosomes 4, 5, 15, and 19 [31]. These subunits comprising the pentameric GABA receptor formation have a coordinated expression [32], and, as revealed by human brain transcriptome analysis, this produces a subject and region-specific expression signature of $GABA_A$ receptor subunits [33]. Enlightened by our results, further studies should be made to elucidate whether MeCP2 acts a transcriptional regulator of these clusters, the mechanisms through which this regulation takes place, and the time during development.

As stated from the beginning of the discussion, our main objective was setting the focus over the evolutionary stages in the disease. As for that, we switched to a model that allowed us to evaluate the different prodromic stages of the disease. We used females Bird's Rett mice model, as they better recapitulate the pathophysiology [34]. We focused our analysis on the following three proteins: $GABA_A$-A1R, KCC2, and GAD67. The first two proteins were selected because of the previous observations, and because they have been proven to be not only crucial in Rett syndrome development, but also potential actionable targets of the disease [21,35]. As shown by our results, a markedly reduced expression of $GABA_A$-A1R and KCC2 was recorded in pre-symptomatic mice, while the GAD67 population (used as a marker for GABAergic neuronal) remained unaltered. These results suggest a reduced GABAergic activity without the affectation of the GABAergic general population, which are aligned with the previously described results. Expanding previous descriptions, our results bring focus to the pre-symptomatic stages of the disease, where the most differences were observed. Furthermore, we did not observe any reliable difference in both proteins' expressions in fully-symptomatic mice (or even an increase in KCC2 expression), once again, enhancing the importance of the time-frame when addressing Rett syndrome. The variability in KCC2 expression, and its activation through phosphorylation, is a field that further studies should explore, especially under the recent scenario in which KCC2 is being addressed as a therapeutic target [36]. Preliminary results (data not shown) have pointed towards an over-phosphorylation of KCC2 in symptomatic mice, drawing a scenario in which KCC2 will be under expressed in early pre-symptomatic stages and inactivated in symptomatic phases—again, shaping different therapeutic strategies on different prodromic stages. KCC2 activation and membrane diffusion has been related with GABAergic activity itself, increasing its therapeutic interest [29,30].

During the preparation of this manuscript, Dr. Ben-Ari's team also pointed out the importance of GABAergic dysfunction during development in Rett syndrome [37], providing neurophysiological evidence of such a temporal alteration. The detected expression to normal extents of $GABA_A$-A1R in mature Rett brains reinforces the idea of time-dependence in the MeCP2 control of the GABAergic cluster during the specific developmental stages.

Completing the observed results, we analyzed the aforementioned targets in two Rett patient's necropsied brains. Backing up our previous description, the results showed an even enhanced expression of $GABA_A$-A1R and KCC2, without alterations on GAD67. The overall analysis of the GABAergic pathway showed a slightly increased expression of almost all of the implicated genes. These results confirm the previous findings, pointing towards the importance of pre-symptomatic damage. These results are in agreement with the dataset reported by Renthal et al. [38]. An increasing body of evidence pointing towards the importance of early intervention has been reported in the last few years, as reviewed by Constentino et al. [39], and has extended from neurotransmission to other therapeutic targets in Rett syndrome, such as energetic dysfunction, as very recently published [40], or inflammatory processes [41,42].

To summarize, our results show a direct relationship between MeCP2 and GABA ionotropic receptors' expression, which is not extensible to other GABAergic proteins such as KCC2, which is altered in the context of MeCP2 dysfunction, but does not appear to be directly regulated by its activity. More important than this, our work points attention towards the importance of the time frame when addressing Rett syndrome, as changes appear to be time-dependent, with greater importance in the pre-symptomatic stages.

Therapeutically, early GABAergic modulation in Rett syndrome may represent a promising strategy. While our results suggest that GABA-A1 R can be a potential therapeutic target, the time window of intervention is, according to our findings, critical. Additionally, the development of novel drugs enhancing GABA-A1 R function (for potential use in the initial clinical stages) and devoid of side effects are required, for an early intervention of Rett syndrome.

4. Materials and Methods

4.1. Cell Lines and Samples Utilization

Immortalized Neuro2a cells (also known N2A cells, a fast-growing mouse neuroblastoma cell line) were grown following standard conditions in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 1% glutamine, 10% fetal bovine serum (FBS), and antibiotics.

For neuronal primary cells cultures, the protocol described in the literature [43] was followed. All of the experimental procedures were carried out according to European Union guidelines (Directive 2010/63/EU) and following protocols that were approved by the Ethics Committee of the Bellvitge Biomedical Research Institute (IDIBELL, Barcelona, Spain). Briefly, mouse embryos (embryonic day E18) were obtained from pregnant CD1 females, the cortexes were isolated and maintained in cold Hank's Balanced Salt Solution supplemented with 0.45% glucose (HBSS-Glucose) and digested mildly with trypsin for 17 min at 37 °C, and dissociated. The cells were washed three times in HBSS and resuspended in a Neurobasal medium supplemented with 2 mM Glutamax (Gibco, MA, USA) before filtering in 70 mm mesh filters (BD Falcon, San Jose, CA). The cells were then plated onto glass coverslips (5×10^4 cells/cm^2) coated with 0.1 mg/mL poly-L-lysine (Sigma, Darmstadt, Germany), and 2 h after seeding, the plating medium was substituted by a complete growth medium, consisting of a Neurobasal medium supplemented with 2% B27 (Invitrogen, MA, USA) and 2 mM Glutamax.

In this study, we used post mortem brain samples from two unrelated Rett patients bearing the *MECP2*_c.763C>T mutation and an intra-assay control for RNAseq (i.e., an extra sample that due to its pathology could not be used as a bona fide control, but that allowed us to run technical comprobations). This mutation is the second most frequent Rett-causative mutation, present in 10.9% of the cases [44]. The patients were between 10 and 15 years old at exitus, which is noteworthy, as life expectancy is not highly reduced in Rett syndrome. In all of the cases, RNA was isolated from two brain regions (frontal and occipital cortex) and the samples were treated according to the informed consent of the legal representatives.

The study was approved by the Ethics Committee of Sant Joan de Déu Hospital, project PI15/01159, 01/2016. We are indebted to the "Biobank de Hospital infantil Sant Joan de Déu per la Investigació" integrated in the "Spanish Biobank Network of ISCIII for the sample and the data procurement".

4.2. Mouse Colony

Cortex samples from one and six-month old Mecp2$^{-/+}$ female mice [45] Bird' model (B6.129P2(C)-*Mecp2*$^{tm1.1Bird}$/J) were obtained after mouse sacrifice and brain dissection. The proteins from cortex were extracted by tissue homogenization with an ice-cold RIPA buffer with protease inhibitors (cOmplete, mini, EDTA-free protease inhibitor cocktail, Merck), for 30 min at 4 °C followed by 15 min of centrifuge at 4 °C. The protein samples were quantified by Bradford assay and stored at −80 °C.

4.3. Plasmids and Mutagenesis

In certain experiments, we attempted to silence the endogenous MECP2 expression and re-express the human gene either in the wild type or mutated form.

For the MECP2 silencing, transient transfection with a vector containing a shRNA targeting mMECP2 was performed. Silencing was done with the MISSION® shRNA technology (Sigma Aldrich, Darmstadt, Germany; Clone TRCN0000304464), and the efficiency was checked at the

protein level. To ensure the sole silencing of the endogenous gene, and not the re-expressed form, we used a shRNA directed to the 3′UTR part of the gene, absent in the transfected cDNA. We used, as a control for transfection, a vector with the same backbone but no shRNA.

The MeCP2 c.763C>T genetic variant was introduced by site-directed mutagenesis using the QuickChange II XL Kit (Agilent Technologies, Santa Clara, CA), in the pEGFP-C1-hMeCP2 (wild-type) mammalian expression vector (kindly provided by Dr. Landsberger). The mutation was confirmed by Sanger sequencing. Both vectors, together with the pEGFP-C1 vector (BD Clontech, Palo Alto, CA, USA) and a mock vector with the same backbone, were used for the experiments that required plasmid transfections.

These were carried out using Lipofectamine 2000 (Thermofisher, MA, USA) following manufacturer recommendations. For neuronal primary cultures, 0.8 ◎g of total DNA was mixed with Lipofectamine 2000 and incubated with cortical neurons (at DIV11). The transient expression was allowed for 96 h, and the neurons were fixed at DIV14. For the N2A cells, 4 ◎g of DNA was transfected over 300,000 cells grown in 10 cm^2 plates, and the cells were collected four days after transfection for RNA analysis.

4.4. RNA Extraction and qRT-PCR

RNA for RNAseq (from post-mortem human brain samples) and for qRT-PCR (from post-mortem human brain samples and N2A cells) was extracted using RNeasy® Fibrous Tissue Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. The total RNA was eluted in 40 µL of RNAse-free water and stored at −80 °C. The RNA concentration was measured using the NanoDrop 2000 Spectrophotometer (ThermoScientific, MA, USA), and integrity was assessed by running the samples through a 1% agarose gel.

qPCRs were carried out following a two-step protocol. First, cDNA was synthesized from a total of 500 ng of RNA per reaction, following the recommendations provided with SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen™). After the RT-PCR reaction, the post-mortem brain cDNA from frontal and occipital cortex samples was pooled.

Second, qPCR was performed in a QuantStudioTM 6 Flex Real Time PCR System (Applied Biosystems™ MA, USA) with PowerUpTM SYBRTM Green Master Mix (Applied Biosystems™). The data were analyzed using a comparative method, correlating the initial template concentration with the cycle threshold (Ct) so as to obtain the relative quantity (RQ) of the RNA. The RQ is defined as $2^{-\Delta\Delta Ct}$, where ΔΔCt is the ΔCt of the patient cell line minus the ΔCt of the control cell line, and ΔCt is the Ct of the target gene minus Ct of the endogenous gene (*RPLP0* and *GUSB*).

The probes design was done through the selection of the exonic areas present in all of the transcript variants of each gene, by the selection of common fragments in the UCSC genome browser, based on GRCh38/hg38 version. Primers for N2A-derived qPCR experiments were, in 5′-3′ sense, as follows: m-Mecp2 (F: ACCATCATCACCACCATCAC; R: GGGCATCTTCTCTTCTTTGC), h-MECP3 (F: AGGAGAGACTGGAAGAAAAGT; R: CTTGAGGGGTTTGTCCTTGA), m-Gabra1 (F: ACCAGTTTCGGACCAGTTTC; R: TACAGCAGAGTGCCATCCTC), m-Gabrb2 (F: TCGCTGGTTAAAGAGACGGT; R: TCTCTCCAGGCTTGCTGAAA) and m-Gabrg2 (F: TGGGCTACTTCACCATCCAG; R: GCCATACTCCACCAAAGCAG). The primers sequence for the brain samples of qPCRs were not included.

4.5. Western Blotting and ICC

Western blot analysis of the cortex protein samples from $Mecp2^{-/+}$ female mice was performed. The proteins were subjected to SDS-PAGE and transferred to a nitrocellulose membrane using the Pierce® Power Station (Thermo Scientific). The membranes were blocked with milk, as follows: PBST buffer 5% for 1 h at room temperature. Incubation with primary antibodies was directed against GABA-A1 (Neuromab, UCDavis, CA, USA, 75-136) at a concentration of 1:500, MeCP2 (ab2828; Abcam, Madrid, Spain) at a concentration of 1:1000, and vinculin (ab129002, Abcam, Madrid, Spain) was performed o/n. The secondary antibodies used were horseradish peroxidase-conjugated goat

anti-rabbit and goat anti-mouse IgG antibodies (Lfe Technologies, MA, USA); these were detected using the Enhanced Chemiluminescence System (GE Healthcare, Berkshire, UK).

Immunochemistry experiments were performed as described in the literature [43]. Anti-GABA A1R (Neuromab, UCDavis, CA, USA, 75-136) was used at a concentration of 1:100, and MeCP2 (ab2828; Abcam, Spain) at a concentration of 1:250. The conjugated secondary antibodies for the confocal microscopy were used.

Fluorescence was visualized with a Leica TCS-SL spectral confocal microscope (Leica Microsystems, Wetzlar, Germany) using a Plan-Apochromat 63×/1.4 N.A. immersion oil objective (Leica Microsystems). To excite the different fluorophores, the confocal system was equipped with excitation laser beams at 488 and 546 nm. The images were analyzed with ImageJ software. For the intensity quantification of the ICCs, pictures from independent primary cultures were used. Regions Of Interest (ROIs) were defined on the green channel (GFP positive neurons) applying a threshold to only select the desired neuron. The same ROI was exported to the red channel pictures, and the mean gray value was measured.

4.6. RNAseq Data Analysis

Origin of data*:* A frontal and occipital cortex paired-end RNAseq was performed for two RTT patients bearing the *MECP2*c.763C>T, p.Arg255* mutation and the mentioned intra-assay control. Technical triplicates (three separate RNA extractions) were run for each of the two brain regions. The RNA samples were sent to the Centre Nacional d'Anàlisi Genòmica (CNAG), where the RNAseq experiment was performed, within the framework of the project FIS PI15/01159 Rett Syndrome (IP: Judith Armstrong, Ph.D.). Both brain areas were sequenced separately, and as no significant differences were observed between them, they were analyzed as a sole data pool. To discard the differences between the brain areas, we performed a principal component analysis (PCA). In such an analysis, we compare the variance between all of the samples (patients and areas). At this point, we also added an "intra-assay control" sample. Principal Component 1 (explaining 75% of the variance), clearly discriminated between the Rett and not-Rett samples, while there was not a principal component separating the brain areas (data not shown).

Because of the lack of true control data, the data from our RNAseq were compared to various public controls' data. We used data from public controls available on the GTEx (The Genotype-Tissue Expression) Portal. Cortex RNAseq data from five controls were used, two of which were female (GTEX-15ER7 and GTEX-T2IS) and three were male (GTEX-12126, GTEX-T5JC, and GTEX-WHSE), all of them with ages ranging 20–29. We also compared our data to the cortex RNAseq data from two public male controls (ages 24 and 39) available on the Allen Human Brain Atlas, and the RNAseq data from three female control samples (ages 15–25) used in a publication by Lin, et al. in 2016 (3). Two of these samples were a pool of frontal and temporal cortex RNA, and one of them was just temporal cortex RNA.

RNAseq analysis pipeline*:* As a result of the low performance of the sequencing experiment, the internal control's occipital cortex data was excluded from the analysis. The RNAseq analysis pipeline was run by the Bioinformatics Unit from the Genetics and Molecular Medicine Service at the Hospital SJD. The FASTQ files passed through a first quality control, after which a trimming was performed and the adapters were removed. Then, low quality bases were eliminated so only reads longer than 70 bp were left to continue the analysis. Here, a second quality control was performed and the reads were mapped with TopHat2 (4). The counting was performed with HTseq (5) and the R package DESeq2 (6) was used for library normalization. The frontal and occipital cortex data from our two RTT patients were averaged.

In order to compare the data obtained from our RNAseq experiment to the public data, an internal normalization over the endogens RPLP0 or GUSB of every patient's and control's data was performed.

A validation of the results was carried out at the Hospital Sant Joan de Déu using qRT-PCR, comparing patients with an internal RNAseq control; that is, a sample that could be used for a later comparison of the results through qPCR and therefore RNAseq technical validation, but could not

be used to biologically validate the results, as it was not a healthy brain, as previously described. Thus, following the RNAseq analysis, the targeted gene expression of a subset of 21 genes corresponding to different nodes of the GABAergic pathway and differentially expressed between patients and the internal control was validated by qRT-PCR, showing a strong correlation (20 out of 21 transcripts deregulated; Figure S1), with an overall coincidence between RQ values (qRT-PCR experiments) and fold-change (RNAseq experiment).

4.7. Data Availability Statement

The present study is not a clinical trial of any kind. All of the data, materials, and methods to conduct the research are available in the manuscript. Patient samples are located at the "Biobank de Hospital Infantil Sant Joan de Déu per la Investigació" integrated in the "Spanish Biobank Network of ISCIII for the sample and the data procurement", to whom we are indebted.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1.

Author Contributions: A.O. designed and conceptualized the study, did the data collection throughout all of the experiments and analysis, and drafted the manuscript for intellectual content. C.X. performed the RNAseq and qPCR analysis and data collection. A.A.C. did the Western blot analysis. C.G. participated in the qPCR data collection, and primary cultures and ICC management. M.O. assessed the patient management and the clinical significance of the results. G.F. did the RNAseq analysis and data collection. S.A. guided with the mouse model-derived experiments. M.P. helped with patient management and the clinical significance of the results. J.A. assessed with the RNAseq and qPCR analysis. X.A. designed and conceptualized the study, and drafted the manuscript for intellectual content. A.G.-C. designed and conceptualized the study, and drafted the manuscript for intellectual content, leading the team. All of the authors have read and approved the final manuscript.

Funding: The study has been supported by Mi Princesa Rett (Patients' association) and FIS (PI15/01159, PI16/00851 and PI15/01082 Instituto de Salud Carlos III: ISCIII and "Fondo Europeo de desarrollo regional" FEDER).

Acknowledgments: The pEGFP-C1-hMeCP2 wild-type mammalian expression vector was kindly provided by Landsberger, to whom we are grateful. We are indebted to the "Biobank de Hospital infantil Sant Joan de Déu per la Investigació" integrated in the "Spanish Biobank Network of ISCIII for the sample and the data procurement". We want to acknowledge the patients' association "Mi Princesa Rett", whose support has gone beyond their generous funding, and has been essential for our work development.

Conflicts of Interest: The authors declare no conflict of interest.

References.

1. Leonard, H.; Cobb, S.; Downs, J. Clinical and biological progress over 50 years in Rett syndrome. *Nat. Rev. Neurol.* **2017**, *13*, 37–51.
2. Liyanage, V.R.; Rastegar, M. Rett syndrome and MeCP2. *Neuromol. Med.* **2014**, *16*, 231–264.
3. Ehrhart, F.; Coort, S.L.; Cirillo, E,; Smeets, E.;Evelo, C.T.; Curfs, L.M. Rett syndrome-biological pathways leading from MECP2 to disorder phenotypes. *Orphanet J. Rare Dis.* **2016**, *11*, 158.
4. Bedogni, F.; et al. Rett syndrome and the urge of novel approaches to study MeCP2 functions and mechanisms of action. *Neurosci. Biobehav. Rev.* **2014**, *46*, 187–201.
5. Gonzales, M.L.; LaSalle, J.M. The role of MeCP2 in brain development and neurodevelopmental disorders. *Curr. Psychiatry Rep.* **2010**, *12*, 127–134.
6. Jung, B.P.; et al. The expression of methyl CpG binding factor MeCP2 correlates with cellular differentiation in the developing rat brain and in cultured cells. *J. Neurobiol.* **2003**, *55*, 86–96.
7. Ip, J.P.K.; Mellios, N.; Sur, M. Rett syndrome: Insights into genetic, molecular and circuit mechanisms. *Nat. Rev. Neurosci.* **2018**, *19*, 368–382.
8. Chapleau, C.A.; et al. Dendritic spine pathologies in hippocampal pyramidal neurons from Rett syndrome brain and after expression of Rett-associated MECP2 mutations. *Neurobiol. Dis.* **2009**, *35*, 219–233.
9. Kang, S.K.; et al. Temporal- and Location-Specific Alterations of the GABA Recycling System in Mecp2 KO Mouse Brains. *J. Cent. Nerv. Syst. Dis.* **2014**, *6*, 21–28.

10. Maezawa, I.; Jin, L.W. Rett syndrome microglia damage dendrites and synapses by the elevated release of glutamate. *J. Neurosci.* **2010**, *30*, 5346–5356.
11. Na, E.S.; et al. The impact of MeCP2 loss- or gain-of-function on synaptic plasticity. *Neuropsychopharmacology* **2013**, *38*, 212–219.
12. Calfa, G.; et al. Excitation/inhibition imbalance and impaired synaptic inhibition in hippocampal area CA3 of Mecp2 knockout mice. *Hippocampus* **2015**, *25*, 159–168.
13. Duarte, S.T.; et al. Abnormal expression of cerebrospinal fluid cation chloride cotransporters in patients with Rett syndrome. *PLoS ONE* **2013**, *8*, e68851.
14. Jacob, T.C.; Moss, S.J.; Jurd, R. GABA(A) receptor trafficking and its role in the dynamic modulation of neuronal inhibition. *Nat. Rev. Neurosci.* **2008**, *9*, 331–343.
15. Tretter, V.; Moss, S.J. GABA(A) Receptor Dynamics and Constructing GABAergic Synapses. *Front. Mol. Neurosci.* **2008**, *1*, 7.
16. McKernan, R.M.; Whiting, P.J. Which GABAA-receptor subtypes really occur in the brain? *Trends Neurosci.* **1996**, *19*, 139–143.
17. Cicek, S.S. Structure-Dependent Activity of Natural GABA(A) Receptor Modulators. *Molecules* **2018**, *23*, 1512.
18. Blue, M.E.; Naidu, S.; Johnston, M.V. Altered development of glutamate and GABA receptors in the basal ganglia of girls with Rett syndrome. *Exp. Neurol.* **1999**, *156*, 345–352.
19. Chen, C.Y.; et al. Defective GABAergic neurotransmission in the nucleus tractus solitarius in Mecp2-null mice, a model of Rett syndrome. *Neurobiol. Dis.* **2018**, *109*, 25–32.
20. Zhang, Z.W.; Zak, J.D.; Liu, H. MeCP2 is required for normal development of GABAergic circuits in the thalamus. *J. Neurophysiol.* **2010**, *103*, 2470–2481.
21. Tang, X.; et al. KCC2 rescues functional deficits in human neurons derived from patients with Rett syndrome. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 751–756.
22. Ure, K.; et al. Restoration of Mecp2 expression in GABAergic neurons is sufficient to rescue multiple disease features in a mouse model of Rett syndrome. *Elife* **2016**, *5*, e14198.
23. Cobb, S.; Guy, J.; Bird, A. Reversibility of functional deficits in experimental models of Rett syndrome. *Biochem. So. Trans.* **2010**, *38*, 498–506.
24. Lombardi, L.M.; Baker, S.A.; Zoghbi, H.Y. MECP2 disorders: From the clinic to mice and back. *J. Clin. Investig.* **2015**, *125*, 2914–2923.
25. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585.
26. El-Khoury, R.; et al. GABA and glutamate pathways are spatially and developmentally affected in the brain of Mecp2-deficient mice. *PLoS ONE* **2014**, *9*, e92169.
27. Zhang, W.; et al. Loss of MeCP2 from forebrain excitatory neurons leads to cortical hyperexcitation and seizures. *J. Neurosci.* **2014**, *34*, 2754–2763.
28. Meng, X.; et al. Manipulations of MeCP2 in glutamatergic neurons highlight their contributions to Rett and other neurological disorders. *Elife* **2016**, *5*, e14199.
29. Come, E.; et al. Reciprocal Regulation of KCC2 Trafficking and Synaptic Activity. *Front. Cell Neurosci.* **2019**, *13*, 48.
30. Heubl, M.; et al. GABAA receptor dependent synaptic inhibition rapidly tunes KCC2 activity via the Cl(-)-sensitive WNK1 kinase. *Nat. Commun.* **2017**, *8*, 1776.
31. Simon, J.; et al. Analysis of the set of GABA(A) receptor genes in the human genome. *J. Biol. Chem.* **2004**, *279*, 41422–41435.
32. Enoch, M.A.; et al. A factor analysis of global GABAergic gene expression in human brain identifies specificity in response to chronic alcohol and cocaine exposure. *PLoS ONE* **2013**, *8*, e64014.
33. Sequeira, A.; et al. Human brain transcriptome analysis finds region- and subject-specific expression signatures of GABAAR subunits. *Commun. Biol.* **2019**, *2*, 153.
34. Ricceri, L.; de Filippis, B.; Laviola, G. Mouse models of Rett syndrome: From behavioural phenotyping to preclinical evaluation of new therapeutic approaches. *Behav. Pharmacol.* **2008**, *19*, 501–517.
35. Moore, Y.E.; et al. Developmental Regulation of KCC2 Phosphorylation Has Long-Term Impacts on Cognitive Function. *Front. Mol. Neurosci.* **2019** *12*, 173.
36. 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.* **2019**, *11*, eaau0164.

37. Lozovaya, N.; et al. Early alterations in a mouse model of Rett syndrome: The GABA developmental shift is abolished at birth. *Sci. Rep*. **2019**, *9*, 9276.
38. Renthal, W.; et al. Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. *Nat. Neurosci.* **2018**, *21*, 1670–1679.
39. Cosentino, L.; et al. Rett syndrome before regression: A time window of overlooked opportunities for diagnosis and intervention. *Neurosci. Biobehav. Rev*. **2019**, *107*, 115–135.
40. Cortelazzo, A.; et al. Brain protein changes in Mecp2 mouse mutant models: Effects on disease progression of Mecp2 brain specific gene reactivation. *J. Proteom.* **2020**, *210*, 103537.
41. Cortelazzo, A.; et al. Persistent Unresolved Inflammation in the Mecp2-308 Female Mutated Mouse Model of Rett Syndrome. *Mediat. Inflamm.* **2017**, *2017*, 9467819.
42. Cortelazzo, A.; et al. Subclinical inflammatory status in Rett syndrome. *Mediat. Inflamm*. **2014**, *2014*, 480980.
43. Soto, D.; Olivella, M.; Grau, C.; Armstrong, J.; Alcon, C.; Gasull, X.; Gómez de Salazar, M.; Gratacòs-Batlle, E.; Ramos-Vicente, D.; Fernández-Dueñas, V.; et al. Rett-like Severe Encephalopathy Caused by a De Novo GRIN2B Mutation Is Attenuated by D-serine Dietary Supplement. *Biol. Psychiatry* **2018**, *83*, 160–172.
44. Bienvenu, T.; et al. Spectrum of MECP2 mutations in Rett syndrome. *Genet Test* **2002**, *6*, 1–6.
45. Guy, J.; et al. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **2001**, *27*, 322–326.

Pascual-Alonso *et al. Human Genomics* (2023) 17:85
https://doi.org/10.1186/s40246-023-00532-1

Human Genomics

RESEARCH

Open Access

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] and Judith Armstrong[2,5,6*]

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 multi-functional 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 at both 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-omics 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.

†Ainhoa Pascual-Alonso and Clara Xiol have contributed equally to this work.

*Correspondence:
Judith Armstrong
judith.armstrong@sjd.es
Full list of author information is available at the end of the article

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

Background

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.

Material and methods

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 and 1 male); 15 male patients with MDS; 12 patients with RTT-like phenotypes and mutations in *CDKL5* (1 female and 3 males), *FOXG1* (1 female and 1 male), *NR2F1* (1 female), *GRIN2B* (1 female), *AHDC1* (1 female) and 3 female patients without molecular diagnosis; and 13 healthy controls (7 females and 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% CO_2 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 (Additional file 4: Table S1).

Table 1 Composition of the studied cohort, which consists of individuals with Rett syndrome (RTT) with mutations in *MECP2*, *MECP2* duplication syndrome (MDS) and Rett-like (RTT-like) with mutations in different genes that are not *MECP2* and healthy controls. 'Age' and 'Duration of disease' are given in years; Mean ± Standard Deviation with available data (Additional file 5: Table S2)

Individuals	Age	Duration of disease	Female	Male	Total
RTT	9 ± 6	12 ± 2	21	1	**22**
MDS	7 ± 6	6 ± 6	–	15	**15**
RTT-like	12 ± 7	11 ± 7	8	4	**12**
CDKL5	9.5 ± 10	9 ± 10.5	1	3	**4**
FOXG1	9.5 ± 3.5	9 ± 3.5	1	1	**2**
NR2F1	9	8.7	1	–	**1**
GRIN2B	9	8.7	1	–	**1**
AHDC1	13	11	1	–	**1**
Unknown mutation	18 ± 5	17 ± 5	3	–	**3**
Healthy controls	18 ± 14	–	7	6	**13**

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] (Additional file 5: 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 (Additional file 2: Fig. 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 (Additional file 2: Fig. S2).

Differential expression 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.

Enrichment and upstream 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 11,904 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.

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].

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).

Results

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 (Additional file 3: Fig. 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 analysing may be extrapolated to biological processes occurring in the brain and therefore may impact neurological phenotypes.

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 (Fig. 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 (Fig. 1C).

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 (Fig. 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 (Fig. 2B, Additional file 8: 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.

Multi-omics: integrating transcriptomics and proteomics data

Proteomics differential expression analysis revealed 224 DEPs, 123 upregulated and 101 upregulated (Fig. 2C, Additional file 6: Table S3a,b). Thirty-three 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 (12,448). 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

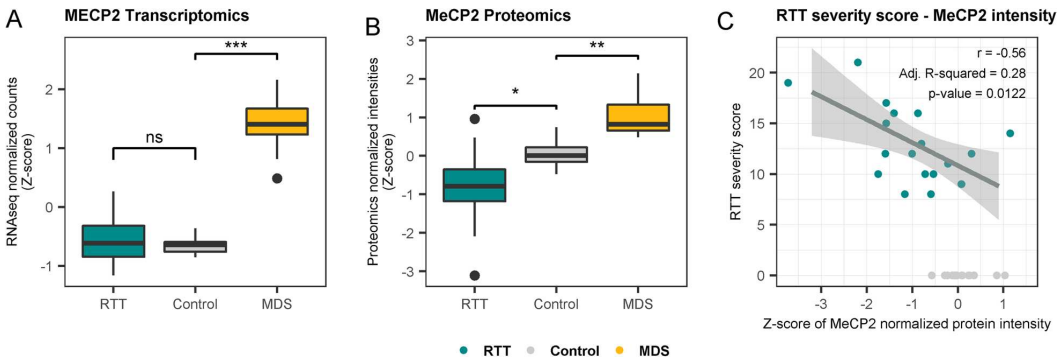


Fig. 1 **A** MECP2 expression levels for RTT, MDS and control individuals obtained by RNAseq. **B** MeCP2 intensity levels for RTT, MDS and control individuals obtained by proteomics. **C** Pearson's correlation between the severity score of RTT patients and MeCP2 levels

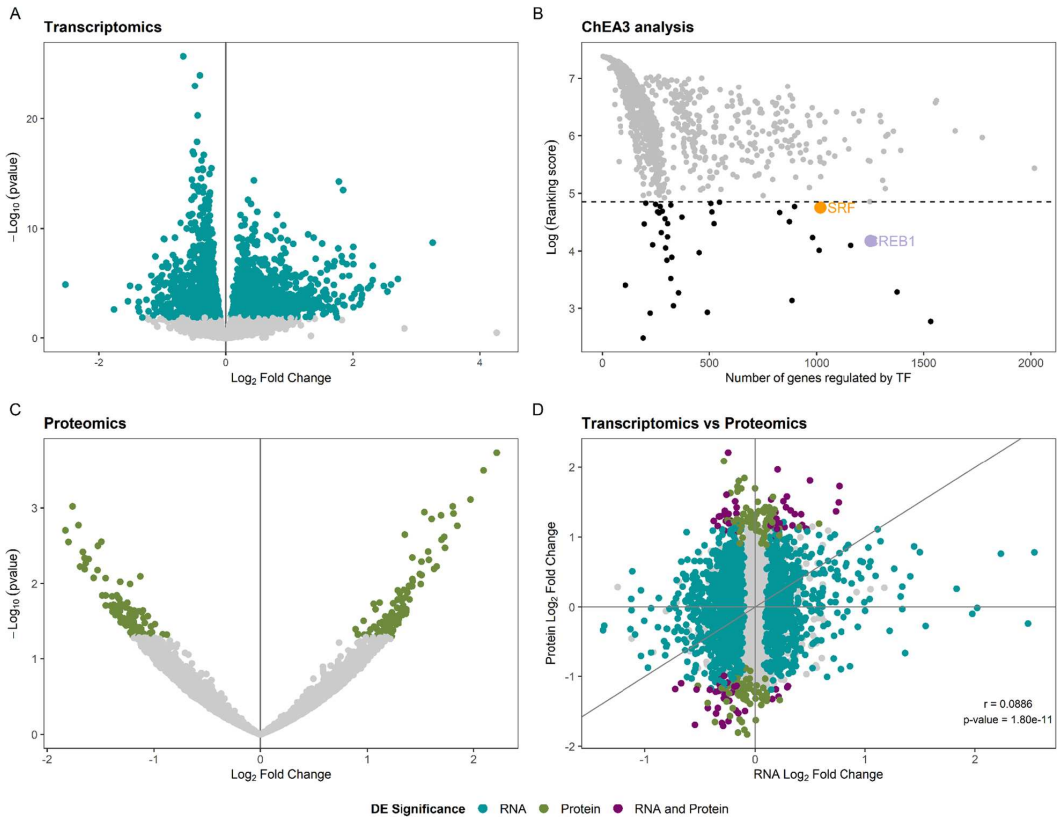


Fig. 2 Summary of the results of the RTT-MECP2 versus healthy controls analysis. **A** RNAseq DE analysis results. The coloured genes are considered differentially expressed, passing a threshold of BH < 0.05. **B** Upstream TF ChEA3 analysis for the DEGs. The 40 TFs that were further studied are coloured 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 analyses are coloured in purple

correlation coefficient = 0.09, $p = 1.8e\text{-}11$, Fig. 2D], we found 75 genes deregulated at both the RNA and protein levels in patients with RTT (Additional file 6: Table S3C). 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 (Additional file 7: 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 (Fig. 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.

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.

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
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

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 (Fig. 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) (Additional file 9: 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 (Fig. 4A,B; Additional file 7: 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 (Fig. 4A, Additional file 7: 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 (Fig. 4C, Additional file 8: 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 (Fig. 4B, Additional file 7: 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* (Fig. 4C, Additional file 8: Table S5c).

Among the shared 12 DEPs, three, *APPL2*, *CNPY4* and *CTSC*, regulate immune response and are downregulated in RTT and upregulated in MDS (Fig. 4E, Additional file 9: 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.

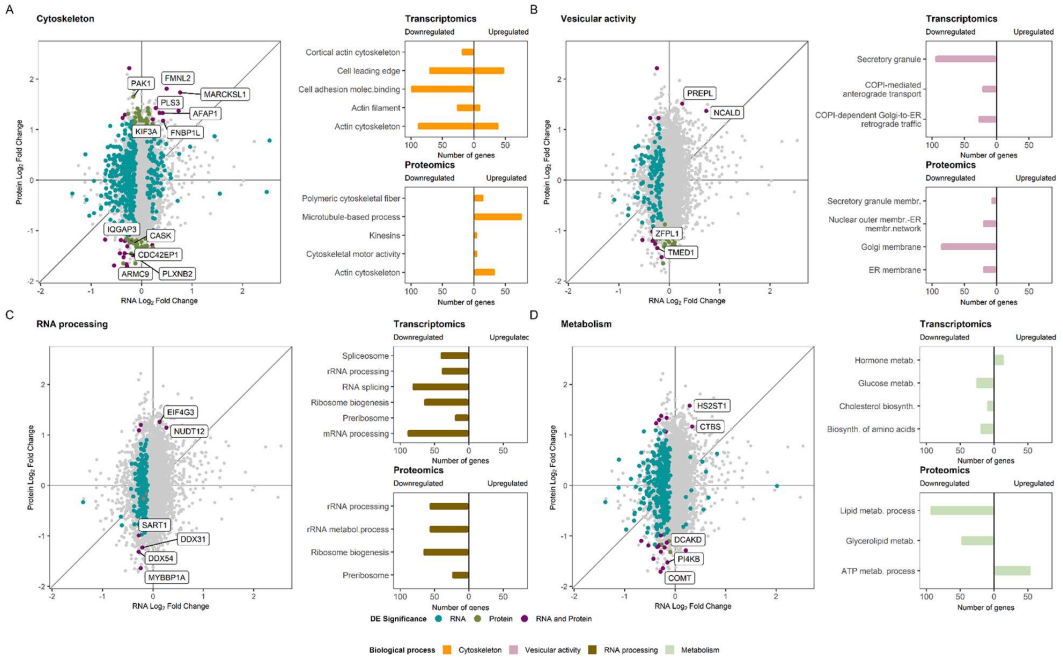


Fig. 3 Summary of the enriched biological processes found in RTT-MECP2 versus healthy control analysis. **A** Significant DEGs and DEPs related to cytoskeleton (coloured in the dot plot) and the enriched terms found for those DEGs (upper bar plot) and DEPs (lower bar plot). **B** Significant DEGs and 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

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) (Fig. 5A, Additional file 9: 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) (Fig. 5C; Additional file 9: Table S6e), but no gene was altered in both 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].

Transcriptomic and proteomic profiles of patients with RTT-like phenotypes are significantly correlated with 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) (Fig. 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 (Fig. 5B and D; Additional file 7: Table S4d-e).

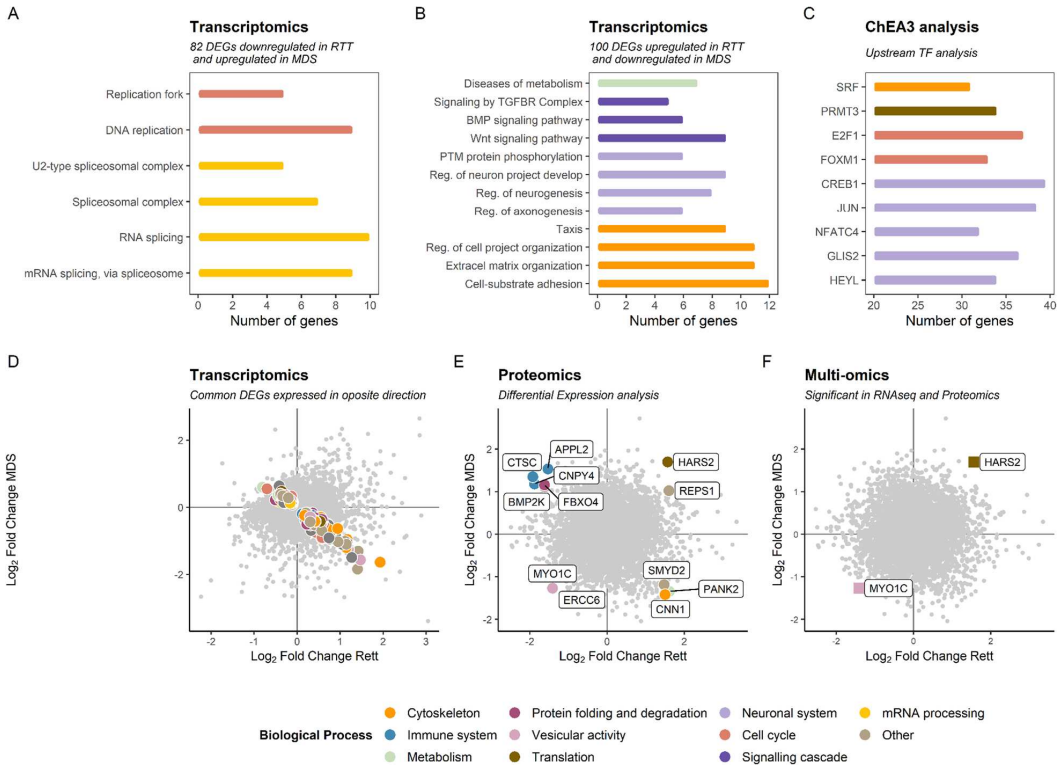


Fig. 4 Summary of the common findings between the analysis of female RTT-MECP2 patients versus female controls and male MDS patients versus male controls. **A** Enrichment analysis results for the shared 82 DEGs downregulated in RTT and upregulated in MDS, coloured by Biological Process (BP). **B** Enrichment 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

Discussion

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, signalling 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

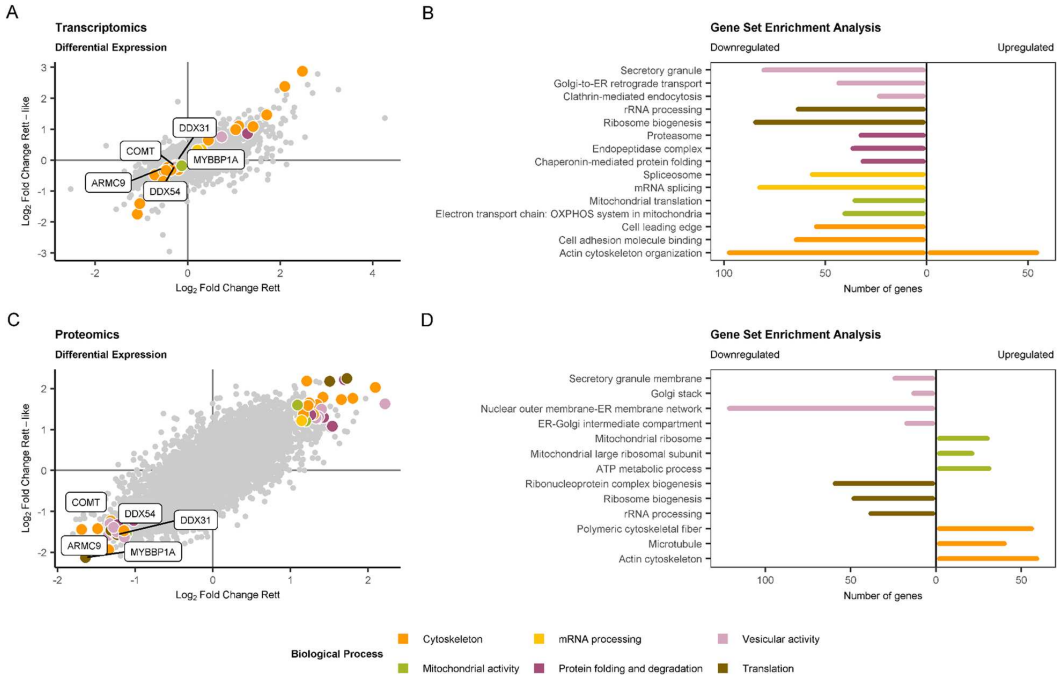


Fig. 5 Summary of the common findings between the analysis of RTT-*MECP2* patients versus healthy controls and RTT-like patients versus healthy controls. **A** Common 63 DEGs (25 upregulated and 38 downregulated), coloured by Biological Process (BP). **B** Gene set enrichment analysis results for the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

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

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

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.

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 signalling 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.

In total, 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 signalling cascades. The Wnt, BMP and TGFß signalling 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 orthopaedic 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], whilst *NFATC4* regulates adult hippocampal neurogenesis and shares a common signalling 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 cell 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.

TT-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 among 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 preprocessing and could explain, at least to some extent, the shared dysregulation of ribosome biogenesis.

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 organisation, 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.

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	Intellectual 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
RNAseq	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

The online version contains supplementary material available at https://doi.org/10.1186/s40246-023-00532-1.

Additional file 1: Fig. S1. Detection of fibroblast samples under oxidative stress conditions.

Additional file 2: Fig. S2. Comparison of the gene expression results between RNAseq and RT-qPCR.

Additional file 3: Fig. S3. Comparison of genes from Genotype-Tissue Expression (GTEx) project and mean TPM (Transcripts per Kilobase Million) in fibroblast cultured cells in RNAseq and RT-qPCR.

Additional file 4: Table S1. Detailed description of the study cohort.

Additional file 5: Table S2. Sequence of the designed primers.

Additional file 6: Table S3. RNAseq and proteomics differential expression results for RTT - MECP2 versus healthy controls.

Additional file 7: Table S4. Enrichment analysis results for the RTT-MECP2 versus healthy controls, RTT versus MDS, RTT versus RTT-like, transcriptomic and proteomic data.

Additional file 8: Table S5. ChEA3 upstream analysis results for RTT patients versus healthy controls, RTT versus MDS, RTT versus RTT-like.

Additional file 9: Table S6. Summary for the DEG and DEP from RTT-MECP2 vs MDS and RTT-MECP2 versus RTT-like.

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 Riccardo from the Institute of Human Genetics of Munich for the raw RNAseq data preparation and 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.

Author contributions
APA, CX and JA conceptualised the study, interpreted the data and wrote the original draft. JA designed the work. CX and APA assisted with the acquisition. APA, CX, DS and RK analysed the data. JA and HP contributed to supervision. All authors took part in writing—reviewing and editing. All authors read and approved the final manuscript.

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.

Availability of data and materials
Our ethics approval and consent agreements allow us to share non-identifiable patient data and analysis data only, as such, we cannot provide BAM or VCF files. 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 RNAseq. They will be available for download without restriction when this article will be published.

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 pseudonymised patient data and analysis data.

Competing interests
The authors declare that they have no competing interests.

Author details
[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.

Received: 18 January 2023 Accepted: 3 September 2023
Published online: 15 September 2023

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.
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;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;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;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;31(2):103–8.
9. Van Esch H. MECP2 duplication syndrome. Mol Syndromol. 2011;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;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;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;1(14):38.
13. Gomathi M, Padmapriya S, Balachandar V. Drug studies on Rett syndrome: from bench to bedside. J Autism Dev Disord. 2020;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):1–24.
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;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;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;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;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;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;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;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;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;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;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;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;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):e93181.
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;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;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;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;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. 2000;288:1796–802.
42. Patel PA, Hegert JV, 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;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;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;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;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;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;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;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;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;41(8):1900045.
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;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;2(14):7678.
59. Knöll B, Nordheim A. Functional versatility of transcription factors in the nervous system: the SRF paradigm. Trends Neurosci. 2009;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;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;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;7(1):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;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;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;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;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;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;29(8):975–87.
72. Li W, Pozzo-Miller L. BDNF deregulation in Rett syndrome. Neuropharmacology. 2014;76:737–46.
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;1(19):256.
74. Xu X, Miller EC, Pozzo-Miller L. Dendritic spine dysgenesis in Rett syndrome. Front Neuroanat. 2014;10(8):1–8.
75. Phillips M, Pozzo-Miller L. Dendritic spine dysgenesis in autism related disorders. Neurosci Lett. 2015;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;10(12):715346.

Publisher's Note

International Journal of
Molecular Sciences

MDPI

Review

Technological Improvements in the Genetic Diagnosis of Rett Syndrome Spectrum Disorders

Clara Xiol [1,2], Maria Heredia [1,2], Ainhoa Pascual-Alonso [1,2], Alfonso Oyarzabal [1,2,3] and Judith Armstrong [2,3,4,*]

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

Abstract: Rett syndrome (RTT) is a severe neurodevelopmental disorder that constitutes the second most common cause of intellectual disability in females worldwide. In the past few years, the advancements in genetic diagnosis brought by next generation sequencing (NGS), have made it possible to identify more than 90 causative genes for RTT and significantly overlapping phenotypes (RTT spectrum disorders). Therefore, the clinical entity known as RTT is evolving towards a spectrum of overlapping phenotypes with great genetic heterogeneity. Hence, simultaneous multiple gene testing and thorough phenotypic characterization are mandatory to achieve a fast and accurate genetic diagnosis. In this review, we revise the evolution of the diagnostic process of RTT spectrum disorders in the past decades, and we discuss the effectiveness of state-of-the-art genetic testing options, such as clinical exome sequencing and whole exome sequencing. Moreover, we introduce recent technological advancements that will very soon contribute to the increase in diagnostic yield in patients with RTT spectrum disorders. Techniques such as whole genome sequencing, integration of data from several "omics", and mosaicism assessment will provide the tools for the detection and interpretation of genomic variants that will not only increase the diagnostic yield but also widen knowledge about the pathophysiology of these disorders.

Keywords: Rett syndrome; Rett-like; NGS; WES; WGS; RNAseq; genetics; MECP2

Citation: Xiol, C.; Heredia, M.; Pascual-Alonso, A.; Oyarzabal, A.; Armstrong, J. Technological Improvements in the Genetic Diagnosis of Rett Syndrome Spectrum Disorders. *Int. J. Mol. Sci.* **2021**, *22*, 10375. https://doi.org/10.3390/ijms221910375

Academic Editor: Nicoletta Landsberger

Received: 30 July 2021
Accepted: 22 September 2021
Published: 26 September 2021

1. Rett Syndrome Spectrum Disorders: Clinical Picture

1.1. Rett Syndrome

Rett syndrome (RTT, OMIM #312750) is a severe neurodevelopmental disorder characterized by a regression of acquired skills, including purposeful hand use and language, after a normal psychomotor development in the first months of life [1]. RTT has an incidence of approximately 1:10,000–20,000 live female births and is the second most common cause of severe intellectual disability in females [2,3]. RTT was first reported in 1966 by the Austrian doctor Andreas Rett, and in 1983, Bengt Hagberg further described the syndrome in a larger cohort of patients [3].

Although belonging to the same clinical entity, patients with RTT show heterogeneous phenotypes, with varying symptoms and severity. In the classic form, patients display a regression in psychomotor development, partial or complete loss of acquired purposeful hand skills and spoken language, gait abnormalities and stereotypic hand movements, which are the required features to diagnose typical RTT. These symptoms are frequently accompanied by breathing disturbances, bruxism, impaired sleep patterns, abnormal muscle tone, and scoliosis, which constitute supportive criteria [4,5]. It is also common that patients with RTT present with acquired microcephaly and epilepsy [3–5].

Presently, the diagnosis of RTT is clinical and follows a set of guidelines published in 2010 [4]. A genetic confirmation typically follows the clinical diagnosis of RTT, but its role is still supportive, since mutations in *MECP2* (the main RTT-associated gene) may cause other phenotypes than RTT, and mutations in other genes have been found in patients with a clinical diagnosis of RTT [4,6]. According to the current diagnostic guidelines, patients who only fulfil two of the four required criteria and five of the 11 supportive criteria are diagnosed with atypical RTT.

There are several atypical forms of RTT. The Zappella (or preserved speech) variant is a generally milder form of RTT characterized by the recovery of some language after the regression, with the ability to speak single words or phrases [4,7]. Other atypical forms of RTT entail more severe phenotypes. The Hanefeld (early-onset seizures) variant is defined by a very prompt onset of epilepsy, and in the Rolando (congenital) variant, there is no clear regression, with the symptoms already being apparent during the neonatal period [4].

Although valid diagnostic guidelines highlight the importance of an apparently normal early development followed by a regression period as a distinct feature for the identification of this disorder, recent reports draw attention towards subtle impairments in motor and communication skills in early stages of development and before the onset of clear RTT features [4,8,9]. Both parents and clinicians have reported feeding difficulties, abnormal crying, and delay in reaching developmental milestones in the first months of life before the clear regression period had begun [10]. Since the identification of these slight symptoms, there has been increasing interest in finding molecular and neurofunctional markers that will enable early detection of RTT.

1.2. A Broader Clinical Entity: RTT Spectrum Disorders

Despite the common characteristic features, patients with RTT display a wide range of phenotypic variation that can be appreciated after a thorough clinical characterization. The idea behind the standardized clinical criteria is to provide guidelines to evaluate all traits related to RTT and aid differential diagnosis. Nevertheless, there are patients with many of the distinct features of RTT that do not fulfil the established clinical criteria for either typical or atypical RTT. Currently, the term "Rett-like" is used to describe patients with these overlapping phenotypes, but there are no consensus clinical criteria for a Rett-like diagnosis [11–13].

Additionally, recent evidence of genetic heterogeneity behind RTT and RTT-like disorders has become available. Over the past few years, massively parallel sequencing technologies used to diagnose RTT and RTT-like disorders have led to the identification of disease-causing variants in many different genes, some of which are novel and others are traditionally associated with other neurodevelopmental disorders or epileptic encephalopathies with a considerable phenotypic overlap with RTT [11,13–15].

In light of the findings regarding the phenotypic and genetic heterogeneity behind this clinical entity, the term "RTT spectrum disorders" can be used to encompass all RTT (typical and atypical) and RTT-like phenotypes [13]. The rationale supporting this idea is that, if the dysfunction of several different genes is causing such similar phenotypes, they must be connected in some way, probably because they share functions or because they are involved in the same molecular pathways [13,16].

2. Single-Gene Genetic Testing for RTT Spectrum Disorders

2.1. MECP2

The genetic cause of RTT remained unknown until 1999 when it was associated with mutations in the *MECP2* gene, which encodes methyl-CpG-binding protein 2 (MeCP2) [17]. MeCP2 is a multifunctional protein involved in transcriptional regulation, chromatin remodeling, micro-RNA processing, and alternative splicing, modulating gene expression levels both at transcriptional and post-transcriptional levels [18–20]. Recent studies show that the loss of MeCP2 alters the expression of many genes, but the effects at an individual gene level are small [1]. This indicates that MeCP2 acts as a global genome regulator of gene

expression and chromatin architecture that mediates cellular changes through activation and repression of a great number of genes genome-wide [20,21].

MECP2 has four exons and produces two different isoforms through alternative splicing. Isoform e1 (498 amino acids) encompasses exons 1, 3, and 4, and has its translation start codon in exon 1 (NM_001110792). Isoform e2 (486 amino acids) encompasses all four exons at the RNA level, but has the translation start codon in exon 2, while exon 1 remains noncoding (NM_004992) [3,22,23]. MeCP2e2 was the form initially described, while MeCP2e1 was not identified until 2004 [22,23]. The two isoforms differ in their N-terminal regions, but share all functional domains [24]. MeCP2e1 is the predominant isoform in the brain, while MeCP2e2 is more abundant in other tissues such as fibroblasts [3,23]. Nonetheless, both isoforms coexist in the brain, where MeCP2e2 has shown a later onset of expression during mouse embryonic development and a more restrictive pattern of expression [24,25]. At a functional level, both proteins differ in binding dynamics, turn-over rates, and interacting partners, suggesting non-overlapping functions of *MECP2* isoforms [26].

According to recent studies, approximately 95% of patients with typical RTT and 75% of patients with atypical RTT (especially those with the Zappella variant) have pathogenic variants in *MECP2* [9,27,28]. So far, 925 different *MECP2* variants, including 535 pathogenic variants and 212 variants of unknown significance (VUSs), have been reported in RettBASE, a specialized database that gathers information related to RTT-related genomic variants [29]. These include missense, nonsense, frameshift, splicing, and intronic variants, and moreover, 280 gross deletions that produce the loss of *MECP2* function have also been described and are included in the Human Gene Mutation Database (HGMD®) [30]. However, there are eight recurrent point mutations (missense and nonsense) that account for approximately 50% of all genetically diagnosed cases of RTT (Table 1) [27,29]. Interestingly, mutations specific to MeCP2e1, the predominant isoform in the brain, have been found in RTT patients, whereas there are no mutations specific to isoform e2 known to cause RTT [31].

Table 1. Most recurrent point mutations in *MECP2* and their frequencies in RTT patients [29].

Coding DNA Variant (NM_004992.4)	Amino Acid Change	Percentage of RTT Patients
c.473C>T	p.Thr158Met	8.74%
c.502C>T	p.Arg168 *	7.57%
c.763C>T	p.Arg255 *	6.64%
c.808C>T	p.Arg270 *	5.74%
c.916C>T	p.Arg306Cys	5.14%
c.880C>T	p.Arg294 *	4.97%
c.397C>T	p.Arg133Cys	4.52%
c.316C>T	p.Arg106Trp	2.79%
		Total = 46.11%

The * in the table represents stop codons according to the current variant nomenclature guidelines of the HGVS (https://varnomen.hgvs.org/recommendations/protein/variant/substitution/).

Since its association with RTT, *MECP2* has been studied by direct sequencing (by Sanger sequencing) and gene dosage analysis (by multiplex ligation-dependent probe amplification (MLPA), qPCR, and FISH) to obtain a molecular confirmation of the clinical diagnosis. At the beginning, the studies did not include exon 1 (because it was thought to be noncoding), but with the description of MeCP2e1, with coding sequences in this exon, it was added to routine mutation screening [23,32]. Taking advantage of the recurrence of the mutations, an electronic DNA microchip was developed to detect seven of the eight most common *MECP2* pathogenic variants in a faster and more economical manner [33].

2.2. CDKL5 and FOXG1

Although most patients carry disease-causing variants in *MECP2*, several other genes have progressively been linked to RTT over the past two decades. In 2004, pathogenic variants in cyclin-dependent kinase-like 5 (*CDKL5*) were associated with the atypical form of RTT known as the early-onset seizure variant. CDKL5 is a protein kinase that directly interacts with MeCP2 and mediates its phosphorylation. This phosphorylation modulates the function of MeCP2 in neurons [34–36]. MeCP2 also binds *CDKL5* at the DNA level and represses its transcription [37].

In 2008, forkhead box G1 (*FOXG1*) was found to be related to the congenital variant of RTT [38–40]. FOXG1 is a brain-specific transcriptional repressor that is coexpressed and colocalized with MeCP2 in the postnatal cortex [40]. FOXG1, like MeCP2, associates with histone deacetylase 1 (HDAC1) to repress transcription [41].

Currently, some authors consider the neurodevelopmental disorders caused by pathogenic variants in *CDKL5* and *FOXG1* distinct clinical entities because of the defining features that differentiate them from typical RTT [42,43]. Nevertheless, the connections and interactions among these three proteins highlight their relationship and indicate that they are in fact involved in common processes, which could explain the overlapping symptoms that arise when the function of any of them is impaired.

When *CDKL5* and *FOXG1* were linked to RTT, single-gene molecular diagnosis techniques were also applied to these genes, particularly in *MECP2*-negative patients with RTT. Combining single-gene approaches in these three genes, about 28% of patients with RTT spectrum disorders were diagnosed with disease-causing variants in *MECP2*, *CDKL5*, and *FOXG1* [44].

3. The Revolution of Next Generation Sequencing

3.1. Gene panels and Exome Sequencing

The arrival of next generation sequencing (NGS) has allowed us to simultaneously sequence multiple genes in the same experiment [45]. NGS for diagnosis is especially useful in genetically heterogeneous disorders, where many successive single-gene approaches may be more expensive and inefficient. Instead of limiting the scope of the genetic study to one single candidate gene, NGS allows us to extend or redirect a genetic analysis if needed [46].

NGS approaches can target anything from a set of specific genes (gene panels) to the whole genome. Presently, the most widely used approach in a medical diagnostic setting is clinical exome sequencing (CES), which targets all exons of the genes currently known to cause monogenic disorders [44,46–48]. Nevertheless, the diagnosis by whole exome sequencing (WES), which targets all exons and canonical splice sites of the ~20,000 known protein-coding genes, is becoming more popular [49,50]. The American College of Medical Genetics and Genomics (ACMG) currently recommends WES as the gold standard of clinical practice in children with intellectual disability (ID), developmental delay, or multiple congenital anomalies, due to the reduction in costs and the increase in diagnostic rate [51].

The overall diagnostic yield when applying WES to patients affected with pediatric rare diseases is 28% on average [46,49,52,53]. Nevertheless, the positive diagnostic rate varies greatly depending on the group of genetic disorders considered [52]. The Deciphering Developmental Disorders (DDD) Study successfully diagnosed 28% of the patients enrolled by applying a combination of WES and exome-focus array comparative genomic hybridization (exome-aCGH) [54]. In a recent systematic review, neurodevelopmental disorders (NDDs), among which we can classify RTT spectrum disorders, were found to have a 23.7% overall diagnostic yield by NGS (22.6% using gene panels and 27.3% using WES). Among NDD subtypes, patients with intellectual disability showed the highest diagnostic yields (28.2%), while patients with autism spectrum disorder (ASD) showed the lowest diagnostic yields (17.1%) [55].

Since a good clinical characterization is critical for variant interpretation, it is not surprising that the best defined clinical entities have the highest diagnostic yields. Therefore, patients with general or imprecise clinical diagnoses, such as ASD or ID, tend to show less successful outcomes than patients with a well-defined, accurate clinical description, such as RTT or Noonan syndrome, despite genetic heterogeneity. Moreover, complex phenotypes that can be caused and modified by several factors, such as ASD, are less likely to be definitively diagnosed than monogenic Mendelian disorders with a clear genetic root.

In terms of technical capacities, NGS technologies are especially sensitive to detecting single nucleotide variants (SNVs) and small insertions or deletions (indels), but copy number variant (CNV) detection is also possible through read depth analysis [46,56]. In fact, the power of CNV detection in WES could be superior to that of low-resolution genomic microarrays [57]. Therefore, WES studies that include CNV analysis usually have a higher diagnostic yield.

Including the parents of the affected child in the WES analysis (trio-WES) has certain advantages. This approach allows us to directly assess the inheritance pattern of candidate variants (de novo or inherited), as well as the phase (whether two variants in the same gene are in the same or in different chromosomes). It also enables directly filtering out rare benign familial variants, which may lead to an up to 10-fold reduction in the number of candidate variants to analyze in families where parents are unaffected [46]. Since de novo mutations are the most common cause of neurodevelopmental disorders such as RTT, the trio-based approach can streamline the genetic diagnosis [58]. Even though the sequencing costs of the experiments are higher when compared to proband-only analyses, the reduction in costs of segregation studies can compensate for this fact. In pediatric rare diseases, diagnostic rates can be increased up to about 40% when applying a trio-WES approach [46].

3.2. The Results of NGS Studies of Patients with RTT Spectrum Disorders

Many diagnoses of patients with RTT spectrum disorders come from studies where they were enrolled together with other probands in wider groups characterized as NDD or ID. Nevertheless, there are several diagnostic NGS studies focused only on RTT spectrum patients (Table 2). WES studies of RTT spectrum cohorts have had diagnostic yields around 65% when used as a first-tier diagnostic test and slightly lower when performed after a negative CES result [15,44,59–63]. Gene panels have a higher variation in the diagnostic rates of RTT spectrum patients, strongly depending on the subset of targeted genes and the characteristics of the studied cohort [14,44,64–66].

The higher diagnostic yield in WES studies focused only on patients with RTT spectrum disorders, usually above 60% (Table 2), when compared to general NGS data, usually around 28–40%, could be explained due to the thorough phenotypic characterization of the patients included in these studies. As mentioned above, an exhaustive clinical characterization is essential to reach a definite diagnosis by NGS, where many candidate variants can be potentially identified and must be interpreted in the phenotypic context of each patient. Due to the problem of phenotypic heterogeneity, researchers working with RTT spectrum cohorts tend to establish rather accurate criteria for inclusion in the studies and are familiar with comprehensive phenotyping, leading to such positive results (see Table S1 in supplementary data for gene list of RTT spectrum genes).

NGS has identified more than 90 novel causative genes of RTT spectrum disorders over the past 7 years (Supplementary Table S1) [11]. Some of these genes had previously been linked to other well-characterized disorders with overlapping features with RTT, such as Pitt–Hopkins syndrome (*TCF4*), Angelman syndrome (*UBE3A*), or Cornelia de Lange syndrome (*SMC1A*), while others are linked to epileptic encephalopathies (such as *STXBP1* and *GRIN2B*), intellectual disability with epilepsy (*IQSEC2* and *MEF2C*), or other NDDs. Moreover, thanks to NGS, novel possible causative genes for RTT spectrum phenotypes, such as *JMJD1C* and *GABBR2,* have also been identified, [13–15,44,60–63,67–71].

The current list of RTT spectrum genes is especially enriched in chromatin modulators (such as *HDAC1*, *MEF2C*, *NCOR2*, or *SATB2*, and including *MECP2*) and genes involved in synaptic function (such as *GABRB2*, *GRIN2B*, *SHANK3*, *IQSEC2*, *STXBP1*, *SLC6A1*, or *SYNGAP1*) [11,62]. These pathways and functions have been found impaired in patients with RTT spectrum disorders, as well as in RTT animal models, and this might be the link between RTT spectrum genes and the reason why patients with RTT spectrum disorders present with overlapping features [72,73].

It is noteworthy that patients with pathogenic variants in several of the abovementioned genes (*TCF4*, *STXBP1*, *SCN2A*, *WDR45*, *KCNQ2*, and *MEF2C*) present with significant phenotypic overlap with patients with typical RTT and meet the four main established criteria for a diagnosis of typical RTT—gait abnormalities, loss or absence of purposeful hand movements, stereotypies and speech loss, or severe deficit [11]. These six genes are the most frequent causative genes in patients with RTT spectrum disorders, apart from *MECP2*, *CDKL5*, and *FOXG1* [14].

Table 2. NGS studies performed on RTT cohorts and their diagnostic yields.

Publication	Type of Genetic Testing	Number of Genes in Test	Number of Patients	Diagnostic Yield
Olson et al., 2015 [59]	Singleton-WES	Whole exome	11	64%
Lucariello et al., 2016 [60]	Trio-WES	Whole exome	21	67%
Lopes et al., 2016 [61]	aCGH and trio-WES	Whole genome (aCGH), whole exome (WES)	19	68.5% (58% due to WES)
Vidal et al., 2017 [44]	Gene panels and WES	17 (custom panel), 4813 (commercial panel), and whole exome (WES)	242 (custom panel), 51 (commercial panel) and 22 (WES)	23% (custom panel), 24% (commercial panel) and 32% (WES)
Sajan et al., 2017 [62]	SNP array-based CNV analysis and trio-WES	Whole genome (CNV analysis) and whole exome (WES)	22	68.4%
Allou et al., 2017 [66]	Gene panel and trio-WES	5 (gene panel) and whole exome (WES)	30 (gene panel) and 2 (trio-WES)	10% (gene panel) and 50% (trio-WES).
Yoo et al., 2017 [63]	Trio-WES	Whole exome	34	67.6%
Iwama et al., 2019 [15]	Singleton-WES	Whole exome	77	61%
Henriksen et al., 2020 [74]	Direct *MECP2* analysis and WES	Whole exome	91	NA

Although the genetic findings of NGS are still validated by molecular genetic techniques in standard clinical practice, this approach is more cost-effective than successive single-gene diagnostic testing. Interestingly, some studies have detected, by NGS, pathogenic variants in *MECP2* that were previously missed by classical molecular genetic testing, and some studies show that NGS can outperform Sanger sequencing in detecting heterozygous changes and mosaic variants [46,59,62,75].

3.3. NGS Data Re-Analysis

Nowadays, big efforts are made to improve our variant interpretation capacities. Evidence of the biological consequences of variants of unknown significance (VUSs) is obtained from functional studies, descriptions of disease phenotypes associated with novel genes and variants are stored in comprehensive databases, and the characterization of new functional genomic elements enables the correct interpretation of variants.

Therefore, the frequent update of all these data makes it feasible to reanalyze negative WES cases prior to proceeding with whole genome sequencing (WGS). A recent study showed that 30% of the positive cases solved by WGS could be identified by reanalyzing the WES raw data [76]. Moreover, several studies reported a diagnostic rate after reanalysis of 10.5–15.3% within a period of approximately 1 year after the first analysis [77–81].

The increase in diagnostic success after reanalysis can be due to several reasons. In some cases, new diagnoses are reached because of recent publications of disease-gene associations or particular phenotypes that were not considered when the former analysis was

conducted [77,78,80,81]. Another common reason is a revision of the patient's phenotype by the clinician that can eventually redirect the analysis towards a new set of candidate genes or reconsider variants that were previously detected and dismissed [77,79,80]. Reclassification of a formerly detected variant can also change the result of a WES analysis. For instance, functional and in silico studies can help, over time, to reclassify a former VUS as either a benign or a pathogenic variant or to identify a synonymous variant as a splicing aberration [77]. Finally, a common reason is an improvement in the bioinformatics pipeline or a database update that allows for correct detection or annotation of variants missed in the prior analysis [77–80].

Considering these outcomes, NGS data reanalysis becomes an interesting diagnostic tool to contemplate in medical genetics until WGS-trio costs decrease and WGS-trio is implemented as ordinary clinical care.

4. Future Perspectives of Genetic Diagnosis for RTT Spectrum Disorders

4.1. The Bigger Picture—WGS

Exome sequencing has been extremely useful in diagnosing rare diseases, such as RTT. It enables the detection of protein-coding and splice-site variants, which constitute 85% of the known disease-causing mutations [82,83]. Nevertheless, this approach targets only 2% of the genome and has certain limitations, and therefore unsolved exome sequencing cases could potentially be elucidated by WGS.

In addition to the protein-coding and splice-site variants detected by exome sequencing, WGS can also identify several types of non-coding variants that can compromise gene function. Introns, which are not sequenced in exome assays, harbor deep intronic mutations that can increase the activity of cryptic splice sites that otherwise seldom produce splicing events, causing an intron inclusion within the transcript. These aberrant splicing events produce dysfunctional transcripts and may lead to disease. Deep intronic variants have been implicated in the pathogenesis of rare diseases, such as ocular albinism due to *GPR143* malfunctioning and hyperammonemia as a result of *OTC* deficiency [84].

Variants located in regulatory elements (promoters, enhancers, and insulators) may modify gene expression levels by changing transcription factor binding affinities to the DNA sequence. In the 3′ untranslated region (3′-UTR), which contributes to the regulation of gene expression by binding to microRNAs, non-coding variants could affect this binding ability, thus modifying transcript stability [85].

Non-coding variation is progressively gaining acknowledgment, but there are still few reports of the implication of these kind of variants in Mendelian disease phenotypes [86]. This may be partly because these regions are not covered by a regular exome sequencing assay, which is the most common type of genetic test nowadays, so pathogenic variants within these sequences simply have not yet been detected. Moreover, the difficulty in deciphering the effects that these variants may produce and the necessity of functional studies to validate these hypotheses hinder the interpretation and report of these variants.

Another advantage of WGS is its power to detect structural variants (SVs) and complex rearrangements. Since exome sequencing must rely almost exclusively on read depth, it can only confidently detect unbalanced alterations (CNVs). WGS, on the other hand, because of its extensive coverage, enables detection of SVs not only by read depth analysis but also by discordantly aligned paired-end reads and split reads, which allow the detection of SV breakpoints (even at nucleotide resolution) [87]. Thus, WGS can detect balanced SVs (inversions and translocations), besides CNVs and insertions. Furthermore, in the case of duplications, it can distinguish whether they are in tandem or inserted elsewhere in the genome. An interesting study by Gilissen et al. identified disease-causing CNVs and SVs by WGS in nine patients with ID that had been previously missed by a comparative genomic hybridization array (aCGH) [69]. One of these variants was a partial duplication of *TENM3* inversely inserted into *IQSEC2*, which is related to a phenotype of intellectual disability and epilepsy that lies within the RTT spectrum [69].

Furthermore, WGS enables reliable identification of runs of homozygosity (ROH), which are long genomic stretches that display identical haplotypes in both homologous chromosomes. WGS is the best strategy to identify these events, since the non-uniform distribution of WES and SNP-array data complicate the detection of these signals, especially the shorter ones [88]. ROH may denote identity by descent (IBD), which means that the homozygosity originates because the two alleles come from the same common ancestor, indicating some degree of consanguinity [89]. These regions may contain pathogenic recessive variants and are especially relevant when suspecting an autosomal recessive inheritance pattern. ROH might also be indicative of uniparental disomy (UPD) [90]. UPD happens when both homologs of a chromosome pair are inherited from the same parent, and it has implications for disease by causing either the lack or the overexpression of genes affected by gender-specific imprinting and by converting recessive pathogenic variants in a heterozygous state in unaffected carrier parents into homozygous variants causing disease in the proband [90,91]. The detection of long regions of homozygosity in NGS data allows us to uncover UPD events, which have been linked to several cases of NDDs [90].

Finally, WGS uniformity in terms of coverage depth and genotype quality enables the detection of variants previously missed by WES, particularly in GC-rich exons [92,93]. In fact, the proportion of false-positive SNVs is 61% lower in WGS compared to WES [92]. While WES has around a 28% diagnostic rate in patients with ID, WGS has been shown to solve up to 42% of cases, even when only the coding region is studied [69].

Despite the advantages of WGS, it is not always the final answer to reaching a diagnosis [46]. The current issue is that the source of the disease will most likely be blended among the huge amount of data generated, and we may not be able to pinpoint it with our still-limited knowledge.

4.2. Multi-Omics

While WES yields 20,000–23,000 variants per individual, WGS reveals 3–5 million [94]. A large proportion of these variants lies outside the coding regions and canonical splice sites, which makes the interpretation of their possible effects challenging. To help prioritize the variants detected at the DNA level, WGS analysis can be coupled with several other technologies focused on studying other molecules, such as RNA sequencing (RNA-seq) and proteomics [94,95].

RNA-seq is an NGS approach that allows us to simultaneously sequence and quantify all transcripts (usually coding transcripts) present in a sample. RNA-seq analyses allow us to not only detect variants in the RNA sequence but also to identify aberrant events that may be caused by variants detected in WGS [96]. Aberrantly expressed transcripts can lead to the identification of both coding variants that trigger nonsense-mediated decay (NMD) and noncoding variants in regulatory regions, such as promoters and enhancers, that hamper transcription. Abnormal splicing events may reveal pathogenic variants in canonical splice sites or within exons or introns that produce dysfunctional transcripts. Finally, finding that one allele is absent among the analyzed transcripts of a given gene (monoallelic expression) helps reconsider the effect that a heterozygous variant in a recessive gene may have if it is the only expressed allele [96–98].

Mass spectrometry-based proteomics enable us to identify and quantify all proteins present in a sample (the proteome) at the same time [99]. As with RNA-seq aberrantly expressed transcripts, expression outliers in proteomics may help reprioritize variants detected by WGS within those genes and might indicate that those variants affect protein stability or post-translational modifications [95,96].

Integrating data from genomics and other "omics" is currently known as "multi-omics." Using multi-omics, several studies have unraveled previously unsolved cases by WGS, WES, or gene panels, increasing diagnostic rates in 10–36% of patients [97,100,101].

The most delicate issue when using multi-omics is the decision of which tissue to study. Since not all transcripts and proteins are produced by all cell types, it would be ideal to select the most relevant tissue according to the condition affecting each patient. In the

case of neuromuscular disorders, for instance, muscle biopsies have proved to generate more robust data for RNA-seq analysis compared to blood and cultured fibroblasts [100]. In the case of RTT spectrum disorders, the most disease-relevant tissue would be the central nervous system, which is unfortunately not easily accessible. However, cultured fibroblasts reliably express almost 70% of the disease-related genes registered in OMIM and 70–75% of the RTT-spectrum-related genes according to GTEx data (Figure 1) [96,102]. This demonstrates that a minimally invasive procedure, such as a skin biopsy could provide a sample that, at least to some extent, could be useful for multi-omics analyses to increase the diagnostic yield in RTT spectrum disorders.

4.3. Mosaicism

The most widely used NGS technologies for the diagnosis of rare diseases are usually designed to detect genetic variants that affect all the cells of an individual (germline variants). Mosaic mutations (those variants only present in a subset of cells of an individual) are more difficult to recognize and call. Since they are usually found at low frequencies, with a nonspecific analysis they can be confused with technical errors or artifacts [103,104].

The effects of a mosaic mutation on an individual strongly depend on the developmental stage at which the mutation occurs, and consequently, of the number of cells that carry the mutation and the tissues to which they belong [105]. These mutations can happen in the first stages of development and be present in several tissues at different frequencies or relatively late in the process and affect only certain tissues or groups of cells [103,105].

In the former case, a deep coverage NGS approach (at least 200×) may be able to recognize and call the mosaic variant, while in the latter, the mutation will only be detectable if the affected tissue is available for analysis [103]. CES or WES approaches may be too comprehensive to achieve such deep coverage. Therefore, the design of a smaller custom gene panel that contains relevant disease-causing genes with sufficient evidence of pathogenicity for mosaic variants might be a possible strategy to streamline resources. On the other hand, single-cell NGS approaches provide the higher resolution and enable a more shallow sequencing coverage to detect mosaicism, but require access to the tissues relevant for the disease, which is a setback when facing neurological phenotypes [103].

The contribution of mosaic variants to neurodevelopmental disorders is becoming increasingly manifest. A 2014 study using WGS to identify de novo mutations in patients with ID validated seven mosaic pathogenic variants in candidate ID genes (6.5% of the presumed germline variants) [69,104]. Moreover, the study of this cohort found that four presumed de novo mutations in patients with ID were in fact mosaic mutations present at very low frequencies (average of 3.54%) in the blood of unaffected parents [104]. In addition, another recent study found that 6.6% of parents of children with epileptic encephalopathies presented with low levels of mosaicism for the disease-causing variant affecting their children [106]. Therefore, when evaluating the implications of mosaic variants for the genetic diagnosis of neurodevelopmental disorders, we must consider both mosaicism in patients and mosaicism in parents.

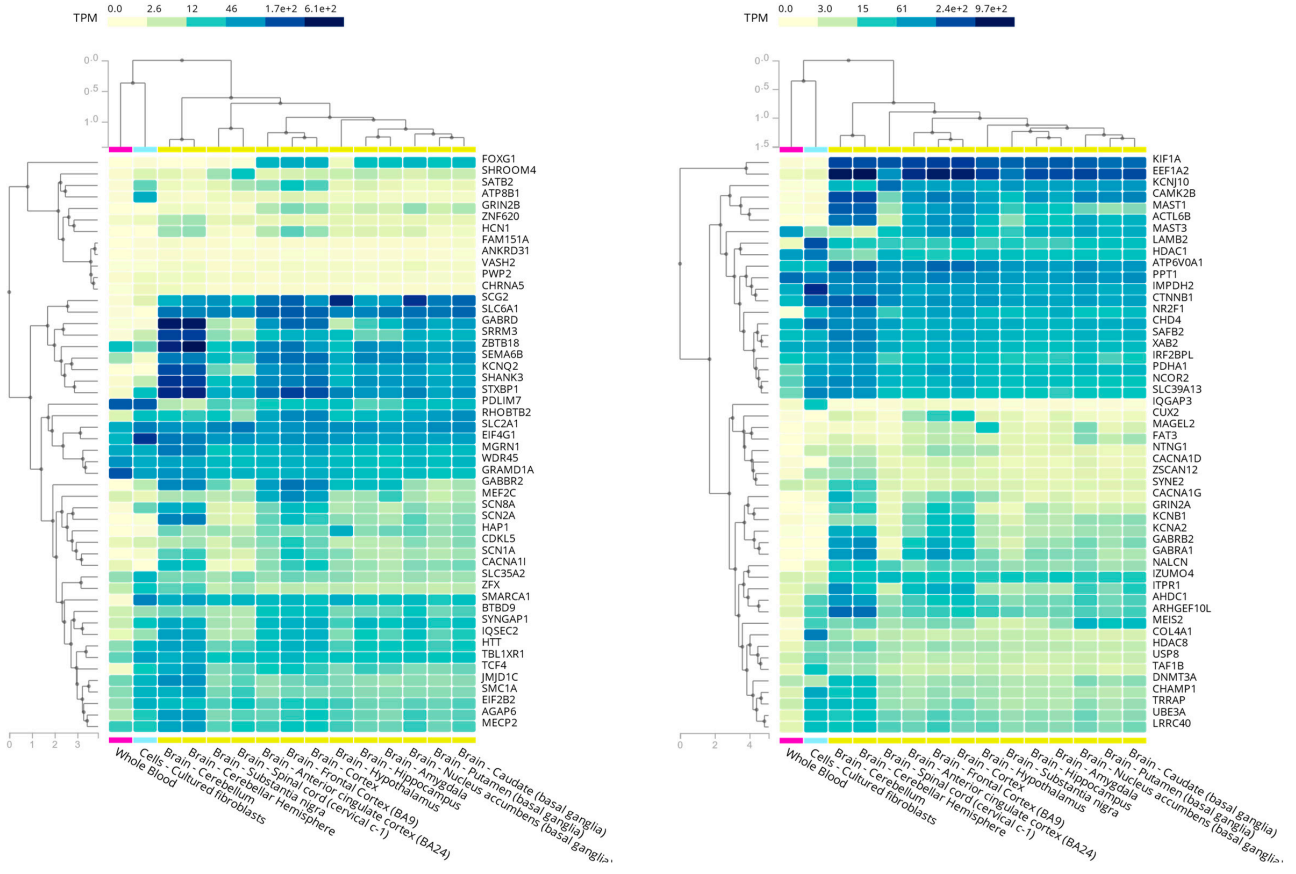


Figure 1. RTT-spectrum-related genes in whole blood, cultured fibroblasts, and different brain regions according to GTEx data (obtained from the GTEx portal on 05/18/21).

4.3.1. Mosaicism in Probands

A mosaic variant in a patient with a neurodevelopmental disorder, such as RTT, could lead to a less specific or a milder phenotype, hindering clinical diagnosis. Furthermore, we must consider that, if the pathogenic variant is present only in the affected tissues (such as the central nervous system), it may be inaccessible and thus remain undetected when sequencing a common peripheral blood sample [105]. Several reports have described male patients with mosaic *MECP2* mutations and different presentations of RTT spectrum disorders, ranging from classical RTT to a mild form resembling the forme fruste [107–113]. A few studies have also implicated *MECP2* mosaicism in female RTT pathogenesis, typically with a phenotypic presentation according to typical or atypical RTT [113,114].

If detected, a low rate of mosaicism may also complicate interpretations of pathogenicity, questioning if a few mutated cells can have a strong enough effect to cause disease. Nevertheless, mosaic mutations with frequencies as low as 1% have been shown to cause focal cortical dysplasia, an epileptogenic neurodevelopmental malformation [115]. *MECP2* mosaic mutations identified in blood samples have been reported to cause disease from frequencies as low as 6.5% in males and 12.28% in females [113].

4.3.2. Mosaicism in Parents

On the other hand, apparently unaffected parents of children with a neurodevelopmental condition can also carry mosaic variants. The mosaicism may be detectable in many tissues but not cause symptoms if there is a low frequency of the pathogenic mutation insufficient to cause disease or causing very mild subclinical manifestations. On the other hand, there are cases where the mutation arises in germ cells and is not present in any other tissue (germline mosaicism). In these cases, the mutation may seem to be de novo if only blood is tested, even though a high proportion of germ cells carry the variant. In these circumstances, correctly detecting parental mosaicism has important implications for adequate genetic counseling, since it increases disease recurrence risk in subsequent pregnancies [105,116,117].

Several familial cases of RTT have been reported where one of the parents had a germline mosaicism that caused them to have more than one RTT case among their offspring, while they remained asymptomatic [17,117–121]. Although most of these case reports implicate maternal germline mosaicism, a recent study in a large cohort of patients with RTT found paternal germline mosaicism in as much as 23.8% of the fathers studied [113]. As these cases raised the awareness on this issue, the detection of parental germline mosaicism as a cause for RTT spectrum disorders has increased considerably in the past few years. Currently, prenatal diagnosis is indicated in any subsequent pregnancy of families with an individual affected by RTT spectrum disorders, even though the detected disease-causing variant is thought to be de novo [105,116].

4.4. Functional Validation of Genomic Variants

The outcome of genomic testing may not always be a conclusive diagnosis. Usually, to reach a final diagnosis directly from the sequencing data, a variant must be identified in a disease-associated gene matching the phenotype of the patient, and this variant should be either a known pathogenic variant previously described and characterized in the literature, or an unknown variant very likely to cause disease [122]. This latter case applies to null variants (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, and single exon or multiexon deletions), which can be assumed to disrupt gene function, when loss of function is a known mechanism of disease for a particular gene [123].

In many other cases, the analysis of sequencing data results in the identification of novel, uncharacterized variants in disease-associated genes, that are classified as VUSs. Additionally, if the analysis is not targeted to previously known disease-genes, it may lead to the identification of genes of unknown significance (GUSs). GUSs are genes with no solid evidence of being disease-causing that may be interesting candidates given their functional roles or molecular interactions.

In order to ascertain the biological consequences of VUSs and reach a definite genetic diagnosis, a functional validation is required [122,124]. A functional assay is an experiment (in vitro or in vivo) that can assess the influence of a VUS on protein function or conformation, and thus help re-classify this variant [125]. In order to re-classify VUSs into either pathogenic or benign variants, the results of functional assays constitute a strong criterion according to the ACMG variant interpretation guidelines [123]. Given the large number of VUSs encountered through genomic testing and the need to validate their functional consequences, the field of "functional genomics" is progressively gaining acknowledgement.

There are diverse approaches to the functional validation of the biological consequences of VUSs [122]. The more extensive, untargeted strategy is the use of multi-omics data in order to find evidence of the malfunctioning at RNA or protein levels of genes carrying VUSs, as discussed above [126]. On the other hand, other functional validation methods are targeted to one or few candidate VUSs previously identified in genomic data. Targeted validation methods include rescue experiments (usually performed with patient-derived cells), where the wild-type allele is introduced to see if the pathogenic phenotype is reverted, and test experiments (usually in model systems), where the VUS is introduced and its consequences assessed [122].

Depending on the type of variant and its predicted biological effect, there are different suitable types of assays [85]. Non-synonymous variants (missense and nonsense) may cause aberrant protein structure and function, leading to decreased gene product levels that can be measured by qRT-PCR or Western blotting. Another possible consequence is erroneous cellular localization, which can be detected by immunocytochemistry assays. To assess the impact of splice-site variants, minigene assays, where the splicing pattern of a subset of exons and introns of a gene is studied, can be used to check the effect on splicing of the candidate variant. Testing the effect of VUSs found in 3′-UTRs and regulatory regions tends to be more complicated. Some assays can be applied though, such as luciferase assays, to compare gene expression levels with and without the candidate variant.

Despite the potential of functional assays to unveil the pathogenicity of VUSs, it must always be taken into account that some biological effects of candidate variants may be tissue or cell-type specific, or they may take place only at a certain developmental stage or under specific environmental conditions [122]. Thus, the choice of a relevant assay and model system (patient-derived material, commercial cell-lines, or animal models, etc.), together with a cautious interpretation of the experimental results, are key to successful outcomes.

Functional validation has been carried out to confirm the pathogenicity of several VUSs associated with RTT spectrum disorders [63,127]. For instance, a specific functional assay demonstrated that a recurrent de novo missense variant in *GABBR2*, which was found in 3 unrelated RTT-like patients, impaired the activity of the mutated protein, thus confirming its pathogenicity [63].

As functional assays can potentially transform a possible diagnosis into a certain diagnosis, it is important to implement functional genomics in a diagnostic setting. The major setbacks for this implementation are high economic costs and long turnaround times [122]. To ensure that the limited resources are spent on the functional validation of the most relevant VUSs, expert clinical geneticists should be involved in the selection of candidate variants. Moreover, as more and more functional studies are performed worldwide, the results of these experiments should be stored in specialized databases, which combined with newly developed machine learning methods, could generate variant pathogenicity predictions with the highest accuracy [128]. The access to this valuable information would streamline variant interpretation and shorten turnaround times in future patients.

5. Conclusions

Over the last 20 years, the concept of RTT spectrum disorders has evolved from a monogenic disease, towards a spectrum of overlapping phenotypes caused by pathogenic

variants in a great number of genes. The genetic diagnosis techniques applied to this clinical entity have concurrently progressed from single-gene genetic testing to genome-wide approaches and integration of different types of data (Figure 2 summarizes the diagnostic approaches discussed in this review).

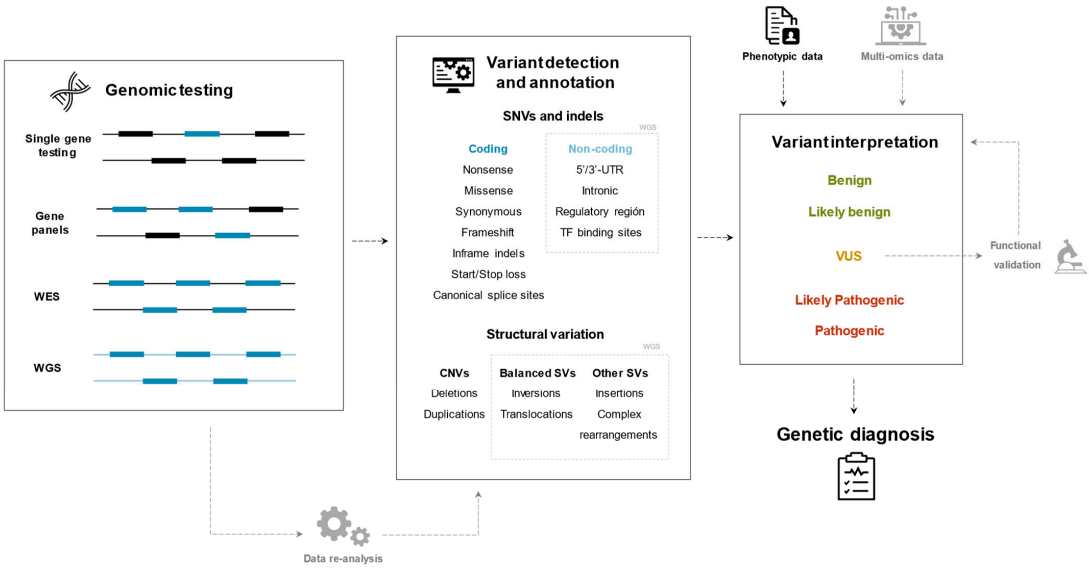


Figure 2. Summary flowchart of the diagnostic approaches presented in this review.

The fact that the dysfunction of different interconnected genes can give rise to such similar phenotypes suggests the implication of common pathways in the pathophysiology of the disease, but these mechanisms currently remain unknown. Moreover, the high phenotypic resemblance between patients with pathogenic variants in different genes and the also high phenotypic variability among patients with mutations in the same gene complicate the clinical diagnosis. That is why single-gene genetic testing can be inefficient and the more cost-effective solution is multiple gene approaches enabled by NGS technologies. The ability to reanalyze, extend, or redirect a genetic analysis can benefit the patients and their families by speeding up the diagnostic process. A recent study concluded that if WES or WGS had been performed at symptom onset, genetic diagnoses of NDDs could have been reached more than 6 years earlier [93].

The field of genetics is evolving rapidly and the key to an efficient genetic diagnosis lies in a combination of technological progress and organized knowledge. On the one hand, cutting-edge technologies, such as WGS and multi-omics, are expanding the boundaries of researchers towards more genomic regions and new functional levels. On the other hand, new knowledge is being generated about regulatory regions, gene expression and chromatin organization, which will enable the correct identification and interpretation of pathogenic variants within the huge amount of data generated. In the case of RTT spectrum disorders, with high genetic heterogeneity and significant phenotypic overlap, a thorough clinical characterization remains crucial for assessing the pathogenicity of the identified variants and their relationship with the phenotypes of patients.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms221910375/s1, Table S1: Genes associated with RTT spectrum disorders in case reports from the HGMD Professional database (June 2021).

Author Contributions: C.X. and J.A. conceptualized the review; C.X. and M.H. wrote the original draft; C.X., M.H., A.P.-A., A.O. and J.A. reviewed and edited the article. J.A. and supervised. All

authors reviewed the article critically for intellectual content and have agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the Spanish Ministry of Health (Instituto de Salud Carlos III/FEDER, PI15/01159 and PI20/00389), an FPU (Formación del Profesorado Universitario) doctoral grant from the Spanish Ministry of Science, Innovation and Universities (FPU18/02152) to Clara Xiol, and funding from Muévete por los que no Pueden (PCP/00282).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no competing interests. 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. Leonard, H.; Cobb, S.; Downs, J. Clinical and biological progress over 50 years in Rett syndrome. *Nat. Rev. Neurol.* **2016**, *13*, 37–51. [CrossRef]
2. Liyanage, V.R.B.; Rastegar, M. Rett syndrome and MeCP2. *Neuromol. Med.* **2014**, *16*, 231–264. [CrossRef]
3. Weaving, L.S.; Ellaway, C.J.; Gecz, J.; Christodoulou, J. Rett syndrome: Clinical review and genetic update. *J. Med. Genet.* **2005**, *42*, 1–7. [CrossRef]
4. 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]
5. Hagberg, B. Clinical manifestations and stages of Rett syndrome. *Ment. Retard. Dev. Disabil. Res. Rev.* **2002**, *8*, 61–65. [CrossRef]
6. Percy, A.K.; Lane, J.; Annese, F.; Warren, H.; Skinner, S.A.; Neul, J.L. When Rett syndrome is due to genes other than MECP2. *Transl. Sci. Rare Dis.* **2018**, *3*, 49–53. [CrossRef]
7. Zappella, M. The Rett girls with preserved speech. *Brain Dev.* **1992**, *14*, 98–101. [CrossRef]
8. 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.* **2019**, *107*, 115–135. [CrossRef] [PubMed]
9. Neul, J.L.; Lane, J.B.; Lee, H.-S.; Geerts, S.; Barrish, J.O.; Annese, F.; Baggett, L.M.; 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] [PubMed]
10. Einspieler, C.; Marschik, P.B. Regression in Rett syndrome: Developmental pathways to its onset. *Neurosci. Biobehav. Rev.* **2019**, *98*, 320–332. [CrossRef] [PubMed]
11. 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**, *20*, 3925. [CrossRef] [PubMed]
12. Ehrhart, F.; Sangani, N.B.; Curfs, L.M. Current developments in the genetics of Rett and Rett-like syndrome. *Curr. Opin. Psychiatry* **2018**, *31*, 103–108. [CrossRef]
13. Schönewolf-Greulich, B.; Bisgaard, A.-M.; Møller, R.; Dunø, M.; Brøndum-Nielsen, K.; Kaur, S.; van Bergen, N.; Lunke, S.; Eggers, S.; Jespersgaard, C.; et al. Clinician's guide to genes associated with Rett-like phenotypes-investigation of a Danish cohort and review of the literature. *Clin. Genet.* **2018**, *95*, 221–230. [CrossRef] [PubMed]
14. Vidal, S.; Brandi, N.; Pacheco, P.; Maynou, J.; Fernandez, G.; Xiol, C.; Pascual-Alonso, A.; Pineda, M.; Armstrong, J.; del Mar, O.M.; 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]
15. Iwama, K.; Mizuguchi, T.; Takeshita, E.; Nakagawa, E.; Okazaki, T.; Nomura, Y.; Iijima, Y.; Kajiura, I.; Sugai, K.; Saito, T.; et al. Genetic landscape of Rett syndrome-like phenotypes revealed by whole exome sequencing. *J. Med. Genet.* **2019**, *56*, 396–407. [CrossRef]
16. Lopergolo, D.; Privitera, F.; Castello, G.; Rizzo, C.L.; Mencarelli, M.A.; Pinto, A.M.; Ariani, F.; Currò, A.; Lamacchia, V.; Canitano, R.; et al. IQSEC2 disorder: A new disease entity or a Rett spectrum continuum? *Clin. Genet.* **2021**, *99*, 462–474. [CrossRef]
17. 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]
18. Ip, J.P.K.; Mellios, N.; Sur, M. Rett syndrome: Insights into genetic, molecular and circuit mechanisms. *Nat. Rev. Neurosci.* **2018**, *19*, 368–382. [CrossRef]
19. Lyst, M.; Bird, A. Rett syndrome: A complex disorder with simple roots. *Nat. Rev. Genet.* **2015**, *16*, 261–275. [CrossRef]
20. Marano, D.; Fioriniello, S.; D'Esposito, M.; Della Ragione, F. Transcriptomic and epigenomic landscape in rett syndrome. *Biomolecules* **2021**, *11*, 967. [CrossRef]
21. Della Ragione, F.; Filosa, S.; Scalabrì, F.; D'Esposito, M. MeCP2 as a genome-wide modulator: The renewal of an old story. *Front. Genet.* **2012**, *3*, 181. [CrossRef] [PubMed]
22. Kriaucionis, S. The major form of MeCP2 has a novel N-terminus generated by alternative splicing. *Nucleic Acids Res.* **2004**, *32*, 1818–1823. [CrossRef] [PubMed]

23. Mnatzakanian, G.N.; Lohi, H.; Munteanu, I.; Alfred, S.E.; Yamada, T.; MacLeod, P.J.M.; Jones, J.R.; Scherer, S.; 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] [PubMed]
24. 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]
25. Dragich, J.M.; Kim, Y.-H.; Arnold, A.P.; Schanen, C. Differential distribution of the Mecp2 splice variants in the postnatal mouse. *Brain. J. Comp. Neurol.* **2007**, *501*, 526–542. [CrossRef]
26. De Paz, A.M.; Khajavi, L.; Martin, H.; Claveria-Gimeno, R.; Dieck, S.T.; 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. *Epigenet. Chromatin* **2019**, *12*, 1–16. [CrossRef]
27. Percy, A.K.; Neul, J.L.; Glaze, D.G.; Motil, K.J.; Skinner, S.A.; Khwaja, O.; Lee, H.-S.; Lane, J.B.; Barrish, J.O.; Annese, F.; et al. Rett syndrome diagnostic criteria: Lessons from the natural history study. *Ann. Neurol.* **2010**, *68*, 951–955. [CrossRef]
28. De Bona, C.; Zappella, M.; Hayek, J.; Meloni, I.; Vitelli, F.; Bruttini, M.; Cusano, R.; Loffredo, P.; Longo, I.; Renieri, A. Preserved speech variant is allelic of classic Rett syndrome. *Eur. J. Hum. Genet.* **2000**, *8*, 325–330. [CrossRef]
29. Krishnaraj, R.; Ho, G.; Christodoulou, J. RettBASE: Rett syndrome database update. *Hum. Mutat.* **2017**, *38*, 922–931. [CrossRef]
30. Stenson, P.D.; Mort, M.; Ball, E.V.; Chapman, M.; Evans, K.; Azevedo, L.; Hayden, M.; Heywood, S.; Millar, D.S.; Phillips, A.D.; et al. The Human Gene Mutation Database (HGMD®): Optimizing its use in a clinical diagnostic or research setting. *Qual. Life Res.* **2020**, *139*, 1197–1207. [CrossRef]
31. Gianakopoulos, P.J.; Zhang, Y.; Pencea, N.; Orlic-Milacic, M.; Mittal, K.; Windpassinger, C.; White, S.-J.; Kroisel, P.M.; Chow, E.W.; Saunders, C.J.; et al. Mutations in MECP2 exon 1 in classical Rett patients disrupt MECP2_e1 transcription, but not transcription of MECP2_e2. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **2012**, *159B*, 210–216. [CrossRef]
32. Huppke, P.; Gärtner, J. Molecular diagnosis of Rett syndrome. *J. Child Neurol.* **2005**, *20*, 732–736. [CrossRef]
33. Thistlethwaite, W.A.; Moses, L.M.; Hoffbuhr, K.C.; Devaney, J.M.; Hoffman, E.P. Rapid genotyping of common MeCP2 mutations with an electronic DNA microchip using serial differential hybridization. *J. Mol. Diagn.* **2003**, *5*, 121–126. [CrossRef]
34. Mari, F.; Azimonti, S.; Bertani, I.; Bolognese, F.; Colombo, E.; Caselli, R.; Scala, E.; Longo, I.; Grosso, S.; Pescucci, C.; et al. CDKL5 belongs to the same molecular pathway of MeCP2 and it is responsible for the early-onset seizure variant of Rett syndrome. *Hum. Mol. Genet.* **2005**, *14*, 1935–1946. [CrossRef]
35. Tao, J.; Hu, K.; Chang, Q.; Wu, H.; Sherman, N.; Martinowich, K.; Klose, R.J.; Schanen, C.; Jaenisch, R.; Wang, W.; et al. Phosphorylation of MeCP2 at serine 80 regulates its chromatin association and neurological function. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4882–4887. [CrossRef]
36. Weaving, L.S.; Christodoulou, J.; Williamson, S.L.; Friend, K.L.; McKenzie, O.L.; Archer, H.; Evans, J.; Clarke, A.; Pelka, G.J.; Tam, P.P.; et al. Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. *Am. J. Hum. Genet.* **2004**, *75*, 1079–1093. [CrossRef]
37. Carouge, D.; Host, L.; Aunis, D.; Zwiller, J.; Anglard, P. CDKL5 is a brain MeCP2 target gene regulated by DNA methylation. *Neurobiol. Dis.* **2010**, *38*, 414–424. [CrossRef] [PubMed]
38. Papa, F.T.; Mencarelli, M.A.; Caselli, R.; Katzaki, E.; Sampieri, K.; Meloni, I.; Ariani, F.; Longo, I.; Maggio, A.; Balestri, P.; et al. A 3 Mb deletion in 14q12 causes severe mental retardation, mild facial dysmorphisms and Rett-like features. *Am. J. Med. Genet. Part A* **2008**, *146A*, 1994–1998. [CrossRef] [PubMed]
39. Mencarelli, M.A.; Spanhol-Rosseto, A.; Artuso, R.; Rondinella, D.; de Filippis, R.; Bahi-Buisson, N.; Nectoux, J.; Rubinsztajn, R.; Bienvenu, T.; Moncla, A.; et al. Novel FOXG1 mutations associated with the congenital variant of Rett syndrome. *J. Med. Genet.* **2009**, *47*, 49–53. [CrossRef] [PubMed]
40. Ariani, F.; Hayek, J.; Rondinella, D.; Artuso, R.; Mencarelli, M.A.; Spanhol-Rosseto, A.; Pollazzon, M.; Buoni, S.; Spiga, O.; Ricciardi, S.; et al. FOXG1 is responsible for the congenital variant of Rett syndrome. *Am. J. Hum. Genet.* **2008**, *83*, 89–93. [CrossRef] [PubMed]
41. Yao, J.; Lai, E.; Stifani, S. The winged-helix protein brain factor 1 interacts with groucho and hes proteins to repress transcription. *Mol. Cell. Biol.* **2001**, *21*, 1962–1972. [CrossRef] [PubMed]
42. Wong, L.-C.; Singh, S.; Wang, H.-P.; Hsu, C.-J.; Hu, S.-C.; Lee, W.-T. FOXG1-related syndrome: From clinical to molecular genetics and pathogenic mechanisms. *Int. J. Mol. Sci.* **2019**, *20*, 4176. [CrossRef] [PubMed]
43. Kadam, S.D.; Sullivan, B.J.; Goyal, A.; Blue, M.E.; Smith-Hicks, C. Rett syndrome and CDKL5 deficiency disorder: From bench to clinic. *Int. J. Mol. Sci.* **2019**, *20*, 5098. [CrossRef] [PubMed]
44. Vidal, S.; Brandi, N.; Pacheco, P.; Gerotina, E.; Blasco, L.; Trotta, J.-R.; Derdak, S.; O'Callaghan, M.D.M.; Garcia-Cazorla, À.; Pineda, M.; et al. The utility of next generation sequencing for molecular diagnostics in Rett syndrome. *Sci. Rep.* **2017**, *7*, 12288. [CrossRef]
45. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **2008**, *24*, 133–141. [CrossRef]
46. Wright, C.F.; FitzPatrick, D.R.; Firth, H.V. Paediatric genomics: Diagnosing rare disease in children. *Nat. Rev. Genet.* **2018**, *19*, 253–268. [CrossRef]
47. Chérot, E.; Keren, B.; Dubourg, C.; Carré, W.; Fradin, M.; Lavillaureix, A.; Afenjar, A.; Burglen, L.; Whalen, S.; Charles, P.; et al. Using medical exome sequencing to identify the causes of neurodevelopmental disorders: Experience of 2 clinical units and 216 patients. *Clin. Genet.* **2017**, *93*, 567–576. [CrossRef]

48. Lee, H.; Deignan, J.L.; Dorrani, N.; Strom, S.P.; Kantarci, S.; Quintero-Rivera, F.; Das, K.; Toy, T.; Harry, B.; Yourshaw, M.; et al. Clinical exome sequencing for genetic identification of rare mendelian disorders. *JAMA* **2014**, *312*, 1880–1887. [CrossRef]
49. Yang, Y.; Muzny, D.M.; Reid, J.G.; Bainbridge, M.N.; Willis, A.; Ward, P.A.; Braxton, A.; Beuten, J.; Xia, F.; Niu, Z.; et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **2013**, *369*, 1502–1511. [CrossRef]
50. Srivastava, S.; Love-Nichols, J.A.; Dies, K.A.; Ledbetter, D.H.; Martin, C.L.; Chung, W.K.; Firth, H.V.; Frazier, T.; Hansen, R.L.; Prock, L.; et al. Meta-analysis and multidisciplinary consensus statement: Exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet. Med.* **2019**, *21*, 2413–2421. [CrossRef]
51. Manickam, K.; McClain, M.R.; Demmer, L.A.; Biswas, S.; Kearney, H.M.; Malinowski, J.; Massingham, L.J.; Miller, D.; Yu, T.W.; Hisama, F.M.; et al. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: An evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **2021**, *2021*, 1–9. [CrossRef]
52. Retterer, K.; Juusola, J.; Cho, M.T.; Vitazka, P.; Millan, F.; Gibellini, F.; Vertino-Bell, A.; Smaoui, N.; Neidich, J.; Monaghan, K.G.; et al. Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* **2016**, *18*, 696–704. [CrossRef]
53. Farwell, K.D.; Shahmirzadi, L.; El-Khechen, D.; Powis, Z.; Chao, E.C.; Davis, B.T.; Baxter, R.M.; Zeng, W.; Mroske, C.; Parra, M.C.; et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model–based analysis: Results from 500 unselected families with undiagnosed genetic conditions. *Genet. Med.* **2015**, *17*, 578–586. [CrossRef]
54. Fitzgerald, T.W.; Gerety, S.S.; Jones, W.D.; van Kogelenberg, M.; King, D.A.; McRae, J.; Morley, K.I.; Parthiban, V.; Al-Turki, S.; Ambridge, K.; et al. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **2015**, *519*, 223–228. [CrossRef]
55. Stefanski, A.; Calle-López, Y.; Leu, C.; Pérez-Palma, E.; Pestana-Knight, E.; Lal, D. Clinical sequencing yield in epilepsy, autism spectrum disorder, and intellectual disability: A systematic review and meta-analysis. *Epilepsia* **2021**, *62*, 143–151. [CrossRef]
56. Cheng, A.Y.; Teo, Y.-Y.; Ong, R.T.-H. Assessing single nucleotide variant detection and genotype calling on whole-genome sequenced individuals. *Bioinformatics* **2014**, *30*, 1707–1713. [CrossRef]
57. De Ligt, J.; Boone, P.; Pfundt, R.; Vissers, L.; Richmond, T.; Geoghegan, J.; O'Moore, K.; de Leeuw, N.; Shaw, C.; Brunner, H.G.; et al. Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum. Mutat.* **2013**, *34*, 1439–1448. [CrossRef]
58. Deciphering Developmental Disorders Study. Prevalence and Architecture of de Novo Mutations in Developmental Disorders. *Nature* **2018**, *542*, 433–438. [CrossRef]
59. Olson, H.E.; Tambunan, D.; LaCoursiere, C.; Goldenberg, M.; Pinsky, R.; Martin, E.; Ho, E.; Khwaja, O.; Kaufmann, W.E.; Poduri, A. Mutations in epilepsy and intellectual disability genes in patients with features of Rett syndrome. *Am. J. Med. Genet. Part A* **2015**, *167*, 2017–2025. [CrossRef]
60. Lucariello, M.; Vidal, E.; Vidal, S.; Saez, M.; Roa, L.; Huertas, D.; Pineda, M.; Dalfó, E.; Dopazo, J.; Jurado, P.; et al. Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype. *Qual. Life Res.* **2016**, *135*, 1343–1354. [CrossRef] [PubMed]
61. Lopes, F.; Barbosa, M.; Ameur, A.; Soares, G.; de Sá, J.; Dias, A.I.; Oliveira, G.; Cabral, P.; Temudo, T.; Calado, E.; et al. Identification of novel genetic causes of Rett syndrome-like phenotypes. *J. Med. Genet.* **2016**, *53*, 190–199. [CrossRef] [PubMed]
62. Sajan, S.A.; Jhangiani, S.N.; Muzny, D.M.; Gibbs, R.A.; Lupski, J.R.; Glaze, D.G.; Kaufmann, W.E.; Skinner, S.A.; Annese, F.; Friez, M.J.; et al. Enrichment of mutations in chromatin regulators in people with Rett syndrome lacking mutations in MECP2. *Genet. Med.* **2017**, *19*, 13–19. [CrossRef] [PubMed]
63. Yoo, Y.; Jung, J.; Lee, Y.-N.; Lee, Y.; Cho, H.; Na, E.; Hong, J.; Kim, E.; Lee, J.S.; Lee, J.S.; et al. GABBR2mutations determine phenotype in Rett syndrome and epileptic encephalopathy. *Ann. Neurol.* **2017**, *82*, 466–478. [CrossRef]
64. Okamoto, N.; Miya, F.; Tsunoda, T.; Kato, M.; Saitoh, S.; Yamasaki, M.; Shimizu, A.; Torii, C.; Kanemura, Y.; Kosaki, K. Targeted next-generation sequencing in the diagnosis of neurodevelopmental disorders. *Clin. Genet.* **2014**, *88*, 288–292. [CrossRef]
65. Wang, J.; Zhang, Q.; Chen, Y.; Yu, S.; Wu, X.; Bao, X.; Wen, Y. Novel MEF2C point mutations in Chinese patients with Rett (-like) syndrome or non-syndromic intellectual disability: Insights into genotype-phenotype correlation. *BMC Med. Genet.* **2018**, *19*, 191. [CrossRef]
66. Allou, L.; Julia, S.; Amsallem, D.; El Chehadeh, S.; Lambert, L.; Thevenon, J.; Duffourd, Y.; Saunier, A.; Bouquet, P.; Pere, S.; et al. Rett-like phenotypes: Expanding the genetic heterogeneity to the KCNA2 gene and first familial case of CDKL5-related disease. *Clin. Genet.* **2016**, *91*, 431–440. [CrossRef]
67. Srivastava, S.; Desai, S.; Cohen, J.; Smith-Hicks, C.; Barañano, K.; Fatemi, A.; Naidu, S. Monogenic disorders that mimic the phenotype of Rett syndrome. *Neurogenetics* **2018**, *19*, 41–47. [CrossRef] [PubMed]
68. Jiang, Y.-H.; Yuen, R.; Jin, X.; Wang, M.; Chen, N.; Wu, X.; Ju, J.; Mei, J.; Shi, Y.; He, M.; et al. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am. J. Hum. Genet.* **2013**, *93*, 249–263. [CrossRef] [PubMed]
69. Gilissen, C.; Hehir-Kwa, J.Y.; Thung, D.T.; van de Vorst, M.; van Bon, B.W.M.; Willemsen, M.H.; Kwint, M.; Janssen, I.M.; Hoischen, A.; Schenck, A.; et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature* **2014**, *511*, 344–347. [CrossRef]
70. Sáez, M.A.; Fernández-Rodríguez, J.; Moutinho, C.; Sanchez-Mut, J.V.; Gomez, A.; Vidal, E.; Petazzi, P.; Szczesna, K.; López-Serra, P.; Lucariello, M.; et al. Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. *Genet. Med.* **2015**, *18*, 378–385. [CrossRef]

71. Vuillaume, M.; Xue, L.; Blesson, S. A novel mutation in the transmembrane 6 domain of GABBR2 leads to a Rett-like phenotype. *Ann. Neurol.* **2018**, *83*, 437–439. [CrossRef] [PubMed]
72. Krishnaraj, R.; Haase, F.; Coorey, B.; Luca, E.; Wong, I.; Boyling, A.; Ellaway, C.; Christodoulou, J.; Gold, W.A. Genome-wide transcriptomic and proteomic studies of Rett syndrome mouse models identify common signaling pathways and cellular functions as potential therapeutic targets. *Hum. Mutat.* **2019**, *40*, 2184–2196. [CrossRef] [PubMed]
73. Ehrhart, F.; Coort, S.L.; Eijssen, L.; Cirillo, E.; Smeets, E.E.; Sangani, N.B.; Evelo, C.T.; Curfs, L.M. Integrated analysis of human transcriptome data for Rett syndrome finds a network of involved genes. *World J. Biol. Psychiatry* **2020**, *21*, 712–725. [CrossRef]
74. Henriksen, M.W.; Breck, H.; Sejersted, Y.; Diseth, T.; von Tetzchner, S.; Paus, B.; Skjeldal, O.H. Genetic and clinical variations in a Norwegian sample diagnosed with Rett syndrome. *Brain Dev.* **2020**, *42*, 484–495. [CrossRef]
75. Beck, T.F.; Mullikin, J.C.; NISC comparative sequencing program. Systematic evaluation of sanger validation of next-generation sequencing variants. *Clin. Chem.* **2016**, *62*, 647–654. [CrossRef]
76. Alfares, A.; Aloraini, T.; Al Subaie, L.; Alissa, A.; Al Qudsi, A.; Alahmad, A.; Al Mutairi, F.; Alswaid, A.; Alothaim, A.; Eyaid, W.; et al. Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genet. Med.* **2018**, *20*, 1328–1333. [CrossRef]
77. Li, J.; Gao, K.; Yan, H.; Xiangwei, W.; Liu, N.; Wang, T.; Xu, H.; Lin, Z.; Xie, H.; Wang, J.; et al. Reanalysis of whole exome sequencing data in patients with epilepsy and intellectual disability/mental retardation. *Gene* **2019**, *700*, 168–175. [CrossRef]
78. Jalkh, N.; Corbani, S.; Haidar, Z.; Hamdan, N.; Farah, E.; Ghoch, J.A.; Ghosn, R.; Salem, N.; Fawaz, A.; Khayat, C.D.; et al. The added value of WES reanalysis in the field of genetic diagnosis: Lessons learned from 200 exomes in the Lebanese population. *BMC Med. Genom.* **2019**, *12*, 1–7. [CrossRef]
79. Al-Nabhani, M.; Al-Rashdi, S.; Al-Murshedi, F.; Al-Kindi, A.; Al-Thihli, K.; Al-Saegh, A.; Al-Futaisi, A.; Al-Mamari, W.; Zadjali, F.; Al-Maawali, A. Reanalysis of exome sequencing data of intellectual disability samples: Yields and benefits. *Clin. Genet.* **2018**, *94*, 495–501. [CrossRef]
80. Ewans, L.J.; Schofield, D.; Shrestha, R.; Zhu, Y.; Gayevskiy, V.; Ying, K.; Walsh, C.; Lee, E.; Kirk, E.P.; Colley, A.; et al. Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. *Genet. Med.* **2018**, *20*, 1564–1574. [CrossRef] [PubMed]
81. Wright, C.F.; McRae, J.F.; Clayton, S.; Gallone, G.; Aitken, S.; FitzGerald, T.W.; Jones, P.; Prigmore, E.; Rajan, D.; Lord, J.; et al. Making new genetic diagnoses with old data: Iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genet. Med.* **2018**, *20*, 1216–1223. [CrossRef]
82. Majewski, J.; Schwartzentruber, J.; Lalonde, E.; Montpetit, A.; Jabado, N. What can exome sequencing do for you? *J. Med. Genet.* **2011**, *48*, 580–589. [CrossRef] [PubMed]
83. Petersen, B.-S.; Fredrich, B.; Hoeppner, M.P.; Ellinghaus, D.; Franke, A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet.* **2017**, *18*, 1–13. [CrossRef] [PubMed]
84. Lionel, A.C.; Costain, G.; Monfared, N.; Walker, S.; Reuter, M.S.; Hosseini, M.; Thiruvahindrapuram, B.; Merico, D.; Jobling, R.; Nalpathamkalam, T.; et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet. Med.* **2018**, *20*, 435–443. [CrossRef] [PubMed]
85. Hitomi, Y.; Tokunaga, K. Significance of functional disease-causal/susceptible variants identified by whole-genome analyses for the understanding of human diseases. *Proc. Jpn. Acad. Ser. B* **2017**, *93*, 657–676. [CrossRef]
86. Ma, M.; Ru, Y.; Chuang, L.-S.; Hsu, N.-Y.; Shi, L.-S.; Hakenberg, J.; Cheng, W.-Y.; Uzilov, A.; Ding, W.; Glicksberg, B.S.; et al. Disease-associated variants in different categories of disease located in distinct regulatory elements. *BMC Genom.* **2015**, *16*, S3. [CrossRef] [PubMed]
87. Cameron, D.L.; di Stefano, L.; Papenfuss, A.T. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat. Commun.* **2019**, *10*, 1–11. [CrossRef]
88. Ceballos, F.C.; Hazelhurst, S.; Ramsay, M. Assessing runs of homozygosity: A comparison of SNP array and whole genome sequence low coverage data. *BMC Genom.* **2018**, *19*, 1–12. [CrossRef]
89. Magi, A.; Tattini, L.; Palombo, F.; Benelli, M.; Gialluisi, A.; Giusti, B.; Abbate, R.; Seri, M.; Gensini, G.F.; Romeo, G.; et al. H^3M^2: Detection of runs of homozygosity from whole-exome sequencing data. *Bioinformatics* **2014**, *30*, 2852–2859. [CrossRef]
90. King, D.; Fitzgerald, T.; Miller, R.; Canham, N.; Clayton-Smith, J.; Johnson, D.; Mansour, S.; Stewart, F.; Vasudevan, P.; Hurles, M.E.; et al. A novel method for detecting uniparental disomy from trio genotypes identifies a significant excess in children with developmental disorders. *Genome Res.* **2013**, *24*, 673–687. [CrossRef]
91. Yamazawa, K.; Ogata, T.; Ferguson-Smith, A.C. Uniparental disomy and human disease: An overview. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **2010**, *154C*, 329–334. [CrossRef]
92. Belkadi, A.; Bolze, A.; Itan, Y.; Cobat, A.; Vincent, Q.B.; Antipenko, A.; Shang, L.; Boisson, B.; Casanova, J.-L.; Abel, L. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5473–5478. [CrossRef]
93. Soden, S.E.; Saunders, C.J.; Willig, L.K.; Farrow, E.G.; Smith, L.D.; Petrikin, J.E.; Lepichon, J.-B.; Miller, N.A.; Thiffault, I.; Dinwiddie, D.L.; et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci. Transl. Med.* **2014**, *6*, 265ra168. [CrossRef]
94. Kremer, L.S.; Wortmann, S.B.; Prokisch, H. "Transcriptomics": Molecular diagnosis of inborn errors of metabolism via RNA-sequencing. *J. Inherit. Metab. Dis.* **2018**, *41*, 525–532. [CrossRef]

95. Karczewski, K.J.; Snyder, M.P. Integrative omics for health and disease. *Nat. Rev. Genet.* **2018**, *19*, 299–310. [CrossRef] [PubMed]
96. 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.* **2020**, *43*, 25–35. [CrossRef]
97. Kremer, L.S.; Bader, D.M.; Mertes, C.; Kopajtich, R.; Pichler, G.; Iuso, A.; Haack, T.B.; Graf, E.; Schwarzmayr, T.; Terrile, C.; et al. Genetic diagnosis of mendelian disorders via RNA sequencing. *Nat. Commun.* **2017**, *8*, 15824. [CrossRef]
98. Ferraro, N.M.; Strober, B.J.; Einson, J.; Abell, N.S.; Aguet, F.; Barbeira, A.N.; Brandt, M.; Bucan, M.; Castel, S.E.; Davis, J.R.; et al. Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science* **2020**, *369*, eaaz5900. [CrossRef] [PubMed]
99. Aebersold, R.; Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537*, 347–355. [CrossRef] [PubMed]
100. Gonorazky, H.D.; Naumenko, S.; Ramani, A.K.; Nelakuditi, V.; Mashouri, P.; Wang, P.; Kao, D.; Ohri, K.; Viththiyapaskaran, S.; Tarnopolsky, M.A.; et al. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease. *Am. J. Hum. Genet.* **2019**, *104*, 466–483. [CrossRef]
101. Cummings, B.B.; Marshall, J.L.; Tukiainen, T.; Lek, M.; Donkervoort, S.; Foley, A.R.; Bolduc, V.; Waddell, L.B.; Sandaradura, S.A.; O'Grady, G.L.; et al. Improving genetic diagnosis in mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* **2017**, *9*, eaal5209. [CrossRef] [PubMed]
102. Amberger, J.S.; Bocchini, C.A.; Schiettecatte, F.; Scott, A.F.; Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **2015**, *43*, D789–D798. [CrossRef] [PubMed]
103. D'Gama, A.M.; Walsh, C.A. Somatic mosaicism and neurodevelopmental disease. *Nat. Neurosci.* **2018**, *21*, 1504–1514. [CrossRef] [PubMed]
104. Hidalgo, R.A.; Bo, T.; Kwint, M.P.; van de Vorst, M.; Pinelli, M.; Veltman, J.; Hoischen, A.; Vissers, L.; Gilissen, C. Post-zygotic point mutations are an underrecognized source of de novo genomic variation. *Am. J. Hum. Genet.* **2015**, *97*, 67–74. [CrossRef] [PubMed]
105. Biesecker, L.G.; Spinner, N.B. A genomic view of mosaicism and human disease. *Nat. Rev. Genet.* **2013**, *14*, 307–320. [CrossRef]
106. Møller, R.S.; Liebmann, N.; Larsen, L.H.G.; Stiller, M.; Hentschel, J.; Kako, N.; Abdin, D.; di Donato, N.; Pal, D.K.; Zacher, P.; et al. Parental mosaicism in epilepsies due to alleged de novo variants. *Epilepsia* **2019**, *60*, e63–e66. [CrossRef]
107. Armstrong, J.; Poo, P.; Pineda, M.; Aibar, E.; Gean, E.; Català, V.; Monrós, E. Classic Rett syndrome in a boy as a result of somatic mosaicism for a mecp2 mutation. *Ann. Neurol.* **2001**, *50*, 692. [CrossRef]
108. Pieras, J.I.; Muñoz-Cabello, B.; Borrego, S.; Marcos, I.; Sanchez, J.; Madruga, M.; Antiñolo, G. Somatic mosaicism for Y120X mutation in the MECP2 gene causes atypical Rett syndrome in a male. *Brain Dev.* **2011**, *33*, 608–611. [CrossRef]
109. Clayton-Smith, J.; Watson, P.; Ramsden, S.; Black, G. Somatic mutation in MECP2 as a non-fatal neurodevelopmental disorder in males. *Lancet* **2000**, *356*, 830–832. [CrossRef]
110. Kleefstra, T.; Yntema, H.G.; Nillesen, W.M.; Oudakker, A.R.; Mullaart, R.A.; Geerdink, N.; van Bokhoven, H.; de Vries, B.B.; Sistermans, E.A.; Hamel, B.C.; et al. MECP2 analysis in mentally retarded patients: Implications for routine DNA diagnostics. *Eur. J. Hum. Genet.* **2003**, *12*, 24–28. [CrossRef]
111. 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]
112. Topçu, M.; Akyerli, C.; Sayı, A.; Törüner, G.A.; Koçoğlu, S.R.; Cimbiş, M.; Özçelik, T. Somatic mosaicism for a MECP2 mutation associated with classic Rett syndrome in a boy. *Eur. J. Hum. Genet.* **2002**, *10*, 77–81. [CrossRef]
113. Zhang, Q.; Yang, X.; Wang, J.; Li, J.; Wu, Q.; Wen, Y.; Zhao, Y.; Zhang, X.; Yao, H.; Wu, X.; et al. Genomic mosaicism in the pathogenesis and inheritance of a Rett syndrome cohort. *Genet. Med.* **2019**, *21*, 1330–1338. [CrossRef] [PubMed]
114. Bourdon, V.; Philippe, C.; Bienvenu, T.; Koenig, B.; Tardieu, M.; Chelly, J.; Jonveaux, P. Evidence of somatic mosaicism for a MECP2 mutation in females with Rett syndrome: Diagnostic implications. *J. Med. Genet.* **2001**, *38*, 867–871. [CrossRef] [PubMed]
115. D'Gama, A.M.; Woodworth, M.B.; Hossain, A.A.; Bizzotto, S.; Hatem, N.E.; LaCoursiere, C.M.; Najm, I.; Ying, Z.; Yang, E.; Barkovich, A.J.; et al. Somatic mutations activating the mTOR pathway in dorsal telencephalic progenitors cause a continuum of cortical dysplasias. *Cell Rep.* **2017**, *21*, 3754–3766. [CrossRef]
116. Mari, F.; Caselli, R.; Russo, S.; Cogliati, F.; Ariani, F.; Longo, I.; Bruttini, M.; Meloni, I.; Pescucci, C.; Schürfeld, K.; et al. Germline mosaicism in Rett syndrome identified by prenatal diagnosis. *Clin. Genet.* **2005**, *67*, 258–260. [CrossRef]
117. Evans, J.; Archer, H.; Whatley, S.; Clarke, A. Germline mosaicism for a MECP2 mutation in a man with two Rett daughters. *Clin. Genet.* **2006**, *70*, 336–338. [CrossRef] [PubMed]
118. Venâncio, M.; Santos, M.; Pereira, S.A.; Maciel, P.; Saraiva, J.M.; Ven, M. An explanation for another familial case of Rett syndrome: Maternal germline mosaicism. *Eur. J. Hum. Genet.* **2007**, *15*, 902–904. [CrossRef]
119. Wan, M.; Lee, S.S.J.; Zhang, X.; Houwink-Manville, I.; Song, H.-R.; Amir, R.E.; Budden, S.; Naidu, S.; Pereira, J.L.P.; Lo, I.; et al. Rett syndrome and beyond: Recurrent spontaneous and familial MECP2 mutations at CpG hotspots. *Am. J. Hum. Genet.* **1999**, *65*, 1520–1529. [CrossRef]
120. Yaron, Y.; Ben Zeev, B.; Shomrat, R.; Bercovich, D.; Naiman, T.; Orr-Urtreger, A. MECP2 mutations in Israel: Implications for molecular analysis, genetic counseling, and prenatal diagnosis in Rett syndrome. *Hum. Mutat.* **2002**, *20*, 323–324. [CrossRef]

121. Villard, L.; Levy, N.; Xiang, F.; Kpebe, A.; Labelle, V.; Chevillard, C.; Zhang, Z.; Schwartz, C.E.; Tardieu, M.; Chelly, J.; et al. Segregation of a totally skewed pattern of X chromosome inactivation in four familial cases of Rett syndrome without MECP2 mutation: Implications for the disease. *J. Med. Genet.* **2001**, *38*, 435–442. [CrossRef] [PubMed]
122. Rodenburg, R.J. The functional genomics laboratory: Functional validation of genetic variants. *J. Inherit. Metab. Dis.* **2018**, *41*, 297–307. [CrossRef] [PubMed]
123. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; 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**, *17*, 405–423. [CrossRef] [PubMed]
124. Raraigh, K.S.; Han, S.; Davis, E.; Evans, T.A.; Pellicore, M.; McCague, A.F.; Joynt, A.T.; Lu, Z.; Atalar, M.; Sharma, N.; et al. Functional assays are essential for interpretation of missense variants associated with variable expressivity. *Am. J. Hum. Genet.* **2018**, *102*, 1062–1077. [CrossRef] [PubMed]
125. Thouvenot, P.; Ben Yamin, B.; Fourriere, L.; Lescure, A.; Boudier, T.; del Nery, E.; Chauchereau, A.; Goldgar, D.E.; Houdayer, C.; Stoppa-Lyonnet, D.; et al. Functional assessment of genetic variants with outcomes adapted to clinical decision-making. *PLoS Genet.* **2016**, *12*, e1006096. [CrossRef]
126. Lappalainen, T.; Scott, A.; Brandt, M.; Hall, I.M. Genomic analysis in the age of human genome sequencing. *Cell* **2019**, *177*, 70–84. [CrossRef]
127. Soto, D.; Olivella, M.; Grau, C.; Armstrong, J.; Alcon, C.; Gasull, X.; Santos-Gómez, A.; Locubiche, S.; de Salazar, M.G.; García-Díaz, R.; et al. L-serine dietary supplementation is associated with clinical improvement of loss-of-function GRIN2B-related pediatric encephalopathy. *Sci. Signal.* **2019**, *12*, eaaw0936. [CrossRef]
128. Starita, L.M.; Ahituv, N.; Dunham, M.J.; Kitzman, J.O.; Roth, F.P.; Seelig, G.; Shendure, J.; Fowler, D.M. Variant interpretation: Functional assays to the rescue. *Am. J. Hum. Genet.* **2017**, *101*, 315–325. [CrossRef]

ANNEX 3

Other publications (collaborations)

Musokhranova, U., Grau, C., Vergara, C., Rodríguez-Pascau, L., **Xiol, C.**, Castells, AA., Alcántara, S., Rodríguez-Pombo, P., Pizcueta, P., Martinell, M., García-Cazorla, A., Oyarzabal, A.. *Mitochondrial modulation with leriglitazone as a potential treatment for Rett syndrome*. Journal of Translational Medicine (Accepted 12/10/2023).

Petazzi, P., Jorge-Torres, O. C., Gomez, A., Scognamiglio, I., Serra-Musach, J., Merkel, A., Grases, D., **Xiol, C.**, O'Callaghan, M., Armstrong, J., Esteller, M., & Guil, S. (2023). *Global Impairment of Immediate-Early Genes Expression in Rett Syndrome Models and Patients Linked to Myelination Defects*. International Journal of Molecular Sciences, 24(2), 1453. https://doi.org/10.3390/ijms24021453.

Siqueira, E., Obiols-Guardia, A., Jorge-Torres, O. C., Oliveira-Mateos, C., Soler, M., Ramesh-Kumar, D., Setién, F., van Rossum, D., Pascual-Alonso, A., **Xiol, C.**, Ivan, C., Shimizu, M., Armstrong, J., Calin, G. A., Pasterkamp, R. J., Esteller, M., & Guil, S. (2021). *Analysis of the circRNA and T-UCR populations identifies convergent pathways in mouse and human models of Rett syndrome*. Molecular Therapy. Nucleic Acids, 27, 621–644. https://doi.org/10.1016/j.omtn.2021.12.030.

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., Brandi, N., Pacheco, P., Ros, C., Del Campo, M., Guillén, E., Ibañez, S., Sánchez, M. J., Lapunzina, P., Nevado, J., Santos, F., … Armstrong, J. (2020). *Molecular characterization of Spanish patients with MECP2 duplication syndrome*. Clinical Genetics, 97(4), 610–620. https://doi.org/10.1111/cge.13718.

Vidal, S., Pascual-Alonso, A., Rabaza-Gairí, M., Gerotina, E., Brandi, N., Pacheco, P., **Xiol, C.**, Pineda, M., Rett Working Group, & Armstrong, J. (2019). *Characterization of large deletions of the MECP2 gene in Rett syndrome patients by gene dosage analysis*. Molecular Genetics & Genomic Medicine, 7(8), e793. https://doi.org/10.1002/mgg3.793.

Vidal, S., Brandi, N., Pacheco, P., Maynou, J., Fernandez, G., **Xiol, C.**, Pascual-Alonso, A., Pineda, M., Rett Working Group, & Armstrong, J. (2019). *The most recurrent monogenic disorders that overlap with the phenotype of Rett syndrome*. European journal of paediatric neurology : EJPN : Official Journal of the European Paediatric Neurology Society, 23(4), 609–620. https://doi.org/10.1016/j.ejpn.2019.04.006.

Review - Vidal, S., **Xiol, C.**, Pascual-Alonso, A., O'Callaghan, M., Pineda, M., & Armstrong, J. (2019). *Genetic Landscape of Rett Syndrome Spectrum: Improvements and Challenges*. International Journal of Molecular Sciences, 20(16), 3925. https://doi.org/10.3390/ijms20163925.

Review - Pascual-Alonso, A., Martínez-Monseny, A. F., **Xiol, C.**, & Armstrong, J. (2021). *MECP2-Related Disorders in Males*. International Journal of Molecular Sciences, 22(17), 9610. https://doi.org/10.3390/ijms22179610.

ATG
TTCA
GTTAGC
CTTCAGCC
TTCCCAAGT
TTTAGCCTACC
GTTTAGCCTAAGTAGC
TAGCCTTCAGCCAGGATCA
GTTAGCCTTCAGCCAGGTCA
TAGCCTTCAGCCAGGAAGCCAG
GTTTAGCCTAAGTTAGCCTAAGTTA
TTTAGCCTACCGCCTACCTACCG
TTCCCAAGTTTAGCCTACC
CTTCAGCCCTACCGCC
GTTAGCCTTCAGCC
TAGCCTTCAGCA
GTTTAGCCTAA
TTTAGCCTAC
CTTCAGCC
GTTAGC
TTCA
TAG