New genes and pathways implicated in Rett syndrome: considerations and future applications

Mario Lucariello

UNIVERSITAT DE BARCELONA

NEW GENES AND PATHWAYS IMPLICATED IN RETT SYNDROME: CONSIDERATIONS AND FUTURE APPLICATIONS

Memoria tesis doctoral

Mario Lucariello

Barcelona, 2016

INSTITUT D'INVESTIGACIÓ BIOMÈDICA DE BELLVITGE

PEBC
Cancer Epigenetics and Biology Program
Programa d'Epigenètica i Biologia del Càncer
Programa de Epigenética y Biología del Cáncer

MARIE CURIE ACTIONS

NEW GENES AND PATHWAYS IMPLICATED IN RETT SYNDROME: CONSIDERATIONS AND FUTURE APPLICATIONS

Memoria presentada por Mario Lucariello para optar al grado de Doctor por la Universidad de Barcelona

UNIVERSITAT DE BARCELONA – FACULTAT DE MEDICINA

PROGRAMA DE DOCTORAT EN BIOMEDICINA 2016

Este trabajo ha sido realizado en el Grupo de Epigenética del Cáncer, dentro del Programa de Epigenética y Biología de Cáncer (PEBC) del Instituto de Investigación Biomèdica de Bellvitge (IDIBELL)

_____

Dr. Manel Esteller Badosa

Director y tutor

_____

Mario Lucariello

Doctorando

ACKNOWLEDGMENTS

ACKNOWLEDGMENTS

My experience in PEBC started with the opportunity of a short Erasmus Placement stay. During this time, I met wonderful people who have accompanied me through the following 3-years PhD. I am very grateful for all that I have achieved throughout these years. The completion of this thesis would not have been possible without the help from many people.

First of all, I would like to express my special appreciation to Dr. **Manel Esteller**, director of my PhD thesis and head of PEBC. Thank you for giving me a chance to work in such a stimulating environment and for the time that you have dedicated to me. Your excellent teachings and guidelines for my scientific career have been really invaluable.

I would like to thank to **Maria Berdasco** as well, brilliant researcher and good friend. Working with you and **Miguel Lopez** during my Erasmus Placement was a great pleasure. You have always believed in me and encouraged me during all the steps of my PEBC experience. Thank you for your important support and your time inside and outside the laboratory.

I would like to give special thanks also to **Mauricio Sáez**. You have been a tremendous mentor for me. I feel very lucky to have had you as a precious guide for my PhD. Your availability and caring towards me, not only during your time in PEBC but also later, is something precious. For me, you are a positive example of a person who loves his work and family and naturally transmits his passion and enthusiasm. Thank you for being an example of life.

Dori, although you supervised me only for a limited period, I really appreciated your smart way to organize a group of people that works together on the same subject. I believe this was the key for the success of the RETT team.

Un ringraziamento particolare per il mio caro compatriota **Paolo**, di una gentilezza ed una disponibilità infinita. Grazie per avermi aiutato quando ne avevo bisogno e per avermi insegnato i segreti dell'immunoprecipitazione della cromatina. La buona

riuscita di questa tesi dottorale è sicuramente anche merito tuo. Te ne sono veramente grato.

Karolina, thank you for your important working suggestions, for our tea times and especially for all the moments of fun we have spent together. I really enjoyed having you as colleagues and friend.

Un agradecimiento de corazón al Dr. **Fernando**. Si pienso en alguien que me haya enseñado con gusto y pasión las cosas y que al mismo tiempo haya sido un óptimo amigo, esa persona eres tú. Gracias por trasmitirme tu infinita experiencia técnica y conocimiento científico.

Miguel L, mi primer supervisor del PEBC. Gracias por ayudarme a hacer los primeros pasos en el laboratorio, por haber estado siempre allí en caso de necesidad y por los momentos de tapeo. Poble Sec mola más!

Paula, your professionalism and your positive energy have represented a lifeline for me in many difficult moments. Thanks for your friendship and for all the fun we have had together.

Catia, my neighbor in the office. Thank you for your support, your kindness and your smiles.

Anna M, thank you for your time, your kindness and our funny moments in Badalona.

I am grateful to all PEBC members for being great co-workers and friends. Thanks to **Yutaka, Davide, Edilene, David, Julia, Alexia, Laia S, Carmen, Miguel V, Juan, Vanesa, Humberto, Laia P, Matthew, Maxime, Michael, Angel, Holger, Sebas, Raul, Montse, Pere, Quique, Toni, Sergi, Anna P, Patri, Laura, Rute, Olga, Aida, Vero, Paola, Sonia, Raquel, Cristina, Marta, Aida**, **Helena**, **Anna V**, **Espe** and **Vero P**.

I would like to give a special mention also to **Judith Armstrong**, **Silvia Vidal** and **Mercedes Pineda** from Hospital Sant Joan de Déu. Your contribution has been fundamental for the completion of exome project. Thank you very much.

Thanks to all the **Marie Curie EpiTrain** partners. Being part of such an enthusiastic network has given me the opportunity to work outside of my country, to join scientific congresses all over Europe, to know and interact with people of the network for an exchange of ideas and discussions. Thanks for contributing to my professional growth.

Grazie a tutti! Gràcies a tots! ¡Gracias a todos! Thank you all!

INDEX

INDEX

INDEX

ABBREVIATIONS

ASDs	Autism spectrum disorders
BF-1	Brain factor 1
CDKL5	Cyclin-dependent kinase-like 5
ChIP	Chromatin immunoprecipitation
CREB1	CAMP Responsive Element Binding Protein 1
c-Ski	Corepressor Sloan-Kettering Institute
DNA	Deoxyribonucleic acid
E/I	Excitatory/Inhibitory
FOXG1	Forkhead box G1
GABA	γ-aminobutyric acid
HDAC	Histone deacetylase
ID	Intellectual disability
IQ	Intelligence quotient
JARID1B	Jumonji/ARID Domain-Containing Protein 1B
Kb	Kilobase
KO	Knockout
LANA	Latency-associated nuclear antigen
LTP	Long-term potentiation
MBD	Methyl-CpG-binding domain
MeCP2	Methyl-CpG binding protein 2
mEPSCs	Miniature excitatory postsynaptic currents
mIPSCs	Miniature inhibitory postsynaptic currents
miRNA	microRNA
mRNA	Messenger RNA
mRNP	Messenger ribonucleoprotein particle
N-CoR1	Nuclear receptor co-repressor 1
NDD	Neurodevelopmental disorder
NLS	Nuclear localization signal
NMDA	N-methyl-D-aspartate receptor
NMJs	Neuromuscular junctions
PDD	Pervasive developmental disorder

Rac1	Ras-Related C3 Botulinum Toxin Substrate 1
RNA	Ribonucleic acid
RNAi	RNA interference
RTT	Rett Syndrome
Sin3A	SIN3 Transcription Regulator Family Member A
STK9	Serine/threonine kinase 9
STP	Short-term plasticity
Suv39H1	Suppressor Of Variegation 3-9 Homolog 1
SWI/SNF	Switch/Sucrose Non-Fermentable
TFIIB	Transcription Factor II B
TRD	Transcriptional repression domain
XCI	X chromosome inactivation
YB1	Factor Y-box-binding protein
SNP	Single-nucleotide polymorphism
Indel	Insertion-deletion
CNV	Copy number variation
NGS	Next-generation sequencing
MPS	massively parallel sequencing
SBS	Sequencing by synthesis
SIFT	Sorting intolerant from tolerant
Polyphen	Polymorphism phenotyping
WGS	Whole-genome sequencing
WES	Whole-exome sequencing

GENERAL INTRODUCTION

1. Rett Syndrome

Rett Syndrome (RTT, OMIM 312750) is a postnatal progressive neurodevelopmental disorder, originally described in the 1960's by Andreas Rett (Rett, 1966), which manifests mainly in girls during early childhood with an incidence of approximately 1:10.000 female live births. It represents the second most common cause of mental retardation in females after Down's syndrome (Hagberg, 1995). RTT is characterized by apparently normal early-development, followed by psychomotor regression with loss of the acquired motor and language skills. Patients present with stereotypic hand movements, acquired microcephaly, autistic-like behaviors, irregular breathing with hyperventilation, and seizures (Neul *et al*., 2010). The history of RTT started in 1965, when two severely disabled young girls were insistently wringing their hands in the laps of their respective mother in the waiting room of a Vienna pediatric clinic. This coincidental event induced Dr. Andreas Rett to search for more patients with the same unusual behavior and, one year later, he described for the first time a new clinical entity in 22 patients that now bears his name. Nevertheless, international medical community had not recognized the disease until Dr. Bengt Hagberg and his colleagues reported 35 cases of RTT in the English language 17 years later (Hagberg *et al*., 1983). A mutation in the gene of *MECP2* was identified as the primary cause of the disease in 1999 by Amir *et al*. Approximately 90% of RTT patients with a classical form of the disease carry a mutation in *MECP2*, as well as 50% of the atypical form. Many of the atypical RTT patients, who deviate from classical clinical presentation of RTT, are mutation-negative for the mentioned gene and carry mutations in *CDKL5*, described in individuals with an early seizure onset variant, or *FOXG1*, responsible for the congenital variant (Evans *et al*., 2005; Philippe *et al*., 2010). However, there remains a subset of patients with a clinical diagnosis of RTT who are mutation-negative for all the aforementioned genes.

1.1 Neurodevelopmental disorders

Human brain development is a complex and orchestrated process that begins in the third gestational week with the differentiation of the neural progenitor cells and extends at least through late adolescence, arguably throughout the lifespan (Stiles and Jernigan, 2010). Brain development requires a series of dynamic and adaptive processes that operate to promote the differentiation of new neural structures and functions and range from the molecular events of gene expression to environmental input. Both gene expression and environmental input are essential for normal brain development, and disruption of either can fundamentally alter the resulting neural outcomes. The incredibly complex circuit that constitutes the substrate of the brain needs an equally network of genes to orchestrate own self-assembly. Therefore, mutations affecting any of a wide range of cellular processes can lead to altered neurodevelopment and result in neurological or psychiatric diseases.

Neurodevelopmental disorders (NDDs) encompass a heterogeneous group of disorders with severely affected behavioral features caused by changes in early brain development. Most NDDs are associated with a lifelong endurance and drastic impact on normal brain function that lead to affected behavior often resulting in severe emotional, physical and economic problems, not only for the individual but also for the family and society as well. The various NDDs are usually characterized by alterations in sensor, motor and language systems, as well as various cognitive impairments in learning and organizational skills (van Loo and Martens, 2007). Autism spectrum disorders (ASDs) and intellectual disability (ID), previously termed mental retardation, are some of the more common NDDs, but also schizophrenia, bipolar disorder and attention deficit hyperactivity disorder are considered NDDs (Harris, 2014).

ASDs are a group of NDDs characterized by the following three core deficits: impaired communication, impaired reciprocal social interaction and restricted, repetitive and stereotyped patterns of behaviors or interests. The presentation of these impairments is variable in range and severity and often changes with the acquisition of other developmental skills (Montes and Halterman, 2006). RTT is classified as an ASD in the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) (American Psychiatric Association, 1994). ASD has a prevalence of almost 0.6% in the general population and is four times more prevalent in boys than in girls (Fombonne *et al*.,

2006). Consequently, RTT is a distinct type of autism due to its prevalent occurrence in females rather than males, as described above. Although the exact cause of autism is still not known, research clearly indicates that the etiology is multifactorial with a strong genetic basis (Bailey *et al*., 1995).

ID occurs in approximately 2-3% of the population in developed countries (Chelly and Mandel, 2001). ID is characterized by significantly impaired intellectual and adaptive functioning. In addition to an intelligence quotient (IQ) lower than 70, significant limitations in adaptive behaviors, represented by conceptual, social, and practical skills, are required for the diagnosis of ID. The underlying causes of ID can be diverse. Inborn causes such as Down's syndrome, Fragile X syndrome and fetal alcohol syndrome are responsible for 30% of the ID cases (Batshaw, 1993), but also malnutrition and problems during pregnancy or birth can increase the risk for ID (Hagberg and Kyllerman, 1983).

1.2 Classical clinical features of RTT

Rett Syndrome, in its classic or typical form, begins to manifest in early childhood and is characterized by neurodevelopmental regression that severely affects motor, cognitive and communication skills. Patients with RTT have an apparently normal post-natal initial phase of development up to 6-18 months of age, achieving appropriate milestones such as the ability to walk and even to say some words. The characteristic clinical features appear successively over several stages, forming a distinctive disease progression pattern (**Figure 1**).

Stage I: developmental stagnation (age of onset: 6-18 months). This stage is characterized by deceleration of head growth, which eventually leads to acquired microcephaly, general growth retardation, loss of weight and muscle hypotonia. Patients are delayed in the acquisition of skills. Although babbling and new world may appear, language skills usually remain poor. Autistic features may begin to display in RTT patients, such as emotional withdrawal, expressionless face, indifference to the surrounding environment and some self-abusing behavior (Nomura, 2005), that follow in the next stage.

Stage II: rapid regression (age of onset: 1-4 years). With the progression of the syndrome, patients lose purposeful hand use, speech and interpersonal contact (Nomura, 2005). Voluntary hand use, such as grasping and reaching out for toys, is replaced with repetitive stereotypic movements of hands, representing the most typical feature of girls affected by RTT (Hagberg, 2002) (**Figure 2**). Patterns consisting of wringing, hand washing, mouthing, clapping, rubbing, squeezing and other automatisms occur during waking hours (Nomura and Segawa, 1992; Hagberg, 1995). Irregular breathing patterns, such as episodes of hyperventilation, breath holding and aerophagia represent the earliest autonomic dysfunction during wakefulness. One of the most invalidating aspects of RTT is the manifestation of seizures, which range from easily controlled to intractable epilepsy, with the most common type being tonic-clonic seizures (Jian *et al*., 2006). However, epilepsy often tends to decrease after adolescence, with lower seizure frequency and less secondary generalized seizures even in those patients who were previously considered quite insusceptible (Krajnc *et al*., 2011).

Stage III: stationary stage (age of onset 4-7 years). This stage can last for years or decades and is characterized by a relative stabilization of the disorder. RTT patients may recover some skills lost in the previous stage. Patients can become more sociable, use eye-pointing as a typical way to express their needs, and may even improve their language skills. Despite improved eye contact and non-verbal communication ability, they follow to suffer from important cognitive impairments and the loss of motor functions further progress in this stage. Breathing dysfunction become prominent, as well as stereotypic hand movements. Many patients develop generalized rigidity, dystonia and scoliosis, which eventually require surgical treatments. Sleeping patterns are often altered, characterized by frequent night time waking and day time sleeping (Young *et al*., 2007).

Stage IV: late motor deterioration (age of onset: 5-15 years). Despite the presence of serious cognitive disablement, older patients with RTT are usually sociable and pleasant with others, in contrast to patients with childhood autism (Mount *et al*.,

2001). Seizures become less frequent and less severe and stereotypic hand movements less intense. Anyhow, motor deteriorations continue with age and many patients lose mobility and become wheelchair-dependent, leading to pronounced muscle wasting and general rigidity. Patients also suffer from additional autonomic dysfunction such as severe constipation, oropharyngeal dysfunction and cardiac abnormalities, including tachycardia, prolonged QT intervals and bradycardia (Acampa and Guideri, 2006). At older age, patients often develop Parkinsonian features (Roze *et al*., 2007). In spite of the great amount of physical debilitation and autonomic dysfunctions, some patients survive up to 60-70 years of life, in severely debilitated physical conditions. The estimated annual death rate from RTT is 1.2%. Approximately 25% of these deaths occur suddenly and are due to autonomic dysfunctions and cardiac abnormalities (Kerr *et al*., 1997; Guideri *et al*., 1999; Kerr and Julu, 1999).

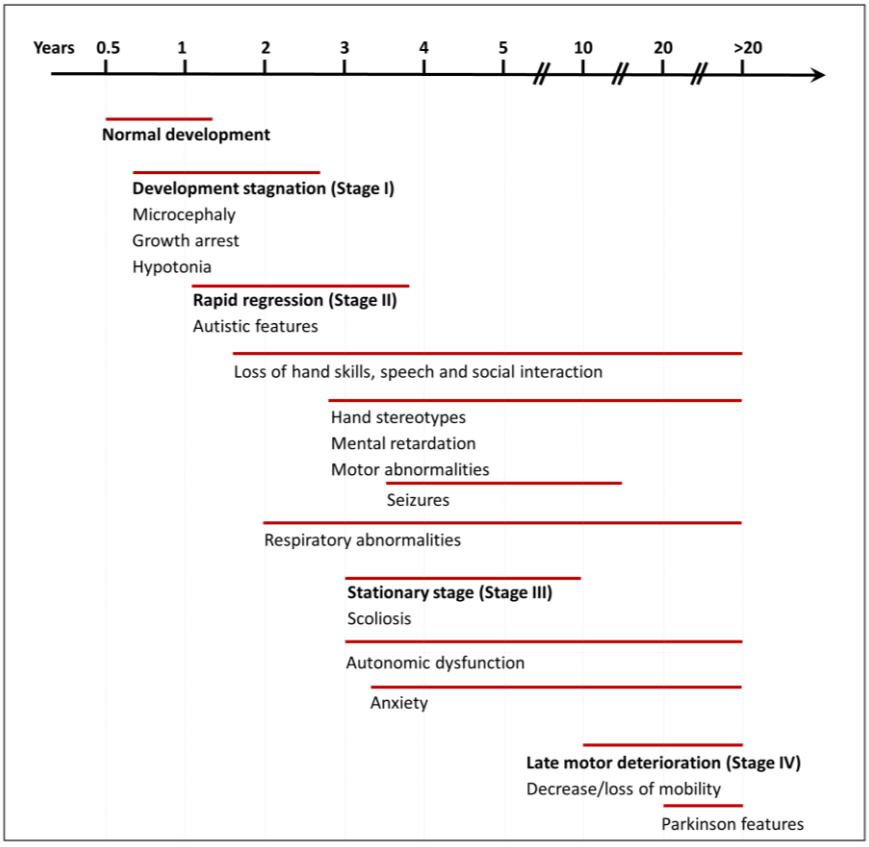


Figure 1 Onset and progression of classic RTT clinical features. Following a period of normal development, an apparently healthy baby girl first manifests a developmental stagnation and then a rapid regression of the acquired motor and language skills. The replacement of purposeful use of hands with restless stereotypies is the hallmark of RTT. Patients also develop social autistic features and breathing dysfunction, as well as seizures. In the latest phases of the disease, patients also suffer from anxiety, autonomic dysfunctions and profound motor impairment. Based on Chahrour and Zoghbi, 2007.

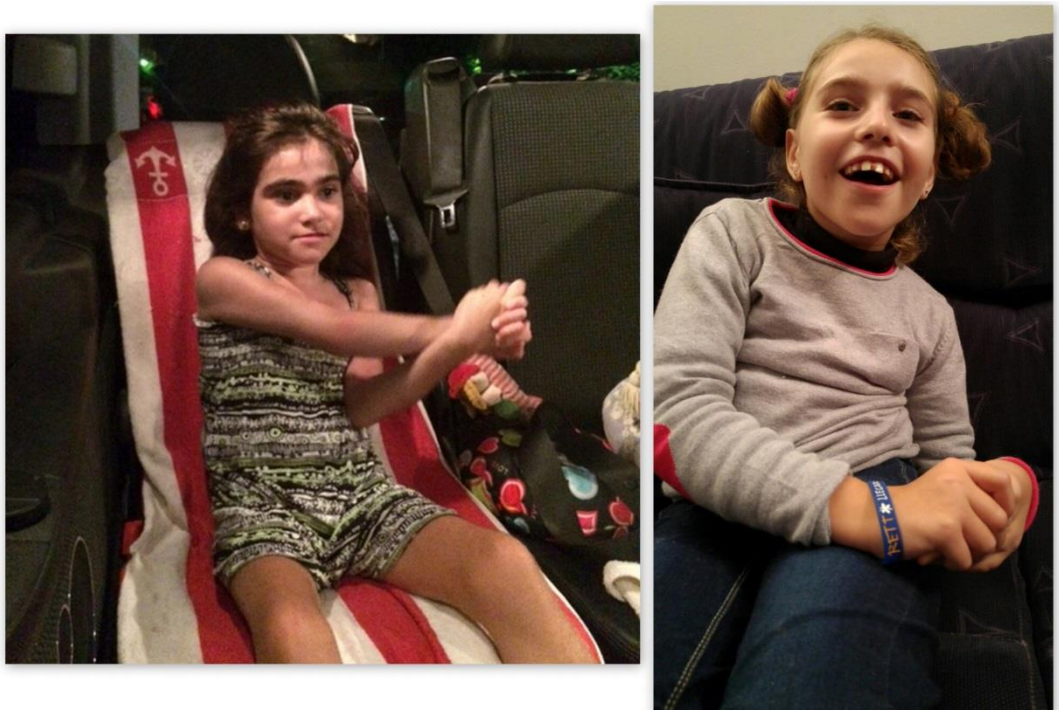


Figure 2 Stereotypic movements of hands represent the most typical feature of girls affected by RTT**.**

1.3 Genetic bases of RTT

The primary cause of RTT has been debated in the literature for long time (Hagberg *et al*., 1983; Martinho *et al*., 1990; Migeon *et al*., 1995). Considering that RTT occurs almost exclusively in female, early reports proposed an X-linked dominant mode of inheritance with crucial consequences in hemizygous males. As more than 99% of RTT cases are sporadic and patients rarely reproduce, it was very difficult to map the disease locus by traditional linkage analysis. However, the use of information from rare familial cases (Schanen *et al*., 1997), including a severely affected male infant (Schanen and Francke, 1998), and the alternative approach of exclusion mapping (Ellison *et al*., 1992) allowed the exclusion of most regions of the X chromosome and focused the search for the RTT gene on band Xq28 (Sirianni *et al*., 1998). Following further screening analysis of plausible disease-causing mutations of candidate genes in this locus, it was found that a mutation in the gene *MECP2,* coding for Methyl-CpG-binding protein 2 (MeCP2), was the primary cause of RTT (Amir *et al*., 1999). Although perturbations in *MECP2* explain around 90% of the classical or typical form of RTT cases, it is important to mention that mutations in other genes, such as the cyclin-dependent kinase-like 5 (*CDKL5*) or in the forkhead box G1 (*FOXG1*), can results in

atypical or variant forms of RTT. Indeed, approximately 8% of classical RTT and 42% of variant RTT patients are *MECP2* mutation negative (Monrós *et al*., 2001; Percy, 2008). After the discovery of the main disease-causing gene, a process of mutation screening in males with related neurological phenotypes started, leading to the breakdown of the dogma by which RTT was thought to occur exclusively in females (Jan *et al*., 1999). Indeed, it has been shown that *MECP2* mutations that cause classic RTT usually lead to neonatal encephalopathy and death in the first year of life in males with a normal karyotype. Although the clinical prognosis of male RTT patients is poor, the reason because male patients are underrepresented in RTT is not based on the likelihood of embryonic lethality. A study performed in sporadic RTT patients and the respective families showed that in 26 out of 27 cases, *de novo* mutations in *MECP2* arise on the paternal X chromosome (Trappe *et al*., 2001). According to this model, males are protected from X-linked dominant diseases because they do not inherit the mutated paternal X chromosome. Thus, two possible hypotheses may explain the high ratio female/male in RTT. The first one is based on the assumption that paternal germ cells have higher mutational incidence in *MECP2* than the maternal ones. On the other hand, an alternative explanation considers that mutation incidence at paternal germ cells is normal, whereas a mutation in *MECP2* may be detrimental for the oocytes that carry the mutated protein, decreasing the likelihood of being fertilized (Trappe *et al*., 2001).

1.3.1 MeCP2 structure

MECP2 spans approximately 76kb and consists of four exons that code for two different isoforms of the protein, due to alternative splicing of exons 2. The MeCP2 splice variants differ only in their N-terminus; the more abundant MeCP2-e1 isoform contains 24 aminoacids encoded by exon 1 and lack the 9 aminoacids encoded by exon 2, whereas the start site for the MeCP2-e2 isoform is in exon 2 (Kriaucionis and Bird, 2004) (**Figure 3**). Both MeCP2 isoforms are nuclear and co-localize with methylated heterochromatic foci in mouse cells. A recent report suggests that MeCP2 translocate to the nucleus upon neuronal differentiation (Miyake and Nagai, 2007). Although the MeCP2 protein is widely expressed, it is relatively more abundant in the brain,

primarily in mature post-migratory neurons (Jung *et al*., 2003). MeCP2 protein levels are low during embryogenesis and increase progressively during post-natal period of neuronal maturation (Balmer *et al*., 2003; Kishi and Macklis, 2004). Since MeCP2 is expressed in mature neurons and its levels increase during postnatal development, MeCP2 may play a role in modulating the activity or plasticity of mature neurons.

MeCP2 is a member of methyl-CpG binding protein family (Hendrich and Bird, 1998) and is composed of three domains: the methyl-CpG-binding domain (MBD), the transcriptional repression domain (TRD) and a C-terminal domain, in addition to two nuclear localization signals (NLS) (**Figure 3**). The MBD specifically binds to methylated CpG dinucleotides, with preference for CpG sequence adjacent to A/T-rich motifs (Klose *et al*., 2005). The TRD domain is implicated in transcriptional repression through recruitments of several co-repressors and chromatin remodeling complexes (Jones *et al*., 1998; Nan *et al*., 1998). The C-terminal region facilitates MeCP2 binding to naked DNA and to the nucleosomal core, as well as to WW domain splicing factors through conserved poly-proline sites (Buschdorf and Strätling, 2004). Despite the C-terminus of MeCP2 is not yet well characterized, it is crucial for protein function as demonstrated by the numerous mutations causing RTT that involve deletions of this region, and the fact that a mouse model lacking the MeCP2 C-terminal domain recapitulates many RTT features (Shahbazian *et al*., 2002). The nuclear localization signals (NLS) mediate translocation of the protein to the nucleus.

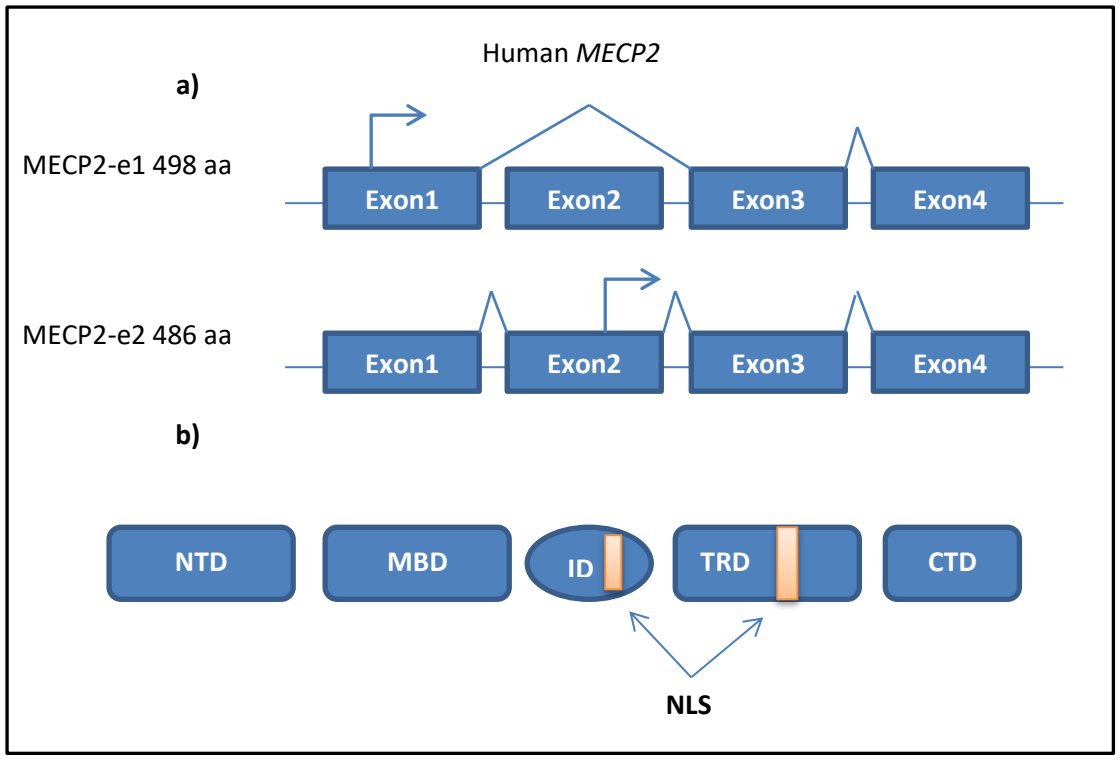


Figure 3 Schematic composition of MECP2 gene structure, splicing patterns and functional domains. **a)** Splicing patterns generate two isoforms of MeCP2, MeCP2-e1 and -e2, with different N-termini due to the use of alternative translation start site (bent arrow) and the absence or presence of exon2 in the transcript. **b)** With the exception of the N-terminus, both MeCP2 isoforms are identical and contains the following domains: NTD, N-terminal domain; MBD, methylated DNA-binding domain; ID, interdomain; TRD, transcription repression domain; CTD, C-terminal domain; NLS, nuclear localization signals. Based on Gadalla *et al.*, 2011.

1.3.2 MeCP2 as an epigenetic multifunctional regulator of gene expression

CpG dinucleotides densely populate the promoter regions of many genes. Approximately 60-90% of cytosines present within CpG dinucleotides are methylated in the adult vertebrate by methylation at carbon in position 5. Methylation of CpG dinucleotides is an important epigenetic mechanism of gene silencing, both in terms of stables silencing of heterochromatin and in the reversible regulation of gene expression (Ng and Bird, 1999; Jones and Wolffe, 1999), and is characterized by DNA modification without changing the genetic code. Transcriptional repression mediated by DNA methylation is thought to occur predominantly through an indirect mechanism in which repressor proteins are recruited to methylated sites (Tate and Bird, 1993), although methylation alone can sometimes directly repress transcription (Siegfried *et*

al., 1999). In this contest, MeCP2 acts as a critical mediator of methylation-dependent repression that couples DNA methylation to silencing machinery.

The original model suggested that MeCP2 is a global transcriptional repressor and it was based on *in vitro* experiments in which MeCP2 specifically inhibited transcription from methylated promoters (Nan *et al*., 1997). When MeCP2 binds to methylated CpG dinucleotides of target gene via its MBD, its TRD recruits the corepressor Sin3A and histone deacetylases (HDACs) 1 and 2 (Jones *et al*., 1998; Nan *et al*., 1998). These interactions involve the condensation of the chromatin by favoring nucleosome clustering, either through recruitment of HDACs, or through direct interaction between its C-terminal domain and chromatin (Nikitina *et al*., 2007). Later on, the identification of several other MeCP2-interacting proteins, such as the catalytic component of the SWI/SNF chromatin-remodeling complex Brahma, the histone methyltransferase Suv39H1, the transcription factors TFIIB and PU.1, the corepressors c-Ski and N-CoR, LANA, and the SWI2/SNF2 DNA helicase/ATPase (Harikrishnan *et al*., 2005; Kaludov and Wolffe, 2000; Kimura and Shiota, 2003; Kokura *et al*., 2001; Nan *et al*., 2007) further reinforced the interplay between MeCP2 and chromatin. The capability of MeCP2 to work as an architectural chromatin protein was additionally supported showing that at high molar ratio to nucleosomes, MeCP2 mediates the formation of a highly compacted chromatin structure (Georgel *et al*., 2003). Noteworthy, this MeCP2 feature does not require additional factors apart from core histones, and appears independent of DNA methylation, thus raising the possibility that MeCP2 might affect the expression/structure of genomic unmethylated regions (Georgel *et al*., 2003).

In addition to the roles in transcriptional repression and modulation of chromatin structure, MeCP2 has been proposed such as a regulator of mRNA splicing. Using co-immunoprecipitation from HeLa cell extracts, MeCP2 was shown to interact with the splicing factor Y-box-binding protein (YB1) (Young *et al*., 2005). YB-1 protein is a highly conserved component of messenger ribonucleoprotein particles (mRNPs) that functions as the main mRNA packaging protein. The interaction among MeCP2 and YB-1 requires the presence of RNA, as co-immunoprecipitation treated with RNase failed to pull down YB-1 with MeCP2. The observation of aberrant alternative splicing patterns in a Mecp2 mutant mouse model of RTT, together with the fact that MeCP2

can form complexes with RNA *in vitro* independently of its MBD (Jeffery and Nakielny, 2004), suggest that MeCP2 might modulate RNA splicing *in vivo*.

Despite all the evidences proposing a chromatin repression role for MeCP2, transcriptional profiling studies have inferred transcription regulator ability rather than a strict silencer. Using an epigenomic ChIP-chip approach, MeCP2 was shown to occupy many active promoters and to bind mainly to nonmethylated sites along intergenic spaces thousands of bases away from transcription start or ending sites (Yasui *et al*., 2007). This model was first reinforced by the identification of an interaction between MeCP2 and the transcriptional activator CREB1 in a pull down experiments, and then by a transcriptional profiling studies performed in hypothalamus and cerebellum of RTT mice that revealed nearly 85% of MeCP2-bound genes are actively transcribed (Chahrour *et al*., 2008; Mellén *et al*., 2012).

Considering the multitude of interacting partners, MeCP2 is clearly a multifunctional protein, with roles in transcriptional regulation, chromatin remodeling, RNA splicing, as well as repression of retrotransposition (Muotri *et al*., 2010) and suppressor of miRNA biogenesis (Cheng *et al*., 2014).

1.3.3 *MECP2* mutations and genotype-phenotype correlation

Mutations in *MECP2* are found in approximately 90% of classic RTT cases in a spectrum of mutation types that includes missense, nonsense and frameshift pathogenic nucleotide changes. Among over 800 different pathogenic reported mutations in *MECP2* (RettBASE: IRSF MECP2 Variation Database), eight missense and nonsense mutations account for around 70% of all mutations (reported in **Table 1**) (Christodoulou *et al*., 2003). These mutations are provoked by the deamination of methylated cytosines that generates C>T transitions, which in turn are responsible for the “hotspot” mutations (Lee *et al*., 2001). Additionally, small C-terminal deletions account for about 10% of all mutations. As result of several phenotype-genotype studies, some general conclusions can be made. Mutations affecting the NLS of MeCP2 or early truncating mutations usually cause more severe phenotypes, whereas C-terminal deletions are associated with milder phenotypes (Smeets *et al*, 2005). Moreover, the R133C mutation causes a global milder phenotype (Kerr *et al*., 2006;

Neul *et al*., 2008), while the R270X mutation is associated with increased mortality (Jian *et al*., 2005). Despite some overall trends, considerable variability in clinical severity is often observed among patients with the same *MECP2* mutation (Bebbington *et al*., 2008; Halbach *et al*., 2012; Scala *et al*., 2007). A major source of phenotypic variability in RTT females is the pattern of X chromosome inactivation (XCI). In females, only one the two X chromosome is active in each cell and the choice of which X chromosome is active is usually random, in a way that half of the cells have the maternal X chromosome active and the other half have the paternal one active. Interestingly, cells expressing the wild type *MECP2* allele divide faster or survive better than cells expressing the mutant allele, resulting in a nonrandom pattern of XCI and improvement of RTT phenotype. Indeed, according to the extent of such favorable skewing, some patients have been reported to present with a milder phenotype (Amir *et al*., 2000; Huppke *et al*., 2006) and some mothers to be asymptomatic carriers (Villard *et al*., 2000; Villard *et al*., 2001). Finally, a different XCI pattern may be a possible explanation for such variations.

Table 1 Frequency of the eight most common point mutations and small deletions in classic RTT cases. Data are from RettBase **(http://mecp2.chw.edu.au/mecp2/index.php).**

Nucleotide change	Aminoacid change	Cases reported	% (pathogenic mutation reported)
316C>T	R106W	59	4,8%
397C>T	R133C	67	5,4%
473C>T	T158M	150	12,2%
502C>T	R168X	147	11,9%
763C>T	R255X	132	10,7%
808C>T	R270X	118	9,6%
880C>T	R294X	101	8,2%
916C>T	R306C	79	6,4%
20-100bp deletions		120	9,7%
Total		**1231**	**78,9%**

1.4 Atypical RTT

In addition to classic or typical RTT, atypical or variant forms of RTT that deviate from the classical clinical presentation have been delineated in individuals that present

many of the clinical features of classic RTT, but do not necessarily manifest all of the features of the disorder (Neul *et al*., 2010). These variants range from milder forms with a later age of onset to more severe manifestations.

The preserved speech variant (Zappella, 1992), or Zappella variant, is a benign form of RTT characterized by the ability of patients to speak a few words, although not necessarily in the context. Patients with this variant have a normal head size and are usually overweight and kyphotic (Zappella *et al*., 2001). Mutations in *MECP2* have been found in the majority of cases of the preserved speech variant (Renieri *et al*., 2009).

The more severe variants include the congenital form (Rolando, 1985), or Rolando variant, that miss the early period of normal development, and the early seizure one (Hanefeld, 1985), or Hanefeld variant, with onset of epileptic episodes before the age of 6 months. In contrast to preserved speech variant, mutations in *MECP2* have only rarely been identified (Huppke *et al*., 2006) in congenital and early seizures forms (Archer *et al*., 2006). Specifically, mutations in the X-linked *CDKL5* (Xp22) gene have been found to be associated with early seizure variant cases (Kalscheuer *et al*., 2003), whereas mutations in *FOXG1* (14q11-q13) are responsible of congenital variant cases (Ariani *et al*., 2008). *CDKL5* codify for the cyclin-dependent kinase-like 5, previously known as serine/threonine kinase 9 (*STK9*), which autophosphorylates itself and phosphorylates MeCP2 *in vitro*. It has been demonstrated that this latter activity is disrupted in pathogenic *CDKL5* mutants (Bertani *et al*., 2006). In a recent paper by Chen and colleagues, it was proposed that CDLK5 is a critical regulator of neuronal morphogenesis: downregulation of CDKL5 by RNA interference (RNAi) in cultured cortical neurons inhibited neurite growth and dendritic arborization, whereas overexpressing CDKL5 had opposite effects (Chen *et al*., 2010). Moreover, they demonstrated that CDKL5 co-localizes and forms a protein complex with Rac1, a critical regulator of actin remodeling and neuronal morphogenesis. All these evidences contribute to confer a critical role to CDLK5 in neuronal development and explain the involvement of *CDKL5* in a congenital RTT-like disorder. The protein product of *FOXG1,* formerly brain factor 1 (BF-1), is a transcriptional factor with expression restricted to fetal and adult brain and testis. FoxG1 interacts with the transcriptional repressor JARID1B and with global transcriptional corepressors of the Groucho family. The

interaction with these proteins is of functional importance for early brain development (Tan *et al*., 2003; Yao *et al*., 2001).

The clinical features and the genetic loci associated with these specific variants of atypical RTT are schematically shown in **Table 2.**

Table 2 Clinical features and the genetic loci associated with variant forms of RTT (Neul *et al*., 2010).

Preserved Speech Variant (Zappella Variant)	**Early Seizure Variant (Hanefeld Variant)**	**Congenital Variant (Rolando Variant)**

Clinical features
• Regression at 1-3 years, prologed plateau
• Milder reduction of hand skills
• Recovery of language after regression
• Milder intellectual disability (IQ up to 50)
• Autistic behaviors common
• Decreased frequency of typical RTT features
  • Rare epilepsy
  • Rare autonomic dysfunction
  • Milder scoliosis and kyphosis
  • Normal head curcumference
  • Normal height and weight in most
	Clinical features
	• Early onset of seizures
	  • Before 5 months of life
	  • Infantile spasms
	  • Refractory myoclonic epilpsy
	  • Seizure onset before regression
	  • Decreased frequency of typical RTT features
		Clinical features
		• Abnormal initial development
		  • Severe psychomotor delay
		  • Inability to walk
		• Severe post-natal microcephaly before 4 months
		• Regression in the first 5 months
		• Lack of typical RTT like gaze
		• Typical RTT autonomic abnormalities present
		  • Small cold hands and feet
		  • Peripheral vasomotor disturbances
		  • Breathing abnormalities while awake
		• Specific movements abnormalities
		  • Tongue stereotypies
		  • Jerky movements of the limbs

Molecular genetics Mutations in *MECP2* found in the majority of cases	**Molecular genetics** Mutations in *MECP2* rarely found Analysis for mutations in *CDKL5* should be performed	**Molecular genetics** Mutations in *MECP2* rarely found Analysis for mutations in *FOXG1* should be performed

1.5 *MECP2* mutations in other neurodevelopmental disorders

Alterations in *MECP2* result in a variety of neurodevelopmental conditions, of which classic and variant RTT represent only a part (**Table 3**). On the milder site, there are females with mild mental retardation, learning disabilities and ASD (Carney *et al*., 2003; Lam *et al*., 2000). *MECP2* have been found mutated also in females with severe mental retardation, epilepsy and Angelman-like syndrome (Milani *et al*., 2005; Watson *et al*., 2001). The situation of *MECP2*-related disorders aggravates with complex forms of severe mental retardation in males that are associated with epilepsy, ataxia, tremor, hyperactivity, autism and bipolar behavior (Klauck *et al*., 2002) or juvenile-onset schizophrenia (Cohen *et al*., 2002). Additionally, *MECP2* mutations were identified in males with mental retardation, psychosis, pyramidal signs, Parkinsonian features and macroorchidism (Klauck *et al*., 2002).

Mutations that are predicted to result in a null allele suggest that the neurodevelopmental abnormalities are a result of MeCP2 loss of function. However, the *MECP2* duplication disorder demonstrates that an increase in protein levels can be equally detrimental to the nervous system (**Table 3**). Indeed, duplications of Xq28 that span the MECP2 locus have been reported in males with progressive neurodevelopmental phenotypes. The patients suffer from mental retardation with facial and axial hypotonia, progressive spasticity, seizures, recurrent respiratory infections, and often premature death (Meins *et al*., 2005; Van Esch *et al*., 2005; Friez *et al*., 2006; Lugtenberg *et al*., 2006). Although these individuals manifest some form of cognitive impairment, they lack features that define RTT, primarily a history of regression. Thus, these patients cannot be given a diagnosis of RTT (Neul *et al*., 2010).

Table 3 MeCP2-related disorders (Chahrour and Zoghbi, 2007).

	Sex-associated syndromes and symptoms	
MeCP2 State	**Female**	**Male**
Loss of function	Classic RTT	Infantile encephalopathy
	Atypical RTT	Classic RTT (47, XXY or somatic mosaic)
	Angelman-like syndrome	Mental retardation with motor deficits
	Mental retardation with seizures	Bipolar disease, mental retardation and tremors
	Mild mental retardation	Juvenile-onset schizophrenia, mental retardation and tremors
	Learning disability	Mental retardation, psychosis, pyramidals signs and macroorcidism
	Autism	
Overexpression	Preserved speech variant of RTT	Severe mental retardation and RTT features
		Non-specific X-linked mental retardation

1.6 Diagnostic criteria

The clinical phenotypes described above clearly emphasize that mutations in *MECP2* are not synonym of RTT and that a mutation in *MECP2* is not sufficient to make diagnosis of RTT. Consequently, although known genetic causes, the diagnosis of RTT remains clinic and is based on several well-defined clinical consensus criteria, recently revised in 2010 (Neul *et al*., 2010), and later confirmed by genetics (**Table 4**). These criteria clearly highlight the differences between classic and variant forms of RTT (required for typical and atypical RTT), as well as simplify and clarify differential diagnosis among RTT and other neuropsychiatric disorders (exclusion criteria). The basic purpose of the exclusion criteria is to exclude other potential causes of neurological disease, such as prematurity leading to intraventricular hemorrhage, or perinatal meningitis leading to diffuse brain damage.

Table 4 Revised diagnostic criteria for RTT (Neul *et al*., 2010).

Required for typical or classic RTT

1. A period of regression followed by recovery or stabilization
2. All main criteria and exclusion criteria
3. Supportive criteria are not required, although often present in typical RTT

Required for atypical or variant RTT

1. A period of regression followed by recovery or stabilization
2. At least 2 out of the 4 main criteria
3. 5 out of 11 supportive criteria

Main criteria

1. Partial or complete loss of acquired purposeful hand skills
2. Partial or complete loss of acquired spoken language
3. Gait abnormalities: impaired or absence of ability
4. Stereotypic hand movements (wringing/squeezing, clapping/tapping, mouthing and washing/rubbing automatisms)

Exclusion criteria for typical RTT

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

Supportive criteria for atypical RTT

1. Breathing disturbances when awake
2. Bruxism when awake
3. Impaired sleep pattern
4. Abnormal muscle tone
5. Peripheral vasomotor disturbances
6. Scoliosis/kyphosis
7. Growth retardation
8. Small cold hands and feet
9. Inappropiate laughing/screaming spells
10. Dimished sensitivity to pain
11. Intense eye communication and aye-pointing behavior

1.7 Neuropathological changes in RTT

Many neuroanatomical, neurophysiological and neurochemical studies have been carried in post-mortem human subjects and in mouse models of RTT to analyze the impact of RTT on brain, highlighting important biological processes and pathways implicated in the physiopathology of RTT.

The identification of *MECP2* as the primary disease-causing gene of RTT led rapidly to the development of mouse models of RTT that recapitulate, to varying degrees, the underlying molecular and genetic defects and symptoms of the human disease. MeCP2-null mice and conditional mutant mouse model with cell-type or area-specific loss of MeCP2 in the brain have been generated by mutating the endogenous *MECP2* gene with a representative RTT mutation. The three most widely used animal models of RTT syndrome are the Mecp2-Bird, Mecp2-Jae and Mecp2-T308A mouse models. The first two mouse models listed have been used to assess the effects caused by the complete absence of MeCP2 function, although the Mecp2-Jae allele still produces a truncated and modified protein product (Guy *et al*., 2001; Chen *et al*., 2001). The Mecp2-T308A mouse model was generated to study the effects of partial MeCP2 loss-of-function and possesses a truncating nonsense mutation at amino acid position 308, which spares the MBD, TRD and NLS (Shahbazian *et al*., 2002).

1.7.1 Structural changes in the RTT brain

Autopsy studies in patients with RTT showed a 12%–34% decrease in brain weight and volume, with the most pronounced effect in the prefrontal, posterior frontal, and anterior temporal regions (Armstrong, 2005). RTT brain shows no clear signs of degeneration, atrophy or cell death, neither signs of gliosis or neuronal migration defects (Jellinger *et al*., 1988; Reiss *et al*., 1993). These observations indicate that RTT is a disorder of postnatal neurodevelopment rather than a neurodegenerative process.

Although no gross structural changes have been detected in RTT brain, mouse models of RTT manifested delayed neuronal maturation and synaptogenesis (Fukuda *et al*., 2005). At the cellular level, neuronal soma size is reduced in the absence of MeCP2 with increased packing density (Armstrong, 2005). Synaptic structural and morphological defects include reduced dendritic branching, spine density, and reduced spine morphology (Kishi and Macklis, 2004; Belichenko *et al*., 2009). Presynaptically,

loss of MeCP2 affects the number of axonal boutons and axonal arborization in general and their targeting (Belichenko *et al*., 2009), which suggests a decrease in the number of synapses in RTT brain.

1.7.2 RTT as a disorder of synaptic and neural circuit maturation

A robust body of evidences exists for changes in synaptic signaling, affecting both excitatory and inhibitory neurotransmission. Analysis of spontaneous miniature excitatory and inhibitory postsynaptic currents (mEPSCs and mIPSCs, respectively) in MeCP2 KO mice indicated a change in the excitatory/inhibitory (E/I) balance, as revealed by increased excitatory and reduced inhibitory synaptic transmission in the hippocampus and cortex (Dani *et al*., 2005; Nelson *et al*., 2006; Chao *et al*., 2007). On the contrary, the E/I balance in transgenic mice overexpressing the human MeCP2 (MeCP2-Tg1) showed opposite effects (Collins *et al*., 2004). Whole-cell patch-clamp recordings from pyramidal neurons in primary somatosensory cortex demonstrated that spontaneous excitatory postsynaptic currents and spontaneous action potential firing are reduced in MeCP2 KO male mice (Tropea *et al*., 2009). With respect to inhibitory connections, dysfunctions in inhibitory synaptic transmission have been shown to be altered in mouse models of RTT, as evidence indicates that there are both pre- and postsynaptic defects of GABAergic neurotransmission in the brainstem (Medrihan *et al*., 2008). Epilepsy is very frequent in RTT patients and often difficult to treat (Steffenburg *et al*., 2001). The imbalance between excitation and inhibition is believed to play a crucial role in the generation of convulsions. Using a combination of voltage-sensitive dye imaging and electrophysiology, it has been argued that loss-of-function mutations in Mecp2 cause impaired balance between excitation and inhibition onto CA3 pyramidal neurons, leading to a hyperactive hippocampal network, likely contributing to limbic seizures in Mecp2 mutant mice and RTT individuals (Calfa *et al*., 2015). In addition to changes in frequency and amplitude of synaptic events, several studies have reported deficits in both short-term plasticity (STP) and long-term potentiation (LTP), cellular mechanisms of synaptic plasticity underlying learning and memory, in mouse models of RTT (Moretti *et al*., 2006; Weng *et al*., 2011). However, cortical LTP induction mechanisms appear to be preserved in early-symptomatic MECP2 KO animals even with fewer and weaker connections, as revealed by a 50%

reduction in connection probability and a 45% reduction in excitatory postsynaptic potential amplitude (Dani and Nelson, 2009). Taken all together, all these evidences indicate that RTT can, at least in part, be considered as a synaptopathy.

1.7.3 Neurotransmitter system in RTT – GABA and glutamate

Human brain needs to have a stringent balance between excitatory and inhibitory neurotransmission to sustain proper neuronal function. Two important classes of neurons finely regulate this process: the excitatory projecting neurons and their local inhibitory neighbors. In adulthood, excitation is mediated by the neurotransmitter glutamate, whereas γ-aminobutyric acid (GABA) has an inhibitory effect (Petroff, 2002). It has been demonstrated that this important equilibrium is spatially and developmentally disrupted in the brain of Mecp2-deficient mice (El-Khoury *et al*., 2014).

Glutamate is the primary excitatory transmitter in the central nervous system. Lack of MeCP2 is largely reported to determine structural and functional changes in glutamatergic synapses. One of the most consistent finding is a reduction of STP and LTP in Mecp2-null mice. In addition to neuron-to-neuron signaling, it is thought that the neurotoxic action of glutamate from glia are pointed to be one of the main reasons for onset and progression of RTT (Maezawa *et al*., 2010). Besides, mice lacking MeCP2 displayed altered NMDA (N-methyl-D-aspartate) subunits distribution within glutamatergic synapses (Maliszewska-Cyna *et al*., 2010). NMDA receptor is an important class of glutamate receptors that play key functions in many forms of synaptic plasticity and excitotoxic processes.

GABA is the major inhibitory neurotransmitter in human brain. Mice lacking Mecp2 specifically from a subset of forebrain GABAergic neurons recapitulate most of the autism-like stereotypies and behaviors presented in RTT, as well as other RTT-like phenotypes including deficits in locomotor activity, motor function and breathing patterns (Chao *et al*., 2010). Moreover, electrophysiological studies performed in thalamic neurons from MeCP2-null mice have suggested that MeCP2 is required for normal development of GABAergic circuits in the thalamus (Zhang *et al*., 2010).

1.7.4 Neurotransmitter system in RTT – monoamines and acetylcholine

Monoamines represent another neurotransmitter system that is affected in RTT. Bioamines levels, including dopamine, noradrenaline and serotonin, are reported to be reduces in post-mortem RTT biopsies from various brain regions such as cortex, basal ganglia and thalamus (Brücke *et al*., 1987; Lekman *et al*., 1989). Monoamines play an important role in the regulation of brainstem function. One of the main functions of biogenic amines is to facilitate the formation and maintenance of synapses in diverse regions of the central nervous system during development and also in adulthood (Okado *et al*., 2001). In terms of therapeutic potential, much interest has surrounded the monoamine systems, since most of the observations made in post-mortem human samples have been confirmed in mouse models (Panayotis *et al*., 2011).

Acetylcholine is another neuromodulator, produced in midbrain, which shows alterations in RTT. Several reports have highlighted impaired cholinergic function, in particular a reduction in the synthesis of choline acetyltransferase enzyme, responsible for the production of Ach (Wenk and Mobley, 1996). Moreover, a very recent work has shown that loss of MeCP2 specifically in cholinergic neurons causes part of RTT-like phenotypes, which could be rescued by re-expressing Mecp2 in the basal forebrain (BF) cholinergic neurons (Zhang *et al*., 2016). The observed cholinergic dysfunction in RTT led to the proposal that a diet enriched in choline, might incentive cholinergic function and be beneficial for the amelioration of RTT phenotype. The testing of this proposal in MeCP2-null mice, fed with an achetylcoline supplemented diet, only resulted in a modest improvement in some locomotor tasks, but no positive effects were detected for learning, disease progression and survival (Nag *et al*., 2007).

2 Next-Generation Sequencing

The identification of the causative mutation for a Mendelian disease allows for a molecular diagnosis of the patient and the family in exam. This is of great importance for patient management, family counseling and starting therapeutic intervention (Antonarakis and Beckmann, 2006). Moreover, the identification of the disease-causing gene importantly contributes to the understanding of gene functions and biological pathways involving health and disease in general (Oti and Brunner, 2007) and, surprisingly, lessons from rare diseases have often contributed to highlight common disease cases (Peltonen *et al*., 2006). The introduction of new technologies that enable the sequencing of DNA, in a high-throughput manner at much lower costs than previously possible, has provided a strong incentive in the discovery of disease-causing gene for Mendelian disorders (Shendure J and Ji H, 2008).

2.1 First-generation sequencing technologies

In 1977, Sanger *et al*. (Sanger *et al*., 1977) introduced the initial sequencing methods that provided the underpinnings for the development of automated sequencing technologies, termed now as 'first-generation sequencing' (Smith *et al*., 1986; Ansorge *et al*., 1987). Technically, standard Sanger sequencing identifies a linear sequence of nucleotides by electrophoretic separation of randomly terminated extension products. Automated strategy relies on the use of fluorescently-labeled terminators, capillary electrophoresis separation and automated laser signal detection to improve nucleotide sequence detection (Prober *et al*., 1987). Sanger sequencing is the most available and accurate method for sequencing, representing the gold standard for DNA sequencing. However, despite such availability and accuracy, Sanger sequencing has restricted applications due to low throughput, represented by the amount of DNA sequence that can be read with each sequencing reaction. More precisely, throughput is related to the number of sequencing reactions that can be run in parallel and the length of sequences read by each reaction, that is, the number of nucleotides per sequence read. A read refers to the sequence of a cluster that is obtained after the end of the

sequencing process that is ultimately the sequence of a section of a unique fragment. The need for electrophoresis to separate DNA fragments and read DNA content is the primary limitation of the method, increasing time and significantly reducing the number of reactions that can be performed in parallel (Hert *et al*., 2008). Each Sanger instrument can only read reactions in parallel, consequently restricting the throughput to approximately 115kb/day (Mardis, 2011) and extending time to accomplish the sequencing of entire genomes. However, automated Sanger sequencing was the core technology of the Human Genome Project and many others animal and plant genomes projects.

2.2 Human genome project: foundation for the development of new sequencing technologies

With the audacious decision to start a research program to characterize in details the complete sequence of human genome, the Human Genome Project officially began on the 1st October of 1990 (Department of Health and Human Services and Department of Energy, 1990). An initial estimated planning estimated 15-years time and $3 billion to complete the entire job (National Research Council Committee on Mapping and Sequencing the Human Genome, 1988). However, improvements in technology, success in achieving early mapping goals and a growing demand for the human DNA sequence made possible the accomplishment of the complete sequencing of 80000 human genes and 3 billion DNA base pairs two years ahead of schedule, in 2003 (Collins *et al*., 1998), coinciding with the 50th anniversary of Watson and Crick's description of the fundamental structure of DNA (Watson and Crick, 1953). This monumental achievement was realized using a technology based on an optimized, robust and highly automated capillary electrophoresis of individual fluorescently labeled Sanger sequencing reactions to support the huge scale required for the Human Genome Project Once the Human Reference Genome was completed (Lander *et al*., 2001), the efforts of the International Human Genome Project teams turned to identify the positions of common single-nucleotide polymorphisms (SNPs) known to exist in the genome. The international SNP discovery efforts were known as the 'HapMap' project since they aimed to map haplotype diversity in the human genome, characterizing

common SNPs (present at 5% or greater allele frequency) in multiple human populations. The HapMap efforts culminated in the identification of more than 8 million common SNP positions genome-wide, most of which were generated by Sanger-based capillary methods. Also the prominent 1000 Genome Project (1000 Genomes Project Consortium *et al*., 2012) added resolution and new information to our knowledge of genome diversity across all levels of genetic variations, including SNPs, indels and copy number variation (CNV) (Sudmant *et al*., 2010). The project concentrated on sequencing and analyzing the genomes of at least one thousand anonymous participants from a number of different ethnic groups, populations already studied in the HapMap projects as well as newly consented ones. Both projects were accompanied by the necessity for extensive sequencing. They led, together with a program aiming at the economic sequencing of complete, high quality, mammal-sized genomes (Service, 2006; Mardis, 2006), to the development of new sequencing technologies.

2.3 Key principles of next-generation sequencing technology

The years following the Human Genome Project has witnessed a considerable explosion in the development of next-generation sequencing (NGS) technologies, also known as 'massively parallel sequencing' (MPS), that has enabled rapid generation of data by sequencing massive amounts of DNA in parallel, in a genome-wide and high-throughput manner. This new technology greatly increased the amount of sequence output per run, lowered costs and read lengths. Although each platform is different in its specific, all NGS devices share certain characteristics (Mardis, 2009; Metzker, 2010). An overview of NGS workflow is schematically illustrated in **Figure 4**. Generally, starting material is represented by double-stranded DNA whose source can vary among genomic DNA, immuno-precipited DNA, reverse-transcribed RNA or cDNA. All starting material is converted into a library of sequencing reaction templates, which undergo common steps of fragmentation, size selection, and adapter ligation (Linnarsson, 2010). Fragmentation and size selection steps break the DNA templates into smaller fragments able to be sequenced, the size of which depend on each sequencing platform’s specifications. Adapter ligation adds platform-specific synthetic

DNAs to the ends of library fragments, and primers designed over this adapter are used for downstream amplification and/or sequencing reactions. Depending on the NGS technology used, a library is either sequenced directly (single-molecule templates) or is amplified and later sequenced (clonally amplified templates). However, most sequencing platforms cannot identify single-molecule reactions and template amplification is therefore necessary to generate sufficient signal for detection of nucleotide addition by the instrument's optical system (Metzker, 2010; Mardis, 2011). Amplification strategies vary between platforms and commonly include emulsion PCR (Dressman *et al*., 2003) or bridging amplification strategies (Fedurco *et al*., 2006). Template generation also aims to spatially separate and immobilize DNA fragment populations for sequencing, typically by attachment to solid surfaces or beads. This activates the downstream sequencing step that operates as millions of microreactions performed in parallel on each spatially distinct template (Natrajan and Reis-Filho, 2011). Sequencing reactions are performed as an organized series of repeating steps that are realized and detected automatically. According to the platform and the library construction approach used, sequencing can start from both ends of linear fragments (`paired end sequencing´) or from both ends of previously circularized fragments (`mate pair sequencing´). Paired end sequencing is the standard by which human genomes is sequenced, mainly because it requires a small amount of DNA. By contrast, mate pair sequencing is quite DNA expensive due to low yield of circularization of large DNA molecules. The benefit of obtaining sequence data from both ends of library fragments is obvious if you consider the highly repetitive nature of human genome. While all sequencing platforms differ in their proprietary chemistries, the use of DNA polymerase or DNA ligase enzyme is a common feature, and these methods have been referred to collectively as 'sequencing by synthesis' (SBS) strategies in the literature (Fuller *et al*., 2009). Finally, ultimate challenging step of NGS is the analysis of the data produced through sequencing process.

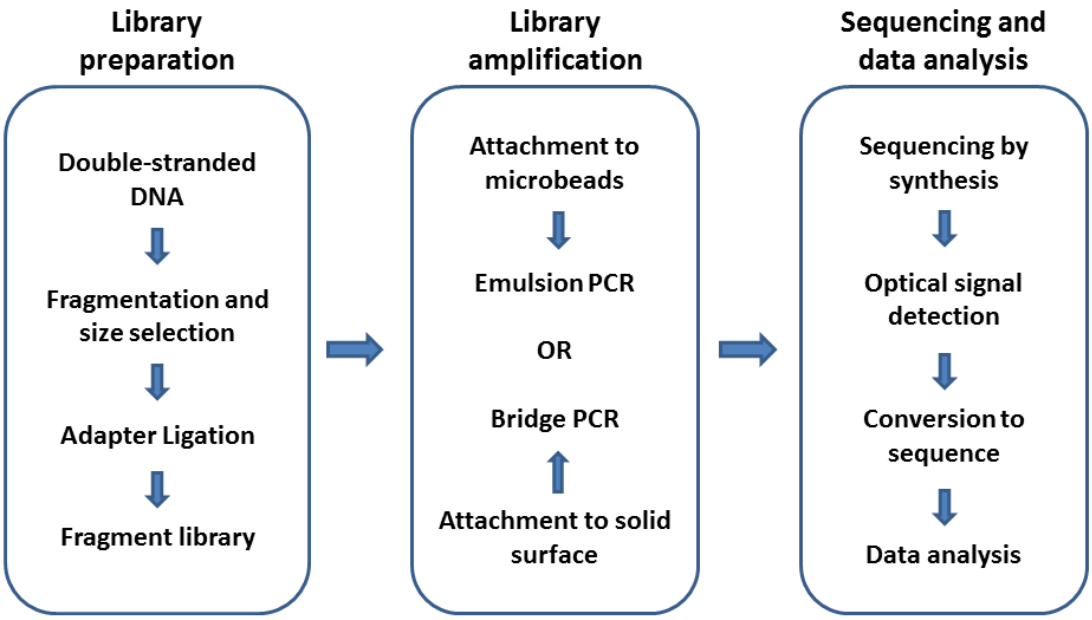


Figure 4 An overview of how NGS works. In general, the starting material for all NGS experiments is double-stranded DNA that is converted into a fragment library through following steps of fragmentation, size selection and adapter ligation. Library amplification varies between platforms and commonly includes emulsion PCR or bridging amplification strategies. During this step, DNA fragment populations are also spatially separated and immobilized for sequencing, typically by attachment to solid surfaces or beads. After that, sequencing step generates millions of reactions performed in parallel on each spatially distinct template and detected automatically by an optical system. Finally, raw sequence data are analyzed in details with specific bioinformatics tools. Based on Raza and Sabahuddin, 2012.

2.4 The promise of whole-exome sequencing in medical research

Since its early days, medical research has striven for identifying the causes of disorders with the ultimate goal of establishing therapeutic treatments and finding cures. Nowadays, whole-genome sequencing (WGS) approaches allow the complete DNA sequence of an organism's genome at a single time and are designed to discover genetic variations contributing to rare or common diseases. Despite the significant improvement in sequencing technology, sequencing whole genomes to a depth sufficient to find pathogenic variants is expensive compared with targeted sequencing. Whole-exome sequencing (WES) is a method that targets only a subset of the genome, significantly reducing the sequencing space and subsequently the cost. The exome has been defined as the portion of genomic sequence including all exons of protein coding

genes that covers between 1 and 2% of the genome, depending on species. Although it represents only a small portion of the genome, WES has revealed a promising approach since it is believed that exome harbors the most functional variation (Botstein and Risch, 2003). This is based on the observation that mutations that cause Mendelian diseases occur primarily in genes (Stenson *et al*., 2003). Mutations that cause amino acid substitutions in their respective genes, including changes to nonsense codons, are the most frequent type of disease mutation (∼60%) (Botstein *et al*., 2003*)*. In addition, small indels in genes account for almost a quarter of the mutations in Mendelian disease (Botstein *et al*., 2003*;* Stenson *et al*., 2003). Meanwhile, less than 1% of Mendelian disease mutations have been found in regulatory regions.

The exome enrichment process is mainly categorized into three steps, namely library preparation, hybridization and sequencing (**Figure 5**). Two main categories of exome capture technology have been developed: solution-based and array-based. In solution-based, DNA samples are fragmented by nebulization or sonication to get desired fragments of about 250 bp and biotinylated oligonucleotides probes (baits) are used to specifically hybridize to target regions. Magnetic streptavidin beads are used to bind the biotinylated probes, the non-targeted portion of the genome is washed away, and PCR is used to amplify and enrich samples for DNA from target regions. The quality and quantity of the exome library is analyzed by high sensitive methods before sequencing step. The sample is then sequenced before proceeding to bioinformatic analysis. Array-based methods are similar except that probes are bound to a microarray. The array-based method was the first to be used in exome capture (Albert *et al*., 2007), but it has been replaced by solution-based methods that require less amount of input DNA and are consequently potentially more efficient. The array-based methods are less scalable due to the limitation of the number of probes that can be accommodated on the array. There are several available exome capture platforms that are constantly updated and improved. The major providers of exome capture platforms are NinbleGen, Agilent and Illumina which concentrated their efforts in developing human kits, although some providers offers also some non-human species kits.

The first successful use of WES to as informative tool for diagnosis and subsequent treatment was the identification of the causal variant of a rare form of inflammatory

bowel disease in an infant (Worthey *et al*., 2011). Conventional diagnosis failed to explain patient's severe symptoms and doctors wanted to understand the underlying causes before selecting a treatment. The analysis of WES data identified two variants located in highly conserved region, one of them with a high null allele frequency that generate a single hemizygous, nonsynonymous variant in the X-linked inhibitor of apoptosis (*XIAP*) gene.

After this first success, WES has been extensively used to diagnose novel diseases and find novel causative mutations known disease phenotypes. Miller syndrome is the first rare Mendelian disorder from which its causal variants were identified by means of WES (Ng *et al*., 2010). Mutations in *DHODH* in three affected pedigrees were found as the cause of the disease after filtering against public SNP databases and eight HapMap exomes. WES is a useful approach in human medicine for diagnosis of particular difficult cases, such as patients who may not yet exhibit a full spectrum of symptoms (Iglesias *et al*., 2014) and prenatal diagnosis (Xu *et al*., 2014) and early diagnosis of debilitating disease (Bras and Singleton, 2011). Moreover, finding the causative mutations also allow for the selection of an appropriate treatment, prevention of invasive tests and accurate prognosis (Grossmann *et al*., 2011; Rabbani *et al*., 2012; Taneri *et al*., 2012). Examples of diseases for which exome sequencing has been used to detect a causative mutation include Leber congenital amaurosis (Wang *et al*., 2011), Alzheimer disease (Sassi *et al*., 2014), maturity-onset diabetes of the young (Johansson *et al*., 2012), high myopia (Shi *et al*., 2011), autosomal recessive polycystic kidney disease (Xu *et al*., 2014), amyotrophic lateral sclerosis (Johnson *et al*., 2010), acromelic frontonasal dystois (Smith *et al*., 2014) and several cancer predisposition mutations (Yan *et al*., 2011; Snape *et al*., 2012; Kiiski *et al*., 2014).

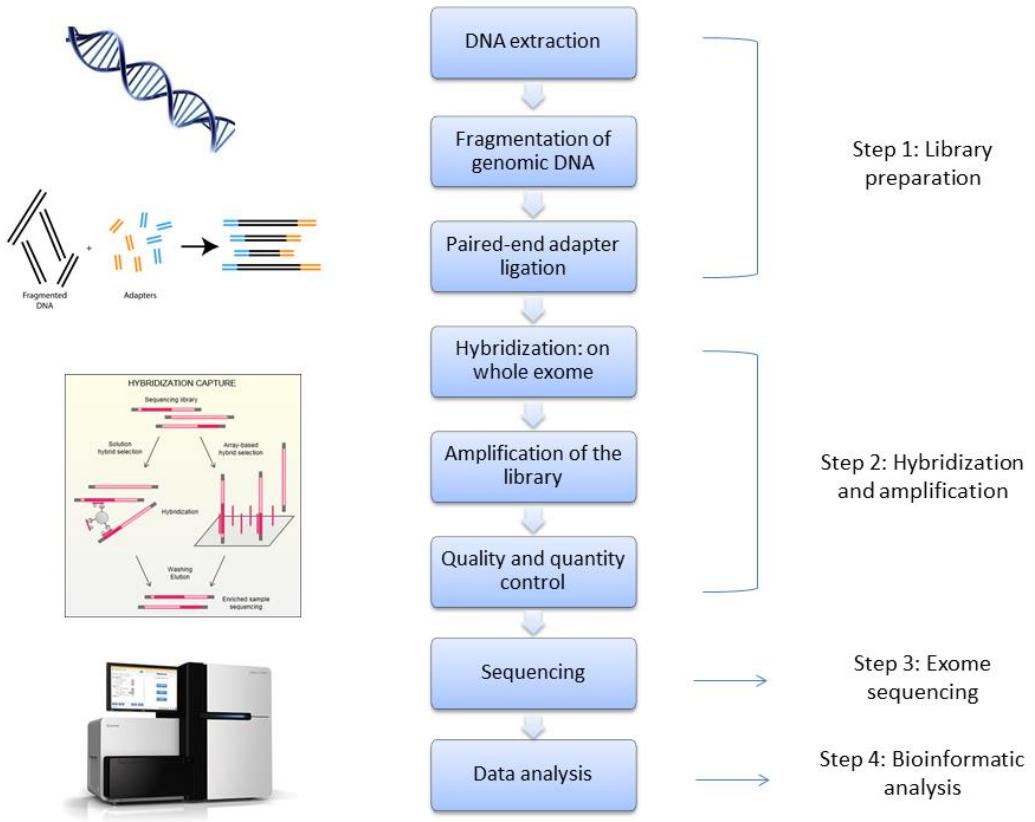


Figure 5 Exome enrichment workflow. Extracted DNA is fragmented by nebulization or sonication to get fragment length of almost 250 bp and paired-end adapters are ligated. In solution-based method, specific target regions hybridize with biotinylated oligonucleotides probes that, in turn, are bound by magnetic streptavidin beads, whereas in array-based method hybridization is realized by probes bound to a microarray. Non-targeted portion of the genome is washed away. An amplification step is used to enrich samples for DNA from target regions. After quality and quantity control, enriched exome sample is ready to be sequenced and data to be analyzed.

2.5 Bioinformatic analysis of next-generation sequencing data

NGS reactions generate huge sequence data sets that range from million megabases to billion gigabases (Natrajan and Reis-Filho, 2011). As consequence, it has been assisted to a revitalization of bioinformatics with data analysis challenges, owing to massive scale of data to be analyzed and interpreted. The difficulties in data analysis are due to many factors, including the size and complexity of the human genome, the variability of read length and the accuracy of NGS data taking in consideration the full length of variant detection. Several bioinformatic tools have been developed and can be either downloaded on local machine or used online for the analysis of NGS data.

After sequencing, the data are processed in three major steps, including mapping (or alignment), variant calling and annotation steps (**Figure 6**). The sequence data are

aligned with Burrow-Wheeler Aligner tool (Li and Durbin, 2009) against a reference genome such as human GRCh37/hg19, alternatively a *de novo* assembly is conducted. In variant calling, previously aligned data could be used by SAMTools (Li *et al*. 2009) for quality control, short read alignment and variant identification, whereas Genome Analysis Toolkit (McKenna *et al*., 2010) to calculate sequencing coverage, that refers to the average number of times a base pair is sequenced in a given experiment. Both the quality and quantity of sequence data derived from NGS experiments will ultimately determine the comprehensiveness and accuracy of downstream analysis (Ajay *et al*., 2011). Qualitatively, individual base calling error rates vary among different NGS platforms, which provide confidence scores for each individual base call and enable researchers to use different quality filter when extracting their sequence data. Quantitatively, the amount of sequence data can be assessed by the metric of sequence coverage. More specifically, the coverage metric is best viewed in the context of the physical location, or distribution, of sequence reads, as NGS reactions may not represent all genomic locations uniformly. In annotation step, ANNOVAR (Wang *et al*., 2010) is a tool for annoting genetic variants based on the function, including information such as gene name, chromosomal position, nucleotide changes, aminoacid changes, SNP database ID, allelic frequency of the SNP in 1000 Genome project and sequence quality. Additionally, SIFT (sorting intolerant from tolerant) (Kumar *et al*., 2009) and Polyphen (polymorphism phenotyping) (Adzhubei *et al*., 2010) values are reported, being useful damage score predictors of the variant impact on the protein function according to sequence homology, protein structure and physicochemical properties of aminoacids. Finally, VAAST (Variant Annotation, Analysis and Search Tool) incorporates previous aminoacid substitution information with annotation and ranks candidate genes with statistical evaluation, which can be useful to generate a list of pathogenic candidate genes and variants (Yandell *et al*., 2011).

Most investigators filter the data based on the nature of variants. Nearly half of variants are synonymous ones, not considered to be deleterious, which are usually filtered out. Although some reports about the causal effect of synonymous variants exists (Sauna *et al*., 2011), the probability is very low. The remaining variants are nonsense, missense, indels, splice mutations and other non-coding RNA transcript.

Although these analytical methods have been refined, it is important and necessary to perform a parallel validation of new discovered variants before reporting them.

A single exome carries about 20.000-30.000 coding SNPs. Filtering steps narrow down the number of putative disease-associated genes. WES takes advantages of the availability of databases of known SNPs, known pathogenic variants and control genomes to find novel variants. Specifically, variants already known to be associated to diseases, found in these databases, can be excluded during bioinformatic analysis, significantly reducing the variant pool. The Exome Aggragation Consortium (ExAC) has created a freely available database containing the exome sequences of more than 60,000 unrelated individuals (Exome Aggregation Consortium, 2014. Available at: http://exac.broadinstitute.org). The exomes have been analyzed using a uniform bioinformatic pipeline and are from individuals with adult-onset diseases. Researchers have a large set of reference exomes that should be free from homozygous variants that cause childhood-onset Mendelian diseases.

According to the knowledge of the affected cases, different analytic frameworks are used to determine causal pathogenic variants (**Figure 6**). A substantial number of *de novo* mutations occurs sporadically with an estimated rate of $7{,}6 \times 10^{-9} – 2{,}2 \times 10^{-8}$ per generation, that is, approximately one in 10^8 base per haploid genome is mutated spontaneously (Roach *et al*., 2010; Lynch, 2010). Spontaneous germline mutations can have serious phenotypic consequences when they affect functionally relevant bases in the genome. In fact, their occurrence may explain why diseases with a severely reduced fecundity remain frequent in the human population, especially when the mutational target is large and comprised of many genes. For *de novo* analysis, the parent-child trios study is practical. Non-pathogenic variants are filtered out and then the variants present in the parents are excluded. It is important to recognize that bias can be introduced in all steps of NGS experiments, particularly during library amplification step as DNA Polymerases are not 100% accurate and can introduce mutations. Consequently, sequencing errors and mapping artefacts might exist and confirmation using Sanger sequencing with high accuracy should be applied (Gilissen *et al*., 2011). Detection of a *de novo* variant is not enough to confirm the causality of the disease. Additional analysis, including replication and functional analysis, should be performed to determine the deleterious or causal variants. Pathogenicity of a variant

not only depends on type and location of the mutation, but also on its functional impact (Cooper *et al*., 2010).

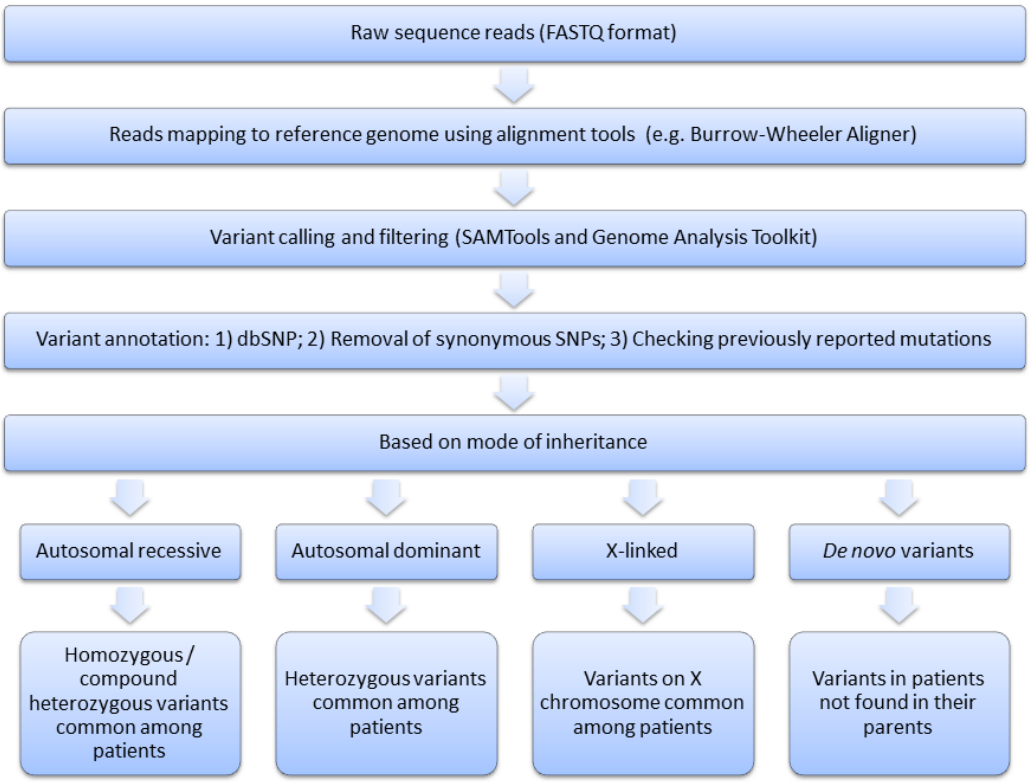


Figure 6 Bioinformatic analysis of WES. Raw sequence data, in FASTQ format, undergo a bioinformatics process mainly divided in three steps: mapping, variant calling and annotation. In the first step, sequence read are mapped to a reference genome such as human GRCh37/hg19 using alignment tools (e.g. Burrow-Wheeler Aligner). In the second one, previously aligned data are used for quality control and variant identification (SAMTools), as well as to calculate technical parameters such as sequence coverage (Genome Analysis Toolkit). In the third one, variant information are annoted, including gene name, chromosomal position, nucleotide changes, aminoacid changes, SNP database ID, allelic frequency of the SNP in online databases and sequence quality (ANNOVAR). Next, depending on the mode of inheritance of the specific cases, different filters can be applied to generate a list of possible disease-associated candidate genes or variants, whose number is narrowed down by filtering steps. Based on Rabbani *et al*., 2012.

2.6 Application of next-generation sequencing technologies for neurodevelopmental disorders

Numerous neurological diseases have well-established evidence of genetic contribution (Urdinguio *et al*., 2009). The complex circuitry that constitutes the substrate of the mind requires an equally complex network of genes to orchestrate its self-assembly. Mutations affecting any of a wide range of cellular processes can lead to altered neurodevelopment and result in neurological or psychiatric disease. The plummeting cost of sequencing (Wetterstrand, 2013) and concomitant introduction of high-throughput sequencing technologies paved the way for studies of neurodevelopmental diseases that do not have structural abnormalities, by which neurodevelopment disorders are often diagnosed, and are of unclear inheritance. Several large-scale studies using whole-exome sequencing have identified a high rate of *de novo* mutations in non-syndromic cognitive disorders, including IDs. Unlike inherited mutations, *de novo* mutations are present in a child but not in that child's parents, and likely arose in the germline of the parent. *De novo* mutations in ID consist of large CNVs that delete one allele of many genes (Cooper *et al*., 2011) and more recently also isolated point mutations (de Ligt *et al*., 2012).

Another example where modern sequencing techniques have elucidated the genetic underpinnings of non-syndromic cognitive disorders is ASDs. ASDs are a complex and genetically heterogeneous group of diseases characterized by impaired social interaction, communication deficits and stereotyped behaviours (Huguet *et al*., 2013). ASDs commonly occur in sporadic individuals without significant family history and would have been impossible to study using traditional methods of linkage analysis and positional cloning (Gusella *et al*. 1983; Riordan *et al*., 1989). Only with the advent of NGS it has been possible to probe the genetic landscape of ASDs by surveying broad regions of the genomes of many individuals. Several recent studies have shown the importance of *de novo* copy-number variations (Levy *et al*., 2011; Pinto *et al*., 2010) and *de novo* point mutations in ASDs (Awadalla *et al*., 2010; Michaelson *et al*., 2012).

A second theme that has emerged from whole-exome sequencing studies is the contribution of recessive hypomorphic mutations to ASDs. These mutations partially inactivate genes that are known to cause more severe phenotypes when completely inactivated. Unlike *de novo* mutations, which inactivate only one of the two copies of

an autosomal gene, leaving one copy of the gene still functional, both alleles are mutated in recessive, hypomorphic mutations causing ASDs. For example, hypomorphic mutations in the aminomethyltransferase *AMT* have been identifies with a range of neurodevelopmental symptoms, including ASDs, language and motor delays, ID and epilepsy (Yu *et al*., 2013). Mutations in *AMT* cause nonketotic hyperglycinemia (NKH), a recessive syndrome resulting from impaired glycine cleavage (Applegarth and Toone, 2004). The biochemical abnormalities that are typically detected in classic NKH often escape clinical detection methods in milder atypical NKH (Applegarth and Toone, 2001; Dinopoulos *et al*., 2005). Functional assays to characterize the ASD *AMT* mutations demonstrated that the enzymatic activity of the mutant AMT was partially reduced, confirming that the affected children had a mild form of undiagnosed NKH presenting as an ASD (Yu *et al*., 2013).

The significant variability in the phenotypes that can be associated with neurodevelopmental syndromes makes genetic diagnosis of the milder atypical forms presenting as ASD and other non-malformation neurodevelopmental diseases challenging. Clinical sequencing should help overcome this challenge in at least a subset of cases, driving genetic rather than phenotypic classification.

3 *Caenorhabditis elegans*

As previously mentioned, the experimental model systems commonly used to understand the detailed mechanism of RTT have been mainly mouse models lacking or overexpressing MeCP2 in various degrees. In the present work, *C. elegans* is introduced for the first time as experimental model system to functionally validate new disease-causing candidate genes, coming from NGS data analysis, potentially implicated in the physiopathology of RTT.

3.1 *Caenorhabditis elegans* as a powerful in vivo model organism

A current challenge in neuroscience is to bridge the gaps between genes, proteins, neurons, neural circuits and behavior in a single animal model. In the second half of the twentieth century Sydney Brenner introduced the soil nematode *Caenorhabditis elegans* (*C. elegans*) as a model organism for the study of development and neurobiology (Brenner, 1974). *C. elegans* has unique features that make it a powerful tool to study the function of novel targets. First, it is easy to culture. Although the natural environment of the worm is the soil and feeds on different bacteria, it can be easily raised in the laboratory on a diet of *Escherichia coli.* Second, it reproduces rapidly and abundantly: within 3.5 days at 20°C it develops from eggs to an adult worm of 1.3 mm in length. This short generation time and having 300-350 descendants per self-fertilizing hermaphrodite allows a large-scale production of animals per day. Third, due to its small size, many assays can be carried out in microtiter plate of 96 wells either on agar or in liquid using more than hundred animals in a single well. Fourth, the worm is transparent and processes such as axon growth, embryogenesis and fat metabolism can be studied in the living animal by using *in vivo* fluorescence markers. Fifth, it is a multicellular animal that forms many different organs and tissues, including muscle, hypodermis, intestine, reproductive system and nervous system (Riddle *et al*., 1997).

C. elegans occurs in two dimorphic sexes, males and self-fertilizing hermaphrodites (**Figure 7**). Both sexes are diploid for five autosomal chromosomes, whereas they differ in sex chromosomes dosage: hermaphrodites have two X chromosomes and males

have a single one, in a way that sex is determined by the X to autosome (X:A) ratio (Zarkower, 2006). *C. elegans* has no Y chromosome and the genotype of males is referred to as XO. The gonad of hermaphrodites forms an ovotestis that first produces haploid amoeboid sperm, stored in a spermatheca over larval development, and then near adulthood the germ line switches fate to produce oocytes, accumulated in the uterus (**Figure 7**). Hermaphrodites are somatically female that can reproduce either by self-fertilization or by mating with males. In the first case, hermaphrodites can produce up to 300 self-progenies that are fertilized by the stored sperm. The majority of offspring produced by self-fertilization are hermaphrodites; males represents only 0.1-0.2% of the progeny due to rare meiotic non-disjunction of the X chromosome. If mated with males, hermaphrodites are capable of producing almost 1000 descendants, indicating that hermaphrodite-produced sperm is a limiting factor in self-fertilization. Adults of both sexes are composed of a precise number of cells: hermaphrodites have 959 somatic nuclei, while males have 1031 (Sulston *et al*., 1977; Sulston *et al*., 1983). Adult males are smaller than aged-matched hermaphrodites, despite the fact that they are composed of more somatic nuclei. In hermaphrodites, the nervous system consists of 302 neurons and 56 glial and support cells, whereas males have 381 neurons and 92 glial and support cells. In both sexes each neuron is uniquely recognizable in different individuals by its characteristic position and morphology (White *et al*., 1986).

The genome of *C. elegans*, comprising about one hundred million base pairs, is completely sequenced and incredibly similar to that of humans; depending on the particular bioinformatics approach used, it is estimated that 60–80% of the genes have a human homolog (Harris *et al*., 2004). Viable mutant strains, as well as strains that overexpress a gene or lack a gene function can be efficiently generated and the resulting phenotypes can be quickly identified (Hariharan *et al*., 2003). Comprehensive information regarding gene structure, expression patterns, protein-protein interactions, mutant or RNA interference (RNAi) phenotypes and microarray data, is available in Wormbase, the online resource for nematode-related information (http://www.wormbase.org/) (Chen *et al*., 2005). Nematode strains can be stored indefinitely in liquid nitrogen allowing large mutant collections and construction of public mutant repositories. *C. elegans* is amenable to forward and reverse genetic

screens and is particularly susceptible to gene inactivation by RNAi (Kamath and Ahringer, 2003; Fire *et al*., 1998). *C. elegans* has emerged as a powerful experimental system to study the molecular and cellular aspects of human neurodegenerative diseases *in vivo*, including Alzheimer's disease, juvenile Parkinson's disease, spinal muscular atrophy, as well as metabolic disorders such as hereditary nonpolyposis colon cancer, and many others age-related disorders (Baumeister and Ge, 2002; Poulin *et al*., 2004; Kenyon, 2005). Modeling a human disease in a simple invertebrate, such as *C. elegans*, allows the dissection of complex molecular pathways into their component parts, thus providing a meaningful insight into the pathogenesis of a complex disease phenotype.

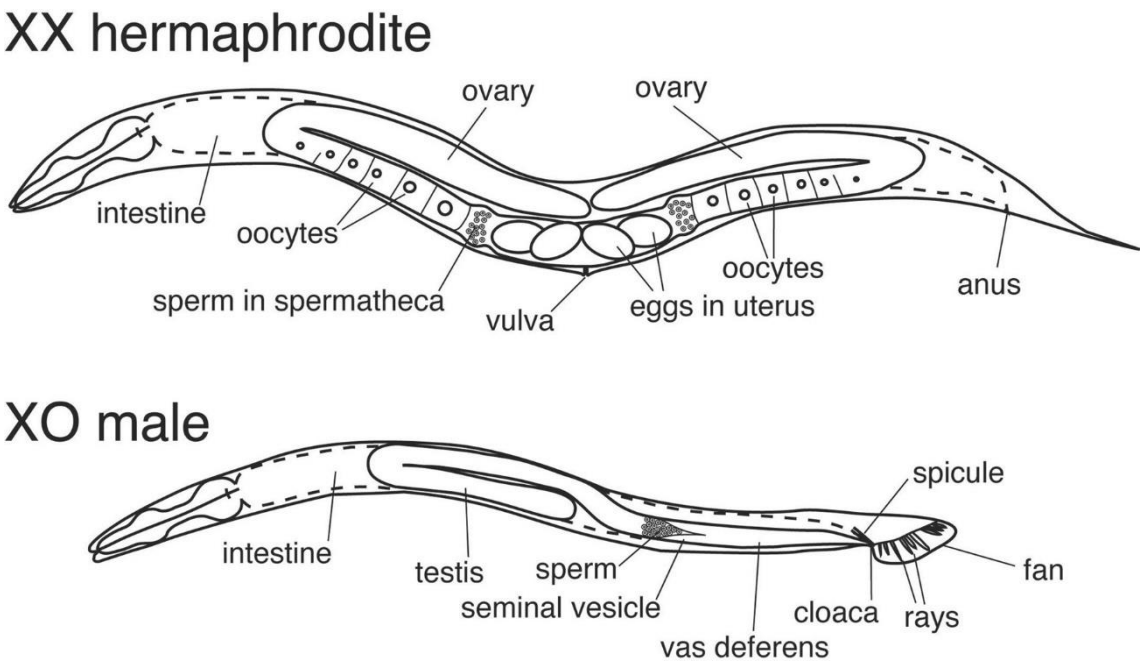


Figure 7 The two sexes of *C. elegans*. Hermaphrodites and males are strikingly different in global body size and structures such as somatic gonad and tail. The gonad of hermaphrodites functions as an ovotestis that produces haploid amoeboid sperm, stored in a spermatheca, as well as much larger oocytes, accumulated in the uterus. Surprisingly, adult males are smaller than aged-matched hermaphrodites, despite that they are composed of more somatic nuclei. Also nervous system is sexually dimorphic. In hermaphrodites, nervous system consists of 302 neurons and 56 glial and support cells, whereas males have 381 neurons and 92 glial and support cells. Figure taken from Wormbook, Zarkower, 2006.

3.2 Wiring of the *C. elegans* nervous system

Referring to the hermaphrodite, *C. elegans* is the only animal whose complete wiring diagram of the nervous system has been reconstructed from serial section electron micrographs, so that the morphology of each neuron and its chemical synapses and gap junctions are known (White *et al*., 1986). The 302 hermaphrodite neurons can be grouped by anatomical criteria into 118 classes. These include 39 classes of predicted sensory neurons, another 27 classes of motor neurons and the remainders are classed as interneurons. Most classes of sensory neurons and interneurons consist of two left/right homologs with similar synaptic connectivity, which are connected by gap junctions and act cooperatively. Together, the 302 neurons make approximately 5000 chemical synapses, 600 gap junctions, and 2000 neuromuscular junctions.

C. elegans contains many of the classic neurotransmitters found in vertebrates, including acetylcholine, glutamate, gamma-aminobutyric acid (GABA), serotonin and dopamine. In addition, *C. elegans* neurons also have dense core vesicles, characteristic of the catecholamine- and neuropeptide-containing vesicles in other animals. The *C. elegans* genome sequence reveals numerous neuropeptides. For example, the Phe-Met-Arg-Phe-amide FMRFamide family of neuropeptides is represented by at least 23 genes that encode 60 peptides (Kim and Li, 2004*;* Li *et al*., 1999).

The *C. elegans* genome sequence encodes nearly 20.000 genes that include homologs of neuronal expressed mammalian genes (Bargmann, 1998). Mutations have been isolated that disrupt genes required for sensory transduction, development of specific neuron classes and neurotransmitter production or release. In many cases these genes represent key point in signaling pathways and neuronal circuit underlying a specific behavior. Many behavioral studies have been driven by the isolation and analysis of mutants. Almost all of them are carried out in the same wild type N2 strain (Bristol), which is homozygous for one allele at all gene loci. Moreover, the development of a genome-wide RNA interference (RNAi) library has also facilitated candidate gene analysis because the function of genes can easily and systematically be knocked down by feeding or injecting RNA (Kamath *et al*., 2003). However, many neural genes are resistant to RNAi and, consequently, mutants are usually required for detailed studies of gene function (Kennedy *et al*., 2004).

3.3 The neuronal circuit of *C. elegans* locomotion

The rhythmic locomotory gait of *C. elegans* has long been an area of interest. Sensory inputs modulate ongoing patterns of neuromuscular activity that regulate motor behavioral outcomes such as locomotion. Worms move by undulatory propulsion in which a multitude of waves, generated by the successive contraction and relaxation of dorsal and ventral longitudinal body wall muscles, is passed along the body. Three forces make movement possible on solid surface: the hydrostatic skeleton of the animal, against which muscles work; surface tension, which pushes the worm against its substrate; and friction between worm and its substrate, which allows body waves to travel along the animal's length without sideslip, therewith producing forward propulsion (Niebur and Erdös, 1993).

The locomotory musculature consists of 95 body-wall muscles cells that receive excitatory inputs at cholinergic neuromuscular junctions (NMJs), and inhibitory inputs at GABAergic ones (Chalfie and White, 1988). Five classes of motor neurons, named A, B, D, VC and AS, are distributed along the length of the animal and innervate a local region of musculature (**Figure 8**). The A, B, and D classes are further divided as the dorsal and ventral muscle innervating subclasses, DA, VA, DB, VB, DD, and VD (Von Stetina *et al*., 2006). The A and B-type neurons make cholinergic excitatory NMJs with muscle cells (Richmond and Jorgensen, 1999). The D-type neurons make GABAergic inhibitory synapses to muscle cells (McIntire *et al*., 1993). The D-type GABAergic neurons are postsynaptic to the cholinergic A and B-type motor neurons that innervate muscles of the opposite side. While the dorsal nerve cord contains only neural processes, the ventral nerve cord also contains many neural cell bodies distributed along its length (White *et al*., 1976), most of which are motor neurons that innervate the body wall muscles.

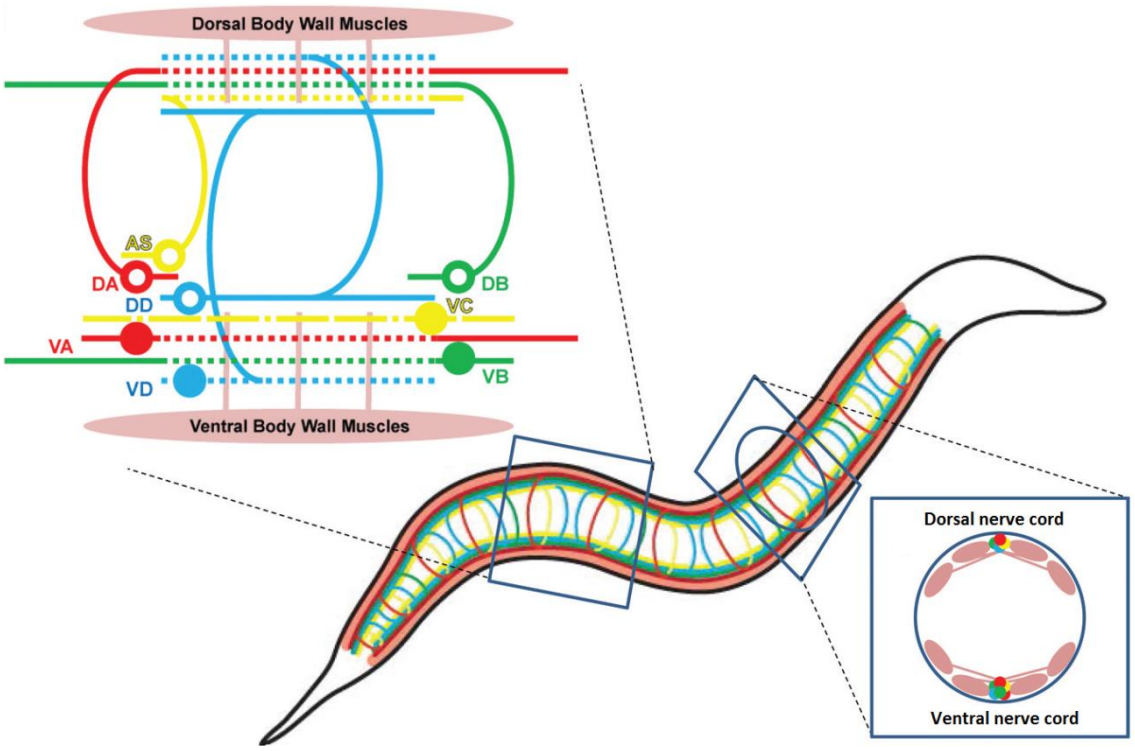


Figure 8 The neuromuscular system of *C. elegans* for locomotion. Five classes of motor neurons (A, B, D, VC and AS) are distributed along the body of the animal and innervate a local region of musculature. Additionally, the A, B, and D classes are divided as the dorsal and ventral muscle innervating subclasses, DA, VA, DB, VB, DD, and VD. The motor neurons and muscles cells repeat along the body to create the complete neuromuscular system. Muscle cells are organized in four quadrants next to the dorsal and ventral nerve cord. Based on Gjorgjieva *et al*., 2016.

One set of excitatory and inhibitory dorsal (DB, DD) and ventral (VB, VD) neurons controls forward movement, and a second set (DA, DD and VA, VD) controls backward movement. The connection of those two sets of neurons guaranties that while a particular region of the ventral musculature contracts, due to excitation by cholinergic motor neurons, the opposite dorsal musculature relaxes, owing to inhibition by GABAergic motor neurons (**Figure 9**). In that manner, the worm sinusoidal bending is controlled at a particular place and time by an elegant mechanism of contralateral inhibition that avoids simultaneous contraction of the dorsal and ventral musculature. Such mutual inhibition is mainly important during rapid escape movements. Indeed, it has been demonstrated that mutants defective in GABAergic signaling manifest a shrinking phenotype rather than coordinated escape movements (McIntire *et al*., 1993).

In addition to motor neurons, a group of interneurons that control forward and backward movement were identified by means electron micrographs reconstruction of

the nervous system with studies of the neural control of touch avoidance. Specialized sensory neurons detect touch to the anterior (AVM, ALM) or posterior (PVM, PLM) parts of *C. elegans*. AVD and a second interneuron called AVA innervate the VA and DA motor neurons, which coordinate backward movement. PVC and AVB interneurons innervate the VB and DB motor neurons, which coordinate forward movement (Chalfie *et al*., 1985; White *et al*., 1986) (**Figure 9**). The anatomical connection among these interneurons mediates the withdrawal response to body touch and control the choice between forward and backward movement. Laser ablation studies support this hypothesis: ablation of AVA and AVD abolishes anterior touch avoidance and backward movement, whereas ablation of AVB and PVC abolishes posterior touch avoidance and forward movement (Chalfie at al., 1985; White *et al*., 1986). Due to their importance for forward and backward movement, AVA, AVB, AVD and PVC are termed command interneurons.

Worms swim in fluids or crawl on agar surfaces with different locomotory gaits (Pierce-Shimomura *et al*., 2008). Swimming is characterized by large amplitude, high frequency, and long wavelength undulations; on the contrary crawling by small amplitude, low frequency short wavelength undulations. When nematodes navigate environments with intermediate characteristics, they display intermediary gaits with continuous variations in amplitude, frequency and wavelength (Berri *et al*., 2009). Being transparent and genetically tractable, the worm was an ideal tool to dissect the motor circuit behind locomotion and define the contribution of individual neurons for motor output. Activity patterns within the muscles of crawling and swimming worms were visualized by calcium imaging in freely moving worms, and it has been shown that different gaits are associated with distinct spatiotemporal pattern of activity in the ventral and dorsal musculature (Pierce-Shimomura *et al*., 2008). Using light-activated channels and electrophysiology, Liu *et al*. measured the electrical current evoked in muscles cells upon depolarization or hyperpolarization of motor neurons, and found that levels of neurotransmission at the neuromuscular junction were continuously graded (Liu *et al*., 2009). These findings indicate that the motor circuit allows continuous variation of contraction and relaxation within muscles cells during the propagation of the waves, adapting locomotory gait to different mechanical environment.

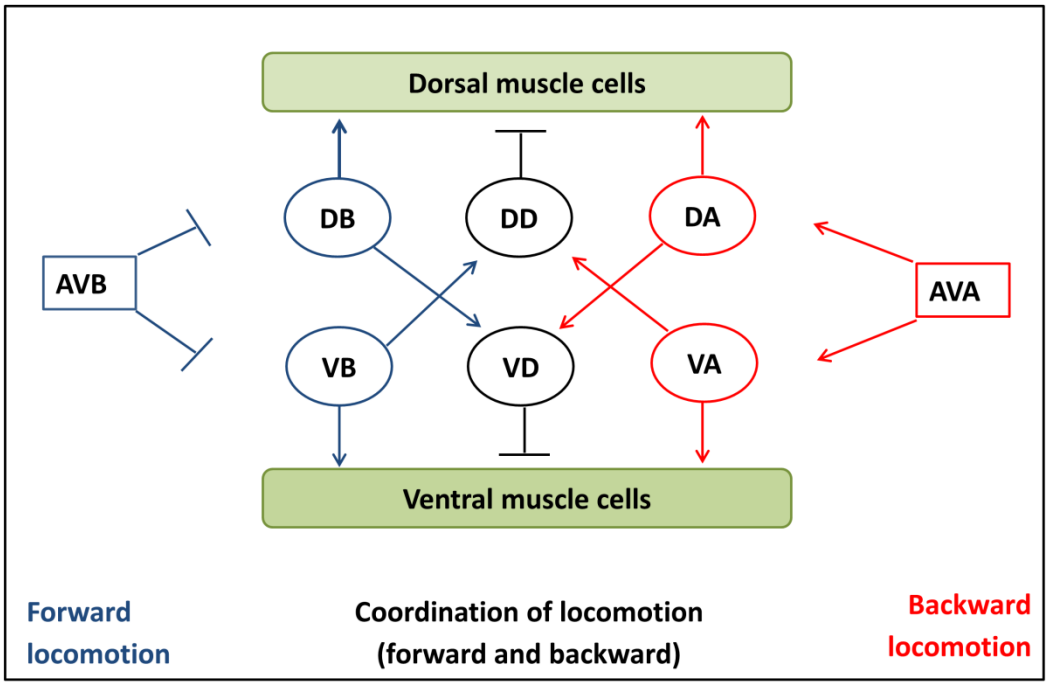


Figura 9 Connectivity between command interneurons (rectangles) and motor neurons (circles). Excitatory connections are shown by arrows, whereas inhibitory ones by bars. All cell bodies lie on the ventral side of the worm. The DB, DD and DA neurons send neuronal processes to the dorsal side where they innervate dorsal body-wall muscle cells. The VB, VD and VA neurons innervate ventral body-wall muscle cells. Figure adapted from Chalfie *et al.*, 1985.

3.4 Deciphering genetic pathways in *C.elegans*

Biological processes are exactly regulated by highly-coordinated gene regulatory networks that are composed of numerous interacting signaling pathways. The basic constituent of signaling pathways are proteins encoded by corresponding genes. Disruptions of these genes can deregulate associated pathways or even the entire network, resulting in an observable outcome referred to as a phenotype. Forward genetic approaches (investigation directed from phenotype to genotype) and reverse genetic approaches (investigation directed from genotype to phenotype) are two powerful ways of elucidating the function of genes that regulate a biological process of interest. Forward genetic screens identify genes in an unbiased manner based on phenotypes of mutants. The screens start by searching for a desired phenotype caused by a mutation that is introduced through a variety of mutagens (Kutscher and Shaham, 2014), including ethyl methanesulfonate (also known as or 1-methylsulfonyloxyethane or EMS) and N-ethyl-N-nitrosourea (also known as 1-Ethyl-1-nitrosourea or ENU). Both

are highly potent alkylating agents that cause preferencially GC-to-AT or AT-to-GC transitions, as well as small deletions all over the genome (Bautz and Freese, 1960; Flibotte *et al*., 2010). Once mutant strains have been obtained, identity of the mutation-harboring genes can be determined by positional cloning or candidate-gene testing. Forward genetic studies in *C. elegans* have made significant contributions to the understanding of a wide range of developmental processes. For example, the forward genetics screens pioneered by Nobel Prize Laureate Robert Horvitz for mutants defective in programmed cell death (PCD) identified the underlying genetic pathways that direct apoptosis, a process conserved among metazoans, including humans (Metzstein *et al*., 1998). Similarly, reverse genetics has provided considerable insights into many biological processes. Reverse genetic approaches begin with a set of genes with known sequences that are of particular interest, such as disease-related genes (Barr *et al*., 2001; Derry *et al*., 2001). Genes are inactivated by target-selected approaches, such as creation of deletion mutants using chemical mutagens or UV light (Jansen *et al*., 1997; Liu *et al*., 1999; Gengyo-Ando and Mitani, 2000), transposon insertion (Rushforth *et al*., 1993), and RNA interference (Fire *et al*., 1998). The availability of the complete and well-annotated genome sequence and RNAi libraries covering over 94% of the predicted genes in *C. elegans* (Ahringer, 2006), allow investigation of nearly any gene in *C. elegans*. For example, Derry *et al*. (2001) used sequence analysis to identify *cep-1*, the C. elegans ortholog of the mammalian p53 tumor suppressor gene (Rubin *et al*., 2000). By generating a *cep-1* deletion mutant and using RNAi for functional assays, they identified and characterized the roles of CEP-1 in regulating apoptosis, stress response in somatic cells, and chromosome segregation in the germ line.

Forward and reverse genetics can often complement each other. A notable example comes from genetic studies involving the *C. elegans* orthologs of two mammalian polycystin proteins *PKD1* and *PKD2,* which are defective in human autosomal dominant polycystic kidney disease. In a forward genetic screen for male mutants defective in the ability to locate the hermaphrodite vulva, Barr and Sternberg (1999) isolated a mutation in *lov-1,* ortholog of the human disease gene *PKD1,* which encodes a large transmembrane receptor-like protein (Harris and Torres, 2009). They found that *lov-1* is exclusively expressed in male-specific sensory neurons and that loss of *lov-1* function

displayed no phenotype in hermaphrodites. Male-specific phenotypes are not often examined, which is likely the reason this gene was not previously identified. Based on that knowledge, they employed a reverse genetic strategy and generated a deletion mutant of *pkd-2*, the *C. elegans* ortholog of the human *PDK2*, which encodes a transient receptor potential channel (Clapham, 2003). Both *lov-1* and *pkd-2* localize to the ciliated endings of male-specific sensory neurons, the site of sensory mechanic transduction, but they are not required for ciliogenesis. This was important, as this study was the first to indicate that these genes may function in sensory transduction in cilia, a location where mammalian *PKD1* and *PKD2* genes were later found to reside and mediate mechanosensation (Yoder *et al*., 2002; Nauli *et al*., 2003).

Treatment of organisms with a chemical mutagen induces nucleotide changes throughout the genome. Following mutagenesis, backcrossing or outcrossing of the mutagenized organism to unmutagenized counterparts is performed to eliminate mutagen-induced mutations (Zuryn *et al*., 2010; Zuryn and Jarriault, 2013) (**Figure 10**). During backcrossing, it is necessary to genotype many animals to determine whether they carry the deletion of interest and whether they are homozygous or heterozygous for the deletion. Alternatively, individuals carrying the phenotype-causing mutation are recognized and retained phenotypically, by selecting only backcrossed individuals showing the phenotype of interest. As a result of several backcrossing steps, mutagen-induced nucleotide changes that are genetically linked to the causal mutation and physically surround it on the chromosome will remain, in contrast to unlinked nucleotide changes. It is important to note that a backcrossing or outcrossing step is necessary for the analysis of mutants obtained from all mutagenesis screens, irrespective of the type of mutant identification strategy used or the type of mutagen or organism used.

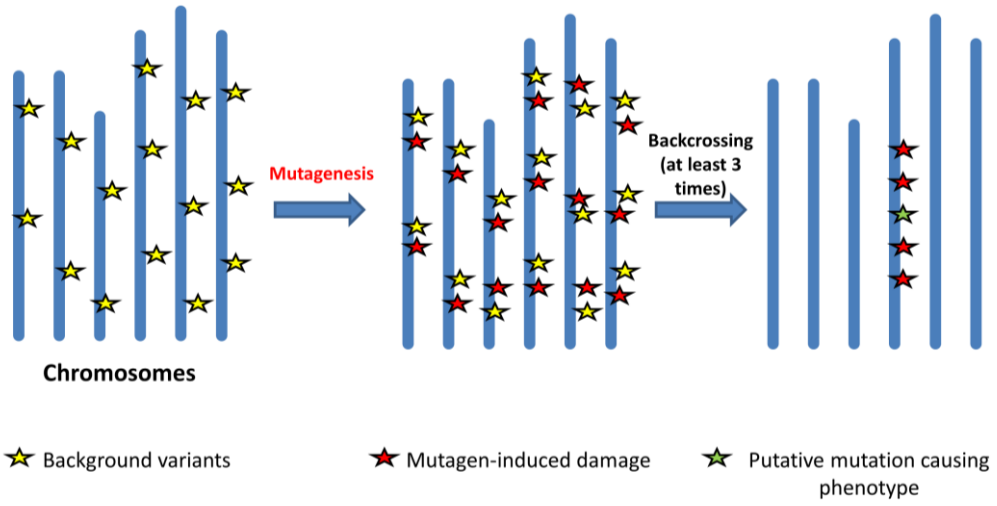


Figure 10 Backcrossing process. Mutagen treatment induces point mutations throughout the genome (red asterisks). Backcrossing to the original unmutated parent strain removes much of the mutagen-induced nucleotide changes except for the putative causal mutation (green asterisk) and those genetically linked to it. Figure adapted from Zuryn *et al.*, 2010.

AIM OF THE STUDY

STUDY I

HYPOTHESIS

Genetic alterations have already been recognized as major etiological factors for ASD and ID, of which RTT is an example. In addition to the contribution of polymorphic variants that confer low or moderate risk of appearance of these neurodevelopmental defects, *de novo* mutations affecting genes in a number of cellular pathways have been reported to be a cause of ASD, ID and associated NDDs (Vissers *et al*., 2010; Ronemus *et al*., 2014; Gilissen *et al*., 2014*;* Deciphering Developmental Disorders Study, 2015). Although the strong correlation between NDDs and genetic factors has been long established, the exact genetic background of ASD and ID remains unclear because of the strong heterogeneity of these disabilities. Genetic research has focused on the use of unbiased genome-wide approaches, including genomic microarrays and, more recently, NGS technology with the use of extensive gene panels, the exome or the whole genome. Consequently, new ASD and ID genes are now being identified in rapid succession. Among these new candidate genes, supportive role for the Jumonji Domain Containing 1C (*JMJD1C*) histone demethylase in ASD and ID has been fostered by both positional cloning strategies (Castermans *et al*., 2007) and exome-sequencing studies (Neale *et al*., 2012; Iossifov *et al*., 2014). Thus, based on the strong genetic architecture underlying NDDs and the emerging evidences for a role of *JMJD1C* as potential candidate genes, in the present PhD thesis we proposed to study the involvement of *JMJD1C* alterations in ASD, ID and RTT and their effects on the protein function.

OBJECTIVES

The specific objectives of study I are as follows:

1. To investigate the occurrence of *JMJD1C* mutations in the mentioned disabilities, a comprehensive mutational analysis was performed in samples

from ASD, ID and RTT, searching for SNVs and indels, as well as larger genetic defects.

2. To address the functional consequences of the identified *de novo* JMJD1C-Pro163Leu mutation, the intracellular localization of the wildtype and mutated JMJD1C protein, as well as the efficiency in demethylating a non-histone target of JMJD1C, MDC1 (mediator of DNA-damage checkpoint 1), was studied in immunofluorescence, fractionation and immunoprecipitation experiments in HEK293 cells.
3. To direct a molecular explanation for the implication of JMJD1C in Rett Syndrome, the interaction between MeCP2 and JMJD1C, in the wildtype and mutated form, was assessed in an immunoprecipitation assay.
4. To study the cellular effects of the disruption of wildtype JMJD1C on a neuronal system, we analyzed the existence of changes in dendritic branching of primary neuron cultures from neonatal mouse hippocampus.

STUDY II

HYPOTHESIS

Mutations in *MECP2* cause most of the classical or typical forms of RTT (Chahrour and Zoghbi, 2007). Approximately 8 % of classic RTT and 42 % of variant RTT patients are *MECP2* mutation-negative (Monrós *et al*. 2001; Percy, 2008). Some of the latter group have mutations in other genes, such as that of *CDKL5*, which is described in individuals with an early seizure onset variant of RTT (Kalscheuer *et al*., 2003) or *FOXG1*, which is responsible for the congenital variant of RTT (Ariani *et al*., 2008). However, there remains a subset of patients with a clinical diagnosis of RTT who are mutation-negative for all the aforementioned genes. In the present PhD thesis, we proposed to identify new candidate genes that could explain the RTT-like phenotype of several clinical cases without mutations in *MECP2*, *CDKL5* and *FOXG1*, using NGS, with the purpose of expanding the knowledge of the impaired biological pathways in RTT.

OBJECTIVES

The specific objectives of study II are as follows:

1. To identify previously undescribed variants potentially implicated in RTT-like phenotype, WES was performed on a cohort of 19 Spanish parent–child trios and one family with two affected daughters presenting features associated with RTT. A bioinformatics process of WES data was realized to filter and select putative pathogenic *de novo* variants absent or present with very low frequency in the control population, and putatively dangerous for protein function.
2. To realize a differential diagnosis among diverse neurodevelopmental disorders with overlapping phenotype, a gene-association analysis was carried out to define a list of variants previously associated with neurodevelopmental disorders and another one with undescribed new variants not previously associated with.
3. To demonstrate a neurological implication for a loss of function of detected candidate genes not previously associated with neurodevelopmental disorders, the model organism *C. elegans* was used to confirm a genotype-phenotype correlation by performing locomotion assays in worm mutants that carry deleterious mutations in the orthologous genes to those human genes with potentially pathogenic mutations in the patients.

RESULTS

DIRECTORS REPORT

To who may concern, we authenticate that the PhD student Mario Lucariello will present his PhD thesis by scientific publications. His contribution for each publication will be next pointed out:

STUDY I

"Mutations in JMJD1C are involved in Rett syndrome and intellectual disability"

Sáez MA, Fernández-Rodríguez J, Moutinho C, Sanchez-Mut JV, Gomez A, Vidal E, Petazzi P, Szczesna K, Lopez-Serra P, **Lucariello M**, Lorden P, Delgado-Morales R, de la Caridad OJ, Huertas D, Gelpí JL, Orozco M, López-Doriga A, Milà M, Perez-Jurado LA, Pineda M, Armstrong J, Lázaro C, Esteller M.

Contribution: In this article, Mario Lucariello collaborated for the genetic screening of 215 patients and 500 healthy volunteers to detect different mutation classes including single-nucleotide changes and indels, using Sanger sequencing. He also participated in manuscript revision.

Journal: Genet Med. 2016 Apr;18(4):378-85. doi: 10.1038/gim.2015.100. Epub 2015 Jul 16. PMID: 26181491.

Impact factor: 7.71

STUDY II

"Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype"

Lucariello M, Vidal E, Vidal S, Saez M, Roa L, Huertas D, Pineda M, Dalfó E, Dopazo J, Jurado P, Armstrong J, Esteller M.

Contribution: In this article, Mario Lucariello was the responsible of experimental design, supervised by Dr. Manel Esteller. In addition, he performed all the experiments, and participated in data analysis and interpretation. The manuscript was written by Mario Lucariello and revised by Dr. Manel Esteller.

Journal: Hum Genet. 2016 Dec;135(12):1343-1354. Epub 2016 Aug 19. doi: 10.1007/s00439-016-1721-3 PMID: 27541642. PMCID: PMC5065581.

Impact factor: 5.138

Dr. Manel Esteller Badosa, M.D, Ph.D.

Cancer Epigentics Group, Leader

Cancer Epigenetics and Biology Program (PEBC),
Bellvitge Biomedical Research Institute (IDIBELL)

Avda. Gran Via 199-203
08908 L'Hospitalet de Llobregat, Barcelona
+34 932607500 ext.7253

mesteller@idibell.cat

STUDY I

"Mutations in JMJD1C are involved in Rett syndrome and intellectual disability"

Mauricio A. Sáez[1], Juana Fernández-Rodríguez[2], Catia Moutinho[1], Jose V. Sanchez-Mut[1], Antonio Gomez[1], Enrique Vidal[1], Paolo Petazzi[1], Karolina Szczesna[1], Paula Lopez-Serra[1], Mario Lucariello[1], Patricia Lorden[1], Raul Delgado-Morales[1], Olga J. de la Caridad[1], Dori Huertas[1], Josep L. Gelpí[2,3], Modesto Orozco[2–4], Adriana López-Doriga[5], Montserrat Milà[6,7], Luís A. Perez-Jurado[7,8], Mercedes Pineda[7,9], Judith Armstrong[7,9], Conxi Lázaro[5], and Manel Esteller[1,4,10]

[1]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Catalonia, Spain; [2]Joint Biomedical Research Institute-Barcelona Supercomputing Center (IRB-BSC) Computational Biology Program, Barcelona, Catalonia, Spain; [3]Department of Biochemistry and Molecular Biology, University of Barcelona, Barcelona, Catalonia, Spain; [4]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain; [5]Hereditary Cancer Program, Catalan Institute of Oncology-Bellvitge Institute for Biomedical Research (ICO-IDIBELL), Barcelona, Catalonia, Spain; [6]Biochemistry and Molecular Genetics Department, Hospital Clínic, Barcelona, Catalonia, Spain; [7]CIBERER (Biomedical Network Research Centre on Rare Diseases, Instituto de Salud Carlos III), Barcelona, Spain; [8]Genetics Unit, University Pompeu Fabra, Barcelona, Catalonia, Spain; [9]Department of Neurology, Hospital Sant Joan de Déu (HSJD), Barcelona, Catalonia, Spain; [10]Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain. Correspondence: Manel Esteller (mesteller@idibell.cat).

Submitted 5 November 2014; accepted 9 June 2015; advance online publication 16 July 2015. doi:10.1038/gim.2015.100

For the sake of clarity and higher figure resolution, I next present the published article in Word format. The published PDF file and Supporting Information can be found in **Annex I**.

ABSTRACT

Purpose: Autism spectrum disorders are associated with defects in social response and communication that often occur in the context of intellectual disability. Rett syndrome is one example in which epilepsy, motor impairment, and motor disturbance may co-occur. Mutations in histone demethylases are known to occur in several of these syndromes. Herein, we aimed to identify whether mutations in the candidate histone demethylase *JMJD1C* (jumonji domain containing 1C) are implicated in these disorders.

Methods: We performed the mutational and functional analysis of JMJD1C in 215 cases of autism spectrum disorders, intellectual disability, and Rett syndrome without a known genetic defect.

Results: We found seven JMJD1C variants that were not present in any control sample (~ 6,000) and caused an amino acid change involving a different functional group. From these, two *de novo* JMJD1C germline mutations were identified in a case of Rett syndrome and in a patient with intellectual disability. The functional study of the JMJD1C mutant Rett syndrome patient demonstrated that the altered protein had abnormal subcellular localization, diminished activity to demethylate the DNA damage-response protein MDC1, and reduced binding to MECP2. We confirmed that JMJD1C protein is widely expressed in brain regions and that its depletion compromises dendritic activity.

Conclusions: Our findings indicate that mutations in *JMJD1C* contribute to the development of Rett syndrome and intellectual disability.

Genet Med advance online publication 16 July 2015

Keywords: autism, intellectual disability, mutational screening, Rett syndrome.

INTRODUCTION

Autism spectrum disorders are a heterogeneous clinical and genetic group of neurodevelopmental defects that are characterized by impaired social communication functions and inappropriate repetitive behavior.[1] This family of disorders is characterized by enormous phenotypic variability, from mild primary deficits in language pragmatics[2] to major neurological phenotypes, such as that of Rett syndrome (OMIM 312750), where it co-occurs with epilepsy, motor impairment, and sleep disturbance.[3] The disabilities associated with autism spectrum disorders are often so severe that affected individuals do not generally reach parenthood, thereby preventing comprehensive familial genetic studies from being undertaken. However, genetic alterations are already recognized as major etiological factors. In this regard, concordance with autism spectrum disorders is higher than with any other cognitive or behavioral disorder.[4,5] In addition to the contribution of polymorphic variants that confer low or moderate risk of the appearance of these neurodevelopmental defects, a cause of autism spectrum disorders can be the occurrence of *de novo* mutations affecting genes in a number of cellular pathways.[6–8] A similar scenario can be proposed for the genetic contribution to the even more heterogeneous group of disorders classified as intellectual disabilities.[9] Among the described genetic defects associated with intellectual disabilities, our attention was caught by a single case report of an autistic patient carrying a *de novo* balanced paracentric inversion 46, XY in (10)(q11.1;q21.3) in which the distal breakpoint disrupted what was at that time known as the TRIP8 gene,[10] which has been characterized as a member of the jmJC domain-containing protein family involved in the methyl group removal and renamed as JMJD1C (jumonji domain containing 1C).[11–13] Support for a role of JMJD1C in autism has been further fostered by large-scale exome sequencing studies in which three *de novo* variants have been identified,[14,15] at least one of which is a loss-of-function mutation.[15] However, two of the three described variants occurred in unaffected siblings. To investigate whether JMJD1C mutations occur in intellectual disability, Rett syndrome, and autism spectrum disorders, we performed a comprehensive screening to identify single nucleotide changes and indels in a large collection of 215 patients in whom a genetic defect had not previously been identified.

MATERIALS AND METHODS

Patients

The samples used in this study consisted of 215 patients of either gender with autism, Rett syndrome, or intellectual disability from the Hospital Clínic, Sant Joan de Deu and Pompeu Fabra University, Barcelona, Catalonia, Spain. A cohort of 500 healthy volunteers was obtained from the same geographic region (Catalonia, Spain). All were Caucasian individuals (such as in the case cohort) and were matched according to gender distribution (case cohort: female, 53% [114/215]; male, 47% [101/215]; control cohort: female, 58% [290/500]; male, 42% [210/500]). Ethical approval for the molecular genetic studies was obtained from each institutional review board. DNA was extracted from peripheral blood leukocytes using standard techniques. We measured DNA concentration with the Quant-iT Picogreen (Invitrogen, Life Technologies – Grand Island, NY) and then normalized all concentrations to 25–50 ng/µl before proceeding with the Access Array amplification.

Fluidigm access array

Forty-eight pairs of primers were designed using the Access Array Amplicon Tagging Assay design service from Fluidigm to cover all 26 exons of the JMJD1C gene (NM_032776.1), including exon-intron boundaries. These primers also contained the sequencing adaptors necessary for subsequent sequencing in the 454 GS Junior Sequencer.

Multiplex ligation-dependent probe amplification

Large rearrangements in the JMJD1C gene were studied using multiplex ligation-dependent probe amplification. We designed nine such probes specific to the JMJD1C gene and six control probes according to the instructions provided by MRC-Holland.

Sanger sequencing

The variants were validated by Sanger sequencing using a BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies, Grand Island, NY) in an Applied Biosystems

3730/DNA Analyzer (Applied Biosystems – Life Technologies – Grand Island, NY). The raw data were analyzed with Codon Code Aligner Software (CodonCode Corporation – Centerville, MA).

Exome sequencing

The patient and healthy parents were analyzed by whole exome sequencing with TruSeq Sample Preparation Kit (Illumina, San Diego, CA). Exomes were captured with TruSeq Exome Enrichment Kit (Illumina) and paired-end 100 × 2 sequenced with the HiScan SQ (Illumina). Format DNA and Protein Sequence Alignment Qual (FASTQ) files were analyzed as described in Supplementary Methods online. The raw data were analyzed in Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Catalonia, Spain.

Cell culture and vectors

JMJD1C coding sequence in the pCMV6-AC-GFP vector was purchased from Origene (Rockville, MD) (RG214878). The mutants were generated with Mutant QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA). Wild-type (WT) WT and Pro163Leu and His2336Ala mutants were subcloned in the pCMV6-entry vector to introduce Myc-DDK-tag. shRNAs against the coding sequence of the mouse Jmjd1c gene were cloned in the pLVX-shRNA2 vector between the BamHI and EcoRI restriction sites (shRNA24 target: CAGAGACTGCTTGAGGAAT). Hek293 cells were cultivated in Dulbecco's modified Eagles medium with 10% fetal bovine serum. To generate stable WT or mutant clones, Hek293 cells were transfected with Lipofectamine 2000 (Invitrogen), selected with G418 antibiotic, and individual clones were isolated 2 weeks later. For transient expression, 6 mg of vector were transfected in 35-mm six-well plates with jetPRIME transfection reagent (Polyplus, NY) following the manufacturer's instructions. Primary cultures of hippocampal neurons were prepared from neonate mice (P0) of either sex. Cultures were infected at 3DIV with lentiviral vectors to express scramble or shRNAs against JMJD1C together with a GFP tracer (pLVX-shRNA2 system). Coverslips were fixed, and protein was extracted at 15DIV.

Immunoprecipitation

Seven hundred fifty micrograms of chromatin fraction were diluted 10-fold in immunoprecipitation (IP) buffer (5 mmol/l Tris–HCl pH 7.6, 15 mmol/l HEPES pH 8.0, 1 mmol/l Ethylene diamine tetraacetic acid, 1 mmol/l ethylene glycol tetraacetic acid, 0.1% sodium dodecyl sulfate (SDS), 1% Triton X-100) incubated with 2 µg of antibodies anti-Me-Lysine (Abcam ab23366, Cambridge, UK) overnight at 4 °C and for 2 hours with PureProteome Protein A/G Magnetic Beads. Beads were washed twice with IP buffer and twice in RBS NP-40 and eluted in laemli buffer in reduction conditions at 70 °C for 10 minutes.

MeCP2 immunoprecipitation

For the MeCP2 immunoprecipitation procedure, anti-JMJD1C and anti-MeCP2 antibodies were coupled to Dynabeads Protein G (Invitrogen). JMJD1C transfected HEK293F cells were transiently transfected with MeCP2-Flag tagged plasmid and the nuclear fraction was obtained by RIPA buffer (10 mmol/l TRIS-Cl pH 8.0, 1 mmol/l EDTA (Ethylene diamine tetraacetic acid), 0.5 mmol/l EGTA (ethylene glycol tetraacetic acid), 1% Triton X-100, 0.1 % sodium deoxycholate, 0.1% SDS, and 140 mmol/l NaCl) supplemented with protease inhibitors (complete; Roche, Rotkreuz, Switzerland) and hybridized with the antibodies at 4 °C for 2 hours. Then, 150 mM NaCl phosphate-buffered saline (PBS) buffer was used for washing. Human IgG was used as negative control. Anti-Flag HRP (M2-SIGMA) antibody was used to visualize binding.

Western blot

Protein extract of Hek293 cells and primary neuronal culture were obtained in radioimmunoprecipitation assay (RIPA) buffer supplemented with a complete protease inhibitor cocktail tablet (Roche) and sonicated. Then, 50 µg of each protein sample were denatured in Laemli buffer 4% β-mercaptoethanol for 10 minutes at 95 °C, separated on a 7.5% or 15% SDS-polyacrylamide gel, and then transferred onto a polyvinylidene fluoride membrane (Immobilon-P; Millipore, Hessen, Germany) by liquid electroblotting. The antibodies and dilutions used were as follows: rabbit anti-JMJD1C 1:2,000 (09-817; Millipore); mouse anti-nucleolin 1:1,000 (SC-8031; Santa Cruz, Dallas, Texas); rabbit anti-H3 1:10,000 (Abcam AB1791); mouse anti-H3 1:4,000

(AbcamAB10799); and rabbit anti-MDC1 1:5,000 (Abcam AB11171). Theblots were developed with Luminata Crescendo Western HRP substrate (MERCK MILLIPORE - Hessen, Germany) or with the LiCor Odyssey System (LI-COR – Bad Homburg, Germany).

Immunofluorescence

Cells were fixed in 4% paraformaldehyde–phosphate-buffered serum (PFA-PBS), quenched in 100 mmol/l glycine-PBS, and permeabilized with 0.25% Triton X-100, 1% bovine serum albumin, and PBS. The cells were blocked with 0.2% gelatin and 0.25% Triton X-100. Antibody dilutions were prepared in 0.25% Triton X-100, 1% bovine serum albumin, and PBS. The dilutions used were: rabbit anti-JMJD1C 1:200; chicken anti-Map2 1:5,000; and anti-β-tubulin 1:1,500 (Abcam AB21058). Nuclei were stained with 2 mg/ml Hoechst 33342. Coverslips were mounted in ProLong Gold antifade reagent (Molecular Probes– Life Technologies – Grand Island, NY). Confocal images were captured with a Leica SP5 confocal microscope (Leica, Buffalo Grove, IL). For fluorescence recovery after photobleaching analysis, the cells were maintained at 37 °C in an atmosphere of 5% CO2. We captured images every 70 seconds at 63×, at 128 × 128 resolution, and at 1,400 Hz with bidirectional acquisition. We captured 25 control images at 3% laser transmission before bleaching and then bleached the region of interest (ROI) inside a nucleus 25 times at a nominal level of 100% laser transmission. For this experiment, 150 images were captured after bleaching. The row data were analyzed with FrapAnalyzer software (Igor Pro 6.1 Software – WaveMetrics – Lake Oswego, OR).

RNA extraction and real-time PCR

Total RNA was extracted from cell lysates using TRIzol Reagent (Invitrogen), purified using the RNeasy Kit (Qiagen, Valencia, CA), and 2 µg were retrotranscribed using the ThermoScriptTM RT-PCR System (Invitrogen). Real-time PCR reactions were performed in triplicate on an Applied Biosystems 7,900HT Fast Real-Time PCR system using 20 ng cDNA, 5 µl SYBR Green PCR Master Mix (Applied Biosystem), and 150 nmol/l specific primers (sequences are available on request) in a final volume of 10 µl.

RESULTS

JMJD1C mutational analyses in samples from autism spectrum disorders, intellectual disability, and Rett syndrome

We first collected blood samples in EDTA tubes for DNA extraction from 215 patients with autism spectrum disorders (n = 69; 58 males and 11 females), intellectual disability (IQ < 70) (n = 85; 43 males and 42 females), or Rett syndrome (n = 61; all females) without mutations in the disease-associated genes MECP2, CDKL5, and FOXG1.[16] Approval for this was obtained from the corresponding institutional review boards. The presence of single-nucleotide changes and indels in the JMJD1C gene was analyzed by sequencing using a GSJUNIOR system with an amplicon library prepared with the Fluidigm Access Array, and the presence of larger genetic defects was assessed by multiplex ligation-dependent probe amplification. We detected no major genomic defects at the JMJD1C gene locus using the multiplex ligation-dependent probe amplification approach. Identified synonymous variants for JMJD1C are shown in **Supplementary Table S1** online and seven previously informed JMJD1C nucleotide variants are described in **Supplementary Table S2** online. Most importantly, the sequencing strategy identified seven nucleotide changes in the exonic regions of JMJD1C that were not present in the NHLBI Exome Sequencing Project (ESP) Exome Variant Server (~6,000 control samples) (**Supplementary Table S3** online). These seven nucleotide changes were also absent in the 1000 Genomes Project[17] and the NCBI database of genetic variation.[18] Furthermore, we sequenced 500 healthy volunteers for these seven missense variants and none of the studied samples showed the described changes (**Supplementary Table S3** online and **Supplementary Figure S1** online). Not finding prior documentation of a variant in a mutation database is not in itself evidence of pathogenicity, and most missense variants are rare due to factors such as rapid population growth and weak purifying selection.[19] Interestingly, four of our seven (57%) missense mutations were clustered in the exon 10 of JMJD1C, where the only two previously reported JMJD1C missense mutations were also found[14,15] (**Figure 1a**). In addition, the seven single-nucleotide shifts caused an amino acid change involving a different functional group (**Supplementary Table S3** online) that could alter

important domains of the JMJD1C protein, such as the nuclear localization signal and the JmjC hydroxylase domain (**Figure 1a**). Unfortunately, for the understanding of the functional consequences of the JMJD1C mutations undergoing study, the 3D structure of this protein is only available for the C-terminal end (amino acid positions from 2,157 to 2,497) corresponding to the JmJC domain.[20] From the described mutations only c.6997A>G (T/A 2333) (**Supplementary Table S3** online) is included in the available structure, and the identified change affects a critical amino acid in one of the beta strands that confers the core beta barrel for histone substrate interaction (**Figure 1b**). Conventional Sanger sequencing confirmed the GSJUNIOR sequencing results (**Figure 1c**). The seven point mutations described occurred in one case of Rett syndrome that was not associated with MECP2, CDKL5, and FOXG1 defects, in three autistic patients, and in three patients with intellectual disability (**Supplementary Table S3** online). Our experimental, genetic, and clinical experiences with Rett syndrome enabled us to select the newly identified JMJD1C nucleotide change in this severe form of autism spectrum disorder and further characterize its functional relevance. The c.488C>T nucleotide change in exon 4 of the JMJD1C gene causes a proline-to-leucine shift in codon 163 of the protein. The potential pathogenic involvement of the JMJD1CPro163Leu change is also suggested because the wild-type proline amino acid is highly conserved in all mammalian JMJD1Cs (**Figure 2a**), and because it is included in the balanced JMJD1C inversion that occurs in the aforementioned autistic patient.[10]

Our patient with the JMJD1C-Pro163Leu mutant is a 29-yearold female who was born at term, weighed 3,800 g, and presented a 9/10 APGAR value. She showed stagnation in head growth and also in normal development at 14 months old and lost social interaction at 18 months old. The patient started to use single words at 20 months old and propositive sentences at 24 months old, but only babbled at 36 month old. The patient also presented gait dyspraxia, hand-washing stereotype, learning impairment, teeth grinding, air swallowing, kyphoscoliosis, and tonic epilepsy. She was diagnosed as having classical Rett syndrome from the clinical standpoint, but without mutations in the known MECP2, CDKL5, and FOXG1 Rett-associated genes. Sanger sequencing of the progenitors showed that neither of the parents was a carrier of the described JMJD1C nucleotide change (Figure 2b). Thus, the described mutation can be considered a *de novo* germline event. In this regard, from the other five identified JMJD1C missense

changes that were not present in the NHLBI Exome Variant Server, we identified by conventional Sanger sequencing an additional germline *de novo* mutation. The c.3559A>G nucleotide change (T/A amino acid change) occurred in a patient with intellectual disability (case 4 in **Supplementary Table S3** online) but it was not present in the parents (**Supplementary Figure S2** online). In a recently developed statistical framework to distinguish disease-relevant mutations from background variation due to random expectation,[21] JMJD1C (NM_032776) was reported as having a mutation probability of 9.0804e-05.21 Thus, out of 215 samples we would expect to find 0.0195 mutated, so discovering 2 *de novo* mutations out of 215 samples is higher than expected by chance (102.4-fold, binomial test, P = 1.9e-4). One possible way to explain the higher rate of observed JMJD1C mutations than that in the described study[21] could be related to the fact that the previously analyzed 5,000 samples corresponded to autism spectrum disorders,[21] whereas our work jointly profiled autism with Rett syndrome and intellectual disability. Most importantly, our two *de novo* mutations were only identified in these last two entities. Thus, the identified JMJD1C mutation rate could match the one reported[21] because we did not find any *de novo* mutation in our 69 cases of autism spectrum disorders. Interestingly, when we applied the PMUT predictor for these two *de novo* JMJD1C mutations,[22] a neural network trained with neutral and pathogenic mutations extracted from the protein database SwissVar, we found that both of them were considered potentially pathogenic. To discard any other pathogenic mutation in our Rett syndrome patient, the patient and the healthy parents were analyzed by whole exome sequencing as described in Methods. The only *de novo* mutation identified in the patient was the described c.488C>T nucleotide change in exon 4 of the JMJD1C. The complete exome sequencing data of these three samples are available at the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the codes SRX667201 (patient), SRX667384 (mother), and SRX667386 (father).

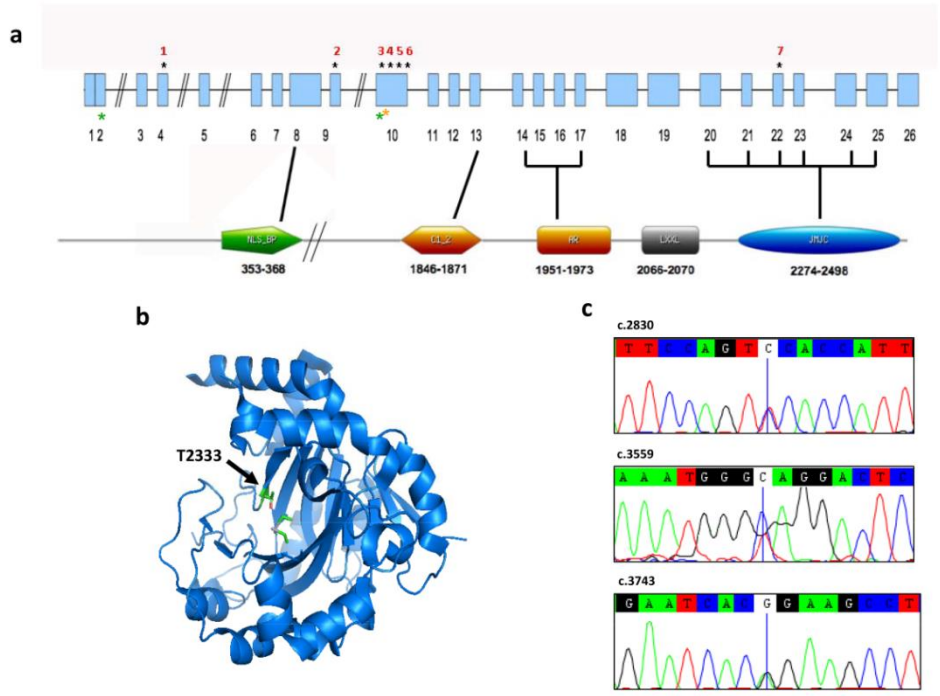


Fig. 1 Diagram of JMJD1C and mutations found. (a) Black asterisks indicate the position of mutations identified in the JMJD1C gene. The three previous mutations identified by Iossifov *et al*. and Neal *et al*. are indicated by green and orange asterisks, respectively. The most important motif and domains are: NLS_Bp, bipartite nuclear localization signal; C1_2, phorbol esther/diacylglycerol-binding domain; AR, androgen receptor-interacting zone; LxxL, motif involved in transcriptional regulation; JMJC, JmjC hydroxylase domain. **(b)** JMJD1C 3D structure for the C-terminal end corresponding to the JmjC domain derived from the Protein Data Bank (PDB) code 2YPD, X-Ray diffraction data for 2.1 Angstroms resolution. The T2333 amino acid is indicated by a black arrow. **(c)** Chromatograms of Sanger sequencing showing c.2830C>T, c.3559A>G, and c.3743A<G mutations.

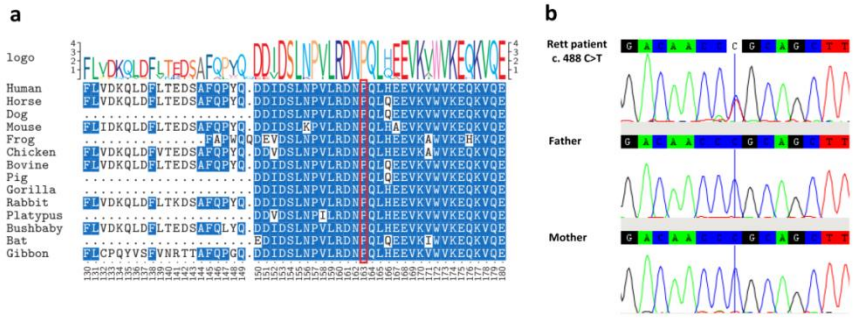


Fig. 2 Characterization of the JMJD1C Pro163Leu mutation. (a) The Pro163Leu missense mutation is in a highly conserved region of JMJD1C. **(b)** Chromatogram of Sanger sequencing confirms c.488 A>T, *de novo* mutation, and its absence in the parents.

Functional analyses of a Rett syndrome–associated JMJD1C-Pro163Leu mutation

To address the effects of the identified JMJD1C-Pro163Leu mutation, we first studied the intracellular localization of the protein. Endogenous JMJD1C was mainly cytoplasmic in Hek293 cells (**Figure 3a**), a pattern that was also observed in the transfected JMJD1C-WT-GFP in stable clones (**Figure 3a**). However, clones expressing the mutant JMJD1C-Pro163Leu-GFP forms show a strong nuclear mark (**Figure 3a**). To characterize the subcellular scenarios in which the mutant JMJD1C protein was located in more detail, we fractionated cytoplasmic, nuclear, and chromatin-bound proteins. We found that if endogenous and transfected JMJD1C proteins were almost absent from the chromatin fraction, then the mutant JMJD1C-Pro163Leu-GFP was markedly enriched in the chromatin (**Figure 3b**). The use of the fluorescence recovery after photobleaching assay in nuclear areas confirmed that mutant JMJD1C-Pro163Leu-GFP had reduced recovery in comparison with the JMJD1C wild-type transfected protein (**Figure 3c**). Most importantly, we also found that the mutant JMJD1C-Pro163Leu protein was less efficient in demethylating a non-histone target of JMJD1C that has been recently identified, MDC1 (mediator of DNA-damage checkpoint 1), a regulator of the RAP80-BRCA1 branch of the DNA damage response pathway.[13] Using

immunoprecipitation with an antibody against methylated lysines followed by western blot with the MDC1 antibody, we observed that Hek293 cells overexpressing the wild-type JMJD1C protein efficiently demethylated the MDC1 protein but that the mutant JMJD1C-Pro163Leu-GFP form showed diminished activity (**Figure 3d**). Thus, the studied JMJD1C mutation might also impair the repair of DNA damage in the studied patient. However, JMJD1C could also exert additional functions. In this regard, using JMJD1C-transfected HEK293F cells that we have co-transfected with MECP2-Flag, we observed the interaction between JMJD1C and MECP2 by using the immunoprecipitation assay (**Figure 3e**). Importantly, the JMJD1C-Pro163Leu mutant protein cannot efficiently bind to MECP2 (**Figure 3e**). These findings could explain the role of JMJD1C in Rett syndrome, a disease mainly associated with *de novo* mutations in MECP2.[3]

Once we had shown the aberrant functions of the mutant JMJD1C protein in the above model, we wondered whether the disruption of the wild-type JMJD1C protein had any cellular effect on a neuronal system. We first confirmed that JMJD1C was expressed in different brain regions by examining the Allen Brain Atlas Database (http://www.brain-map.org/) (**Figure 4a**).[23] We also validated the widespread expression of JMJD1C throughout distinct regions of the mouse (**Figure 4b**) and human (**Figure 4c**) brain by quantitative RT-PCR. Then, we used the short hairpin RNA (shRNA) approach to analyze the impact of its loss in neurons. Lentiviral shRNA-mediated depletion of JMJD1C was performed in primary neuronal cultures obtained from neonatal mouse hippocampus and JMJD1C protein downregulation was observed using western blot (**Figure 5a**) and immunofluorescence (**Figure 5b**). We studied the existence of changes in the dendrites in these JMJD1C-impaired cells. A significant reduction in the complexity of the dendritic process is known to be a common hallmark in Rett syndrome.[24] In this regard, and using immunofluorescence for the dendrite marker Map2, we found that JMJD1C shRNA-depleted neurons showed diminished dendritic branching in comparison with shRNAscrambled control neurons (**Figure 5c, d**). Thus, a functional JMJD1C protein seems to be required for a correct dendritic pattern, and a defect in this gene could be associated with the dendritic impairment observed in autistic patients.

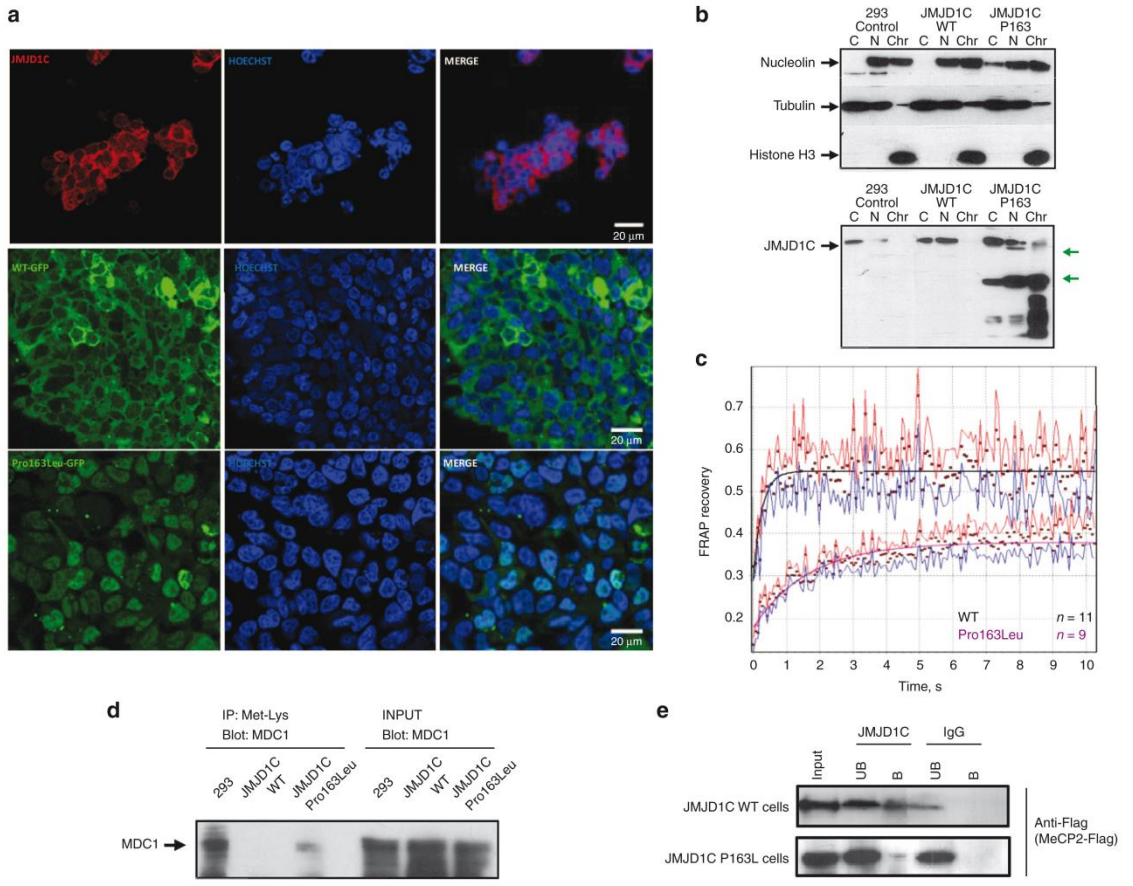


Figure 3 Differential subcellular localization, processing, and chromatin binding of JMJD1C-Pro163Leu. **(a)** Upper panels show immunofluorescence with anti-JMJD1C antibody in untransfected cells showing that the endogenous protein is localized predominantly in the cytoplasm; middle panels show clones expressing JMJD1C-WT-GFP, which is also localized in the cytoplasm; bottom panels show that clones expressing JMJD1C-P163L-GFP have a nuclear signal. **(b)** Clones and control cells were fractionated to separate the cytoplasm (C), nuclear (N), and chromatin (Chr) fractions. Proteins were processed for western blot and blotted with compartment markers (upper panel) and with anti-JMJD1C (lower panel). JMJD1C-WT-GFP looks like endogenous protein (black arrow), but JMJD1C-P163L-GFP shows several processed bands in nucleus and chromatin (green arrows). **(c)** Normalized fluorescence recovery after photobleaching assay (FRAP) in clones overexpressing JMJD1C-WT-GFP and JMJD1C-P163L-GFP. The bleached area was situated above the nuclei. The mutant P163L recovers more slowly than WT and has a larger immobile fraction. **(d)** Chromatin fractions from Hek293 control cells and Hek293 clones expressing JMJD1C-WT-GFP and JMJD1C-Pro163Leu-GFP were immunoprecipitated with anti-Methyl-lysine antibodies and blotted with anti-MDC1. **(e)** JMJD1C transfected HEK293F cells were co-transfected with MeCP2-Flag and subjected to immunoprecipitation assay with anti-JMJD1C antibody, anti-MeCP2 antibody, and control rabbit IgG. The western blot using anti-Flag shows the interaction between JMJD1C and MeCP2 when JMJD1C is immunoprecipitated (top). JMJD1C-Pro163Leu mutant protein cannot efficiently bind to MECP2 (below). B, bound fraction; UB, unbound fraction.

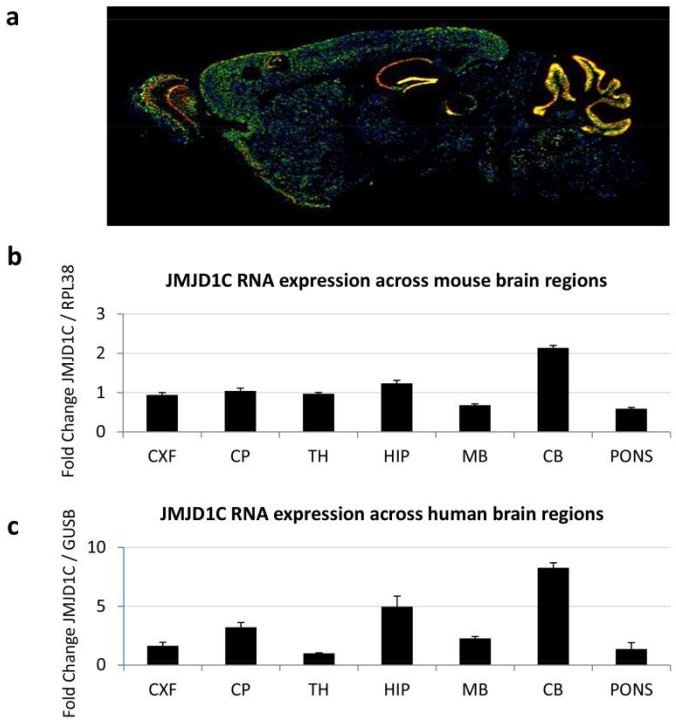


Fig. 4 JMJD1C expression in human and mouse brain regions. CXF refers to frontal cortex; CP, Caudatus-Putamen; TH, thalamus; HIP, hippocampus; MB, midbrain; CB, cerebellum. **(a)** ISH analysis performed using a 2-month-old male mouse sample extracted from the Allen Brain Atlas Database (©2014 Allen Institute for Brain Science. Allen Mouse Brain Atlas [Internet]. Available at http://mouse.brain-map.org/). **(b)** Quantitative RT-PCR using three 18-month-old female samples. The y axis shows the fold-change in JMJD1C RNA expression, normalized with respect to thalamus expression, in relation to the RPL38 housekeeping gene. Error bars represent the SEM across different samples. **(c)** Quantitative RT-PCR using a sample from a 64-year-old male. The y axis shows the fold-change in JMJD1C RNA expression, normalized with respect to thalamus expression, in relation to the GUSB housekeeping gene. Error bars represent the SEM across different experiments.

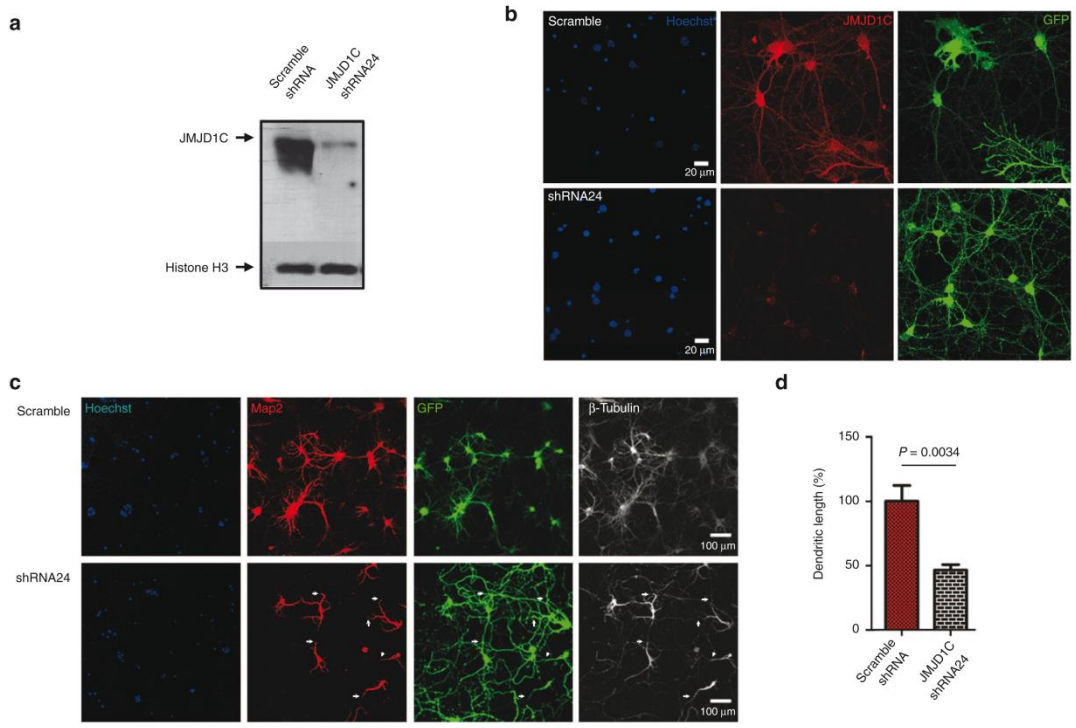


Figure 5 JMJD1C expression in primary neuron culture and shRNA knockdown. 3DIV neuron primary cultures were infected with lentiviral vectors expressing shRNAs against JMJD1C or scramble shRNA. Neurons were maintained in 14DIV culture. **(a)** Western blot shows the bands of JMJD1C and the effect of a scramble shRNA and shRNA against JMJD1C on the exon 24. Only the shRNA24 reduces the expression. Anti-H3 histone antibodies were used for normalization. **(b)** Immunofluorescence against JMJD1C. Neurons infected with shRNA24 reduce the expression of JMJD1C; infected neurons were identified with GFP tracer under CMV promoter in viral vector and nuclei stained with Hoechst3342. **(c)** Map2 immunostaining showing dendritic branching in primary cell culture neurons transfected with JMJD1C shRNA24 or scramble shRNA (SCR; Control). JMJD1C downregulation via shRNA24 induced a marked reduction of the Map2 signal, demonstrating a decrease in the length of neuronal dendrites (red). **(d)** The dendritic length (Map2 staining) in primary cell culture neurons transfected with JMJD1C shRNA24 or scramble shRNA was calculated using IMAGEJ software. Five fields per condition were analyzed and the results are plotted as a bar graph. Bar errors represent SEM across different fields and P-values are indicated. shRNA-mediated downregulation of JMJD1C significantly decreased dendritic length.

DISCUSSION

Our findings suggest that mutations in the candidate histone demethylase JMJD1C contribute to the development of intellectual disability, including well-defined clinical entities such as Rett syndrome in those cases in which the usual mutations in *MECP2*, *CDKL5*, and *FOXG1* are not present. Importantly, from a functional standpoint, a JMJD1C mutant protein is unable to correctly demethylate the MDC1 DNA repair-

response protein but the wild-type JMJD1C protein plays a key role in dendritic activity. It is interesting to note that many of the genes mutated in intellectual disability have an epigenetic component, [25, 26] and JMJD1C can now be included in this growing list. Epigenetics can be broadly defined as the inheritance of gene activity that does not depend on the "naked" DNA sequence, and it includes the control of gene transcription by several chemical marks, such as DNA methylation and histone modifications. Examples include the methyl-CpG-binding protein MeCP2 in Rett syndrome, the chromatin remodeling protein CHD7 in CHARGE syndrome, and the histone methyltransferase NSD1 in Sotos syndrome.[25, 26] The latter disorder is a representative case of how the importance of post-translational modifications of histone N-terminal tails in the genetic origin of neurodevelopmental disorders is becoming increasingly recognized. In this context, two mental retardation syndromes have been linked to mutations in the histone lysine demethylases KDM5C (previously known as JARID1C)[27] and PHF8 (also known as JHDM1F).[28] Thus, it is possible that mutations in many other members of the large histone demethylase family are involved in the development of intellectual disability.[29] In this regard, although further research is required to understand the molecular pathways that involve JMJD1C, our work highlights the increasing contribution of the genetic disruption of epigenetic genes to human neurodevelopmental disorders.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

ACKNOWLEDGMENTS

This study was supported by the European Community's Seventh Framework Program (FP7/2007–2013) under grant agreement PITN-GA-2012–316758 of the EPITRAIN project and PITN-GA-2009–238242 of DISCHROM; ERC grant agreement 268626 of the EPINORC project; the E-RARE EuroRETT network (Carlos III Health Institute project PI071327); the Fondation Lejeune (France); MINECO projects SAF2011-22803 and CSD2006-00049; the Cellex Foundation; the Botín Foundation; the Catalan Association for Rett Syndrome; Fundación Alicia Koplowitz 2011 Grant AKOPLOWITZ11_006; the

FIS project PI1002512; Grants PI10/01422, PI13/00285, CA10/01474, RD06/0020/1050, RD12/0036/008, and 2009-SGR293; and the Health and Science Departments of the Catalan government (Generalitat de Catalunya). K.S. and P.P. are EPITRAIN Research Fellows. M.E. is an ICREA Research Professor.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Publishing: Arlington, VA, 2013.

2. Anderson DK, Lord C, Risi S, *et al*. Patterns of growth in verbal abilities among children with autism spectrum disorder. *J Consult Clin Psychol* 2007;75:594–604.

3. Bedogni F, Rossi RL, Galli F, *et al*. Rett syndrome and the urge of novel approaches to study MeCP2 functions and mechanisms of action. *NeurosciBiobehav Rev* 2014;46:187–201.

4. Bailey A, Le Couteur A, Gottesman I, *et al*. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 1995;25:63–77.

5. Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Arch Pediatr Adolesc Med* 2009;163:907–914.

6. Zafeiriou DI, Ververi A, Dafoulis V, Kalyva E, Vargiami E. Autism spectrum disorders: the quest for genetic syndromes. *Am J Med Genet B Neuropsychiatr Genet* 2013;162B:327–366.

7. Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol* 2014;10:74–81.

8. Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet* 2014;15:133–141.

9. Flore LA, Milunsky JM. Updates in the genetic evaluation of the child with global developmental delay or intellectual disability. *Semin Pediatr Neurol* 2012;19:173–180.

10. Castermans D, Vermeesch JR, Fryns JP, *et al*. Identification and characterization of the TRIP8 and REEP3 genes on chromosome 10q21.3 as novel candidate genes for autism. *Eur J Hum Genet* 2007;15:422–431.

11. Wolf SS, Patchev VK, Obendorf M. A novel variant of the putative demethylase gene, s-JMJD1C, is a coactivator of the AR. *Arch Biochem Biophys* 2007;460:56–66.

12. Kim SM, Kim JY, Choe NW, *et al*. Regulation of mouse steroidogenesis by WHISTLE and JMJD1C through histone methylation balance. *Nucleic Acids Res* 2010;38:6389–6403.

13. Watanabe S, Watanabe K, Akimov V, *et al*. JMJD1C demethylates MDC1 to regulate the RNF8 and BRCA1-mediated chromatin response to DNA breaks.*Nat Struct Mol Biol* 2013;20:1425–1433.

14. Neale BM, Kou Y, Liu L, *et al*. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012;485:242–245.

15. Iossifov I, O'Roak BJ, Sanders SJ, *et al*. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014;515:216–221.

16. Guerrini R, Parrini E. Epilepsy in Rett syndrome, and CDKL5- and FOXG1-gene-related encephalopathies. *Epilepsia* 2012;53:2067–2078.

17. Consortium T 1000 GP. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.

18. Sherry ST, Ward MH, Kholodov M, *et al*. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–311.

19. Tennessen JA, Bigham AW, O'Connor TD, *et al*.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 012;337:64–69.

20. Vollmar M, Johansson C, Krojer T, *et al*. Crystal structure of the Jumonji domain of human Jumonji domain containing 1C protein. Protein Data Base PDB Id: 2YPD. 2012.

21. Samocha KE, Robinson EB, Sanders SJ, *et al*. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014;46:944–950.

22. Ferrer-Costa C, Gelpí JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics* 2005;21:3176–3178.

23. Lein ES, Hawrylycz MJ, Ao N, *et al*. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 2007;445:168–176.

24. Percy AK. Rett syndrome. Current status and new vistas. *Neurol Clin* 2002;20:1125–1141.

25. Sanchez-Mut JV, Huertas D, Esteller M. Aberrant epigenetic landscape in intellectual disability. *Prog Brain Res* 2012;197:53–71.

26. Urdinguio RG, Sanchez-Mut JV, Esteller M. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol* 2009;8:1056–1072.

27. Jensen LR, Amende M, Gurok U, *et al*. Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum Genet* 2005;76:227–236.

28. Laumonnier F, Holbert S, Ronce N, *et al*. Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. *J Med Genet* 2005;42:780–786.

29. De Rubeis S, He X, Goldberg AP, *et al*.; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014;515:209–215.

INVENTORY OF SUPPLEMENTARY DATA

SUPPLEMENTARY METHODS

DNA samples

The samples used in this study consisted of 215 autism, Rett syndrome and intellectual disability patients from the Hospital Clínic, Sant Joan de Deu and Pompeu Fabra University, Barcelona, Spain. DNA was extracted from peripheral blood leukocytes using standard techniques. We measured DNA concentration with the Quant-iT Picogreen (Invitrogen) and then normalized all concentrations to 25-50 ng/µl before proceeding with the Access Array amplification.

Primer design

48 pairs of primers were designed using the Access Array Amplicon Tagging Assay design service from Fluidigm to cover all the 26 exons of the JMJD1C gene (NM_032776.1), including exon-intron boundaries. These primers generated 48

fragments varying in size from 360-489 bp. Amplicons were designed of approximately the same length to obtain an optimal sequencing result on the Junior 454. We also used the 96 Access Array Barcode Library (Fluidigm) to identify all the sequences in a pool of samples. This also contained the sequencing adaptors necessary for subsequent sequencing in the 454 GS Junior Sequencer.

Fluidigm access array

The Fluidigm Access Array is a microfluidic chip on which 48 patient samples and 48 primer pairs can be loaded. The outcome is a pool of 48 fragments per patient sample. By incorporating a unique identifier or barcode for each sample and the necessary sequencing adaptors, it is possible to pool the samples on the sequencing platform. Three access arrays were used to amplify the DNA samples in this study. For each experiment, 25-50 ng DNA per sample was used as input for the system. Experiments were performed according to the manufacturer's 4-Primer Amplicon Tagging protocol. Briefly, the target-specific primers were injected into the primer inlets and the sample-specific primers with their unique MID were loaded into the sample inlets along with the DNA samples and the PCR reagents. The primers and DNA mixture were then combined in the reaction chambers in the chip. After PCR, 10 µl of the samples were collected from their original wells, now containing a pool of 48 amplicons.

Verification and quantitation of harvested pcr products

Before running the samples on the GS Junior 454, we verified the amplification of the fragments using an Agilent 2100 BioAnalyzer with DNA 1000 chips, following the manufacturer's instructions. We ran 1 µl from all the amplified samples to ensure that the amplicon size and distribution were within the expected range. We also checked that primer dimer contamination was less than 25%. In addition, we obtained a concentration value used to ensure equimolar pools of amplified samples. 11 pools were obtained, which were sequenced on a 454 Titanium PicoTiterPlate device before purifying the pooled samples with Agencourt AMPure XP system (Beckman Coulter Genomics), following the manufacturer's instructions. This consists of magnetic beads that allow a high level of recovery of amplicons, efficient removal of unincorporated dNTPs, primers, primer dimers and salts.

Multiplex ligation-dependent probe amplification (MLPA)

All samples had been screened for large rearrangements in *JMJD1C*. We designed 9 MLPA probes specific to the JMJD1C gene and 6 control probes according to the instructions provided by MRC-Holland (www.mrcholland.com/pages/support_desing_synthetic_probespag.html). Probes are available from the authors on request. Unique sequences were identified using the BLAT program from UCSC (www.genome.ucsc.edu), and care was taken to avoid the presence of known sequence variants in the probe annealing site. Probes were designed to produce PCR products differing by 3 bp to allow correct separation by size. Oligonucleotides were obtained from Sigma–Aldrich (Haverhill, UK). The signal of each probe was adjusted after visual examination of preliminary results by raising or lowering the concentration in the probe mix. PCR products were analyzed on an ABI 3100 capillary sequencer using Gene Mapper software (Applied Biosystems, Foster City, CA, USA). The proportion of each peak relative to the height of all peaks was calculated for each sample and then compared with the proportions of the corresponding peak averaged over a set of at least ten normal DNA samples. Ratios between 0.8-1.2 were considered to have a normal copy number (2n).

Sanger sequencing

The variants were validated by Sanger sequencing using a *BigDye*® Terminator v3.1 Cycle Sequencing Kit in an Applied Biosystems 3730/DNA Analyzer. The raw data were analyzed with Codon Code Aligner Software.

Exome sequencing

The patient and healthy parents were analyzed by whole exome sequencing with TruSeq Sample Preparation Kit (Illumina). Exomes were captured with TruSeq Exome Enrichment Kit (Illumina) and paired-end 100x2 sequenced with the equip HiScan SQ. The raw data were analyzed in Centre Nacional d'Anàlisi Genòmica (CNAG), in Barcelona, Catalonia, Spain. FASTQ files were analyzed as is it follows:

1. Alignment and variant calling. Sequence reads were aligned to the human reference genome build GRCh37 (hg19) by using the Burrows-Wheeler Aligner

(BWA) (Li & Durbin, 2009). Properly mapped reads were filtered with SAMtools [Li, 2009], which was also used for sorting and indexing mapping files. GATK [McKenna, 2012] was used to realign the reads around known indels and for base quality score recalibration. Once a satisfactory alignment was achieved, identification of single nucleotide variants and indels was performed using GATK standard multisample variant calling protocol, including variant recalibration (DePristo *et al* 2011). For the final exome sequencing analysis report we used the ANNOVAR [Wang *et al* 2010] annotation tool to provide additional variant information to ease the final selection of candidates. In particular, minor allele frequency (MAF), obtained from dbSNP (Sherry 2001) and 1000 Genomes project was provided to help to select previously undescribed variants in healthy population.

2. SNV. To identified *de novo* single nucleotide variations, the patient´s variant were filtered first for the parental variants and then for the variants of a pool of controls made up by all healthy parents included in the study. Also SIFT (Kumar 2009) and Polyphen (Adzhubei 2010) damage scores were computed to predict putative impact over protein structures. The successive application of quality control filters and the prioritization by the parameters with potential functional impact was used to construct a list of candidate genes (and variants) ranked by its uniqueness in the cases (or very low frequency in the control population, as derived from the MAFs) and the putative potential impact. The variants were validated by Sanger sequencing using BigDye® Terminator v3.1 Cycle Sequencing Kit in an Applied Biosystems 3730/DNA Analyzer.
3. CNV. To identify Copy Number Variation, we used the C++ software XHMM (eXome-Hidden Markov Model) (Poultney 2013). The CNV events were filtered by DGV Data Base to remove common CNV and validated by Quantitative PCR using for normalization Type-it CNV KIT primers from QIAGEN and 2 multicopy amplicons with similar results.

Cell culture and vectors

JMJD1C coding sequence in pCMV6-AC-GFP vector was purchased from Origene (RG214878). The mutants were generated with Mutant QuikChange™ Site-Directed

Mutagenesis Kit. WT and Pro163Leu, His2336Ala mutant was subcloned in pCMV6-Entry vector to introduce Myc-DDK-tag. shRNAs against the coding sequence of mouse Jmjd1c gene were cloned in pLVX-shRNA2 vector between the BamHI and EcoRI restriction sites (shRNA24 target: CAGAGACTGCTTGAGGAAT). Hek293 cells were cultivated in DMEM 10% FBS. To generate stable WT or mutant clones, Hek293 cells were transfected with Lipofectamine 2000 (Invitrogen), selected with G418 antibiotic, and individual clones were isolated 2 weeks later. For transient expression, 6 mg of vector were transfected in 35 mm 6-well plates with jetPRIME™ transfection reagent following the manufacturer's instructions. Primary cultures of hippocampal neurons were prepared from neonate mice (P0). Brains were dissected out with forceps on ice and placed in petri dishes containing ice-cold Hibernate® media, meninges were removed under a dissecting microscope and hippocampuses were removed and saved in Hibernate. The tissues were transferred to 15-ml tubes containing trypsin-EDTA (0.025% in PBS). The tubes were incubated in a 37°C chamber for 15 minutes and agitated every 5 minutes. After stopping trypsinization with 5 ml 20% FBS, these tissues were triturated 15 times with a Pasteur pipette. The suspension was filtered through a 75 µm pore-sized filter and centrifuged in a tube for 2 minutes at 1000 *g*. The cells were resuspended in DMEM 10% FSB, Glutamax-Pyruvate and seeded at 2x106 cell/cm2. Three hours later, the medium was replaced with Neurobasal medium (GIBCO) containing 2% B27, 0.5 mM glutamine, and 50 U/ml penicillin/streptomycin, AraC 5 µM and cultured at 37°C, 5% CO2. Cultures were infected at 3DIV with lentiviral vectors to express scramble or shRNAs against JMJD1C together with a GFP tracer (pLVX-shRNA2 system). Coverslips were fixed and protein was extracted at 15DIV.

Cellular fractionation

Cell were harvested by scratching, washed in PBS buffer, and incubated 5 minutes in RBS buffer (10mM Tris HCl pH7.6, 10mM NaCl, 1.5mM MgCl2, 0.1% NP40) and centrifuge 1 minute at max speed. Supernatant was considerate cytoplasm fraction. The pellet was washed in once in RBS buffer and then incubated 5 minutes with RIPA (50 mM Tris HCl pH 8.0, 150mM NaCl, 1.0% NP-40, 0.5% Sodium Deoxycholate), the supernatant was considerate nuclear fraction, the pellet containing the chromatin fraction was resuspended en RIPA and sonicated to fragment the DNA.

Immunoprecipitation

750µg of chromatin fraction was diluted 10 fold in IP buffer (5mM Tris-HCl pH 7.6, 15mM HEPES pH 8.0, 1mM EDTA, 1mM EGTA, 0.1% SDS, 1% Triton X- 100), incubated with 2µg of antibodies anti-Me-Lysine (Abcam ab23366) overnight at 4°C and 2 hours with PureProteome Protein A/G Magnetic Beads. Beads were washed twice with IP buffer, twice in RBS NP-40 and eluted in laemli buffer in reduction condition at 70°C by 10 minutes. For the MeCP2 immunoprecipitation procedure, anti-JMJD1C and anti-MeCP2 antibodies were coupled to Dynabeads Protein G (Invitrogen). JMJD1C transfected HEK293F cells were transiently transfected with MeCP2-Flag tagged plasmid and the nuclear fraction was obtained by RIPA buffer (10 mM TRIS-Cl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 1% TRITON x-100, 0.1 % Sodium deoxycholate, 0.1% SDS and 140 mM NaCl) supplemented with protease inhibitors (complete, ROCHE) and hybridized with the antibodies at 4°C for 2h. 150 mM NaCl RBS buffer was used for washing. Human IgG was used as negative control. Anti- Flag HRP (M2-SIGMA) antibody was used to visualize the binding.

Western blot

Protein extract of Hek293 cells and primary neuronal culture was obtained in RIPA buffer supplemented with cOmplete^{TM} Protease Inhibitor Cocktails tablet (Roche) and sonicated. Protein concentrations were determined using a DC Protein Assay kit from Bio-RAD. 50 µg of each protein sample were denatured in Laemli buffer 4% β-mercaptoethanol for 10 minutes at 95°C and separated on a 7.5% or 15% SDS-polyacrylamide gel, then transferred onto a PVDF membrane (Immobilon-P, Millipore) by liquid electroblotting for 90 minutes at 100 V. The membrane was blocked in 5% nonfat dry milk in TBS-0.05% Tween 20. The antibodies and dilutions used are as follows: rabbit anti-JMJD1C 1:2000 (Millipore 09-817), mouse anti-nucleolin 1:1000 (Santa-Cruz SC-8031), rabbit anti-MeCP2 1:5000 (Sigma M9317), rabbit anti-H3 1:10000 (Abcam AB1791); mouse anti-H3 1:4000 (Abcam AB10799), rabbit anti-H3K9Me2 1:4000 (Abcam AB32521), rabbit anti-MDC1 1:5000 (Abcam AB11171). The blots were developed with Luminata^{TM} Crescendo Western HRP Substrate or with the LiCor Odyssey System.

Immunofluorescence

Cells were fixed in 4% PFA-PBS, quenched in 100 mM glycine-PBS, permeabilized with 0.25% Triton X-100, 1% BSA, PBS. The cells were blocked with 0.2% gelatin, 0.25% triton X-100. Antibody dilutions were prepared in 0.25% Triton X-100, 1% BSA, PBS. The dilutions used were: rabbit anti-JMJD1C 1:200, chicken anti-Map2 1:5000, anti β-tubulin 1:1500 (Abcam AB21058). Nuclei were stained with 2 mg/ml Hoechst 33342. Coverslips were mounted in ProLong® Gold antifade reagent.

Microscopy

Confocal images were captured with a Leica SP5 confocal microscope. For FRAP analysis the cells were maintained at 37°C in an atmosphere of 5% CO2. We captured images every 70 µs at 63x, 128x128 resolution, at 1400Hz with bidirectional acquisition. We captured 25 control images at 3% laser transmission before bleaching, and then bleached the ROI inside a nucleus 25 times at nominal level of 100% laser transmission. For this experiment, 150 images were captured after bleaching. The raw data were analyzed with FrapAnalyzer Software (http://actinsim.uni.lu/eng/Downloads/FRAPAnalyser).

RNA extraction and real-time PCR

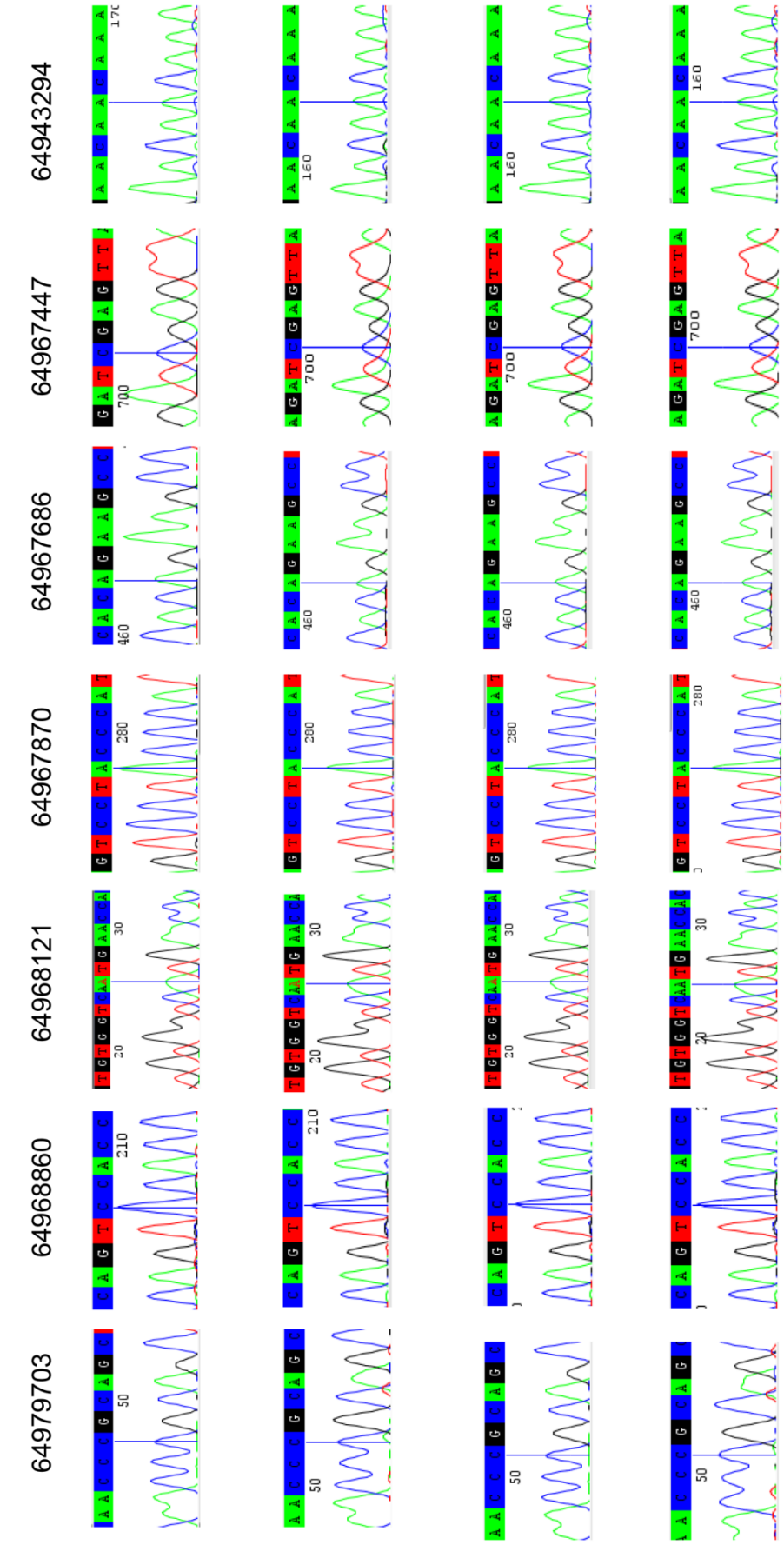
Total RNA was extracted from cell lysates using TRIzol Reagent (Invitrogen), purified using the RNeasy Kit (Qiagen) and 2 µg were retrotranscribed using the ThermoScript™ RT-PCR System (Invitrogen). Real-time PCR reactions were performed in triplicate on an Applied Biosystems 7,900HT Fast Real-Time PCR system using 20 ng cDNA, 5 µl SYBR Green PCR Master Mix (Applied Biosystem) and 150 nM specific primers (sequences are available upon request) in a final volume of 10 µl.

Supplementary Table S1. JMJD1C synonymous variants found in the studied patients

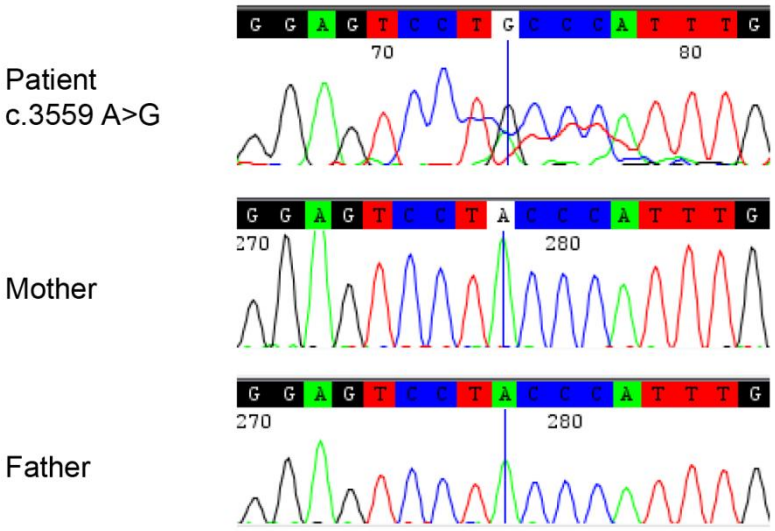
	Chr 10 position	Nucleotide Change	Exon	Amino acid	Protein Position
1	64974871	c. 1056A>G	8	R	352
2	64973860	c. 2067T>A	8	I	689
3	64973890	c.2037T>C	8	H	679
4	64968828	c. 2862T>C	9	H	954
5	64967478	c. 3951T>C	10	S	1317
6	64967445	c. 3984T>A	10	R	1146
7	64966764	c. 4665C>A	10	L	1555
8	64966860	c. 4569C>A	10	I	1523
9	64952834	c.5940G>A	16	P	1980
10	64945364	c. 6789C>T	20	D	2263

	Supplementary Table S2. Previously informed JMJD1C variants in studied patients					
	Chr10 position	**Nucleotide Change**	**Exon/Intron**	**Aminoacid change**	**Protein Position**	**Exome Variant Server**	**Patient**
1	64975327	c.808G>A	Exon 8	V/I	270	1/6019	Intellectual Disability
2	64974224	c.1703A>T	Exon 8	D/V	568	15/5953	1 Intellectual Disability 1 Autism Spectrum Disorder
3	64973978	c.1949C>T	Exon 8	T/I	650	114/5981	1 Autism Spectrum Disorder 1 Intellectual Disability
4	64967249	c.4180A>T	Exon 10	T/S	1394	8/6058	1 Intellectual Disability 1 Autism Spectrum Disorder
5	64966621	c.4808T>C	Exon 10	I/T	1603	2/5954	Intellectual Disability
6	64945336	c.6817A>G	Exon 20	M/V	2273	1/5918	Autism Spectrum Disorder
7	64948925	c.6570+3G>A	Intron 18	-	-	107/5896	Autism Spectrum Disorder

	Supplementary Table S3. JMJD1C mutational status in studied patients							
	Chr 10 position	**Nucleotide Change**	**Exon**	**Amino acid change**	**Protein Position**	**Exome Variant Server**	**Patient**	**Gender**	**Parental Status**
1	64979703	c.488C>T	4	P/L	163	0/5953	Rett syndrome	Female	Wild Type
2	64968860	c.2830C>T	9	P/S	944	0/5943	Autism Spectrum Disorder	Male	Not Available
3	64968121	c.3308A>G	10	N/S	1103	0/5989	Intellectual Disability	Female	Not Available
4	64967870	c.3559A>G	10	T/A	1187	0/6086	Intellectual Disability	Male	Wild Type
5	64967686	c.3743A>G	10	Q/R	1248	0/5946	Intellectual Disability	Female	Not Available
6	64967447	c.3982C>G	10	R/G	1328	0/6000	Autism Spectrum Disorder	Male	Not Available
7	64943294	c.6997A>G	22	T/A	2333	0/5880	Autism Spectrum Disorder	Male	Not Available



Supplementary Figure S1. Illustrative chromatograms of Sanger sequencing from control population (n=500). Numbers represent chromosome position and nucleotides of interest are marked with a vertical line.



Supplementary Figure S2. Chromatograms of Sanger sequencing showing the *de novo* status of the c.3559 A>G JMJD1C mutation. Parents are wild type for the amino acid.

STUDY II

"Whole exome sequencing of Rett syndrome‐like patients reveals the mutational diversity of the clinical phenotype"

Mario Lucariello[1] · Enrique Vidal[1] · Silvia Vidal[2] · Mauricio Saez[1] · Laura Roa[1] · Dori Huertas[1] · Mercè Pineda[3] · Esther Dalfó[4] · Joaquin Dopazo[5,6,7] · Paola Jurado[1] · Judith Armstrong[2,8,11] · Manel Esteller[1,9,10]

1 Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet, 08908 Barcelona, Catalonia, Spain

2 Servei de Medicina Genètica i Molecular, Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Esplugues De Llobregat, Catalonia, Spain

3 Fundació Hospital Sant Joan de Déu (HSJD), Barcelona, Catalonia, Spain

4 Genetics Department, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Catalonia, Spain

5 Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), 46012 Valencia, Spain

6 Bioinformatics of Rare Diseases (BIER), CIBER de Enfermedades Raras (CIBERER), Valencia, Spain

7 Functional Genomics Node (INB) at CIPF, 46012 Valencia, Spain

8 CIBER Enfermedades Raras, Barcelona, Catalonia, Spain

9 Department of Physiological Sciences, School of Medicine and Health Sciences, University of Barcelona, Barcelona, Catalonia, Spain

10 Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

11 Department of Neurology, Hospital Sant Joan de Déu (HSJD), Barcelona, Catalonia, Spain

For the sake of clarity and higher figure resolution, I next present the published article in Word format. The published PDF file and Supporting Information can be found in **Annex II**.

ABSTRACT

Classical Rett syndrome (RTT) is a neurodevelopmental disorder where most of cases carry MECP2 mutations. Atypical RTT variants involve mutations in *CDKL5* and *FOXG1*. However, a subset of RTT patients remains that do not carry any mutation in the described genes. Whole exome sequencing was carried out in a cohort of 21 female probands with clinical features overlapping with those of RTT, but without mutations in the customarily studied genes. Candidates were functionally validated by assessing the appearance of a neurological phenotype in *Caenorhabditis elegans* upon disruption of the corresponding ortholog gene. We detected pathogenic variants that accounted for the RTT-like phenotype in 14 (66.6 %) patients. Five patients were carriers of mutations in genes already known to be associated with other syndromic neurodevelopmental disorders. We determined that the other patients harbored mutations in genes that have not previously been linked to RTT or other neurodevelopmental syndromes, such as the ankyrin repeat containing protein *ANKRD31* or the neuronal acetylcholine receptor subunit alpha-5 (*CHRNA5*). Furthermore, worm assays demonstrated that mutations in the studied candidate genes caused locomotion defects. Our findings indicate that mutations in a variety of genes contribute to the development of RTT-like phenotypes.

INTRODUCTION

Rett syndrome (RTT, MIM 312750) is a postnatal progressive neurodevelopmental disorder (NDD), originally described in the 1960s by Andreas Rett (Rett, 1966), that most frequently manifests itself in girls during early childhood, with an incidence of approximately 1 in 10,000 live births (Chahrour and Zoghbi, 2007). RTT patients are asymptomatic during the first 6–18 months of life, but gradually develop severe motor, cognitive, and behavioral abnormalities that persist for life. It is the second most common cause of intellectual disability in females after Down's syndrome (Chahrour and Zoghbi, 2007). Around 90 % of the cases are explained by more than 800 reported mutations in the methyl CpG-binding protein 2 gene (MECP2) (RettBASE: MECP2 Variation Database) (Christodoulou *et al*., 2003), which is located in the X chromosome and which causes most of the classical or typical forms of RTT (Chahrour and Zoghbi,

2007), and it was originally identified as encoding a protein that binds to methylated DNA (Lewis *et al*., 1992). Individuals affected by atypical or variant RTT present with many of the clinical features of RTT, but do not necessarily have all of the classic characteristics of the disorder (Neul *et al*., 2010). Approximately 8 % of classic RTT and 42 % of variant RTT patients are MECP2 mutation negative (Monrós E *et al*., 2001; Percy AK, 2008). Some of the latter group have mutations in other genes, such as that of the cyclin-dependent kinase-like 5 (CDKL5), which is described in individuals with an early seizure onset variant of RTT (Kalscheuer *et al*., 2003) or the forkhead box G1 (FOXG1), which is responsible for the congenital variant of RTT (Ariani *et al*. 2008). However, there remains a subset of patients with a clinical diagnosis of RTT who are mutation-negative for all the aforementioned genes. Next generation sequencing (NGS) has emerged as a potentially powerful tool for the study of such genetic diseases (Zhu *et al*., 2015). Herein, we report the use of a family based exome sequencing approach in a cohort of 20 families with clinical features of RTT, but without mutations in the usually studied genes. We establish the neurological relevance of the newly identified candidate genes by assessing them in *Caenorhabditis elegans* model.

MATERIALS AND METHOD

Patient samples

A cohort of 19 Spanish parent–child trios and one family with two affected daughters who exhibited clinical features associated with RTT were recruited at Sant Joan de Deu Hospital in Barcelona, Catalonia, Spain. These patients had been diagnosed on the basis of the usual clinical parameters (Monros *et al*., 2001), and according to the recently revised RettSearch International Consortium criteria and nomenclature (Neul *et al*., 2010), but were found to be mutation-negative for *MECP2*, *CDKL5* and *FOXG1* in the original single-gene screening. The parents were clinically evaluated and it was not observed any evidence of intellectual disability. Genomic DNA from these patients was extracted from peripheral blood leukocytes using standard techniques, and analyzed by exome sequencing at the Cancer Epigenetics and Biology Program (PEBC) in Barcelona, Catalonia, Spain. Ethical approval for the molecular genetic studies was obtained from each institutional review board.

Whole exome sequencing and Sanger validation

Coding regions were captured using the TruSeq DNA Sample Preparation and Exome Enrichment Kit (Illumina, San Diego, California). Paired-end 100 × 2 sequences were sequenced with the Illumina HiScan SQ system at the National Center for Genomic Analysis in Barcelona. We also included the exome sequencing data of an *MECP2*, a *CDKL5* and a *JMJD1C* (Sáez *et al*., 2016) RTT-associated family for data processing to improve the *de novo* single nucleotide variant calling. The complete exome sequencing data of all the studied samples are available from the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) with the ID: SRP073424 (private link for the reviewer until publication: http://www.ncbi.nlm.nih.gov/sra/SRP073424). The overall coverage statistics for each individual of the families, considering the regions captures using Exome Enrichment Kit, and number of reads in the position of the variation is shown in **Supplementary Table 1**. The identified variants were validated by Sanger sequencing using a BigDye® Terminator v3.1 Cycle Sequencing Kit in an Applied Biosystems 3730/DNA Analyzer. The raw data were analyzed with Codon Code Aligner Software. The primers used for Sanger sequencing are shown in **Supplementary Table 2**.

Caenorhabditis elegans handling

The techniques used for the culture of *Caenorhabditis elegans* were essentially as described (Brenner, 1974). The worms were backcrossed at least three times to avoid background mutations. The behavior of three sets of ten animals was independently assessed in locomotion assays without food that were performed at 20 °C, as previously described (Sawin *et al*., 2000).

RESULTS

Clinical criteria for selecting RTT trios

The 21 patients (derived from the 20 families studied) included in this study fulfilled the recently revised clinical criteria for the diagnosis of RTT following the usual clinical parameters (Monros *et al*., 2001), and the RettSearch International Consortium criteria and nomenclature (Neul *et al*., 2010). Specifically, all patients presented stereotypic

hand movements, 90.5 % of them (19/21) showed microcephaly and also presented onset of the first signs of the disease before the age of 12 months. 66.7 % of patients (14/21) acquired motor skills, while a further seven (33.3 %), who had a more severe phenotype, never walked. Language skills were progressively lost in 28.6 % of the patients and 71.4 % of them (15/21) never acquired them. Additionally, important episodes of epilepsy were experienced by 81.0 % of the patients (17/21), and 57.1 % of them (12/21) manifested apneas and/or hyperventilation.

Bioinformatic process for filtering and selecting pathogenic variants

Before their inclusion in this study, patients underwent an extensive clinical and genetic work-up to detect genetic alterations in *MECP2*, *CDKL5*, and *FOXG1*. However, no molecular diagnosis could be established. We performed whole exome sequencing (WES) on the 61 individuals (20 pairs of healthy parents and 21 affected daughters) separately by subjecting whole blood derived genomic DNA to exome enrichment and sequencing. We focused our analysis on *de novo* single nucleotide variants (SNVs) due to their known relevance in autism and mental retardation-related diseases (Vissers *et al*., 2010). On average, WES gave rise to 419,045 variants, including SNVs and indels, of which 19,951 non-synonymous variants per family (4.7 %) were predicted to have a functional impact on the genomic sequence. To select variants that had not previously been described in the healthy population, we filtered out the variants with an allele frequency of 1 % or higher (the classic definition of a polymorphism) formerly observed in the Single Nucleotide Polymorphism database (dbSNP) and the 1000 Genomes Pilot Project data. Afterwards, to focus on *de novo* inheritance, patients' variants were filtered first against variants found in their own parents and then against a pool of controls comprising all the healthy parents included in the study. Following this process, we achieved an average of 106 SNVs per family, which corresponded to 81 mutated genes per family. *De novo* candidate variants were selected on the basis of the quality of the alignments, damage score predictors and the conservation level of each of the genes during evolution. The complete exome sequencing data of all the studied samples are available from the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra). The global yield of genomic analysis following the bioinformatics process described herein enabled 22 coding *de novo* mutations to be

identified in 66.7 % (14 of 21) of Rett-like patients: 20 SNVs and 2 indels. The identified variants and their *de novo* status were confirmed by conventional Sanger sequencing. Illustrative samples are shown in **Fig. 1**. Interestingly, in seven (33.3 %) of the studied RTT probands, exome sequencing did not detect any genetic change relative to their respective parents. The clinical characteristics of these seven patients without obvious pathogenic variants are summarized in **Table 1**. In one of the families, there were two affected children, and an analysis of potentially relevant recessive variants was performed. For the recessive analysis, and following the same criteria to define a variant as deleterious, we selected the variants with homozygous recessive genotype, and then at the gene level, we also selected the genes presenting more than one heterozygotic variant in the same gene (compound heterozygosity). We did not find any candidate gene consistent with the phenotype of the family with the two affected sisters.

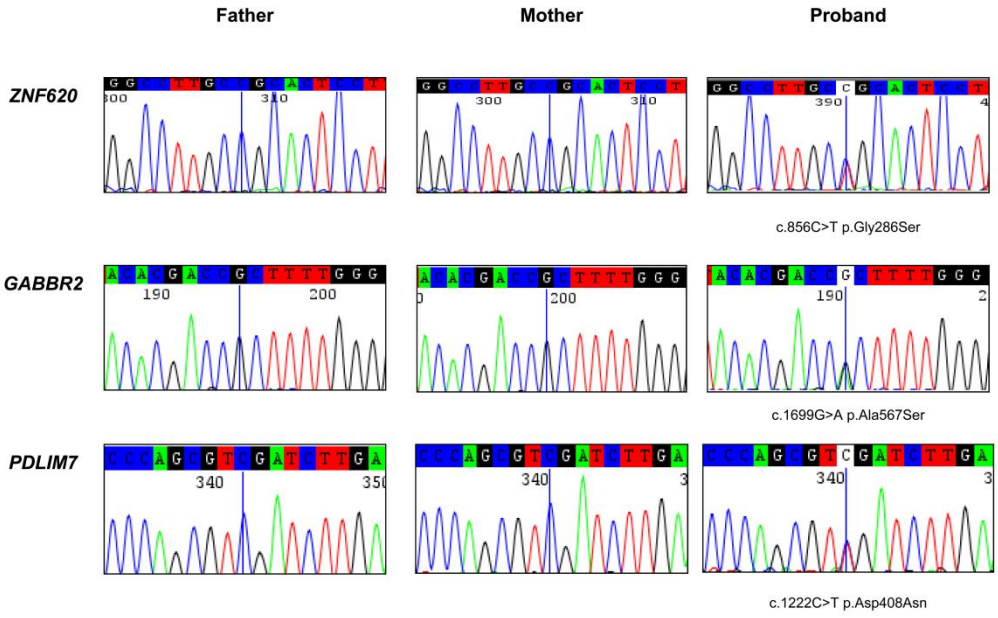


Fig. 1 Sanger sequencing validation of the *de novo* variants identified by exome sequencing. Illustrative examples for *ZNF620* (c.856C > T p.Gly286Ser), *GABBR2* (c.1699G > A p.Ala567Ser) and *PDLIM7* (c.1222C > T p.Asp408Asn) are shown.

Variants in genes previously associated with neurodevelopmental disorders

Of the 22 identified coding *de novo* mutations in the assessed RTT-like patients, five (22.7 %) occurred in genes previously associated with neurodevelopmental disorders that presented a clinicopathological phenotype with features coinciding with those of Rett syndrome (**Table 2**). In particular, we identified four mutations in genes such as *HCN1* (Nava *et al*., 2014) and *GRIN2B* (Endele *et al*., 2010; Lemke *et al*., 2014), which are associated with early infantile epileptic encephalopathy; *SLC6A1*, which is associated with epilepsy and myoclonic-atonic seizures (Carvill *et al*., 2015); *TCF4*, which is associated with Pitt–Hopkins syndrome (Sweatt, 2013); and *SCN1A*, which is associated with Dravet syndrome (Brunklaus and Zuberi, 2014) (**Table 2**). The clinical characteristics of these five patients with variants in genes previously associated with neurodevelopmental phenotypes are summarized in **Table 3**. A comparison of the clinical features of our RTT-like patients, where we have identified mutations in candidate genes previously associated with other neurodevelopmental disorders, with those observed for these diseases is summarized in **Table 4**.

Variants in genes previously not associated with neurodevelopmental disorders

Of the 22 identified coding *de novo* variants in the RTT-like patients assessed here, 17 (77.3 %) occurred in genes that had not previously been associated with neurodevelopmental disorders (**Table 5**). However, two of these variants were associated with non-neurodevelopmental disorders: a *BTBD9* variant linked to restless leg syndrome (Kemlink *et al*., 2009), and an *ATP8B1* SNV associated with familial cholestasis (Klomp *et al*., 2004), respectively. Interestingly, the *BTBD9* variant was detected in the same patient that carried the *SCN1A* variant associated with Dravet syndrome (**Table 2**). The other 15 potentially pathogenic variants identified occurred in genes that had not been linked to any genetic disorder of any type. However, there was an enrichment of genes with a potential role in neuronal biology and functionality, such as the gamma-aminobutyric type B receptor subunit 2 (*GABBR2*), the neuronal acetylcholine receptor subunit alpha-5 (*CHRNA5*), the Huntington-associated protein 1 (*HAP1*), the axon guider semaphorin 6B (*SEMA6B*), the ankyrin repeat containing proteins *ANKRD31* and *AGAP6*, and the neuronal voltage-gated calcium channel *CACNA1* (**Table 5**). Proband 14 was a particularly interesting case in which four

potential pathogenic variants were present, affecting zinc finger (*ZNF620*), a nucleolar complex (*NOC3L*), G patch domain (*GPATCH2*) and GRAM domain (*GRAMD1A*)-related proteins (**Table 5**). The clinical characteristics of these patients with variants in genes previously not associated with neurodevelopmental disorders are summarized in **Table 6.**

Neurological phenotype of candidate genes in C. elegans

To demonstrate a neurological effect for a loss of function of the detected genes that had not previously been associated with neurodevelopmental disorders (**Table 5**), we used the model organism *C. elegans* to confirm the genotype-phenotype correlation. We obtained all the available *C. elegans* mutants that carry deleterious mutations in the orthologous genes to those human genes with potentially pathogenic mutations in the patients. In this model, backcrossing is a commonly used procedure to obtain a specific mutant strain without any secondary mutations from its genetic composition. Under these conditions, we were able to test six available mutant strains that were backcrossed at least three times to prove that any observed phenotype was really associated to specific mutations in the orthologous genes. To this end, we studied the *C. elegans* mutants carrying deleterious mutations in the gene orthologs of the human genes *PDLIM7*, *ANKRD31*, *ZNF620*, *CHRNA5*, *MGRN1* and *GABBR2* described in **Table 7**. Considering that the loss of normal movement and coordination is one of the clearest signs shown by Rett patients, we performed a locomotion assay of the nematodes as previously described (Sawin *et al*., 2000), using the wild-type N2 strain as a control (**Supplementary Video 1**). We observed that in 83.3 % (5 of 6) of the cases the mutation of the ortholog of the human exome sequencing identified genes in *C. elegans* exhibited a locomotion defective phenotype (**Fig.2**). The most severe phenotypes were represented by *alp-1*, *unc-44* and *pag-3*, with mutations in the orthologs of PDZ and LIM domain protein 7 (*PDLIM7*), ankyrin repeat containing protein *ANKRD31* and the zinc protein *ZNF620*, respectively (**Fig. 2** and **Supplementary Videos 2**, **3** and **4**). The case of *alp-1* was particularly interesting, because mutant worms were not only thinner than usual and completely locomotion defective, but they exhibited transitory spasms. Significant defects, such as slower locomotion and uncoordinated movement, were also observed in the mutants of *unc-63* and *C11H1.3*,

the *C. elegans* orthologs of the genes coding for the neuronal acetylcholine receptor subunit alpha-5 (*CHRNA5*) and mahogunin RING finger protein 1 (*MGRN1*), respectively. Although we did not find a clear locomotion defect in the *gbb-2* mutant (the ortholog of *GABBR2*) (**Fig. 2**), it occurs in the *gbb-1;gbb-2* double mutant (Dittman and Kaplan, 2008), being *gbb-1* the *C. elegans* ortholog of *GABBR1* (gamma-aminobutyric acid type B receptor subunit 1). The clinical picture of the particular RTT cases with mutations in the genes studied in *C. elegans* is shown in **Table 6**.

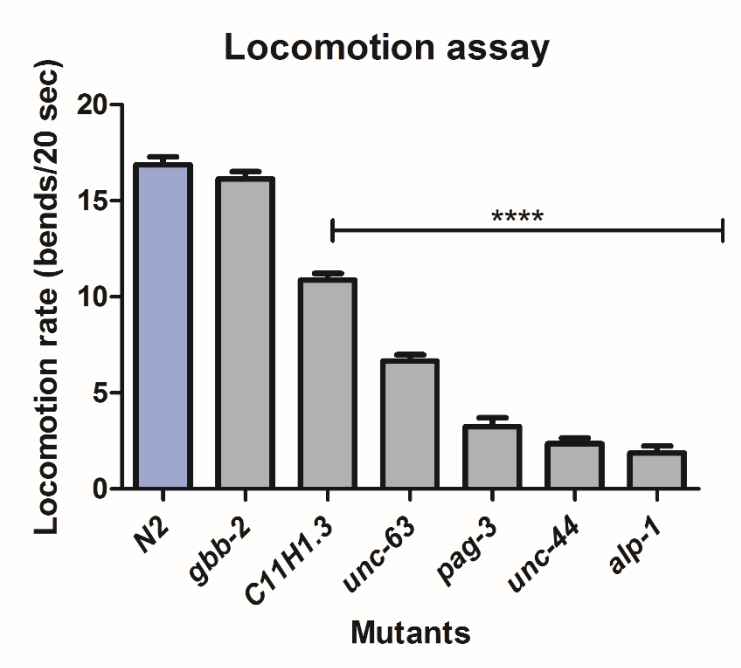


Fig. 2 Locomotion assay in *Caenorhabditis elegans*. Functional validation of mutations was performed by measuring the locomotion rate, expressed in average of measuring, in *C. elegans*. Each mutant strain was compared to a wild-type N2 control strain by measuring worm body bends during 20 sec in three independent sets of experiments. Locomotion rates of mutants, represented by *C11H1.3* (*MGRN1*), *unc-63* (*CHRNA5*), *pag-3* (*ZNF620*), *unc-44* (*ANKRD31*) and *alp-1* (*PDLIM7*) are significantly lower compared to that of the N2 control strain ($p < 0.0001$), on the contrary *gbb-2* (*GABBR2*) mutant move similarly. Standard error of the mean (SEM) values is shown. p values obtained according to Student's t test. ***$p < 0.0001$.

Table 1 Clinical summary of patients without exome candidates

Proband	Age (years)	Onset of signs	Microcephaly	Sitting alone	Ambulation	Respiratory function	Epilepsy	Hand use	Stereotypies	Language	Total score
1	15	1	1	1	1	0	1	1	1	1	8
2	28	3	1	1	0	1	2	2	3	2	15
5.1	7	3	1	3	4	1	0	3	2	2	19
5.2	5	3	1	1	2	0	0	2	3	2	14
7	16	2	1	0	0	0	2	1	2	1	9
15	8	3	0	1	2	0	2	1	1	1	10
18	3.5	3	1	1	1	0	0	1	3	0	10

Clinical scores of our series of patients according to Pineda scale. Severity classification ranges from 0 to 4 as follows: age of onset of first signs (1: >24 months; 2: 12-24 months; 3: 0-12 months), microcephaly (0: absent; 1: present), sitting alone (1: acquired <8 months; 2: seat and mantains; 3: seat and lost), ambulation (0: acquired <18 months, 1: acquired < 30 months; 2: acquired > 30 months; 3: lost acquisition; 4: never acquired), respiratory function (0: no dysfunction; 1: hyperventilation and/or apnea), epilepsy (0: absent; 1: present and controlled; 2: uncontrolled or early epilepsy), hands use (0: acquired and conserved; 1: acquired and partially conserved; 2: acquired and lost; 3: never acquired), onset of stereotypies (1: >10 years, 2: >36 months; 3: 18-36 months) and languages (0: preserved and propositive; 1: lost; 2: never acquired). The total score is the sum of the scores of each clinical feature.

Table 2 List of patients with variants found in genes previously associated with neurodevelopmental phenotype

Proband	Gene	Protein	NM number	Variant: genomic coordinates	cDNA change	Protein change	Gene-disease association
4	*HCN1*	Hyperpolarization Activated Cyclic Nucleotide Gated Potassium Channel 1	NM_021072.3	5:45396665	c.1159G>T	p.Ala387Ser	Early infantile epileptic encephalopathy 24
8	*SCN1A*	Sodium Channel Protein Type I Subunit Alpha	NM_001165963.1	2:166866266	c.3965C>G	p.Arg1322Thr	Dravet syndrome
10	*TCF4*	Transcription Factor 4	NM_001243236.1	18:52901827	c.958delC	p.Gln320Ser_fs8X	Pitt-Hopkins syndrome
11	*GRIN2B*	Glutamate receptor ionotropic, NMDA 2B	NM_000834.3	12:13764782	c.1657C>A	p.Pro553Thr	Autosomal Dominant Mental Retardation 6; Early infantile epileptic encephalopathy 27
17	*SLC6A1*	Solute Carrier Family 6 Member 1	NM_003042.3	3:11067528	c.919G>A	p.Gly307Arg	Myoclonic-atonic epilepsy and schizophrenia

Table 3 Clinical summary of the patients with variants in genes previously associated with neurodevelopmental phenotype

Proband	Age (years)	Onset of the signs	Microcephaly	Sitting alone	Ambulation	Respiratory function	Epilepsy	Hand use	Stereotypies	Language	Total score
4	24	3	1	1	3	1	2	3	1	2	16
8	7	3	1	1	4	1	2	1	3	2	18
10	16	3	1	1	2	1	2	1	1	2	14
11	3	2	0	1	3	0	1	1	1	2	10
17	36	3	0	1	1	0	1	1	1	0	7

Clinical scores of our series of patients according to Pineda scale. Severity classification ranges from 0 to 4 as follows: age of onset of first signs (1: >24 months; 2: 12-24 months; 3: 0-12 months), microcephaly (0: absent; 1: present), sitting alone (1: acquired <8 months; 2: seat and mantains; 3: seat and lost), ambulation (0: acquired <18 months, 1: acquired < 30 months; 2: acquired > 30 months; 3: lost acquisition; 4: never acquired), respiratory function (0: no dysfunction; 1: hyperventilation and/or apnea), epilepsy (0: absent; 1: present and controlled; 2: uncontrolled or early epilepsy), hands use (0: acquired and conserved; 1: acquired and partially conserved; 2: acquired and lost; 3: never acquired), onset of stereotypies (1: >10 years, 2: >36 months; 3: 18-36 months) and languages (0: preserved and propositive; 1: lost; 2: never acquired). The total score is the sum of the scores of each clinical feature.

Table 4 Comparison among clinical features described in OMIM and those observed in patients with variants in genes known to be disease causing

Disease	**Rett**		**Atypical Rett**		**Pitt Hopkins**		**Dravet**		**EEIE27**		**MAE**		**EEIE24**	
GENE	***MECP2***		***CDKL5***		***TCF4***		***SCN1A***		***GRIN2B***		***SLC6A1***		***HCN1***	
OMIM / Patient	**312750**	**6**	**308350**	**16**	**602272**	**10**	**182390**	**8**	**616139**	**11**	**616421**	**17**	**615871**	**4**
Onset age	6-18m	12-24m	1-3m	0-12m	6-12m	0-12m	2d-7m	0-12m	0-24m	12-24m	0-24m	0-12m	0-24m	0-12m
Microcephaly	Yes	Yes	Yes	Yes	Yes	Yes	+/-	Yes	+/-	No	+/-	No	NA	Yes
Hypotonia	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	+/-	Yes	+/-	Yes	NA	Yes
Epilepsy	80%	Yes	Yes	Yes	Yes	Yes	Yes	Yes	+/-	Yes	Yes	Yes	Yes	Yes
Respiratory dysfunction	80%	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	NA	Yes
Expressive language dysfunction	Yes	Yes	Yes	Yes	Yes	Yes	+/-	Yes	Yes	Yes	No	No	NA	Yes
Preserved use of hands	No	No	+/-	No	No	No	Yes	No	+/-	No	+/-	No	NA	No
Stereotypies	Yes	Yes	Yes	Yes	Yes	Yes	+/-	Yes	+/-	Yes	+/-	Yes	NA	Yes
Inheritance	XL		XL		AD		AD		AD		AD		AD	

EEIE, Epileptic Encephalopathy, Early Infantile; MAE, *Myoclonic*-Atonic Epilepsy; m, months; XL, X linkage; AD, Autosomal Dominant; NA, unavailable.

Table 5 List of patients with variants in new candidate disease genes

Proband	Gene	Protein	Function	NM number	Variant: genomic coordinates	cDNA change	Protein change	ExAC	SIFT	Polyphen2	PROVEAN	Mutation Taster2	Conservation
3	*AGAP6*	ArfGAP with GTPase domain, ankyrin repeat and PH domain 6	Putative GTPase-activating protein	NM_001077665.2	10:51748528	c.53insC	p.Asp18Ala_fs10X	Not present	NA	NA	B	P	405
8	*MGRN1*	Mahogunin RING Finger Protein 1	E3 ubiquitin-protein ligase	NM_001142290.2	16:4723583	c.880C>T	p.Arg294Cys	0,000077	P	P	P	P	573
8	*BTBD9*	BTB (POZ) Domain Containing 9	Putative protein-protein interactor	NM_001099272.1	6:38256093	c.1409C>T	p.Ala470Val	Not present	B	P	P	P	512
11	*SEMA6B*	Semaphorin-6B	Role in axon guidance	NM_032108.3	19:4555540	c.508G>A	p.Gly170Ser	Not present	P	P	P	P	510
12	*VASH2*	Vasohibin 2	Angiogenesis inhibitor	NM_001301056.1	1:213161902	c.1044A>C	p.Glu348Asp	Not present	B	B	B	B	473
13	*CHRNA5*	Neuronal acetylcholine receptor subunit alpha-5	Excitator of neuronal activity	NM_000745.3	15:78882481	c.748C>A	p.Pro250Thr	Not present	B	P	P	P	519
14	*ZNF620*	Zinc Finger Protein 620	Transcriptional regulator	NM_175888.3	3:40557941	c.856G>A	p.Gly286Ser	Not present	P	P	P	P	317
14	*GRAMD1A*	GRAM Domain Containing 1A	Not described	NM_020895.3	19:35506764	c.1106G>A	p.Arg369His	Not present	P	P	P	P	358
14	*NOC3L*	Nucleolar complex protein 3 homolog	Regulator of adipogenesis	NM_022451.10	10:96097586	c.2137G>A	p.Ala713Thr	Not present	B	B	B	B	0
14	*GPATCH2*	G patch domain-containing protein 2	Regulator of cell proliferation	NM_018040.3	1:217784371	c.878G>A	p.Gly293Asp	Not present	B	P	P	P	304
19	*GABBR2*	Gamma-aminobutyric acid type B receptor subunit 2	Inhibitor of neuronal activity	NM_005458.7	9:101133817	c.1699G>A	p.Ala567Thr	Not present	P	P	P	P	412
19	*ATP8B1*	Phospholipid-transporting ATPase IC	Aminophospholipid translocator	NM_005603.4	18:55328507	c.2606C>T	p.Thr869Ile	Not present	P	P	P	P	361
20	*HAP1*	Huntingtin-Associated Protein 1	Vesicular transporter	NM_177977.2	17:39890655	c.232G>A	p.Ala78Thre	Not present	P	B	B	B	0
21	*PDLIM7*	PDZ and LIM domain protein 7	Scaffold protein	NM_005451.4	5:176910933	c.1222G>A	p.Asp408Asn	Not present	P	P	B	P	515
21	*SRRM3*	Serine/Arginine Repetitive Matrix 3	Splicing activator	NM_001291831.1	7:75890878	c.655C>G	p.Ser218Cys	Not present	P	P	P	P	491
22	*ANKRD31*	Ankyrin Repeat Domain 31	Not described	NM_001164443.1	5:74518166	c.196A>T	p.Ile66Phe	Not present	P	P	B	B	401
23	*CACNA1I*	Voltage-Gated Calcium Channel Subunit Alpha 1I	Calcium signaling in neurons	NM_021096.3	22:40066855	c.4435C>T	p.Leu1479Phe	Not present	B	P	B	P	695

ExAC, frequency of the identified variants in the Exome Aggregation Consortium. Four *in silico* prediction tools of functional mutation impact were used: 'Sorting Tolerant From Intolerant' (**SIFT**), 'Polymorphism Phenotyping v2' (**Polyphen2**); 'Protein Variation Effect Analyzer' (**PROVEAN**) and **Mutation Taster2**. The output results were classified as: likely pathogenic (P), likely benign (B) and not available (NA). **Conservation** scores refer to conservation level of nucleotide at position of variant between 46 species of vertebrates (Homo Sapiens, Pan troglodytes, Gorilla Gorilla Gorilla, Mus Musculus, Rattus Norvegicus etc.), based on PhastCons (Siepel A *et al*., 2005). It ranges from 0 to 1000: the highest, the more conserved during evolution.

Table 6 Clinical summary of patients with variants in new candidate disease genes

Proband	Age (years)	Onset of the signs	Microcephaly	Sitting alone	Ambulation	Respiratory function	Epilepsy	Hand use	Stereotypies	Language	Total score
3	14	3	1	2	4	1	1	3	2	2	19
8	7	3	1	1	4	1	2	1	3	2	18
11	3	2	0	1	3	0	1	1	1	2	10
12	11	3	1	1	4	1	0	2	2	1	15
13	10	3	1	2	4	1	1	2	3	2	19
14	2	3	1	1	2	0	1	3	2	2	15
19	2	3	1	1	4	0	0	3	2	2	16
20	24	3	1	1	1	1	0	2	2	1	12
21	5	3	1	3	4	0	1	2	1	2	17
22	17	3	0	1	2	0	1	2	3	2	14
23	1/8	3	1	1	2	1	0	3	3	2	16

Clinical scores of our series of patients according to Pineda scale. Severity classification ranges from 0 to 4 as follows: age of onset of first signs (1: >24 months; 2: 12-24 months; 3: 0-12 months), microcephaly (0: absent; 1: present), sitting alone (1: acquired <8 months; 2: seat and mantains; 3: seat and lost), ambulation (0: acquired <18 months, 1: acquired < 30 months; 2: acquired > 30 months; 3: lost acquisition; 4: never acquired), respiratory function (0: no dysfunction; 1: hyperventilation and/or apnea), epilepsy (0: absent; 1: present and controlled; 2: uncontrolled or early epilepsy), hands use (0: acquired and conserved; 1: acquired and partially conserved; 2: acquired and lost; 3: never acquired), onset of stereotypies (1: >10 years, 2: >36 months; 3: 18-36 months) and languages (0: preserved and propositive; 1: lost; 2: never acquired). The total score is the sum of the scores of each clinical feature.

Table 7 Phenotype in Caenorhabditis elegans

Human gene	Ortholog in C.elegans	Similarity (%)	Identity (%)	Mutation in C.elegans	Locomotion phenotype	Neurological phenotypes	Other phenotypes
GABBR2	*gbb-2*	53	34	Deletion	wildtype	Hypersensitivity to aldicarb	-
MGRN1	*C11H1.3*	58	41	Deletion	locomotion defective	-	-
CHRNA5	*unc-63*	58	40	Deletion	locomotion defective	Uncoordinated locomotion with strong levamisole resistance	-
ZNF620	*pag-3*	65	47	Deletion	locomotion defective	Altered neurosecretion and up-regulation of DCV (Dense Core Vescicles) components	-
ANKRD31	*unc-44*	59	39	Deletion	locomotion defective	Asymmetric dynamics of axonal and dendritic microtubules defects	-
PDLIM7	*alp-1*	65	47	Deletion	locomotion defective	-	Defects in actin filament organization in muscle cells

DISCUSSION

Our results indicate that the existence of *de novo* variants in genes with potential neurological functionalities, such as neuronal receptors (*GABBR2* and *CHRNA5*), axon guiders (*SEMA6B*), synaptic ionic channels (*CACNA1I*) and others, contribute to the development of RTT-like clinical phenotypes in the context of wild-type sequences for standard Rett genes such as *MECP2* and *FOXG1*. These patients share most of the clinicopathological features of classic RTT syndrome, such as stereotypic hand movements, relative microcephaly, and onset of the disease after the age of 12 months. Thus, exome sequencing is a powerful tool for genetically characterizing these enigmatic cases. In this regard, once a new candidate gene has been identified, it is now possible to design specific sequencing strategies for the molecular screening of this particular target in larger populations of patients with intellectual disability. The strategy based on exome sequencing patients who have RTT features, but no known mutations in the usual genes, has recently been used in other smaller series of patients (Grillo *et al*., 2013; Okamoto *et al*., 2015; Hara *et al*., 2015; Olson *et al*., 2015; Lopes *et al*., 2016). Most importantly, our study and the aforementioned previous reports strengthen the concept that a mutational heterogeneous profile hitting shared neurological signaling pathways contributes to RTT-like syndromes. Examples of confluence in the same molecular crossroads include the gamma-aminobutyric type B receptor subunit 2 (*GABBR2*) *de novo* variant, described here, and the formerly identified variant in the gamma-aminobutyric acid receptor delta gene (*GABRD*) (Hara *et al*., 2015). Interestingly, a second RTT-like patient has been identified as being a carrier of a *de novo GABBR2* variant (Lopes *et al*., 2016), highlighting the likelihood that this gene and pathway contribute to the clinical entity. Another example of similarly targeted genes in RTT-like patients is that of the proteins containing ankyrin-repeats that are involved in postsynaptic density (Durand *et al*., 2007). This study has revealed *de novo* variants in the ankyrin repeat containing proteins *AGAP6* and *ANKRD31* in RTT-like patients, and the presence of *de novo* variant of the SH3 and multiple Ankyrin repeat domain3 protein (*SHANK3*) (Hara *et al*., 2015) and ankyrin-3 (*ANK3*) (Grillo *et al*., 2013) has been reported in two RTT-like patients. A final example of the convergence of cellular pathways to provide a common RTT-like phenotype is represented by the disruption of the ionic channels. We found the existence of a voltage-gated calcium channel subunit alpha 11 (*CANA1I*) *de novo* variant in an RTT-

like patient. Additionally, the presence of *de novo* variants in the calcium release channel *RYR1* (Grillo *et al*., 2013) and the sodium voltage-gated channel alpha subunit 2 (*SCN2A*) (Baasch *et al*., 2014) in two other RTT-like probands have been reported. It is also intriguing that in our study a variant in HAP was found, whereas in similar series heterozygous variants in huntingtin (HTT) have been described (Lopes *et al*., 2016; Rodan *et al*., 2016), further reinforcing the links between Huntington's disease and Rett syndrome (Roux *et al*., 2012). Another interesting case is provided by TCF4, which is associated with Pitt–Hopkins syndrome (Sweatt, 2013), where in addition to our study, others have found mutations in RTT-like patients (Lopes *et al*., 2016). This observation could be of interest for clinicians due to phenotypic similitudes such as intellectual disability, stereotypic movement, apneas and seizures (Marangi *et al*., 2012). Our findings also suggest that a substantial degree of clinical overlap can exist between the features associated with RTT and those of other neurodevelopmental disorders. Our exome sequencing effort indicated that probands originally diagnosed as RTT-like patients were, in fact, carriers of well-known pathogenic *de novo* mutations linked to Dravet Syndrome (*SCN1A*), myoclonic-atonic epilepsy (*SCLC6A1*), or early infantile epileptic encephalopathies 24 (*HCN1*) and 27 (*GRIN2B*). The purely clinical classification of these patients, without a thorough genetic study, can be difficult because some of these patients are composites that carry at least two pathogenic variants. For example, in our cases, the Dravet syndrome patient also had a *de novo* variant in *BTBD9* associated with the development of restless leg syndrome. In addition, among the newly identified candidate genes associated with RTT-like features, a few of these patients simultaneously carried two *de novo* variants (e.g., probands 8, 19 and 21), further complicating the tasks of correctly diagnosing and managing these individuals. Finally, the studies performed in *C. elegans* validate the functional relevance for nervous system function of the newly proposed candidate genes. Future studies would be necessary to assess the role of the specific variants identified, such as rescuing the defects with the expression of normal cDNAs versus cDNAs containing the mutation, ideally using cDNAs of human origin to prove similar function of the gene in the two species. It is also relevant to mention that for some of the newly reported mutated genes in our RTT-like patients, there are mice models targeting the described loci that show neurological phenotypes such as *BTBD9* (motor restlessness and sleep disturbances) (DeAndrade *et al*., 2012), *MGRN1* (spongiform

neurodegeneration) (He *et al*., 2003), *SEMA6B* (aberrant mossy fibers) (Tawarayama *et al*., 2010), *CHRNA5* (alterations in the habenulo-interpeduncular pathway) (Fowler *et al*., 2011), *GABBR2* (anxiety and depression-related behavior) (Mombereau *et al*., 2005) and *HAP1* (depressive-like behavior and reduced hippocampal neurogenesis) (Chan *et al*., 2002; Xiang *et al*., 2015).

CONCLUSIONS

Overall, this study demonstrates the genetic mutational diversity that underlies the clinical diagnosis of patients with clinical features that resemble RTT cases. Once the recognized *MECP2*, *CDKL5* and *FOXG1* mutations have been discarded, exome sequencing emerges as a very useful strategy for the more accurate classification of these patients. The *de novo* variants identified by this approach can modify the first diagnostic orientation towards another neurodevelopmental disorder, or pinpoint new genes involved in the onset of RTT-like features. Interestingly, most of these new targets are involved in the same functional networks associated with correct neuronal functionality. Further research is required to understand the role of these proteins in the occurrence of neurodevelopmental diseases. Additional functional experiments, such as the *C. elegans* assays used in this study, would be extremely helpful for this purpose.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2012 under REA Grant agreement PITN-GA-2012-316758 of the EPITRAIN project and PITN-GA-2009–238242 of DISCHROM; the E-RARE EuroRETT network (Carlos III Health Institute project PI071327); the Foundation Lejeune (France); the Cellex Foundation; the Botín Foundation; the Finestrelles Foundation; the Catalan Association for Ret Syndrome; and the Health and Science Departments of the Catalan government (Generalitat de Catalunya). M.E. is an ICREA Research Professor.

COMPLIANCE WITH ETHICAL STATEMENT

Conflict of interest

The authors declare that there is no conflict of interest associated with this manuscript.

REFERENCES

Ariani F *et al* (2008) FOXG1 is responsible for the congenital variant of Rett syndrome. Am J Hum Genet 83:89–93

Baasch AL *et al* (2014) Exome sequencing identifies a de novo SCN2A mutation in a patient with intractable seizures, severe intellectual disability, optic atrophy, muscular hypotonia, and brain abnormalities. Epilepsia 55:e25–e29

Brenner S (1974) The genetics of Caenorhabditis elegans. Genetics 77:71–94

Brunklaus A, Zuberi SM (2014) Dravet syndrome–from epileptic encephalopathy to channelopathy. Epilepsia 55:979–984

Carvill GL *et al* (2015) Mutations in the GABA transporter SLC6A1 cause epilepsy with myoclonic-atonic seizures. Am J Hum Genet 96:808–815

Chahrour M, Zoghbi HY (2007) The story of Rett syndrome: from clinic to neurobiology. Neuron 56:422–437

Chan EY *et al* (2002) Targeted disruption of Huntingtin-associated protein-1 (Hap1) results in postnatal death due to depressed feeding behavior. Hum Mol Genet 11:945–959

Christodoulou J, Grimm A, Maher T, Bennetts B (2003) RettBASE: the IRSA MECP2 Variation Database—a new mutation database in evolution. Hum Mutat 21:466–472

DeAndrade MP *et al* (2012) Motor restlessness, sleep disturbances, thermal sensory alterations and elevated serum iron levels in Btbd9 mutant mice. Hum Mol Genet 21:3984–3992

Dittman JS, Kaplan JM (2008) Behavioral impact of neurotransmitter-activated G-protein-coupled receptors: muscarinic and GABAB receptors regulate Caenorhabditis elegans locomotion. J Neurosci 28:7104–7112

Durand CM *et al* (2007) Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet 39:25–27

Endele S *et al* (2010) Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat Genet 42:1021–1026

Fowler CD, Lu Q, Johnson PM, Marks MJ, Kenny PJ (2011) Habenular α5 nicotinic receptor subunit signalling controls nicotine intake. Nature 471:597–601

Grillo E *et al* (2013) Revealing the complexity of a monogenic disease: Rett syndrome exome sequencing. PLoS One 8:e56599

Hara M, Ohba C, Yamashita Y, Saitsu H, Matsumoto N, Matsuishi T (2015) De novo SHANK3 mutation causes Rett syndrome-like phenotype in a female patient. Am J Med Genet A 167:1593–1596

He L *et al* (2003) Spongiform degeneration in mahoganoid mutant mice. Science 299:710–712

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

Kemlink D *et al* (2009) Replication of restless legs syndrome loci in three European populations. J Med Genet 46:315–318

Klomp LW *et al* (2004) Characterization of mutations in ATP8B1 associated with hereditary cholestasis. Hepatology 40:27–38

Lemke JR *et al* (2014) GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. Ann Neurol 75:147–154

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 69:905–914

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

Marangi G *et al* (2012) Proposal of a clinical score for the molecular test for Pitt-Hopkins syndrome. Am J Med Genet A 158A:1604–1611

Mombereau C, Kaupmann K, Gassmann M, Bettler B, van der Putten H, Cryan JF (2005) Altered anxiety and depression-related behaviour in mice lacking GABAB(2) receptor subunits. NeuroReport 16:307–310

Monrós E, Armstrong J, Aibar E, Poo P, Canós I, Pineda M (2001) Rett syndrome in Spain: mutation analysis and clinical correlations. Brain Dev 23:S251–S253

Nava C *et al* (2014) De novo mutations in HCN1 cause early infantile epileptic encephalopathy. Nat Genet 46:640–645

Neul JL *et al* (2010) Rett syndrome: revised diagnostic criteria and nomenclature. Ann Neurol 68:944–950

Okamoto N *et al* (2015) Targeted next-generation sequencing in the diagnosis of neurodevelopmental disorders. Clin Genet 88:288–292

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

Percy AK (2008) Rett syndrome: recent research progress. J Child Neurol 23:543–549 Rett A (1966) On an unusual brain atrophy syndrome in hyperammonemia in childhood. Wien Med Wochenschr 116:723–772

Rodan LH *et al* (2016) A novel neurodevelopmental disorder associated with compound heterozygous variants in the huntingtin gene. Eur J Hum Genet. doi:10.1038/ejhg.2016.74

Roux JC, Zala D, Panayotis N, Borges-Correia A, Saudou F, Villard L (2012) Modification of Mecp2 dosage alters axonal transport through the Huntingtin/Hap1 pathway. Neurobiol Dis 45:786–795

Sáez MA *et al* (2016) Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. Genet Med 18:378–385

Sawin ER, Ranganathan R, Horvitz HR (2000) C. elegans locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. Neuron 26:619–631

Sweatt JD (2013) Pitt-Hopkins syndrome: intellectual disability due to loss of TCF4-regulated gene transcription. Exp Mol Med 45:e21

Tawarayama H, Yoshida Y, Suto F, Mitchell KJ, Fujisawa H (2010) Roles of semaphorin-6B and plexin-A2 in lamina-restricted projection of hippocampal mossy fibers. J Neurosci 30:7049–7060

Vissers LE *et al* (2010) A de novo paradigm for mental retardation. Nat Genet 42:1109–1112

Xiang J, Yan S, Li SH, Li XJ (2015) Postnatal loss of hap1 reduces hippocampal neurogenesis and causes adult depressive-like behavior in mice. PLoS Genet 11:e1005175

Zhu X *et al* (2015) Whole-exome sequencing in undiagnosed genetic disorders: interpreting 119 trios. Genet Med 17:774–781

INVENTORY OF SUPPLEMENTARY DATA

SUPPLEMENTARY TABLE 1

SUPPLEMENTARY TABLE 2

Supplementary Table 1 Technical statistics of WES. Overall coverage statistics for each individual of the families, considering the regions captures using Exome Enrichment Kit, and number of reads in the position of the variation.

Gene	IDs	Average of coverage	Numer of reads	Variant: genomic coordinates
AGAP6	**Patient 3**	51,7222	30	10:51748528
	Mother	19,1904	59	
	Father	19,8947	49	
HCN1	**Patient 4**	37,9988	33	5:45396665
	Mother	21,2688	11	
	Father	53,6885	22	
SCN1A	**Patient 8**	63,9124	98	2:166866266
	Mother	55,4485	98	
	Father	66,9241	91	
MGRN1	**Patient 8**	63,9124	4	16:4723583
	Mother	55,4485	1	
	Father	66,9241	3	
BTBD9	**Patient 8**	63,9124	100	6:38256093
	Mother	55,4485	101	
	Father	66,9241	105	
TCF4	**Patient 10**	57,353	44	18:52901827
	Mother	45,2247	87	
	Father	58,4836	77	
SEMA6B	**Patient 11**	38,1451	21	19:4555540
	Mother	43,3847	16	
	Father	31,635	31	
GRIN2B	**Patient 11**	38,1451	45	12:13764782
	Mother	43,3847	44	
	Father	31,635	41	
VASH2	**Patient 12**	56,4905	84	1:213161902
	Mother	68,9685	84	
	Father	63,1402	109	
CHRNA5	**Patient 13**	63,6187	53	15:78882481
	Mother	65,3851	64	
	Father	58,9521	67	
ZNF620	**Patient 14**	55,267	27	3:40557941
	Mother	57,7491	34	

	Father	60,002	28	
GRAMD1A	**Patient 14**	55,267	56	19:35506764
	Mother	57,7491	42	
	Father	60,002	43	
NOC3L	**Patient 14**	55,267	26	10:96097586
	Mother	57,7491	52	
	Father	60,002	36	
GPATCH2	**Patient 14**	55,267	96	1:217784371
	Mother	57,7491	111	
	Father	60,002	104	
SLC6A1	**Patient 17**	62,2434	43	3:11067528
	Mother	56,8694	53	
	Father	61,0401	66	
GABBR2	**Patient 19**	67,6492	46	9:101133817
	Mother	35,6295	30	
	Father	65,7731	65	
ATP8B1	**Patient 19**	67,6492	61	18:55328507
	Mother	35,6295	38	
	Father	65,7731	70	
HAP1	**Patient 20**	57,3956	11	17:39890655
	Mother	63,086	19	
	Father	52,3069	15	
PDLIM7	**Patient 21**	63,5277	56	5:176910933
	Mother	59,2951	28	
	Father	58,6697	20	
SRRM3	**Patient 21**	63,5277	52	7:75890878
	Mother	59,2951	59	
	Father	58,6697	53	
ANKRD31	**Patient 22**	58,6697	75	5:74518166
	Mother	56,1542	66	
	Father	57,2364	65	
CACNA1I	**Patient 23**	20,1538	32	22:40066855
	Mother	9,48446	9	
	Father	56,6495	75	

Supplementary Table 2 Primers used for Sanger sequencing

Gene	Primers	Sequence	Genomic position (GRCh37/hg19)
AGAP6	F	AGCGGGAAGACCATCTCTG	chr10:51748364-51748795
	R	GAAAGGAGCTCGAAGTGTGG	
HCN1	F	CAGCAGACTGTTTCCACTTCA	chr5:45396198-45397021
	R	CATGCACAATAGCTGCCTGT	
SCN1A	F	TTTTGTGTGTGCAGGTTTCATT	chr2:166866053-166866365
	R	AGGCCTATTTCTCTTGCATATCA	
MGRN1	F	CTGGATTTGAGCCTGGTGAT	chr16:4723322+4723803
	R	CCCACGTTCAGCACAGACTA	
BTBD9	F	TCCTGATGCCAAATCTTGTT	chr6:38255873-38256270
	R	GCACGCTATATCTCGTTGTTG	
TCF4	F	TCAGCGCCCTCTAGTGAAAC	chr18:52901609-52902050
	R	TTAGCGGGCGAAGTTCTAAA	
SEMA6B	F	TGGCCAAGGTCACACAGTAA	chr19:4555127-4555699
	R	AAGGCAGGCAAGAGATGAG	
GRIN2B	F	TACAATCTAACCTAGGCCCTGG	chr12:13764558-13764935
	R	TGGATATGCTAGGGAAAATGCAG	
VASH2	F	GCAAGGTTCAAGAGTACTGGGT	chr1:213161536+213162249
	R	TGGTGAGGCATAATGTTCAAAGC	
CHRNA5	F	GAGCAGGGTCCCTATGTAGC	chr15:78881983+78882780
	R	CGCCATGGCATTATGTGTTGA	
ZNF620	F	TAGCGTCAGCACACAACTCA	chr3:40557616+40558235
	R	GCTGGTGCTGAATCAGGGT	
GRAMD1A	F	CTCACCCCTGAACCAATTGC	chr19:35506400+35507085
	R	AGAAGGAGAAACTGAGGCACA	
NOC3L	F	TGTAGAAAATAGAAGTGGCAGGT	chr10:96097185-96097950
	R	CACATGAAGCACCTATAGCCA	
GPATCH2	F	TGCTGGCAGTTCTTAGAGTCT	chr1:217784024-217784774
	R	TCAATGAGCCTAGCAAGAAAGC	
SLC6A1	F	CTGTCTGACTCCGAGGTGAG	chr3:11067255+11067936
	R	GACGATGATGGAGTCCCTGA	
GABBR2	F	GTGACCTGGGTCTGGTAAGTG	chr9:101133405-101134056
	R	TTCCTTTGACAAGGTCCCCAG	
ATP8B1	F	AATCTTGGGAATGGTACTCCTGG	chr18:55328163-55328763
	R	ACCTTATTTTCCTCTTCGCATCC	
HAP1	F	CATCCGGAACTTGCACTCG	chr17:39890418-39891088
	R	CTTCCTCCAGCTCCCGAATA	
PDLIM7	F	GTGTTCCCGTGACCCAGG	chr5:176910565-176911259
	R	AGCCCTACCCAGAAATGCAG	
SRRM3	F	GTCACTCTGTACAAGGGACCT	chr7:75890525+75891203
	R	CTGCTTGTCTAACTGGCACC	
ANKRD31	F	GGACGCATCAATAGGTGCAG	chr5:74517744-74518408
	R	GCTTCCAGTCAACAGTAGGC	
CACNA1I	F	CCACTGCCAACCTGAGTGA	chr22:40066439+40067235
	R	CACAGTCATTGCCACCCATG	

F, Forward; R, Reverse.

RESULTS SYNTHESIS

STUDY I

Mutations in JMJD1C are involved in Rett syndrome and intellectual disability

- We analyzed the presence of SNPs and indels in *JMJD1C* by sequencing using a GS Junior system with an amplicon library prepared with the Fluidigm Access Array. The sequencing strategy identified 10 synonymous variants, 7 previously informed *JMJD1C* variants and, most importantly, 7 different heterozygous non-synonymous missense changes in the exonic regions of *JMJD1C* that were not present in the dbSNP, 1000 Genomes Project, or Exome Variant Server databases, neither in 500 healthy control volunteers. Among these latter 7 missense changes, 2 of them occurred *de novo* in one woman with a clinical diagnosis of RTT (c. 488C>T; Pro163Leu), and in a boy with intellectual disability (c.3559A>G; Thr1187Ala). Parental samples were not available from the other 5 patients, consequently it could not be determined whether an unaffected parent was a carrier or the mutation occurred *de novo* in those patients.
- We detected no large genetic rearrangements at the *JMJD1C* gene locus using multiplex ligation-dependent probe amplification approach.
- Focusing on the functional effects of the identified JMJD1C-Pro163-Leu on intracellular and subcellular localization, we showed that endogenous JMJD1C, as well as transfected JMJD1C-WT-GFP in stable clones, was mainly cytoplasmic in HEK293 cells, conversely clones expressing the mutant JMJD1C-Pro163Leu-GFP form manifested a strong nuclear mark. Likewise, we found that endogenous and transfected JMJD1C protein were almost absent from the chromatin fraction, whereas the mutant JMJD1C-Pro163Leu was markedly enriched in the chromatin.
- We demonstrated that the mutant JMJD1C-Pro163Leu protein is less efficient in demethylating the non-histone target of JMJD1C, MDC1, compared to wild-type form of JMJD1C.
- We observed the molecular interaction between JMJD1C and MeCP2 using immunoprecipitation assay. The JMJD1C-Pro163Leu mutant protein cannot efficiently bind to MeCP2, suggesting that this finding could explain the role of JMJD1C in RTT.

- Regarding cellular effects on neuronal system, we observed that depletion of JMJD1C in primary neuronal culture leads to a significant reduction in the complexity of dendritic process, which is a common hallmark in RTT.

STUDY II

Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype

- On average, WES gave rise to 419,045 variants, including SNVs and indels, of which 19,951 non-synonymous variants per family (4.7 %) were predicted to have a functional impact on the genomic sequence. To select previously undescribed variants in the healthy population, we filtered out the variants with an allele frequency of 1 % or higher formerly observed in dbSNP and the 1000 Genomes Pilot Project. Subsequently, to focus on *de novo* inheritance, patients' variants were filtered first against variants found in their own parents and then against a pool of controls comprising all the healthy parents included in the study. Following this process, we achieved an average of 106 SNVs per family, which corresponded to 81 mutated genes per family.
- The global yield of genomic analysis allowed the identification of 22 coding *de novo* mutations in 66.7 % (14 of 21) of Rett-like patients: 20 SNVs and 2 indels. Contrarily, in seven (33.3 %) of the studied RTT probands, WES did not detect any genetic change relative to their respective parents.
- Of the 22 identified coding *de novo* mutations, five (22.7 %) occurred in genes previously associated with neurodevelopmental disorders with a clinicopathological phenotype strongly overlapping RTT. Particularly, we identified 4 mutations in *HCN1* and *GRIN2B*, which are associated with early infantile epileptic encephalopathy; *SLC6A1*, which is associated with epilepsy and myoclonic-atonic seizures; *TCF4*, which is associated with Pitt–Hopkins syndrome; and *SCN1A*, which is associated with Dravet syndrome.
- Of the 22 identified coding *de novo* variants, 17 (77.3 %) occurred in genes not previously associated with neurodevelopmental disorders. However, two of these variants were associated with non-neurodevelopmental disorders: a *BTBD9* variant linked to restless leg syndrome, and an *ATP8B1* SNV associated

with familial cholestasis. The other 15 potentially pathogenic variants identified occurred in genes that had not been linked to any genetic disorder of any type. These variants are: *AGAP6, MGRN1, SEMA6BA, VASH2, CHRNA5, ZNF620, GRAMD1A, NOC3L, GPATCH2, GABBR2, HAP1, PDLIM7, SRRM3, ANKRD31* and *CACNA1I*. Among them, we identified an enrichment of genes with a potential role in neuronal biology and functionality.

- Focusing on the functional implication of the identified genes in a neurological phenotype, we performed locomotion assays in *C. elegans* mutants carrying deleterious mutations in the gene *alp-1*, *unc-44*, *pag-3, unc-63, C11H1.3* and *gbb-2,* orthologs of the human genes *PDLIM7*, *ANKRD31*, *ZNF620*, *CHRNA5*, *MGRN1* and *GABBR2,* respectively. We observed that in 83.3 % (5 of 6) of the cases the mutation of the ortholog of the human WES-identified genes exhibited a locomotion defective phenotype *in vivo*. The most severe phenotypes were represented by *alp-1*, *unc-44* and *pag-3*. Slower locomotion and uncoordinated movement were also observed in the mutants of *unc-63* and *C11H1.3.* Differently, we did not find a clear locomotion defect in the *gbb-2* mutant, the ortholog of *GABBR2*.

DISCUSSION

The application of NGS technology has brought an unprecedented era of rare disease gene discovery. Disease-causing genes that had eluded discovery because of their rarity, clinical heterogeneity and the paucity of families with multiple individuals affected by a specific disease are being increasingly identified by NGS approaches. The genetic heterogeneity of NDDs, along with the fact that the causative mutations are usually *de novo,* has been the significant limiting factor for gene discovery in the past. The trio-WES approach, that entails simultaneous sequencing of the proband and parents, enables the detection of all coding de *novo* mutations in an individual and has rapidly facilitated gene discovery. The trio-WES approach to identify *de novo* mutations is very efficient because, on average, each individual carries from one to two *de novo* sequence variants per exome (Frank, 2010). Many of the genes with *de novo* changes identified in a cohort-based WES studies are plausible candidates based on gene function and/or evolutionary conservation. Hence, a major challenge comes with determining which *de novo* change is really pathogenic. Among these new candidate genes, supportive role for the Jumonji Domain Containing 1C (*JMJD1C*) histone demethylase in ASD and ID has been fostered by both positional cloning strategies (Castermans *et al*., 2007) and exome-sequencing studies (Neale *et al*., 2012; Iossifov *et al*., 2014), in which four *de novo* variants have been identified.

RTT is defined as a NDD characterized by normal early-development, followed by psychomotor regression with loss of the acquired motor and language skills. Patients present with stereotypic hand movements, acquired microcephaly, autistic-like behaviors, irregular breathing with hyperventilation, and seizures (Neul *et al*., 2010). A mutation in the gene of *MECP2* was identified as the primary cause of the disease in 1999 by Amir *et al*. Approximately 90% of RTT patients with a classical form of the disease carry a mutation in *MECP2*, as well as 50% of the atypical one. Many of the atypical RTT patients, who deviate from classical clinical presentation of RTT, are mutation-negative for the mentioned gene and carry mutations in *CDKL5,* described in individuals with an early seizure onset variant, or *FOXG1,* responsible for the congenital variant (Evans *et al*., 2005; Philippe *et al*., 2010). However, there remains a subset of patients with a clinical diagnosis of RTT who are mutation-negative for all the aforementioned genes.

This thesis has been developed with the aim of identifying news genes implicated in the complex and as yet uncharacterized molecular pathways related to RTT and NDDs in general. This has a clear diagnostic value, and is also of great importance to recognize new proteins that represent new possible therapeutic targets. To this aim, we applied NGS technology to study I and II, followed by functional validation of new candidate genes. In study I, we specifically investigated the contribution of *JMJD1C* mutations in NDDs, with special regard to the well-defined clinical entities of RTT, in that cases in which the usual mutations in *MECP2*, *CDKL5* and *FOXG1* are not present. In study II, we sought to identify new candidate genes to explain the RTT-like phenotype of a large cohort of clinical cases that present clinical features strongly overlapping classic RTT, but without mutations in the *MECP2*, *CDKL5* and *FOXG1.*

STUDY I

Mutations in JMJD1C are involved in Rett syndrome and intellectual disability

JMJD1C, also known as TRIP8 or JHDM2C, was originally described as a thyroid-hormone receptor-interacting protein by Lee *et al* (Lee *et al*., 1995) and later characterized as a member of the jmJC domain-containing protein family involved in histone demethylation activity (Wolf *et al*., 2007; Kim *et al*., 2010; Watanabe *et al*., 2007). A first association between JMJD1C and intellectual disability was realized by Castermans *et al*. (2007), who reported the identification of a *de novo* balanced paracentric inversion 46,XY,inv(10) with a breakpoint in chromosome 10q21.3 located within the first intron of the JMJD1C gene in a boy with autism. Additional support for a role of JMJD1C in autism has been further fostered by large-scale exome sequencing studies in which three *de novo* variants have been identified (Neale *et al*., 2012; Iossifov *et al*., 2014), at least one with a loss of function mutation (Iossifov *et al*., 2014).

In this study, we performed a *JMJD1C* mutational analysis in an unrelated cohort of 215 patients, including 69 with ASD, 85 with ID and 61 females with a clinical diagnosis of RTT, without mutations in *MECP2*, *CDKL5* and *FOXG1*. In 7 unrelated patients with neurodevelopmental disability disorders, we identified 7 different heterozygous non-synonymous missense changes in the exonic regions of *JMJD1C* that were not present in the dbSNP, 1000 Genomes Project, or Exome Variant Server databases, neither in

500 healthy control volunteers. Among them, Sanger sequencing of the parents of 2 of the patients confirmed that the mutations occurred *de novo* in one woman with a clinical diagnosis of RTT, carrying a Pro163Leu variant, and a boy with intellectual disability, with a Thr1187Ala variant. Parental samples were not available from the other 5 patients, so it could not be determined whether an unaffected parent was a carrier or the mutation occurred *de novo* in those patients. Using i*n vitro* functional expression studies in HEK293 cells, we showed that the JMJD1C-Pro163Leu variant protein had an abnormal subcellular localization and was enriched in the chromatin fraction compared to JMJD1C wildtype form. The variant protein also had a decreased activity in demethylating the DNA damage-response protein MDC1 and, importantly, reduced binding to MeCP2. In addition, siRNA-mediated knockdown of *JMJD1C* in mouse primary neuronal hippocampal cells resulted in a significant reduction in the complexity and branching of dendritic processes. These findings could explain the role of JMJD1C in RTT, a disease mainly associated with *de novo* mutations in *MECP2* and characterized by alterations in the numerical density, size, and shape of dendritic spine arborization (Schüle *et al*., 2008; Xu *et al*., 2014).

It is interesting to note that many of the genes mutated in neurological disorders have an epigenetic component (Sanchez-Mut *et al*., 2012; Urdinguio *et al*., 2009) and *JMJD1C* can now be included in this growing list. Epigenetic regulation of gene expression involves two major mechanisms, DNA methylation at the CpG islands of gene promoters and chromatin remodeling. Chromatin folding and its switch from heterochromatin, transcriptionally silent, to euchromatin, transcriptionally active, and vice versa, are finely tuned by three main categories of proteins: writers, erasers and readers. The first ones are represented by enzymes that add chemical marks to histone substrates, such as histone acetyltransferases and histone methyltransferases; the second ones are enzymes that remove this chemical covalent modification from histone substrates, such as histone deacetylases and histone demethylases; the third ones are proteins that react to specific modified histone residues at epigenetic code level. The category of erasers appears to represent a fundamental piece of equipment in the cells chromatin toolbox, permitting a fast and robust regulation of transcriptional programs of developmental processes. Mutations in genes encoding demethylases that regulate the methylation of lysine 4 of histone H3 have been identified in several neurodevelopmental disorders, including autism and RTT (Wynder

et al., 2010). Moreover, two mental retardation syndromes have been linked to mutations in the histone lysine demethylases KDM5C, previously known as JARID1C (Jensen *et al*., 2005), and PHF8, also known as JHDM1F (Laumonnier *et al*., 2005). Taken together, these examples indicate the importance of post-translational modifications of histone N-terminal tails in the genetic origin of neurodevelopmental disorders, as well as the possibility that several other members of the large histone demethylase family are involved in the development of such disorders. In this regard, although further research is required to understand the molecular pathway that involve JMJD1C, our work highlight the increasing contribution of the genetic disruption of epigenetic genes to human neurodevelopmental disorders.

Finally, the involvement of histone demethylases in essential cellular processes and their implication in human diseases enables these enzymes to constitute attractive drug targets. Recently, several studies have reported the inhibitory effect of various small molecules on the histone demethylase LSD1 (Culhane *et al*., 2006; Lee *et al*., 2006; Yang *et al*., 2006; Huang *et al*., 2007). Significantly, some of these molecules were shown to reactivate transcription of otherwise silenced tumor suppressor genes, raising hopes for the success of such compounds for cancer therapies (Huang *et al*., 2007). Correspondingly, inhibitors targeting specific members of the JmjC group of histone demethylases may be envisioned with potential for treatment and are likely to be explored in the context of neurological disease in the not too distant future (Peter *et al*., 2011).

STUDY II

Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype

In this study, we applied WES to a cohort of 19 trios and one family with two affected daughters to genetically characterize 21 RTT-like cases in the context of wild-type sequences for the standard RTT genes. The identification of 22 coding *de novo* mutations in 14 of the 21 RTT-like patients resulted in a global genomic yield of 66.7%, which enabled both the revaluation of a correct clinical diagnosis and the detection of new potential candidates in a RTT-like context.

In 9 of 21 (42.85%) patients, we identified 17 novel variants in genes that had not been previously associated with a neurodevelopmental disorder and represent good disease-causing candidate genes (Table 5). Our findings indicate that the existence of *de novo* variants in genes with potential neurological functionalities, such as neuronal receptors (*GABBR2* and *CHRNA5*), axon guiders (*SEMA6B*), synaptic ionic channels (*CACNA1I*) and others, contribute to the development of RTT-like clinical phenotypes. These patients share most of the clinicopathological features of classic RTT syndrome, such as stereotypic hand movements, relative microcephaly, and onset of the disease after the age of 12 months (Table 6). Thus, exome sequencing is a powerful tool for genetically characterizing these enigmatic cases. Earlier studies have employed WES as a strategy to analyze RTT-like cases in a smaller series of patients without mutations in the usual RTT genes. All these works reinforce the concept that a mutational heterogeneous profile hitting shared neurological signaling pathways contributes to RTT-like syndromes involving different neuronal types in various brain regions. Examples of confluence in the same molecular crossroads include the gamma-aminobutyric type B receptor subunit 2 (*GABBR2*) *de novo* variant, described here, and the formerly identified variant in the gamma-aminobutyric acid receptor delta gene (*GABRD*) (Hara *et al*., 2015). Interestingly, a second RTT-like patient has been identified as being a carrier of a *de novo GABBR2* variant (Lopes *et al*., 2016), highlighting the likelihood that this gene and pathway contribute to a Rett-like clinical entity. GABAergic neurotransmission is the most abundant inhibitory signaling pathway in the central nervous system of vertebrates and an important balance between inhibitory and excitatory stimuli exists in human brain for a proper brain function (Petroff, 2002). The GABAaergic pathway has been largely described as altered in RTT. Indeed, experimental evidences have shown that MeCP2 is highly expressed in GABAergic and glutamatergic neurons in primate brain and that the mentioned neuronal pathways are spatially and developmentally affected in the brain of MeCP2-deficient mice (Akbarian *et al*., 2001; El-Khoury *et al*., 2014). In line with the existence of an imbalance between excitation and inhibition in Mecp2-deficient mice and a role for GABA in RTT neuropathology, it has been described how the selective removal of Mecp2 in GABAergic neurons in *Viaat-Mecp2* mice lead to the development of Rett-like features including repetitive stereotypic hindlimb clasping, progressive motor dysfunction, impaired social interaction, decreased memory skills, altered respiratory

pattern and neuronal excitability (Chao *et al*., 2010). GABA is also an important neurotransmitter in *C. elegans*; however, in contrast to vertebrates where GABA acts at synapses of the central nervous system, in nematodes GABA acts primarily at neuromuscular synapses. Specifically, GABA acts to relax the body muscles during locomotion and foraging and to contract the enteric muscles during defecation (Jorgensen, 2005). Surprisingly, despite the role that GABAergic system plays in motor control, we did not observe any deficit in our locomotion assays performed in the *C. elegans* mutant of the GABBR2 ortholog, *gbb-2*. However, it occurs in the *gbb-1;gbb-2* double mutant (Dittman and Kaplan, 2008), being *gbb-1* the *C. elegans* ortholog of *GABBR1* (gamma-aminobutyric acid type B receptor subunit 1). Another example of similarly targeted genes in RTT-like patients is that of the proteins containing ankyrin-repeats that are involved in postsynaptic density (Durand *et al*., 2007). This study has revealed two *de novo* variants in the ankyrin repeat containing proteins *AGAP6* and *ANKRD31* in RTT-like patients. The presence of *de novo* variants of the SH3 and multiple ankyrin repeat domain 3 protein (*SHANK3*) (Hara *et al*. 2015) and ankyrin-3 (*ANK3*) (Grillo *et al*., 2013) has been previously reported in two RTT-like patients. Our functional studies in *C. elegans* suggest a further contribution to the involvement of the ankyrin repeat containing proteins in a possible RTT-like phenotype, as the mutant *unc-44*, ortholog for *ANKRD31,* resulted significantly locomotion defective. A final example of the convergence of cellular pathways to provide a common RTT-like phenotype is represented by the disruption of the ionic channels. We found the existence of a voltage-gated calcium channel subunit alpha 11 (*CANA1I*) *de novo* variant in a RTT-like patient. Additionally, the presence of *de novo* variants in the calcium release channel *RYR1* (Grillo *et al*. 2013) and the sodium voltage-gated channel alpha subunit 2 (*SCN2A*) (Baasch *et al*., 2014) in two other RTT-like probands have been previously reported. It is also intriguing that in our study a variant in *HAP1* was found, whereas in similar series heterozygous variants in huntingtin (*HTT*) have been described (Lopes *et al*. 2016; Rodan *et al*., 2016), further reinforcing the links between Huntington's disease and RTT (Roux *et al*., 2012). *HTT* encodes a protein that directly interacts with MeCP2 (McFarland *et al*., 2014) and MeCP2-deficient mice have reduced expression of HTT in the entire brain, leading to a defect in the axonal transport of BDNF (Roux *et al*., 2012). RTT and Huntington disease seem to share features at different levels: at a molecular level, due to nuclear interaction among MeCP2 and HTT

for transcriptional regulation and axonal trafficking through BDNF (McFarland *et al*., 2014); at a neuropathological level, considering the existence of striatum atrophy in both disorders (Reiss *et al*., 1993); and at a clinical level, because of the presentation of compulsive movement disorder plus cognitive dysfunction. Noteworthy, we also detected a novel variant in the neuronal acetylcholine receptor subunit alpha-5 (*CHRNA5*), which had not been previously found in a RTT-like patient. Accumulated evidences suggested that dysfunction of cholinergic system may underlie autism-related behavioral symptoms and the etiology of anxiety-related behaviors (Sarter *et al*., 2005*;* Picciotto *et al*., 2012). Moreover, a previous study showed that cholinergic hypofunction in MeCP2 null mice was associated with milder phenotype that could be improved by post-natal choline supplementation (Nag *et al*., 2008). More recently, it has been demonstrated that conditional deletion of MeCP2 from cholinergic neurons in *Chat-Mecp2* mice resulted in several RTT-like phenotypes such as impaired social interaction, altered anxiety-related behavior and epilepsy susceptibility, and that reactivation of MeCP2 in basal forebrain cholinergic neurons improved phenotypic severity of *Chat-Mecp2* mice. These evidences suggest the involvement of the cholinergic systems in the pathogenesis of RTT. Our functional studies showed that mutant *unc-63*, ortholog for *CHRNA5*, is significantly defective for locomotion. These results are consistent with the excitatory role of acetylcholine in the nematodes locomotion and the opposite inhibitory function of GABA. Indeed, the release of acetylcholine to the ventral dorsal body muscles leads to the contraction of the body wall muscle on one side and stimulates GABA release onto muscles at the opposite side. This stimulation and the contralateral inhibition cause the worm body to bend and lead to a coordinated locomotion pattern (Jorgensen, 2005).

We successfully provided a clinical diagnosis in 5 of the 21 cases (23.80%), in which five variants in genes previously associated with NDDs have been detected (Table 2). Specifically, we identified four mutations in genes such as *HCN1* (Nava *et al*., 2014) and *GRIN2B* (Endele *et al*., 2010; Lemke *et al*., 2014), which are associated with early infantile epileptic encephalopathy; *SLC6A1*, which is associated with epilepsy and myoclonic-atonic seizures (Carvill *et al*., 2015); *TCF4*, which is associated with Pitt–Hopkins syndrome (Sweatt, 2013); and *SCN1A*, which is associated with Dravet syndrome (Brunklaus and Zuberi, 2014) (Table 2). When analyzing the phenotypes associated with genes previously implicated in NDDs, they overlap with those observed

in our series of RTT-like patients (Table 3 and 4). Indeed, there seems to be a spectrum of clinical presentation that enables for the delineation of a shared core phenotype that can make differential diagnosis among RTT and other NDDs often difficult and challenging. Although to varying degrees, clinical features such as epilepsy, microcephaly, stereotypies and hypotonia, as well as language and respiratory dysfunctions and preserved use of hands, appear to be shared among RTT, Pitt-Hopkins, early infantile epileptic encephalopathy, epilepsy and myoclonic-atonic seizures. For instance, Pitt-Hopkins syndrome is clinically characterized by the association of intellectual deficit, characteristic facial dysmorphism and problems of abnormal and irregular breathing (Pitt and Hopkins, 1978). Patients also suffer from psychomotor disorders, hypotonia, late-acquired unstable walking, absence of language, microcephaly, postnatal underdevelopment, as well as tonic-clonic epileptic attacks (Zweier *et al*., 2007). Due to the clinical similitudes between Pitt-Hopkins and RTT, the criteria to differentiate these two diseases and the inclusion of the mentioned genes in panels for genetic screenings has been discussed in the past years (Armani *et al*., 2012; Marangi *et al*., 2012). The purely clinical classification of these patients, without a genetic study, can be difficult because some of these patients are composites that carry at least two pathogenic variants. For example, among our cases, the Dravet syndrome patient also had a *de novo* variant in *BTBD9,* associated with the development of restless leg syndrome. Although this must be proved, we speculate that a possible explanation for the differential phenotype of this patient compared to the canonical signs of Dravet is that additional mutations could add a possible phenotype to this individual (table 4). In addition, among the newly identified candidate genes associated with RTT-like features, a few of these patients simultaneously carried two *de novo* variants (e.g., probands 8, 19 and 21), further complicating the tasks of correctly diagnosing and managing these individuals. In this context, WES enables a correct and robust diagnosis that can significantly improve patient management, even when no new treatments results from the diagnosis. The potential effectiveness of such genotype-driven treatments is often deducible from the assayed biological consequence of the disease-causing mutation and known mechanisms of action of candidate drugs, and is further testable from the behavior of mutant proteins in *in vivo* assays in presence or absence of the candidate drugs.

With the advent of NGS, and in particular of WES, the dramatic increase in the pace of rare gene discovery has led to the necessity of understanding the molecular and cellular bases of rare diseases. This brings a real and immediate need for model organism research platforms to put potentially disease-causing genes into a biological context to understand the underlying molecular pathways. In a simpler perspective, human neuronal cell lines can be an interesting model to study the function of genes identified as associated with human neurodevelopmental disorders, given the presence of the major neuronal molecular components. However, lack of integration among functional circuits and scarsity of interaction with other cell types, as well as absence of behavioral output, lead to the need to use a whole organism approach. Mice have been used for this purpose with very encouraging results. The structure of human and murine nervous systems bears significant resemblance and behavioral paradigms have been developed to analyze traits that are thought to be parallel between these two spices. Considering the greater evolutionary proximity to humans, rat models are even more advanvantageous, particularly in cognitive and social studies, but the tools for genetic manipulation have lagged behind. Globally, disadvantages of the use of rodents consist in their relatively high maintenance costs and difficulty and time consumption of their genetic manipulation (Bryda, 2013; Vandamme, 2014). Moreover, the complex structure of their nervous system, which is certainly advantageous for some studies, presents serious limitations when trying to dissect molecular events underlying disease. In this scenario, organisms with simpler nervous systems and genetic amenability represent an elegant and useful tool for the study of neuronal gene function. *Drosophila melanogaster* is a model organism that has been increasingly used in in genetic, biochemical, anatomical, physiological and behavioral studies of neurodevelopmental disorders (Gatto and Broadie, 2011). It has a brain-like structure and complex behaviors that can be analyzed through the use of highly developed genetic tools. Zebrafish is another simple model organism widely used for the study of neurological disorders, with the main advantages of being a vertebrate with a brain and a high degree of genetic homology with mammals (Kabashi *et al*., 2010). Finally, for very simple functional studies, yeast and fungi can also be used because of their simplicity and ease of genetic manipulation but with clear limitations for the study of neuronal gene function. *C.elegans* provides a good compromise among complexity of vertebrates such as mouse and extreme simplicity

of yeast. Additionally, this model organism presents key advantages that make it unique in the field of neuroscience. First, the well-described neuronal lineage and interconnectivity offer a set up for the study of neuronal mechanisms. Second, amenability to genetic manipulation allows the identification of genes important for neuronal functions. Third, its transparency in combinations with the existence of *in vivo* monitoring techniques enables the correlation of patterns of neuronal activity with behavioral outcomes. Our studies performed in *C. elegans* validate the functional relevance for nervous system function of the newly proposed candidate genes. We studied the *C. elegans* mutants carrying deleterious mutations in the gene orthologs of the human genes *PDLIM7*, *ANKRD31*, *ZNF620*, *CHRNA5*, *MGRN1* and *GABBR2*. Considering that the loss of normal movement and coordination is one of the clearest signs shown by Rett patients, we performed a locomotion assay of the nematodes, using the wild-type N2 strain as a control. We observed that in 83.3 % (5 of 6) of the cases the mutation of the ortholog of the human exome sequencing identified genes in *C. elegans* exhibited a locomotion defective phenotype (Fig.2). Future studies would be necessary to assess the role of the specific variants identified, such as rescuing the defects with the expression of normal cDNAs versus cDNAs containing the mutation, ideally using cDNAs of human origin to prove similar function of the gene in the two species. It is also relevant to mention that for some of the newly reported mutated genes in our RTT-like patients, there are mice models targeting the described loci that show neurological phenotypes such as *BTBD9* (motor restlessness and sleep disturbances) (DeAndrade *et al*., 2012), *MGRN1* (spongiform neurodegeneration) (He *et al*., 2003), *SEMA6B* (aberrant mossy fibers) (Tawarayama *et al*., 2010), *CHRNA5* (alterations in the habenulo-interpeduncular pathway) (Fowler *et al*., 2011), *GABBR2* (anxiety and depression-related behavior) (Mombereau *et al*., 2005) and *HAP1* (depressive-like behavior and reduced hippocampal neurogenesis) (Chan *et al*., 2002; Xiang *et al*., 2015).

Discovery of variants that underlie both Mendelian and complex traits will naturally lead to a much deeper understanding of the implicated molecular pathways and general disease mechanisms that should, in turn, facilitate development of improved diagnostics, prevention strategies and targeted therapeutics (Blakemore *et al*., 2010*;* Dietz, 2010). Once a new candidate gene has been identified, it is now possible to design specific sequencing strategies for the molecular screening of this particular

target in larger populations of patients with intellectual disability. Considering the strong phenotypic variability and the absence of described biomarkers for the diagnosis of RTT, it is clear that molecular genetics needs to be used to confirm clinical diagnosis. Moreover, after a reliable validation for a disease-causing effect of the newly identified mutations, the possibility of a prenatal diagnostic procedure could be offered to all cases of pregnancies in families with a previously diagnosed RTT patient and an identified disease-causing mutation. Severity of RTT clinical manifestations has devastating implications both for the patient and the family. Parental suffering is high, and many couples with a first affected daughter decide not to have more children to avoid recurrence. Offering the parents prenatal detection of the daughter's mutation has a great calming effect on their anxiety and encourages them to have more children. Recurrence within families is rare in RTT, comprising approximately 1% of the cases (Schanen *et al.*, 1997). Genetic analysis of several reported families cases has shown that recurrence can be due to vertical dominant transmission from asyntomatic carrier females or to germline mosaicism in one of the parents (Amir *et al*., 1999; Wan *et al*., 1999; Villard *et al*., 2000; Venâncio *et al.*, 2007). With the aim of giving a proper genetic counseling to the family, the identification of a mutation in a RTT patient should be followed by DNA analysis of the mother and other healthy siblings. The analysis of the mother will test the presence of a non-penetrant constitutive mutation with skewing of X-inactivation towards the wild-type allele. A negative result in the detection of a mutation in maternal DNA extracted from blood samples does not exclude the risk for recurrence within the family. Indeed, recurrence would still exist due to germline mosaicism for the mutation. Thus, the option to test future pregnancies should always be offered to the parents of a child with RTT and an identified disease-causing mutation. The experimental procedure would consist of direct detection of the patient mutation in fetal DNA sample. A recurrence risk equal to the prevalence of RTT, approximately 1:12.000, can be attributed to the family when the fetus is not a carrier of the mutation identified in the sibling (Armstrong *et al.*, 2002). Finally, the involvement of a specific molecular pathway in a RTT-like context enables the development and testing of small molecules as therapeutic agents, such as agonists or antagonists of specific receptors or activator and inhibitor of protein with an enzymatic activity. For instance, the application of PNU282987, a drug that acts as a potent and selective agonist for the α7 subtype of neural nicotinic

acetylcholine receptors, or nicotine significantly rescued the seizure susceptibility in mice with a conditional deletion of *MECP2* in cholinergic neurons (Zhang *et al*., 2016).

In those cases in which WES was uninformative, an integrated analysis using WGS and RNA sequencing transcriptome profiling in peripheral blood lymphocytes can provide a more comprehensive knowledge that would include also non-coding parts of the genome as well as regulatory regions of main genes involved in RTT. The incorporation of peripheral blood transcriptomic data would provide an important additive value to the understanding of transcriptomic consequences of rare genetic variations and to the identification of molecular biomarkers of RTT. Despite blood is not the ideal target tissue to study the RTT transcriptome, it is commonly used in RTT and other neurodevelopmental disorders since it can be easily available for diagnostic testing. Additionally, transcriptomic studies suggest that between 35% and 80% of known transcripts are present in both brain and blood tissue samples (Tylee *et al.*, 2013). This approach could be of foundamental importance for 7 of the 21 RTT probands (33.3%) without exome candidates, in particular for the familial case with two affected daughters in which neither *de novo* nor recessive variants have been identified. Importantly, given to a more complete coverage of the protein-coding portion of the genome, it is well recognized that WGS can detect hundreds of potentially damaging coding SNVs per sample that would have been missed by WES (Belkadi *et al*., 2015). In a study performed by Gilissen *et al*., 50 patients with severe intellectual disability underwent trio WGS following uninformative WES and microarray analysis (Gilissen *et al*., 2014). To their surprise, 21 of the 50 (42%) received diagnoses in known ID genes through analysis for *de novo* and recessive coding mutations. Although it is unclear whether the identified coding mutations had been captured or missed in the previously uninformative WES data, this example demonstrated that WGS is slightly more efficient than WES for detecting mutations in the targeted exome. Thus, we cannot exlude that a similar approach could improve the success rate of our analysis for those cases in which WES was uninformative. However, taking advantages of these more overall data for disease gene discovery and molecular diagnostics in patients crucially depends on the development of analytical strategies to make sense of non-coding variants. Improved technical, statistical and bioinformatics methods are needed to reduce the rate of false-positive and false-negative variant calls, prioritize candidate causal variants and predict the potential functional impact for disease gene discovery

or molecular diagnostics. We can look towards a future in which NGS technology will address the neuronal circuitry underlying the behavioral symptoms and comorbidities, the cell types playing critical roles in these circuits, and common intercellular signaling pathways that link diverse genes implicated in NDDs and RTT. A more comprehensive knowledge of genetic and molecular basis of RTT will help translational research to achieve safe and effective therapies.

CONCLUSIONS

Based on the findings of this PhD thesis, we can conclude:

STUDY I

Mutations in JMJD1C are involved in Rett syndrome and intellectual disability

- Mutations in *JMJD1C* are associated with the development of RTT and intellectual disability.
- In the presence of Pro163Leu mutation, JMJD1C has a predominantly nuclear localization with a markedly enrichment in the chromatin, when compared to wildtype JMJD1C form.
- In the presence of Pro163Leu mutation, demethylating activity of the JMJD1C toward its non-histone target MDC1 is altered, when compared to wildtype JMJD1C form.
- Interaction between mutated JMJD1C-Pro163Leu form and MeCP2 is impaired, when compared to wildtype JMJD1C form.
- Depletion of JMJD1C in neurons dramatically drives to a reduction in the complexity of dendritic branching *in vitro*.

STUDY II

Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype

- WES is a powerful tool for the identification of novel *de novo* variants potentially associated with RTT in enigmatic cases without mutations in the usual RTT genes.
- WES is a useful tool to achieve a differential diagnosis among NDDs with overlapping clinical features.
- The presence of *de novo* variants in genes with potential neurological functionalities, such as neuronal receptors (*GABBR2* and *CHRNA5*), axon guiders (*SEMA6B*), synaptic ionic channels (*CACNA1I*), contribute to the development of RTT-like clinical phenotypes in the context of wild-type sequences for standard RTT genes.
- A mutational heterogeneous profile hitting shared neurological signaling pathways contributes to the development of RTT-like syndromes.

- *C. elegans* is a convenient and practical *in vivo* system to validate the functional relevance for nervous system function of the newly proposed candidate genes.

REFERENCES

1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).

Acampa, M. & Guideri, F. Cardiac disease and Rett syndrome. *Arch. Dis. Child.* **91,** 440–443 (2006).

Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (2010).

Ahringer, J. Reverse genetics. *WormBook* (2006). doi:10.1895/wormbook.1.47.1

Ajay, S. S., Parker, S. C. J., Abaan, H. O., Fajardo, K. V. F. & Margulies, E. H. Accurate and comprehensive sequencing of personal genomes. *Genome Res.* **21,** 1498–1505 (2011).

Akbarian, S. *et al.* Expression pattern of the Rett syndrome gene MeCP2 in primate prefrontal cortex. *Neurobiol. Dis.* **8,** 784–791 (2001).

Albert, T. J. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nat. Methods* **4,** 903–905 (2007).

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th ed.). Washington, DC: American Psychiatric Association (1994).

Amir, R. E. *et al.* Influence of mutation type and X chromosome inactivation on Rett syndrome phenotypes. *Ann. Neurol.* **47,** 670–679 (2000).

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).

Ansorge, W., Sproat, B., Stegemann, J., Schwager, C. & Zenke, M. Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.* **15,** 4593–4602 (1987).

Antonarakis, S. E. & Beckmann, J. S. Mendelian disorders deserve more attention. *Nat. Rev. Genet.* **7,** 277–282 (2006).

Applegarth, D. A. & Toone, J. R. Glycine encephalopathy (nonketotic hyperglycinaemia) : review and update. *J. Inherit. Metab. Dis.* **27,**417–422 (2004).

Applegarth, D. A. & Toone, J. R. Nonketotic hyperglycinemia (glycine encephalopathy): laboratory diagnosis. *Mol. Genet. Metab.* **74,** 139–146 (2001).

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).

Ariani, F. *et al.* FOXG1 is responsible for the congenital variant of Rett syndrome. *Am. J. Hum. Genet.* **83,** 89–93 (2008).

Armani, R. *et al.* Transcription factor 4 and myocyte enhancer factor 2C mutations are not common causes of Rett syndrome. *Am. J. Med. Genet. A* **158A,** 713–719 (2012).

Armstrong, D. D. Neuropathology of Rett syndrome. *J. Child Neurol.***20,** 747–753 (2005).

Armstrong, J. *et al.* Prenatal diagnosis in Rett syndrome. *Fetal. Diagn. Ther.* **17,** 200–204 (2002).

Awadalla, P. *et al.* Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am. J. Hum. Genet.* **87,** 316–324 (2010).

Baasch, A.-L. *et al.* Exome sequencing identifies a de novo SCN2A mutation in a patient with intractable seizures, severe intellectual disability, optic atrophy, muscular hypotonia, and brain abnormalities. *Epilepsia* **55,** e25–29 (2014).

Bailey, A. *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* **25,** 63–77 (1995).

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

Bargmann, C. I. Neurobiology of the Caenorhabditis elegans genome. *Science* **282,** 2028–2033 (1998).

Barr, M. M. & Sternberg, P. W. A polycystic kidney-disease gene homologue required for male mating behaviour in C. elegans. *Nature* **401,** 386–389 (1999).

Barr, M. M. *et al.* The Caenorhabditis elegans autosomal dominant polycystic kidney disease gene homologs lov-1 and pkd-2 act in the same pathway. *Curr. Biol.* **11,** 1341–1346 (2001).

Batshaw, M. L. Mental retardation. *Pediatr. Clin. North Am.* **40,**507–521 (1993).

Baumeister, R. & Ge, L. The worm in us - Caenorhabditis elegans as a model of human disease. *Trends Biotechnol.* **20,** 147–148 (2002).

Bautz, E. & Freese, E. ON THE MUTAGENIC EFFECT OF ALKYLATING AGENTS. *Proc. Natl. Acad. Sci. U.S.A.* **46,** 1585–1594 (1960).

Bebbington, A. *et al.* Investigating genotype-phenotype relationships in Rett syndrome using an international data set. *Neurology* **70,**868–875 (2008).

Belichenko, P. V. *et al.* Widespread changes in dendritic and axonal morphology in Mecp2-mutant mouse models of Rett syndrome: evidence for disruption of neuronal networks. *J. Comp. Neurol.* **514,**240–258 (2009).

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).

Berri, S., Boyle, J. H., Tassieri, M., Hope, I. A. & Cohen, N. Forward locomotion of the nematode C. elegans is achieved through modulation of a single gait. *HFSP J* **3,** 186–193 (2009).

Bertani, I. *et al.* Functional consequences of mutations in CDKL5, an X-linked gene involved in infantile spasms and mental retardation. *J. Biol. Chem.* **281,** 32048–32056 (2006).

Blakemore, A. I. F. & Froguel, P. Investigation of Mendelian forms of obesity holds out the prospect of personalized medicine. *Ann. N. Y. Acad. Sci.* **1214,** 180–189 (2010).

Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* **33 Suppl,** 228–237 (2003).

Bras, J. M. & Singleton, A. B. Exome sequencing in Parkinson's disease. *Clin. Genet.* **80,** 104–109 (2011).

Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77,**71–94 (1974).

Brücke, T., Sofic, E., Killian, W., Rett, A. & Riederer, P. Reduced concentrations and increased metabolism of biogenic amines in a single case of Rett-syndrome: a postmortem brain study. *J. Neural Transm.* **68,** 315–324 (1987).

Brunklaus, A. & Zuberi, S. M. Dravet syndrome--from epileptic encephalopathy to channelopathy. *Epilepsia* **55,** 979–984 (2014).

Bryda, E. C. The Mighty Mouse: The Impact of Rodents on Advances in Biomedical Research. *Mo Med* **110,** 207–211 (2013).

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

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).

Carney, R. M. *et al.* Identification of MeCP2 mutations in a series of females with autistic disorder. *Pediatr. Neurol.* **28,** 205–211 (2003).

Carvill, G. L. *et al.* Mutations in the GABA Transporter SLC6A1 Cause Epilepsy with Myoclonic-Atonic Seizures. *Am. J. Hum. Genet.* **96,**808–815 (2015).

Castermans, D. *et al.* Identification and characterization of the TRIP8 and REEP3 genes on chromosome 10q21.3 as novel candidate genes for autism. *Eur. J. Hum. Genet.* **15,** 422–431 (2007).

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

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

Chalfie, M. & White, J. 11 The Nervous System. *Cold Spring Harbor Monograph Archive* **17,** 337–391 (1988).

Chalfie, M. *et al.* The neural circuit for touch sensitivity in Caenorhabditis elegans. *J. Neurosci.* **5,** 956–964 (1985).

Chan, E. Y. W. *et al.* Targeted disruption of Huntingtin-associated protein-1 (Hap1) results in postnatal death due to depressed feeding behavior. *Hum. Mol. Genet.* **11,** 945–959 (2002).

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

Chao, H.-T., Zoghbi, H. Y. & Rosenmund, C. MeCP2 controls excitatory synaptic strength by regulating glutamatergic synapse number. *Neuron* **56,** 58–65 (2007).

Chelly, J. & Mandel, J. L. Monogenic causes of X-linked mental retardation. *Nat. Rev. Genet.* **2,** 669–680 (2001).

Chen, N. *et al.* WormBase: a comprehensive data resource for Caenorhabditis biology and genomics. *Nucleic Acids Res.* **33,** D383–389 (2005).

Chen, Q. *et al.* CDKL5, a protein associated with rett syndrome, regulates neuronal morphogenesis via Rac1 signaling. *J. Neurosci.***30,** 12777–12786 (2010).

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).

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).

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

Clapham, D. E. TRP channels as cellular sensors. *Nature* **426,** 517–524 (2003).

Cohen, D. *et al.* MECP2 mutation in a boy with language disorder and schizophrenia. *Am J Psychiatry* **159,** 148–149 (2002).

Collins, A. L. *et al.* Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* **13,**2679–2689 (2004).

Collins, F. S. *et al.* New goals for the U.S. Human Genome Project: 1998-2003. *Science* **282,** 682–689 (1998).

Cooper, D. N. *et al.* Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Hum. Mutat.* **31,**631–655 (2010).

Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nat. Genet.* **43,** 838–846 (2011).

Culhane, J. C. *et al.* A mechanism-based inactivator for histone demethylase LSD1. *J. Am. Chem. Soc.* **128,** 4536–4537 (2006).

Dani, V. S. & Nelson, S. B. Intact long-term potentiation but reduced connectivity between neocortical layer 5 pyramidal neurons in a mouse model of Rett syndrome. *J. Neurosci.* **29,** 11263–11270 (2009).

Dani, V. S. *et al.* Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 12560–12565 (2005).

de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367,** 1921–1929 (2012).

DeAndrade, M. P. *et al.* Motor restlessness, sleep disturbances, thermal sensory alterations and elevated serum iron levels in Btbd9 mutant mice. *Hum. Mol. Genet.* **21,** 3984–3992 (2012).

Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519,**223–228 (2015).

Department of Health and Human Services, Department of Energy. *Understanding our genetic inheritance: the U.S. Human Genome Project: the first five years: FY 1991-1995.* (Government Printing Office, 1990).

Derry, W. B., Putzke, A. P. & Rothman, J. H. Caenorhabditis elegans p53: role in apoptosis, meiosis, and stress resistance. *Science* **294,**591–595 (2001).

Dietz, H. C. New therapeutic approaches to mendelian disorders. *N. Engl. J. Med.* **363,** 852–863 (2010).

Dinopoulos, A., Matsubara, Y. & Kure, S. Atypical variants of nonketotic hyperglycinemia. *Mol. Genet. Metab.* **86,** 61–69 (2005).

Dittman, J. S. & Kaplan, J. M. Behavioral impact of neurotransmitter-activated GPCRs: Muscarinic and GABAB receptors regulate C. elegans locomotion. *J Neurosci* **28,** 7104–7112 (2008).

Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 8817–8822 (2003).

Durand, C. M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* **39,** 25–27 (2007).

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).

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. Hum. Genet.* **50,** 278–287 (1992).

Endele, S. *et al.* Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nat. Genet.* **42,** 1021–1026 (2010).

Evans, J. C. *et al.* Early onset seizures and Rett-like features associated with mutations in CDKL5. *Eur. J. Hum. Genet.* **13,** 1113–1120 (2005).

Exome Aggregation Consortium (ExAC), ExAC Browser. Cambridge, MA (2014). Available at: http://exac.broadinstitute.org.

Fedurco, M., Romieu, A., Williams, S., Lawrence, I. & Turcatti, G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.***34,** e22 (2006).

Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391,** 806–811 (1998).

Flibotte, S. *et al.* Whole-Genome Profiling of Mutagenesis in Caenorhabditis elegans. *Genetics* **185,** 431–441 (2010).

Fombonne, E., Zakarian, R., Bennett, A., Meng, L. & McLean-Heywood, D. Pervasive developmental disorders in Montreal, Quebec, Canada: prevalence and links with immunizations. *Pediatrics***118,** e139–150 (2006).

Fowler, C. D., Lu, Q., Johnson, P. M., Marks, M. J. & Kenny, P. J. Habenular α5 nicotinic receptor subunit signalling controls nicotine intake. *Nature* **471,** 597–601 (2011).

Frank, S. A. Evolution in health and medicine Sackler colloquium: Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *Proc. Natl. Acad. Sci. U.S.A.* **107 Suppl 1,**1725–1730 (2010).

Friez, M. J. *et al.* Recurrent infections, hypotonia, and mental retardation caused by duplication of MECP2 and adjacent region in Xq28. *Pediatrics* **118,** e1687–1695 (2006).
Fukuda, T., Itoh, M., Ichikawa, T., Washiyama, K. & Goto, Y. Delayed maturation of neuronal architecture and synaptogenesis in cerebral cortex of Mecp2-deficient mice. *J. Neuropathol. Exp. Neurol.* **64,**537–544 (2005).

Fuller, C. W. *et al.* The challenges of sequencing by synthesis. *Nat. Biotechnol.* **27,** 1013–1023 (2009).

Gatto, C. L. & Broadie, K. Drosophila modeling of heritable neurodevelopmental disorders. *Curr. Opin. Neurobiol.* **21,** 834–841 (2011).

Gengyo-Ando, K. & Mitani, S. Characterization of mutations induced by ethyl methanesulfonate, UV, and trimethylpsoralen in the nematode Caenorhabditis elegans. *Biochem. Biophys. Res. Commun.* **269,** 64–69 (2000).

Georgel, P. T. *et al.* Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J. Biol. Chem.* **278,** 32181–32188 (2003).

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

Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Unlocking Mendelian disease using exome sequencing. *Genome Biol.* **12,** 228 (2011).

Gjorgjieva, J., Biron, D. & Haspel, G. Neurobiology of Caenorhabditis elegans Locomotion: Where Do We Stand? *Bioscience* **64,** 476–486 (2014).

Grillo, E. *et al.* Revealing the complexity of a monogenic disease: rett syndrome exome sequencing. *PLoS ONE* **8,** e56599 (2013).

Grossmann, V. *et al.* Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. *Blood* **118,** 6153–6163 (2011).

Guideri, F., Acampa, M., Hayek, G., Zappella, M. & Di Perri, T. Reduced heart rate variability in patients affected with Rett syndrome. A possible explanation for sudden death. *Neuropediatrics***30,** 146–148 (1999).

Gusella, J. F. *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306,** 234–238 (1983).

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).

Hagberg, B. & Kyllerman, M. Epidemiology of mental retardation--a Swedish survey. *Brain Dev.* **5,** 441–449 (1983).

Hagberg, B. Clinical manifestations and stages of Rett syndrome. *Ment Retard Dev Disabil Res Rev* **8,** 61–65 (2002).

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

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).

Halbach, N. S. J. *et al.* Genotype-phenotype relationships as prognosticators in Rett syndrome should be handled with care in clinical practice. *Am. J. Med. Genet. A* **158A,** 340–350 (2012).

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

Hara, M. *et al.* De novo SHANK3 mutation causes Rett syndrome-like phenotype in a female patient. *Am. J. Med. Genet. A* **167,** 1593–1596 (2015).

Hariharan, I. K. & Haber, D. A. Yeast, flies, worms, and fish in the study of human disease. *N. Engl. J. Med.* **348,** 2457–2463 (2003).

Harikrishnan, K. N. *et al.* Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat. Genet.* **37,** 254–264 (2005).

Harris, J. C. New classification for neurodevelopmental disorders in DSM-5. *Curr Opin Psychiatry* **27,** 95–97 (2014).

Harris, P. C. & Torres, V. E. Polycystic kidney disease. *Annu. Rev. Med.***60,** 321–337 (2009).

Harris, T. W. *et al.* WormBase: a multi-species resource for nematode biology and genomics. *Nucleic Acids Res.* **32,** D411–417 (2004).

He, L. *et al.* Spongiform degeneration in mahoganoid mutant mice. *Science* **299,** 710–712 (2003).

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

Hert, D. G., Fredlake, C. P. & Barron, A. E. Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis***29,** 4618–4626 (2008).

Huang, Y. *et al.* Inhibition of lysine-specific demethylase 1 by polyamine analogues results in reexpression of aberrantly silenced genes. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 8023–8028 (2007).

Huguet, G., Ey, E. & Bourgeron, T. The genetic landscapes of autism spectrum disorders. *Annu Rev Genomics Hum Genet* **14,** 191–213 (2013).

Huppke, P. *et al.* Very mild cases of Rett syndrome with skewed X inactivation. *J. Med. Genet.* **43,** 814–816 (2006).

Huppke, P., Laccone, F., Krämer, N., Engel, W. & Hanefeld, F. Rett syndrome: analysis of MECP2 and clinical characterization of 31 patients. *Hum. Mol. Genet.* **9,** 1369–1375 (2000).

Iglesias, A. *et al.* The usefulness of whole-exome sequencing in routine clinical practice. *Genet. Med.* **16,** 922–931 (2014).

Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515,** 216–221 (2014).

Jan, M. M., Dooley, J. M. & Gordon, K. E. Male Rett syndrome variant: application of diagnostic criteria. *Pediatr. Neurol.* **20,** 238–240 (1999).

Jansen, G., Hazendonk, E., Thijssen, K. L. & Plasterk, R. H. Reverse genetics by chemical mutagenesis in Caenorhabditis elegans. *Nat. Genet.* **17,** 119–121 (1997).

Jeffery, L. & Nakielny, S. Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J. Biol. Chem.***279,** 49479–49487 (2004).

Jellinger, K., Armstrong, D., Zoghbi, H. Y. & Percy, A. K. Neuropathology of Rett syndrome. *Acta Neuropathol.* **76,** 142–158 (1988).

Jensen, L. R. *et al.* Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am. J. Hum. Genet.* **76,** 227–236 (2005).

Jian, L. *et al.* p.R270X MECP2 mutation and mortality in Rett syndrome. *Eur. J. Hum. Genet.* **13,** 1235–1238 (2005).

Jian, L. *et al.* Predictors of seizure onset in Rett syndrome. *J. Pediatr.***149,** 542–547 (2006).

Johansson, S. *et al.* Exome sequencing and genetic testing for MODY. *PLoS ONE* **7,** e38050 (2012).

Johnson, J. O. *et al.* Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* **68,** 857–864 (2010).

Jones, P. L. & Wolffe, A. P. Relationships between chromatin organization and DNA methylation in determining gene expression. *Semin. Cancer Biol.* **9,** 339–347 (1999).

Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19,** 187–191 (1998).

Jorgensen, E. M. GABA. *WormBook* (2005). doi:10.1895/wormbook.1.14.1

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.* **55,** 86–96 (2003).

Kabashi, E., Champagne, N., Brustein, E. & Drapeau, P. In the swim of things: recent insights to neurogenetic disorders from zebrafish. *Trends Genet.* **26,** 373–381 (2010).

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).

Kaludov, N. K. & Wolffe, A. P. MeCP2 driven transcriptional repression in vitro: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. *Nucleic Acids Res* **28,** 1921–1928 (2000).

Kamath, R. S. & Ahringer, J. Genome-wide RNAi screening in Caenorhabditis elegans. *Methods* **30,** 313–321 (2003).

Kamath, R. S. *et al.* Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421,** 231–237 (2003).

Kennedy, S., Wang, D. & Ruvkun, G. A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. *Nature***427,** 645–649 (2004).

Kenyon, C. The plasticity of aging: insights from long-lived mutants. *Cell* **120,** 449–460 (2005).

Kerr, A. M. & Julu, P. O. Recent insights into hyperventilation from the study of Rett syndrome. *Arch. Dis. Child.* **80,** 384–387 (1999).

Kerr, A. M., Archer, H. L., Evans, J. C., Prescott, R. J. & Gibbon, F. People with MECP2 mutation-positive Rett disorder who converse. *J Intellect Disabil Res* **50,** 386–394 (2006).

Kerr, A. M., Armstrong, D. D., Prescott, R. J., Doyle, D. & Kearney, D. L. Rett syndrome: analysis of deaths in the British survey. *Eur Child Adolesc Psychiatry* **6 Suppl 1,** 71–74 (1997).

Kiiski, J. I. *et al.* Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 15172–15177 (2014).

Kim, K. & Li, C. Expression and regulation of an FMRFamide-related neuropeptide gene family in Caenorhabditis elegans. *J. Comp. Neurol.* **475,** 540–550 (2004).

Kim, S.-M. *et al.* Regulation of mouse steroidogenesis by WHISTLE and JMJD1C through histone methylation balance. *Nucleic Acids Res.***38,** 6389–6403 (2010).

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

Kishi, N. & Macklis, J. D. MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Mol. Cell. Neurosci.* **27,** 306–321 (2004).

Klauck, S. M. *et al.* A mutation hot spot for nonspecific X-linked mental retardation in the MECP2 gene causes the PPM-X syndrome. *Am. J. Hum. Genet.* **70,** 1034–1037 (2002).

Klose, R. J. *et al.* DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell***19,** 667–678 (2005).

Kokura, K. *et al.* The Ski protein family is required for MeCP2-mediated transcriptional repression. *J. Biol. Chem.* **276,** 34115–34121 (2001).

Krajnc, N., Župančič, N. & Oražem, J. Epilepsy treatment in Rett syndrome. *J. Child Neurol.* **26,** 1429–1433 (2011).

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

Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4,** 1073–1081 (2009).

Kutscher, L. M. & Shaham, S. Forward and reverse mutagenesis in C. elegans. *WormBook* 1–26 (2014). doi:10.1895/wormbook.1.167.1.

Lam, C. W. *et al.* Spectrum of mutations in the MECP2 gene in patients with infantile autism and Rett syndrome. *J. Med. Genet.* **37,**E41 (2000).

Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).

Laumonnier, F. *et al.* Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. *J. Med. Genet.* **42,** 780–786 (2005).

Lee, J. W., Choi, H. S., Gyuris, J., Brent, R. & Moore, D. D. Two classes of proteins dependent on either the presence or absence of thyroid hormone for interaction with the thyroid hormone receptor. *Mol. Endocrinol.* **9,** 243–254 (1995).

Lee, M. G., Wynder, C., Schmidt, D. M., McCafferty, D. G. & Shiekhattar, R. Histone H3 lysine 4 demethylation is a target of nonselective antidepressive medications. *Chem. Biol.* **13,** 563–567 (2006).

Lee, S. S., Wan, M. & Francke, U. Spectrum of MECP2 mutations in Rett syndrome. *Brain Dev.* **23 Suppl 1,** S138–143 (2001).

Lekman, A. *et al.* Rett syndrome: biogenic amines and metabolites in postmortem brain. *Pediatr. Neurol.* **5,** 357–362 (1989).

Lemke, J. R. *et al.* GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. *Ann. Neurol.* **75,** 147–154 (2014).

Levy, D. *et al.* Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* **70,** 886–897 (2011).

Li, C., Nelson, L. S., Kim, K., Nathoo, A. & Hart, A. C. Neuropeptide gene families in the nematode Caenorhabditis elegans. *Ann. N. Y. Acad. Sci.* **897,** 239–252 (1999).

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

Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

Linnarsson, S. Recent advances in DNA sequencing methods - general principles of sample preparation. *Exp. Cell Res.* **316,** 1339–1343 (2010).

Liu, L. X. *et al.* High-throughput isolation of Caenorhabditis elegans deletion mutants. *Genome Res.* **9,** 859–867 (1999).

Liu, Q., Hollopeter, G. & Jorgensen, E. M. Graded synaptic transmission at the Caenorhabditis elegans neuromuscular junction. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 10823–10828 (2009).

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

Lugtenberg, D. *et al.* Chromosomal copy number changes in patients with non-syndromic X linked mental retardation detected by array CGH. *J. Med. Genet.* **43,** 362–370 (2006).
Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 961–968 (2010).

Maezawa, I. & Jin, L.-W. Rett syndrome microglia damage dendrites and synapses by the elevated release of glutamate. *J. Neurosci.* **30,**5346–5356 (2010).

Maliszewska-Cyna, E., Bawa, D. & Eubanks, J. H. Diminished prevalence but preserved synaptic distribution of N-methyl-D-aspartate receptor subunits in the methyl CpG binding protein 2(MeCP2)-null mouse brain. *Neuroscience* **168,** 624–632 (2010).

Marangi, G. *et al.* Proposal of a clinical score for the molecular test for Pitt-Hopkins syndrome. *Am. J. Med. Genet. A* **158A,** 1604–1611 (2012).

Mardis, E. R. A decade's perspective on DNA sequencing technology. *Nature* **470,** 198–203 (2011).

Mardis, E. R. Anticipating the 1,000 dollar genome. *Genome Biol.* **7,**112 (2006).

Mardis, E. R. New strategies and emerging technologies for massively parallel sequencing: applications in medical research. *Genome Med* **1,** 40 (2009).

Martinho, P. S. *et al.* In search of a genetic basis for the Rett syndrome. *Hum. Genet.* **86,** 131–134 (1990).

McFarland, K. N. *et al.* MeCP2: a novel Huntingtin interactor. *Hum. Mol. Genet.* **23,** 1036–1044 (2014).

McIntire, S. L., Jorgensen, E. & Horvitz, H. R. Genes required for GABA function in Caenorhabditis elegans. *Nature* **364,** 334–337 (1993).

McIntire, S. L., Jorgensen, E., Kaplan, J. & Horvitz, H. R. The GABAergic nervous system of Caenorhabditis elegans. *Nature* **364,**337–341 (1993).

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

Medrihan, L. *et al.* Early defects of GABAergic synapses in the brain stem of a MeCP2 mouse model of Rett syndrome. *J. Neurophysiol.***99,** 112–121 (2008).

Meins, M. *et al.* Submicroscopic duplication in Xq28 causes increased expression of the MECP2 gene in a boy with severe mental retardation and features of Rett syndrome. *J. Med. Genet.* **42,** e12 (2005).
Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151,** 1417–1430 (2012).

Metzker, M. L. Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11,** 31–46 (2010).

Metzstein, M. M., Stanfield, G. M. & Horvitz, H. R. Genetics of programmed cell death in C. elegans: past, present and future. *Trends Genet.* **14,** 410–416 (1998).

Michaelson, J. J. *et al.* Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* **151,** 1431–1442 (2012).

Migeon, B. R., Dunn, M. A., Thomas, G., Schmeckpeper, B. J. & Naidu, S. Studies of X inactivation and isodisomy in twins provide further evidence that the X chromosome is not involved in Rett syndrome. *Am. J. Hum. Genet.* **56,** 647–653 (1995).

Milani, D., Pantaleoni, C., D'Arrigo, S., Selicorni, A. & Riva, D. Another patient with MECP2 mutation without classic Rett syndrome phenotype. *Pediatr. Neurol.* **32,** 355–357 (2005).

Miyake, K. & Nagai, K. Phosphorylation of methyl-CpG binding protein 2 (MeCP2) regulates the intracellular localization during neuronal cell differentiation. *Neurochem. Int.* **50,** 264–270 (2007).

Mombereau, C. *et al.* Altered anxiety and depression-related behaviour in mice lacking GABAB(2) receptor subunits. *Neuroreport***16,** 307–310 (2005).

Monrós, E. *et al.* Rett syndrome in Spain: mutation analysis and clinical correlations. *Brain Dev.* **23 Suppl 1,** S251–253 (2001).

Montes, G. & Halterman, J. S. Characteristics of school-age children with autism. *J Dev Behav Pediatr* **27,** 379–385 (2006).

Moretti, P. *et al.* Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome. *J. Neurosci.* **26,** 319–327 (2006).

Mount, R. H., Hastings, R. P., Reilly, S., Cass, H. & Charman, T. Behavioural and emotional features in Rett syndrome. *Disabil Rehabil***23,** 129–138 (2001).

Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468,** 443–446 (2010).

Nag, N. & Berger-Sweeney, J. E. Postnatal dietary choline supplementation alters behavior in a mouse model of Rett syndrome. *Neurobiol. Dis.* **26,** 473–480 (2007).

Nag, N., Mellott, T. J. & Berger-Sweeney, J. E. Effects of postnatal dietary choline supplementation on motor regional brain volume and growth factor expression in a mouse model of Rett syndrome. *Brain Res.* **1237,** 101–109 (2008).

Nan, X. *et al.* Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 2709–2714 (2007).

Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393,**386–389 (1998).

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

National Research Council (US) Committee on Mapping and Sequencing the Human Genome. *Mapping and Sequencing the Human Genome*. (National Academies Press (US), 1988).

Natrajan, R. & Reis-Filho, J. S. Next-generation sequencing applied to molecular diagnostics. *Expert Rev. Mol. Diagn.* **11,** 425–444 (2011).

Nauli, S. M. *et al.* Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. *Nat. Genet.* **33,** 129–137 (2003).

Nava, C. *et al.* De novo mutations in HCN1 cause early infantile epileptic encephalopathy. *Nat. Genet.* **46,** 640–645 (2014).

Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485,** 242–245 (2012).

Nelson, E. D., Kavalali, E. T. & Monteggia, L. M. MeCP2-dependent transcriptional repression regulates excitatory neurotransmission. *Curr. Biol.* **16,** 710–716 (2006).

Neul, J. L. *et al.* Rett syndrome: revised diagnostic criteria and nomenclature. *Ann. Neurol.* **68,** 944–950 (2010).

Neul, J. L. *et al.* Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. *Neurology* **70,** 1313–1321 (2008).

Ng, H. H. & Bird, A. DNA methylation and chromatin modification. *Curr. Opin. Genet. Dev.* **9,** 158–163 (1999).

Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42,** 30–35 (2010).

Niebur, E. & Erdös, P. Theory of the locomotion of nematodes: control of the somatic motor neurons by interneurons. *Math Biosci* **118,** 51–82 (1993).

Nikitina, T. *et al.* Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Mol. Cell. Biol.* **27,** 864–877 (2007).

Nomura, Y. & Segawa, M. Motor symptoms of the Rett syndrome: abnormal muscle tone, posture, locomotion and stereotyped movement. *Brain Dev.* **14 Suppl,** S21–28 (1992).

Nomura, Y. Early behavior characteristics and sleep disturbance in Rett syndrome. *Brain Dev.* **27 Suppl 1,** S35–S42 (2005).

Okado, N., Narita, M. & Narita, N. A biogenic amine-synapse mechanism for mental retardation and developmental disabilities. *Brain Dev.* **23 Suppl 1,** S11–15 (2001).

Oti, M. & Brunner, H. G. The modular nature of genetic diseases. *Clin. Genet.* **71,** 1–11 (2007).

Panayotis, N., Ghata, A., Villard, L. & Roux, J.-C. Biogenic amines and their metabolites are differentially affected in the Mecp2-deficient mouse brain. *BMC Neurosci* **12,** 47 (2011).

Peltonen, L., Perola, M., Naukkarinen, J. & Palotie, A. Lessons from studying monogenic disease for common disease. *Hum. Mol. Genet.***15 Spec No 1,** R67–74 (2006).

Percy, A. K. Rett syndrome. Current status and new vistas. *Neurol Clin* **20,** 1125–1141 (2002).

Percy, A. K. Rett syndrome: recent research progress. *J. Child Neurol.* **23,** 543–549 (2008).

Peter, C. J. & Akbarian, S. Balancing histone methylation activities in psychiatric disorders. *Trends Mol Med* **17,** 372–379 (2011).

Petroff, O. A. C. GABA and glutamate in the human brain. *Neuroscientist* **8,** 562–573 (2002).

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

Picciotto, M. R., Higley, M. J. & Mineur, Y. S. Acetylcholine as a neuromodulator: cholinergic signaling shapes nervous system function and behavior. *Neuron* **76,** 116–129 (2012).

Pierce-Shimomura, J. T. *et al.* Genetic analysis of crawling and swimming locomotory patterns in C. elegans. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 20982–20987 (2008).

Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466,** 368–372 (2010).

Pitt, D. & Hopkins, I. A syndrome of mental retardation, wide mouth and intermittent overbreathing. *Aust Paediatr J* **14,** 182–184 (1978).

Poulin, G., Nandakumar, R. & Ahringer, J. Genome-wide RNAi screens in Caenorhabditis elegans: impact on cancer research. *Oncogene* **23,** 8340–8345 (2004).

Prober, J. M. *et al.* A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238,**336–341 (1987).

Rabbani, B., Mahdieh, N., Hosomichi, K., Nakaoka, H. & Inoue, I. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J. Hum. Genet.* **57,** 621–632 (2012).

Reiss, A. L. *et al.* Neuroanatomy of Rett syndrome: a volumetric imaging study. *Ann. Neurol.* **34,** 227–234 (1993).

Renieri, A. *et al.* Diagnostic criteria for the Zappella variant of Rett syndrome (the preserved speech variant). *Brain Dev.* **31,** 208–216 (2009).

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

Richmond, J. E. & Jorgensen, E. M. One GABA and two acetylcholine receptors function at the C. elegans neuromuscular junction. *Nat. Neurosci.* **2,** 791–797 (1999).

Riddle DL, Blumenthal T, Meyer BJ, *et al*., editors. C. elegans II. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press (1997). Available at: https://www.ncbi.nlm.nih.gov/books/NBK19997/

Riordan, J. R. *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245,** 1066–1073 (1989).

Roach, J. C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328,** 636–639 (2010).

Rodan, L. H. *et al.* A novel neurodevelopmental disorder associated with compound heterozygous variants in the huntingtin gene. *Eur. J. Hum. Genet.* **24,** 1826–1827 (2016).

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

Ronemus, M., Iossifov, I., Levy, D. & Wigler, M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat. Rev. Genet.* **15,** 133–141 (2014).

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

Roze, E. *et al.* Rett syndrome: an overlooked diagnosis in women with stereotypic hand movements, psychomotor retardation, Parkinsonism, and dystonia? *Mov. Disord.* **22,** 387–389 (2007).

Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287,** 2204–2215 (2000).

Rushforth, A. M., Saari, B. & Anderson, P. Site-selected insertion of the transposon Tc1 into a Caenorhabditis elegans myosin light chain gene. *Mol. Cell. Biol.* **13,** 902–910 (1993).

Sanchez-Mut, J. V., Huertas, D. & Esteller, M. Aberrant epigenetic landscape in intellectual disability. *Prog. Brain Res.* **197,** 53–71 (2012).

Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5463–5467 (1977).

Sarter, M. & Parikh, V. Choline transporters, cholinergic transmission and cognition. *Nat. Rev. Neurosci.* **6,** 48–56 (2005).

Sassi, C. *et al.* Exome sequencing identifies 2 novel presenilin 1 mutations (p.L166V and p.S230R) in British early-ons*et al*zheimer's disease. *Neurobiol. Aging* **35,** 2422.e13–16 (2014).

Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* **12,**683–691 (2011).

Scala, E. *et al.* MECP2 deletions and genotype-phenotype correlation in Rett syndrome. *Am. J. Med. Genet. A* **143A,** 2775–2784 (2007).

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. *Am. J. Hum. Genet.* **63,** 267–269 (1998).

Schanen, N. C. *et al.* A new Rett syndrome family consistent with X-linked inheritance expands the X chromosome exclusion map. *Am. J. Hum. Genet.* **61,** 634–641 (1997).

Schüle, B., Armstrong, D. D., Vogel, H., Oviedo, A. & Francke, U. Severe congenital encephalopathy caused by MECP2 null mutations in males: central hypoxia and reduced neuronal dendritic structure. *Clin. Genet.* **74,** 116–126 (2008).

Service, R. F. Gene sequencing. The race for the $1000 genome. *Science* **311,** 1544–1546 (2006).

Shahbazian, M. *et al.* Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* **35,** 243–254 (2002).

Shahbazian, M. *et al.* Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* **35,** 243–254 (2002).

Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26,** 1135–1145 (2008).

Shi, Y. *et al.* Exome sequencing identifies ZNF644 mutations in high myopia. *PLoS Genet.* **7,** e1002084 (2011).

Siegfried, Z. *et al.* DNA methylation represses transcription in vivo. *Nat. Genet.* **22,** 203–206 (1999).

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).

Smeets, E. *et al.* Rett syndrome in females with CTS hot spot deletions: a disorder profile. *Am. J. Med. Genet. A* **132A,** 117–120 (2005).

Smith, J. D. *et al.* Exome Sequencing Identifies a Recurrent De Novo ZSWIM6 Mutation Associated with Acromelic Frontonasal Dysostosis. *Am J Hum Genet* **95,** 235–240 (2014).

Smith, L. M. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321,** 674–679 (1986).

Snape, K. *et al.* Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. *Breast Cancer Res. Treat.* **134,** 429–433 (2012).

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

Stenson, P. D. *et al.* Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* **21,** 577–581 (2003).

Stiles, J. & Jernigan, T. L. The basics of brain development. *Neuropsychol Rev* **20,** 327–348 (2010).

Sudmant, P. H. *et al.* Diversity of human copy number variation and multicopy genes. *Science* **330,** 641–646 (2010).

Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev. Biol.* **56,** 110–156 (1977).

Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.* **100,** 64–119 (1983).

Sweatt, J. D. Pitt-Hopkins Syndrome: intellectual disability due to loss of TCF4-regulated gene transcription. *Exp. Mol. Med.* **45,** e21 (2013).

Tan, K. *et al.* Human PLU-1 Has transcriptional repression properties and interacts with the developmental transcription factors BF-1 and PAX9. *J. Biol. Chem.* **278,** 20507–20513 (2003).

Taneri, B., Asilmaz, E. & Gaasterland, T. Biomedical impact of splicing mutations revealed through exome sequencing. *Mol. Med.***18,** 314–319 (2012).

Tate, P. H. & Bird, A. P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* **3,** 226–231 (1993).

Tawarayama, H., Yoshida, Y., Suto, F., Mitchell, K. J. & Fujisawa, H. Roles of semaphorin-6B and plexin-A2 in lamina-restricted projection of hippocampal mossy fibers. *J. Neurosci.* **30,** 7049–7060 (2010).

Trappe, R. *et al.* MECP2 mutations in sporadic cases of Rett syndrome are almost exclusively of paternal origin. *Am. J. Hum. Genet.* **68,** 1093–1101 (2001).

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).

Tylee, D. S., Kawaguchi, D. M. & Glatt, S. J. On the outside, looking in: a review and evaluation of the comparability of blood and brain '-omes'. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **162B,** 595–603 (2013).

Urdinguio, R. G., Sanchez-Mut, J. V. & Esteller, M. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol* **8,** 1056–1072 (2009).

Van Esch, H. *et al.* Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am. J. Hum. Genet.* **77,** 442–453 (2005).

van Loo, K. M. J. & Martens, G. J. M. Genetic and environmental factors in complex neurodevelopmental disorders. *Curr. Genomics* **8,**429–444 (2007).

Vandamme, T. F. Use of rodents as models of human diseases. *J Pharm Bioallied Sci* **6,** 2–9 (2014).

Venâncio, M., Santos, M., Pereira, S. A., Maciel, P. & Saraiva, J. M. An explanation for another familial case of Rett syndrome: maternal germline mosaicism. *Eur. J. Hum. Genet.* **15,** 902–904 (2007).

Villard, L. *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.***38,** 435–442 (2001).

Villard, L. *et al.* Two affected boys in a Rett syndrome family: clinical and molecular findings. *Neurology* **55,** 1188–1193 (2000).

Vissers, L. E. L. M. *et al.* A de novo paradigm for mental retardation. *Nat. Genet.* **42,** 1109–1112 (2010).

Von Stetina, S. E., Treinin, M. & Miller, D. M. The motor circuit. *Int. Rev. Neurobiol.* **69,** 125–167 (2006).

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).

Wang, H. *et al.* Exome capture sequencing identifies a novel mutation in BBS4. *Mol. Vis.* **17,** 3529–3540 (2011).

Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).

Watanabe, S. *et al.* JMJD1C demethylates MDC1 to regulate the RNF8 and BRCA1-mediated chromatin response to DNA breaks. *Nat. Struct. Mol. Biol.* **20,** 1425–1433 (2013).

Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171,** 737–738 (1953).

Watson, P. *et al.* Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein. *J. Med. Genet.* **38,** 224–228 (2001).

Weng, S.-M., McLeod, F., Bailey, M. E. S. & Cobb, S. R. Synaptic plasticity deficits in an experimental model of rett syndrome: long-term potentiation saturation and its pharmacological reversal. *Neuroscience* **180,** 314–321 (2011).

Wenk, G. L. & Mobley, S. L. Choline acetyltransferase activity and vesamicol binding in Rett syndrome and in rats with nucleus basalis lesions. *Neuroscience* **73,** 79–84 (1996).

Wetterstrand KA. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP) (2013). Available at: http://www.genome.gov/sequencingcosts.

White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **314,** 1–340 (1986).

White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the ventral nerve cord of Caenorhabditis elegans. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **275,** 327–348 (1976).

Wolf, S. S., Patchev, V. K. & Obendorf, M. A novel variant of the putative demethylase gene, s-JMJD1C, is a coactivator of the AR. *Arch. Biochem. Biophys.* **460,** 56–66 (2007).

Worthey, E. A. *et al.* Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* **13,** 255–262 (2011).

Wynder, C., Stalker, L. & Doughty, M. L. Role of H3K4 demethylases in complex neurodevelopmental diseases. *Epigenomics* **2,** 407–418 (2010).

Xiang, J., Yan, S., Li, S.-H. & Li, X.-J. Postnatal loss of hap1 reduces hippocampal neurogenesis and causes adult depressive-like behavior in mice. *PLoS Genet.* **11,** e1005175 (2015).

Xu, X., Miller, E. C. & Pozzo-Miller, L. Dendritic spine dysgenesis in Rett syndrome. *Front Neuroanat* **8,** 97 (2014).

Xu, Y. *et al.* A novel mutation identified in PKHD1 by targeted exome sequencing: guiding prenatal diagnosis for an ARPKD family. *Gene***551,** 33–38 (2014).

Yan, X.-J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat. Genet.* **43,** 309–315 (2011).

Yandell, M. *et al.* A probabilistic disease-gene finder for personal genomes. *Genome Res.* **21,** 1529–1542 (2011).

Yang, M. *et al.* Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol. Cell***23,** 377–387 (2006).

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).

Yasui, D. H. *et al.* Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 19416–19421 (2007).

Yoder, B. K., Hou, X. & Guay-Woodford, L. M. The polycystic kidney disease proteins, polycystin-1, polycystin-2, polaris, and cystin, are co-localized in renal cilia. *J. Am. Soc. Nephrol.* **13,** 2508–2516 (2002).

Young, D. *et al.* Sleep problems in Rett syndrome. *Brain Dev.* **29,**609–616 (2007).

Young, J. I. *et al.* Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 17551–17558 (2005).

Yu, T. W. *et al.* Using whole-exome sequencing to identify inherited causes of autism. *Neuron* **77,** 259–273 (2013).

Zappella, M. The Rett girls with preserved speech. *Brain Dev.* **14,**98–101 (1992).

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

Zarkower, D. Somatic sex determination. *WormBook* (2006). doi:10.1895/wormbook.1.84.1

Zhang, Y. *et al.* Loss of MeCP2 in cholinergic neurons causes part of RTT-like phenotypes via α7 receptor in hippocampus. *Cell Res.* **26,**728–742 (2016).

Zhang, Z.-W., Zak, J. D. & Liu, H. MeCP2 is required for normal development of GABAergic circuits in the thalamus. *J. Neurophysiol.***103,** 2470–2481 (2010).

Zuryn, S. & Jarriault, S. Deep sequencing strategies for mapping and identifying mutations from genetic screens. *Worm* **2,** (2013).

Zuryn, S., Le Gras, S., Jamet, K. & Jarriault, S. A Strategy for Direct Mapping and Identification of Mutations by Whole-Genome Sequencing. *Genetics* **186,** 427–430 (2010).

Zweier, C. *et al.* Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am. J. Hum. Genet.* **80,** 994–1001 (2007).

ANNEX I

Genetics inMedicine | ORIGINAL RESEARCH ARTICLE

Open

Mutations in JMJD1C are involved in Rett syndrome and intellectual disability

Mauricio A. Sáez, PhD[1], Juana Fernández-Rodríguez, PhD[2], Catia Moutinho, PhD[1], Jose V. Sanchez-Mut, PhD[1], Antonio Gomez, PhD[1], Enrique Vidal, PhD[1], Paolo Petazzi, PhD[1], Karolina Szczesna, PhD[1], Paula Lopez-Serra, PhD[1], Mario Lucariello, PhD[1], Patricia Lorden, PhD[1], Raul Delgado-Morales, PhD[1], Olga J. de la Caridad, PhD[1], Dori Huertas, PhD[1], Josep L. Gelpí, PhD[2,3], Modesto Orozco, PhD[2–4], Adriana López-Doriga, PhD[5], Montserrat Milà, PhD[6,7], Luís A. Perez-Jurado, MD, PhD[7,8], Mercedes Pineda, MD, PhD[7,9], Judith Armstrong, PhD[7,9], Conxi Lázaro, PhD[5], and Manel Esteller, MD, PhD[1,4,10]

Purpose: Autism spectrum disorders are associated with defects in social response and communication that often occur in the context of intellectual disability. Rett syndrome is one example in which epilepsy, motor impairment, and motor disturbance may co-occur. Mutations in histone demethylases are known to occur in several of these syndromes. Herein, we aimed to identify whether mutations in the candidate histone demethylase JMJD1C (jumonji domain containing 1C) are implicated in these disorders.

Methods: We performed the mutational and functional analysis of JMJD1C in 215 cases of autism spectrum disorders, intellectual disability, and Rett syndrome without a known genetic defect.

Results: We found seven JMJD1C variants that were not present in any control sample (~ 6,000) and caused an amino acid change involving a different functional group. From these, two *de novo* JMJD1C germline mutations were identified in a case of Rett syndrome and in a patient with intellectual disability. The functional study of the JMJD1C mutant Rett syndrome patient demonstrated that the altered protein had abnormal subcellular localization, diminished activity to demethylate the DNA damage-response protein MDC1, and reduced binding to MECP2. We confirmed that JMJD1C protein is widely expressed in brain regions and that its depletion compromises dendritic activity.

Conclusions: Our findings indicate that mutations in JMJD1C contribute to the development of Rett syndrome and intellectual disability.

Genet Med advance online publication 16 July 2015

Key Words: autism; intellectual disability; mutational screening; Rett syndrome

INTRODUCTION

Autism spectrum disorders are a heterogeneous clinical and genetic group of neurodevelopmental defects that are characterized by impaired social communication functions and inappropriate repetitive behavior.[1] This family of disorders is characterized by enormous phenotypic variability, from mild primary deficits in language pragmatics[2] to major neurological phenotypes, such as that of Rett syndrome (OMIM 312750), where it co-occurs with epilepsy, motor impairment, and sleep disturbance.[3] The disabilities associated with autism spectrum disorders are often so severe that affected individuals do not generally reach parenthood, thereby preventing comprehensive familial genetic studies from being undertaken. However, genetic alterations are already recognized as major etiological factors. In this regard, concordance with autism spectrum disorders is higher than with any other cognitive or behavioral disorder.[4,5] In addition to the contribution of polymorphic variants that confer low or moderate risk of the appearance of these neurodevelopmental defects, a cause of autism spectrum disorders can be the occurrence of de novo mutations affecting genes in a number of cellular pathways.[6–8] A similar scenario can be proposed for the genetic contribution to the even more heterogeneous group of disorders classified as intellectual disabilities.[9]

Among the described genetic defects associated with intellectual disabilities, our attention was caught by a single case report of an autistic patient carrying a de novo balanced paracentric inversion 46, XY in (10)(q11.1;q21.3) in which the distal breakpoint disrupted what was at that time known as the TRIP8 gene,[10] which has been characterized as a member of the jmJC domain-containing protein family involved in the methyl

[1]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Catalonia, Spain; [2]Joint Biomedical Research Institute-Barcelona Supercomputing Center (IRB-BSC) Computational Biology Program, Barcelona, Catalonia, Spain; [3]Department of Biochemistry and Molecular Biology, University of Barcelona, Barcelona, Catalonia, Spain; [4]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain; [5]Hereditary Cancer Program, Catalan Institute of Oncology-Bellvitge Institute for Biomedical Research (ICO-IDIBELL), Barcelona, Catalonia, Spain; [6]Biochemistry and Molecular Genetics Department, Hospital Clínic, Barcelona, Catalonia, Spain; [7]CIBERER (Biomedical Network Research Centre on Rare Diseases, Instituto de Salud Carlos III), Barcelona, Spain; [8]Genetics Unit, University Pompeu Fabra, Barcelona, Catalonia, Spain; [9]Department of Neurology, Hospital Sant Joan de Déu (HSJD), Barcelona, Catalonia, Spain; [10]Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain. Correspondence: Manel Esteller (mesteller@idibell.cat)

Submitted 5 November 2014; accepted 9 June 2015; advance online publication 16 July 2015. doi:10.1038/gim.2015.100

group removal and renamed as JMJD1C (jumonji domain containing 1C).[11–13] Support for a role of JMJD1C in autism has been further fostered by large-scale exome sequencing studies in which three de novo variants have been identified,[14,15] at least one of which is a loss-of-function mutation.[15] However, two of the three described variants occurred in unaffected siblings. To investigate whether JMJD1C mutations occur in intellectual disability, Rett syndrome, and autism spectrum disorders, we performed a comprehensive screening to identify single nucleotide changes and indels in a large collection of 215 patients in whom a genetic defect had not previously been identified.

MATERIALS AND METHODS

Patients

The samples used in this study consisted of 215 patients of either gender with autism, Rett syndrome, or intellectual disability from the Hospital Clínic, Sant Joan de Deu and Pompeu Fabra University, Barcelona, Catalonia, Spain. A cohort of 500 healthy volunteers was obtained from the same geographic region (Catalonia, Spain). All were Caucasian individuals (such as in the case cohort) and were matched according to gender distribution (case cohort: female, 53% [114/215]; male, 47% [101/215]; control cohort: female, 58% [290/500]; male, 42% [210/500]). Ethical approval for the molecular genetic studies was obtained from each institutional review board. DNA was extracted from peripheral blood leukocytes using standard techniques. We measured DNA concentration with the Quant-iT Picogreen (Invitrogen, Life Technologies – Grand Island, NY) and then normalized all concentrations to 25–50 ng/μl before proceeding with the Access Array amplification.

Fluidigm access array

Forty-eight pairs of primers were designed using the Access Array Amplicon Tagging Assay design service from Fluidigm to cover all 26 exons of the JMJD1C gene (NM_032776.1), including exon-intron boundaries. These primers also contained the sequencing adaptors necessary for subsequent sequencing in the 454 GS Junior Sequencer.

Multiplex ligation-dependent probe amplification

Large rearrangements in the JMJD1C gene were studied using multiplex ligation-dependent probe amplification. We designed nine such probes specific to the JMJD1C gene and six control probes according to the instructions provided by MRC-Holland.

Sanger sequencing

The variants were validated by Sanger sequencing using a BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies, Grand Island, NY) in an Applied Biosystems 3730/DNA Analyzer (Applied Biosystems – Life Technologies – Grand Island, NY). The raw data were analyzed with Codon Code Aligner Software (CodonCode Corporation – Centerville, MA).

Exome sequencing

The patient and healthy parents were analyzed by whole exome sequencing with TruSeq Sample Preparation Kit (Illumina, San Diego, CA). Exomes were captured with TruSeq Exome Enrichment Kit (Illumina) and paired-end 100×2 sequenced with the HiScan SQ (Illumina). Format DNA and Protein Sequence Alignment Qual (FASTQ) files were analyzed as described in **Supplementary Methods** online. The raw data were analyzed in Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Catalonia, Spain.

Cell culture and vectors

JMJD1C coding sequence in the pCMV6-AC-GFP vector was purchased from Origene (Rockville, MD) (RG214878). The mutants were generated with Mutant QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA). Wild-type (WT) WT and Pro163Leu and His2336Ala mutants were subcloned in the pCMV6-entry vector to introduce Myc-DDK-tag. shRNAs against the coding sequence of the mouse Jmjd1c gene were cloned in the pLVX-shRNA2 vector between the *BamH*I and *EcoR*I restriction sites (shRNA24 target: CAGAGACTGCTTGAGGAAT). Hek293 cells were cultivated in Dulbecco's modified Eagles medium with 10% fetal bovine serum. To generate stable WT or mutant clones, Hek293 cells were transfected with Lipofectamine 2000 (Invitrogen), selected with G418 antibiotic, and individual clones were isolated 2 weeks later. For transient expression, 6 mg of vector were transfected in 35-mm six-well plates with jetPRIME transfection reagent (Polyplus, NY) following the manufacturer's instructions. Primary cultures of hippocampal neurons were prepared from neonate mice (P0) of either sex. Cultures were infected at 3DIV with lentiviral vectors to express scramble or shRNAs against JMJD1C together with a GFP tracer (pLVX-shRNA2 system). Coverslips were fixed, and protein was extracted at 15DIV.

Immunoprecipitation

Seven hundred fifty micrograms of chromatin fraction were diluted 10-fold in immunoprecipitation (IP) buffer (5 mmol/l Tris–HCl pH 7.6, 15 mmol/l HEPES pH 8.0, 1 mmol/l Ethylene diamine tetraacetic acid, 1 mmol/l ethylene glycol tetraacetic acid, 0.1% sodium dodecyl sulfate (SDS), 1% Triton X-100) incubated with 2 μg of antibodies anti-Me-Lysine (Abcam ab23366, Cambridge, UK) overnight at 4 °C and for 2 hours with PureProteome Protein A/G Magnetic Beads. Beads were washed twice with IP buffer and twice in RBS NP-40 and eluted in laemli buffer in reduction conditions at 70 °C for 10 minutes.

MeCP2 immunoprecipitation

For the MeCP2 immunoprecipitation procedure, anti-JMJD1C and anti-MeCP2 antibodies were coupled to Dynabeads Protein G (Invitrogen). JMJD1C transfected HEK293F cells were transiently transfected with MeCP2-Flag tagged plasmid and the nuclear fraction was obtained by RIPA buffer (10 mmol/l TRIS-Cl pH 8.0, 1 mmol/l EDTA (Ethylene diamine tetraacetic acid), 0.5 mmol/l EGTA (ethylene glycol tetraacetic acid), 1% Triton X-100, 0.1 % sodium deoxycholate, 0.1% SDS, and 140 mmol/l NaCl) supplemented with protease inhibitors (complete; Roche, Rotkreuz, Switzerland) and hybridized with the antibodies at 4 °C for 2 hours. Then, 150 mM NaCl

phosphate-buffered saline (PBS) buffer was used for washing. Human IgG was used as negative control. Anti-Flag HRP (M2-SIGMA) antibody was used to visualize binding.

Western blot

Protein extract of Hek293 cells and primary neuronal culture were obtained in radioimmunoprecipitation assay (RIPA) buffer supplemented with a complete protease inhibitor cocktail tablet (Roche) and sonicated. Then, 50 µg of each protein sample were denatured in Laemli buffer 4% β-mercaptoethanol for 10 minutes at 95 °C, separated on a 7.5% or 15% SDS-polyacrylamide gel, and then transferred onto a polyvinylidene fluoride membrane (Immobilon-P; Millipore, Hessen, Germany) by liquid electroblotting. The antibodies and dilutions used were as follows: rabbit anti-JMJD1C 1:2,000 (09-817; Millipore); mouse anti-nucleolin 1:1,000 (SC-8031; Santa Cruz, Dallas, Texas); rabbit anti-H3 1:10,000 (Abcam AB1791); mouse anti-H3 1:4,000 (Abcam AB10799); and rabbit anti-MDC1 1:5,000 (Abcam AB11171). The blots were developed with Luminata Crescendo Western HRP substrate (MERCK MILLIPORE - Hessen, Germany) or with the LiCor Odyssey System (LI-COR – Bad Homburg, Germany).

Immunofluorescence

Cells were fixed in 4% paraformaldehyde–phosphate-buffered serum (PFA-PBS), quenched in 100 mmol/l glycine-PBS, and permeabilized with 0.25% Triton X-100, 1% bovine serum albumin, and PBS. The cells were blocked with 0.2% gelatin and 0.25% Triton X-100. Antibody dilutions were prepared in 0.25% Triton X-100, 1% bovine serum albumin, and PBS. The dilutions used were: rabbit anti-JMJD1C 1:200; chicken anti-Map2 1:5,000; and anti-β-tubulin 1:1,500 (Abcam AB21058). Nuclei were stained with 2 mg/ml Hoechst 33342. Coverslips were mounted in ProLong Gold antifade reagent (Molecular Probes – Life Technologies – Grand Island, NY). Confocal images were captured with a Leica SP5 confocal microscope (Leica, Buffalo Grove, IL). For fluorescence recovery after photobleaching analysis, the cells were maintained at 37 °C in an atmosphere of 5% CO_2. We captured images every 70 seconds at 63×, at 128 × 128 resolution, and at 1,400 Hz with bidirectional acquisition. We captured 25 control images at 3% laser transmission before bleaching and then bleached the region of interest (ROI) inside a nucleus 25 times at a nominal level of 100% laser transmission. For this experiment, 150 images were captured after bleaching. The row data were analyzed with FrapAnalyzer software (Igor Pro 6.1 Software – WaveMetrics – Lake Oswego, OR).

RNA extraction and real-time PCR

Total RNA was extracted from cell lysates using TRIzol Reagent (Invitrogen), purified using the RNeasy Kit (Qiagen, Valencia, CA), and 2 µg were retrotranscribed using the ThermoScript™ RT-PCR System (Invitrogen). Real-time PCR reactions were performed in triplicate on an Applied Biosystems 7,900HT Fast Real-Time PCR system using 20 ng cDNA, 5 µl SYBR Green PCR Master Mix (Applied Biosystem), and 150 nmol/l specific primers (sequences are available on request) in a final volume of 10 µl.

RESULTS

JMJD1C mutational analyses in samples from autism spectrum disorders, intellectual disability, and Rett syndrome

We first collected blood samples in EDTA tubes for DNA extraction from 215 patients with autism spectrum disorders ($n = 69$; 58 males and 11 females), intellectual disability (IQ < 70) ($n = 85$; 43 males and 42 females), or Rett syndrome ($n = 61$; all females) without mutations in the disease-associated genes MECP2, CDKL5, and FOXG1.[16] Approval for this was obtained from the corresponding institutional review boards. The presence of single-nucleotide changes and indels in the JMJD1C gene was analyzed by sequencing using a GSJUNIOR system with an amplicon library prepared with the Fluidigm Access Array, and the presence of larger genetic defects was assessed by multiplex ligation-dependent probe amplification. We detected no major genomic defects at the JMJD1C gene locus using the multiplex ligation-dependent probe amplification approach. Identified synonymous variants for JMJD1C are shown in **Supplementary Table S1** online and seven previously informed JMJD1C nucleotide variants are described in **Supplementary Table S2** online. Most importantly, the sequencing strategy identified seven nucleotide changes in the exonic regions of JMJD1C that were not present in the NHLBI Exome Sequencing Project (ESP) Exome Variant Server (~6,000 control samples) (**Supplementary Table S3** online). These seven nucleotide changes were also absent in the 1000 Genomes Project[17] and the NCBI database of genetic variation.[18] Furthermore, we sequenced 500 healthy volunteers for these seven missense variants and none of the studied samples showed the described changes (**Supplementary Table S3** online and **Supplementary Figure S1** online). Not finding prior documentation of a variant in a mutation database is not in itself evidence of pathogenicity, and most missense variants are rare due to factors such as rapid population growth and weak purifying selection.[19] Interestingly, four of our seven (57%) missense mutations were clustered in the exon 10 of JMJD1C, where the only two previously reported JMJD1C missense mutations were also found[14,15] (**Figure 1a**). In addition, the seven single-nucleotide shifts caused an amino acid change involving a different functional group (**Supplementary Table S3** online) that could alter important domains of the JMJD1C protein, such as the nuclear localization signal and the JmjC hydroxylase domain (**Figure 1a**). Unfortunately, for the understanding of the functional consequences of the JMJD1C mutations undergoing study, the 3D structure of this protein is only available for the C-terminal end (amino acid positions from 2,157 to 2,497) corresponding to the JmJC domain.[20] From the described mutations only c. 6997A>G (T/A 2333) (**Supplementary Table S3** online) is included in the available structure, and the identified change affects a critical amino acid in one of the beta strands that confers the core beta barrel for histone substrate interaction (**Figure 1b**). Conventional Sanger sequencing confirmed the GSJUNIOR sequencing results (**Figure 1c**). The seven point mutations described occurred in one case of Rett syndrome that was not associated with MECP2, CDKL5, and FOXG1 defects,

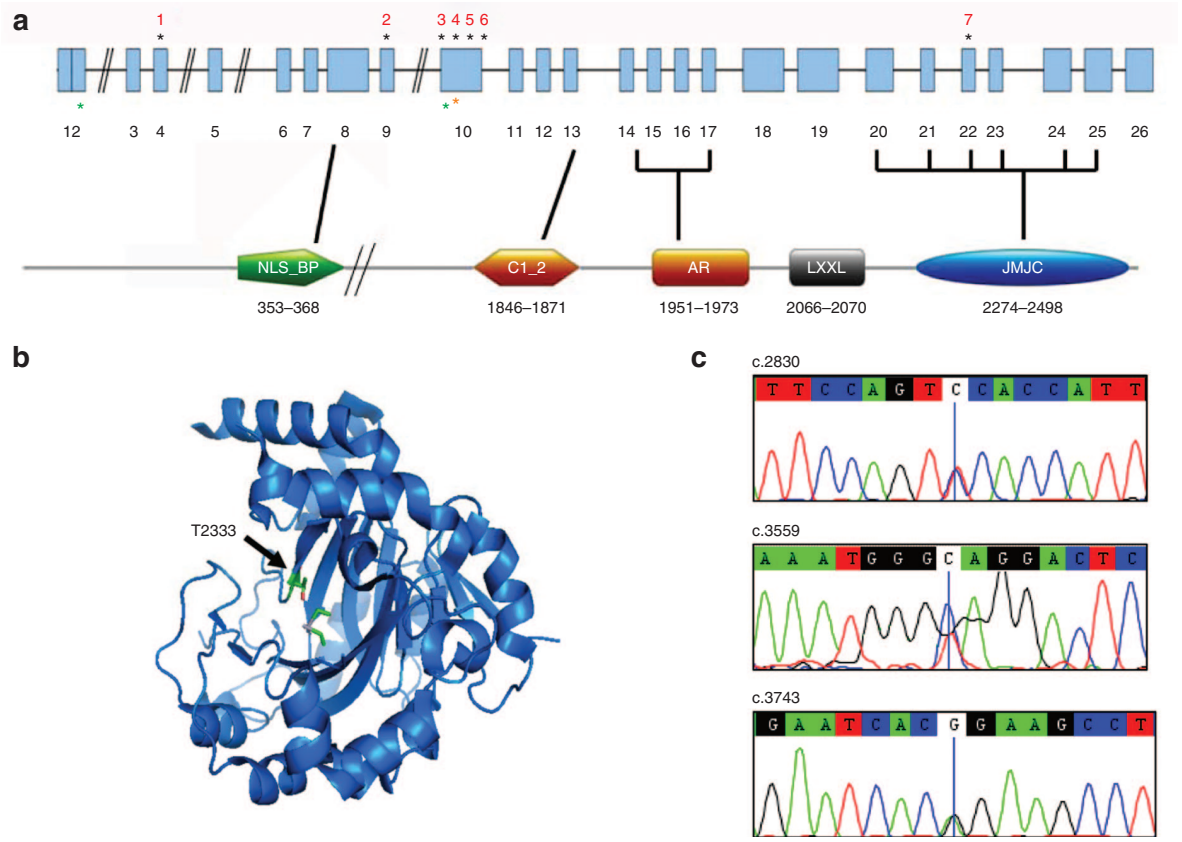


Figure 1 Diagram of JMJD1C and mutations found. (**a**) Black asterisks indicate the position of mutations identified in the JMJD1C gene. The three previous mutations identified by Iossifov et al. and Neal et al. are indicated by green and orange asterisks, respectively. The most important motif and domains are: NLS_Bp, bipartite nuclear localization signal; C1_2, phorbol esther/diacylglycerol-binding domain; AR, androgen receptor-interacting zone; LxxL, motif involved in transcriptional regulation; JMJC, JmjC hydroxylase domain. (**b**) JMJD1C 3D structure for the C-terminal end corresponding to the JmjC domain derived from the Protein Data Bank (PDB) code 2YPD, X-Ray diffraction data for 2.1 Angstroms resolution. The T2333 amino acid is indicated by a black arrow. (**c**) Chromatograms of Sanger sequencing showing c.2830C>T, c.3559A>G, and c.3743A<G mutations.

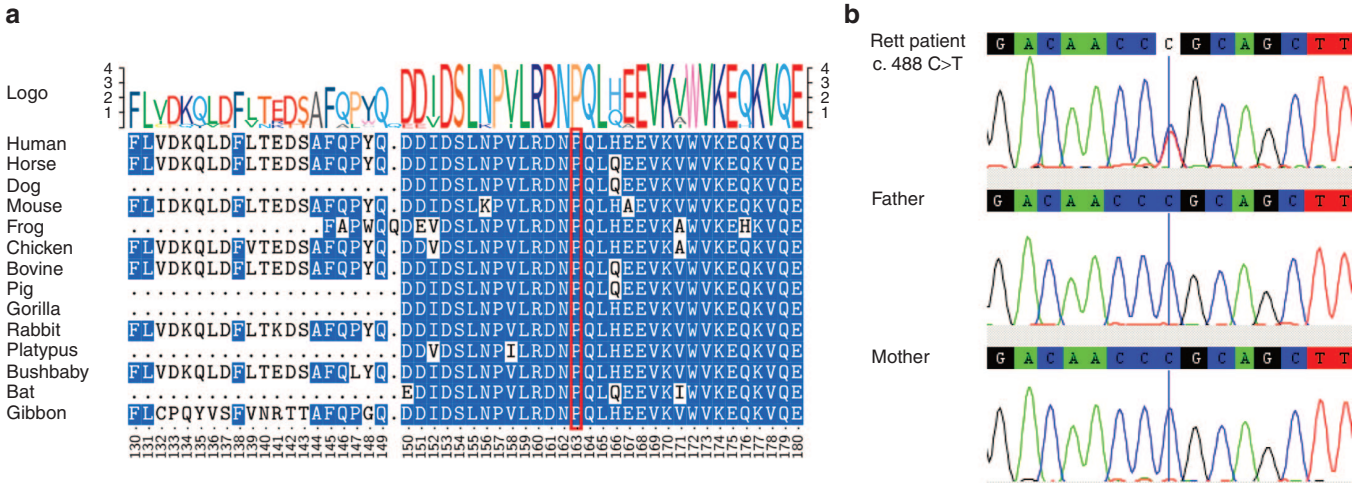


Figure 2 Characterization of the JMJD1C Pro163Leu mutation. (**a**) The Pro163Leu missense mutation is in a highly conserved region of JMJD1C. (**b**) Chromatogram of Sanger sequencing confirms c.488 A>T, de novo mutation, and its absence in the parents.

in three autistic patients, and in three patients with intellectual disability (**Supplementary Table S3** online).

Our experimental, genetic, and clinical experiences with Rett syndrome enabled us to select the newly identified JMJD1C nucleotide change in this severe form of autism spectrum disorder and further characterize its functional relevance. The c. 488C>T nucleotide change in exon 4 of the JMJD1C gene causes a proline-to-leucine shift in codon 163 of the

protein. The potential pathogenic involvement of the JMJD1C-Pro163Leu change is also suggested because the wild-type proline amino acid is highly conserved in all mammalian JMJD1Cs (**Figure 2a**), and because it is included in the balanced JMJD1C inversion that occurs in the aforementioned autistic patient.[10]

Our patient with the JMJD1C-Pro163Leu mutant is a 29-year-old female who was born at term, weighed 3,800 g, and presented a 9/10 APGAR value. She showed stagnation in head growth and also in normal development at 14 months old and lost social interaction at 18 months old. The patient started to use single words at 20 months old and propositive sentences at 24 months old, but only babbled at 36 month old. The patient also presented gait dyspraxia, hand-washing stereotype, learning impairment, teeth grinding, air swallowing, kyphoscoliosis, and tonic epilepsy. She was diagnosed as having classical Rett syndrome from the clinical standpoint, but without mutations in the known MECP2, CDKL5, and FOXG1 Rett-associated genes. Sanger sequencing of the progenitors showed that neither of the parents was a carrier of the described JMJD1C nucleotide change (**Figure 2b**). Thus, the described mutation can be considered a de novo germline event. In this regard, from the other five identified JMJD1C missense changes that were not present in the NHLBI Exome Variant

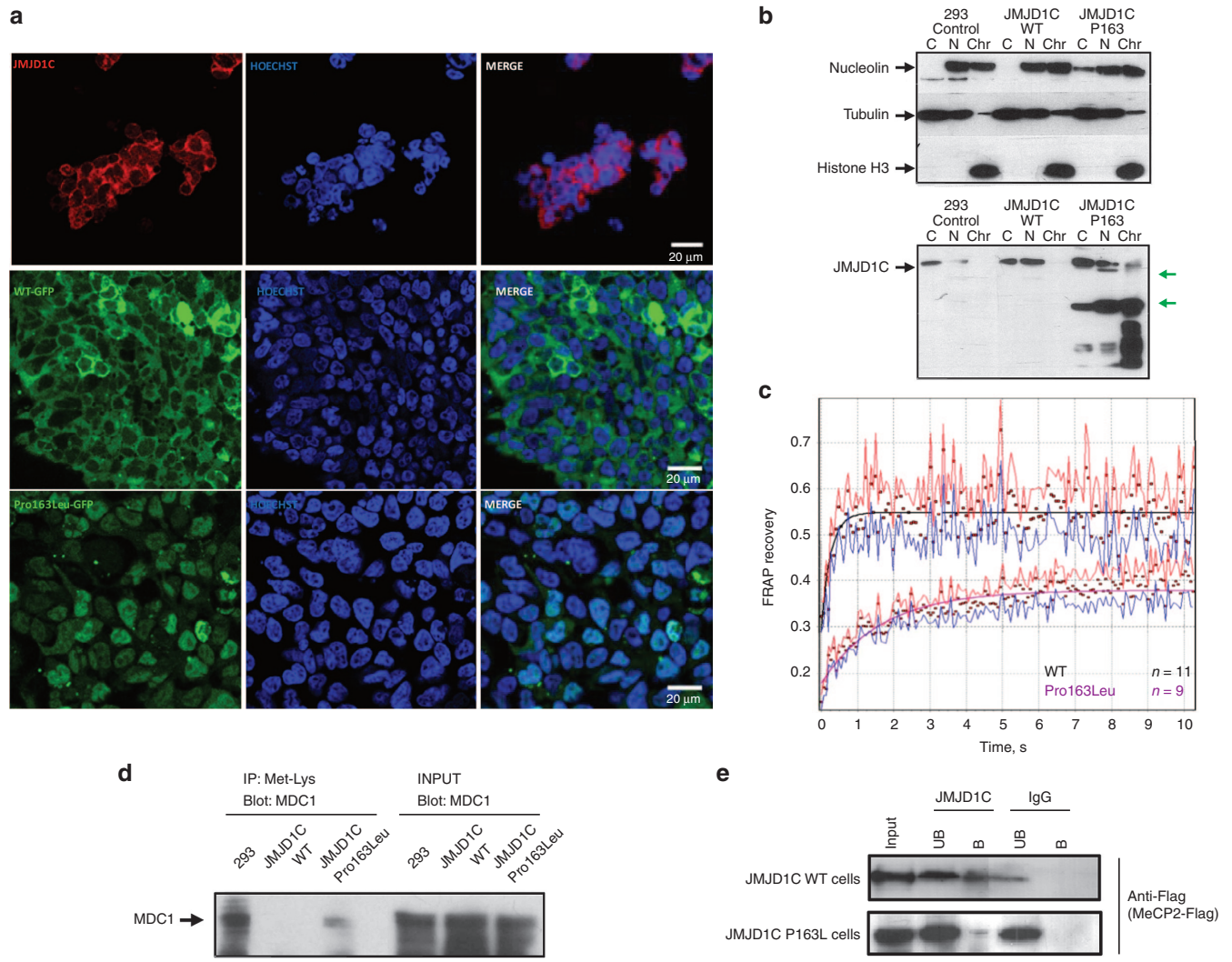


Figure 3 Differential subcellular localization, processing, and chromatin binding of JMJD1C-Pro163. (**a**) Upper panels show immunofluorescence with anti-JMJD1C antibody in untransfected cells showing that the endogenous protein is localized predominantly in the cytoplasm; middle panels show clones expressing JMJD1C-WT- GFP, which is also localized in the cytoplasm; bottom panels show that clones expressing JMJD1C-P163-GFP have a nuclear signal. (**b**) Clones and control cells were fractionated to separate the cytoplasm (C), nuclear (N), and chromatin (Chr) fractions. Proteins were processed for western blot and blotted with compartment markers (upper panel) and with anti-JMJD1C (lower panel). JMJD1C-WT-GFP looks like endogenous protein (black arrow), but JMJD1C-P163-GFP shows several processed bands in nucleus and chromatin (green arrows). (**c**) Normalized fluorescence recovery after photobleaching assay (FRAP) in clones overexpressing JMJD1C-WT-GFP and JMJD1C-P163-GFP. The bleached area was situated above the nuclei. The mutant P163L recovers more slowly than WT and has a larger immobile fraction. (**d**) Chromatin fractions from Hek293 control cells and Hek293 clones expressing JMJD1C-WT-GFP and JMJD1C-Pro163Leu-GFP were immunoprecipitated with anti-Methyl-lysine antibodies and blotted with anti-MDC1. (**e**) JMJD1C transfected HEK293F cells were co-transfected with MeCP2-Flag and subjected to immunoprecipitation assay with anti-JMJD1C antibody, anti-MeCP2 antibody, and control rabbit IgG. The western blot using anti-Flag shows the interaction between JMJD1C and MeCP2 when JMJD1C is immunoprecipitated (top). JMJD1C-Pro163Leu mutant protein cannot efficiently bind to MECP2 (below). B, bound fraction; UB, unbound fraction.

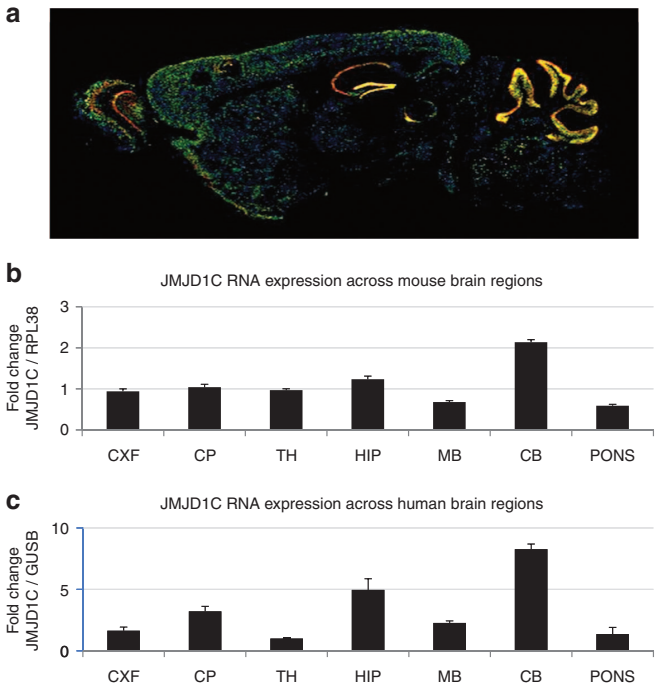


Figure 4 JMJD1C expression in human and mouse brain regions. CXF refers to frontal cortex; CP, Caudatus-Putamen; TH, thalamus; HIP, hippocampus; MB, midbrain; CB, cerebellum. (**a**) ISH analysis performed using a 2-month-old male mouse sample extracted from the Allen Brain Atlas Database (©2014 Allen Institute for Brain Science. Allen Mouse Brain Atlas [Internet]. Available at http://mouse.brain-map.org/). (**b**) Quantitative RT-PCR using three 18-month-old female samples. The y axis shows the fold-change in JMJD1C RNA expression, normalized with respect to thalamus expression, in relation to the RPL38 housekeeping gene. Error bars represent the SEM across different samples. (**c**) Quantitative RT-PCR using a sample from a 64-year-old male. The y axis shows the fold-change in JMJD1C RNA expression, normalized with respect to thalamus expression, in relation to the GUSB housekeeping gene. Error bars represent the SEM across different experiments.

Server, we identified by conventional Sanger sequencing an additional germline de novo mutation. The c. 3559A>G nucleotide change (T/A amino acid change) occurred in a patient with intellectual disability (case 4 in **Supplementary Table S3** online) but it was not present in the parents (**Supplementary Figure S2** online). In a recently developed statistical framework to distinguish disease-relevant mutations from background variation due to random expectation,[21] JMJD1C (NM_032776) was reported as having a mutation probability of 9.0804e-05.[21] Thus, out of 215 samples we would expect to find 0.0195 mutated, so discovering 2 de novo mutations out of 215 samples is higher than expected by chance (102.4-fold, binomial test, P = 1.9e-4). One possible way to explain the higher rate of observed JMJD1C mutations than that in the described study[21] could be related to the fact that the previously analyzed 5,000 samples corresponded to autism spectrum disorders,[21] whereas our work jointly profiled autism with Rett syndrome and intellectual disability. Most importantly, our two de novo mutations were only identified in these last two entities. Thus, the identified JMJD1C mutation rate could match the one reported[21] because we did not find any de novo mutation in our 69 cases of autism spectrum disorders. Interestingly, when we applied the PMUT predictor for these two de novo JMJD1C mutations,[22] a neural network trained with neutral and pathogenic mutations extracted from the protein database SwissVar, we found that both of them were considered potentially pathogenic. To discard any other pathogenic mutation in our Rett syndrome patient, the patient and the healthy parents were analyzed by whole exome sequencing as described in Methods. The only de novo mutation identified in the patient was the described c. 488C>T nucleotide change in exon 4 of the JMJD1C. The complete exome sequencing data of these three samples are available at the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the codes SRX667201 (patient), SRX667384 (mother), and SRX667386 (father).

Functional analyses of a Rett syndrome–associated JMJD1C-Pro163Leu mutation

To address the effects of the identified JMJD1C-Pro163Leu mutation, we first studied the intracellular localization of the protein. Endogenous JMJD1C was mainly cytoplasmic in Hek293 cells (**Figure 3a**), a pattern that was also observed in the transfected JMJD1C-WT-GFP in stable clones (**Figure 3a**). However, clones expressing the mutant JMJD1C-Pro163Ala-GFP forms show a strong nuclear mark (**Figure 3a**). To characterize the subcellular scenarios in which the mutant JMJD1C protein was located in more detail, we fractionated cytoplasmic, nuclear, and chromatin-bound proteins. We found that if endogenous and transfected JMJD1C proteins were almost absent from the chromatin fraction, then the mutant JMJD1C-Pro163Ala-GFP was markedly enriched in the chromatin (**Figure 3b**). The use of the fluorescence recovery after photobleaching assay in nuclear areas confirmed that mutant JMJD1C-Pro163Ala-GFP had reduced recovery in comparison with the JMJD1C wild-type transfected protein (**Figure 3c**). Most importantly, we also found that the mutant JMJD1C-Pro163Ala protein was less efficient in demethylating a non-histone target of JMJD1C that has been recently identified, MDC1 (mediator of DNA-damage checkpoint 1), a regulator of the RAP80-BRCA1 branch of the DNA damage-response pathway.[13] Using immunoprecipitation with an antibody against methylated lysines followed by western blot with the MDC1 antibody, we observed that Hek293 cells overexpressing the wild-type JMJD1C protein efficiently demethylated the MDC1 protein but that the mutant JMJD1C-Pro163Ala-GFP form showed diminished activity (**Figure 3d**). Thus, the studied JMJD1C mutation might also impair the repair of DNA damage in the studied patient. However, JMJD1C could also exert additional functions. In this regard, using JMJD1C-transfected HEK293F cells that we have co-transfected with MECP2-Flag, we observed the interaction between JMJD1C and MECP2 by using the immunoprecipitation assay (**Figure 3e**). Importantly, the JMJD1C-Pro163Leu mutant protein cannot efficiently bind to MECP2 (**Figure 3e**). These findings could explain the role of JMJD1C in Rett syndrome, a disease mainly associated with de novo mutations in MECP2.[3]

Once we had shown the aberrant functions of the mutant JMJD1C protein in the above model, we wondered whether

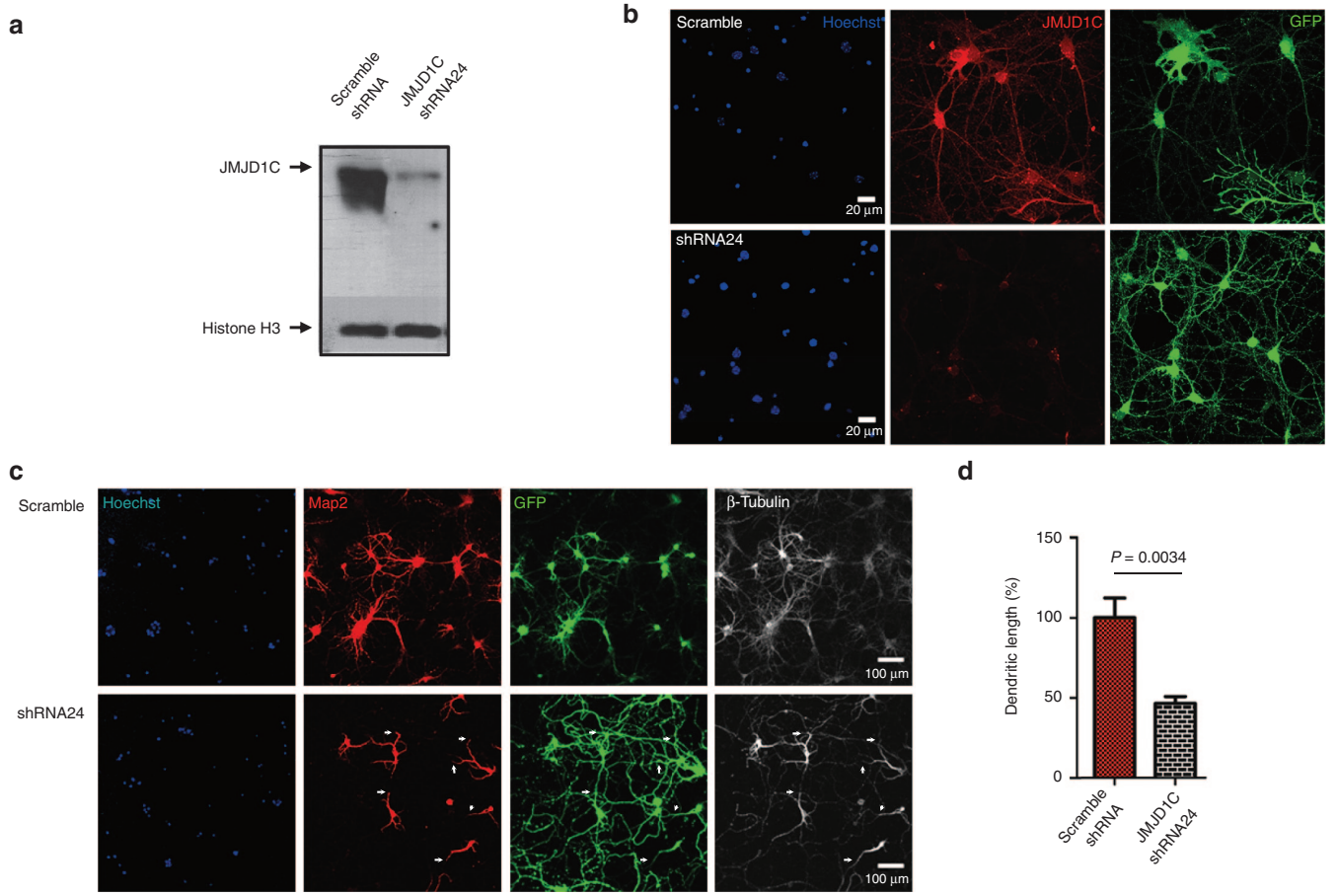


Figure 5 JMJD1C expression in primary neuron culture and shRNA knockdown. 3DIV neuron primary cultures were infected with lentiviral vectors expressing shRNAs against JMJD1C or scramble shRNA. Neurons were maintained in 14DIV culture. (**a**) Western blot shows the bands of JMJD1C and the effect of a scramble shRNA and shRNA against JMJD1C on the exon 24. Only the shRNA24 reduces the expression. Anti-H3 histone antibodies were used for normalization. (**b**) Immunofluorescence against JMJD1C. Neurons infected with shRNA24 reduce the expression of JMJD1C; infected neurons were identified with GFP tracer under CMV promoter in viral vector and nuclei stained with Hoechst3342. (**c**) Map2 immunostaining showing dendritic branching in primary cell culture neurons transfected with JMJD1C shRNA24 or scramble shRNA (SCR; Control). JMJD1C downregulation via shRNA24 induced a marked reduction of the Map2 signal, demonstrating a decrease in the length of neuronal dendrites (red). (**d**) The dendritic length (Map2 staining) in primary cell culture neurons transfected with JMJD1C shRNA24 or scramble shRNA was calculated using IMAGEJ software. Five fields per condition were analyzed and the results are plotted as a bar graph. Bar errors represent SEM across different fields and *P* values are indicated. shRNA-mediated downregulation of JMJD1C significantly decreased dendritic length.

the disruption of the wild-type JMJD1C protein had any cellular effect on a neuronal system. We first confirmed that JMJD1C was expressed in different brain regions by examining the Allen Brain Atlas Database (http://www.brain-map.org/) (**Figure 4a**).[23] We also validated the widespread expression of JMJD1C throughout distinct regions of the mouse (**Figure 4b**) and human (**Figure 4c**) brain by quantitative RT-PCR. Then, we used the short hairpin RNA (shRNA) approach to analyze the impact of its loss in neurons. Lentiviral shRNA-mediated depletion of JMJD1C was performed in primary neuronal cultures obtained from neonatal mouse hippocampus and JMJD1C protein downregulation was observed using western blot (**Figure 5a**) and immunofluorescence (**Figure 5b**). We studied the existence of changes in the dendrites in these JMJD1C-impaired cells. A significant reduction in the complexity of the dendritic process is known to be a common hallmark in Rett syndrome.[24] In this regard, and using immunofluorescence for the dendrite marker Map2, we found that JMJD1C shRNA-depleted neurons showed diminished dendritic branching in comparison with shRNA-scrambled control neurons (**Figure 5c,d**). Thus, a functional JMJD1C protein seems to be required for a correct dendritic pattern, and a defect in this gene could be associated with the dendritic impairment observed in autistic patients.

DISCUSSION

Our findings suggest that mutations in the candidate histone demethylase JMJD1C contribute to the development of intellectual disability, including well-defined clinical entities such as Rett syndrome in those cases in which the usual mutations in MECP2, CDKL5, and FOXG1 are not present. Importantly, from a functional standpoint, a JMJD1C mutant protein is unable to correctly demethylate the MDC1 DNA repair-response protein but the wild-type JMJD1C protein plays a key role in dendritic activity.

It is interesting to note that many of the genes mutated in intellectual disability have an epigenetic component,[25,26] and JMJD1C can now be included in this growing list. Epigenetics can be

broadly defined as the inheritance of gene activity that does not depend on the "naked" DNA sequence, and it includes the control of gene transcription by several chemical marks, such as DNA methylation and histone modifications. Examples include the methyl-CpG-binding protein MeCP2 in Rett syndrome, the chromatin remodeling protein CHD7 in CHARGE syndrome, and the histone methyltransferase NSD1 in Sotos syndrome.[25,26] The latter disorder is a representative case of how the importance of post-translational modifications of histone N-terminal tails in the genetic origin of neurodevelopmental disorders is becoming increasingly recognized. In this context, two mental retardation syndromes have been linked to mutations in the histone lysine demethylases KDM5C (previously known as JARID1C)[27] and PHF8 (also known as JHDM1F).[28] Thus, it is possible that mutations in many other members of the large histone demethylase family are involved in the development of intellectual disability.[29] In this regard, although further research is required to understand the molecular pathways that involve JMJD1C, our work highlights the increasing contribution of the genetic disruption of epigenetic genes to human neurodevelopmental disorders.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

ACKNOWLEDGMENTS

This study was supported by the European Community's Seventh Framework Program (FP7/2007–2013) under grant agreement PITN-GA-2012–316758 of the EPITRAIN project and PITN-GA-2009–238242 of DISCHROM; ERC grant agreement 268626 of the EPINORC project; the E-RARE EuroRETT network (Carlos III Health Institute project PI071327); the Fondation Lejeune (France); MINECO projects SAF2011-22803 and CSD2006-00049; the Cellex Foundation; the Botín Foundation; the Catalan Association for Rett Syndrome; Fundación Alicia Koplowitz 2011 Grant AKOPLOWITZ11_006; the FIS project PI1002512; Grants PI10/01422, PI13/00285, CA10/01474, RD06/0020/1050, RD12/0036/008, and 2009-SGR293; and the Health and Science Departments of the Catalan government (Generalitat de Catalunya). K.S. and P.P. are EPITRAIN Research Fellows. M.E. is an ICREA Research Professor.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Publishing: Arlington, VA, 2013.
2. Anderson DK, Lord C, Risi S, et al. Patterns of growth in verbal abilities among children with autism spectrum disorder. *J Consult Clin Psychol* 2007;75:594–604.
3. Bedogni F, Rossi RL, Galli 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.
4. Bailey A, Le Couteur A, Gottesman I, et al. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 1995;25:63–77.
5. Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Arch Pediatr Adolesc Med* 2009;163:907–914.
6. Zafeiriou DI, Ververi A, Dafoulis V, Kalyva E, Vargiami E. Autism spectrum disorders: the quest for genetic syndromes. *Am J Med Genet B Neuropsychiatr Genet* 2013;162B:327–366.
7. Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol* 2014;10:74–81.
8. Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet* 2014;15:133–141.
9. Flore LA, Milunsky JM. Updates in the genetic evaluation of the child with global developmental delay or intellectual disability. *Semin Pediatr Neurol* 2012;19:173–180.
10. Castermans D, Vermeesch JR, Fryns JP, et al. Identification and characterization of the TRIP8 and REEP3 genes on chromosome 10q21.3 as novel candidate genes for autism. *Eur J Hum Genet* 2007;15:422–431.
11. Wolf SS, Patchev VK, Obendorf M. A novel variant of the putative demethylase gene, s-JMJD1C, is a coactivator of the AR. *Arch Biochem Biophys* 2007;460:56–66.
12. Kim SM, Kim JY, Choe NW, et al. Regulation of mouse steroidogenesis by WHISTLE and JMJD1C through histone methylation balance. *Nucleic Acids Res* 2010;38:6389–6403.
13. Watanabe S, Watanabe K, Akimov V, et al. JMJD1C demethylates MDC1 to regulate the RNF8 and BRCA1-mediated chromatin response to DNA breaks. *Nat Struct Mol Biol* 2013;20:1425–1433.
14. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012;485:242–245.
15. Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014;515:216–221.
16. Guerrini R, Parrini E. Epilepsy in Rett syndrome, and CDKL5- and FOXG1-gene-related encephalopathies. *Epilepsia* 2012;53:2067–2078.
17. Consortium T 1000 GP. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
18. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–311.
19. Tennessen JA, Bigham AW, O'Connor TD, et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 2012;337:64–69.
20. Vollmar M, Johansson C, Krojer T, et al. Crystal structure of the Jumonji domain of human Jumonji domain containing 1C protein. Protein Data Base PDB Id: 2YPD. 2012.
21. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014;46:944–950.
22. Ferrer-Costa C, Gelpí JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics* 2005;21:3176–3178.
23. Lein ES, Hawrylycz MJ, Ao N, et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 2007;445:168–176.
24. Percy AK. Rett syndrome. Current status and new vistas. *Neurol Clin* 2002;20:1125–1141.
25. Sanchez-Mut JV, Huertas D, Esteller M. Aberrant epigenetic landscape in intellectual disability. *Prog Brain Res* 2012;197:53–71.
26. Urdinguio RG, Sanchez-Mut JV, Esteller M. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol* 2009;8:1056–1072.
27. Jensen LR, Amende M, Gurok U, et al. Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum Genet* 2005;76:227–236.
28. Laumonnier F, Holbert S, Ronce N, et al. Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. *J Med Genet* 2005;42:780–786.
29. De Rubeis S, He X, Goldberg AP, et al.; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014;515:209–215.

INVENTORY OF SUPPLEMENTARY DATA

SUPPLEMENTARY METHODS

DNA samples

The samples used in this study consisted of 215 autism, Rett syndrome and intellectual disability patients from the Hospital Clínic, Sant Joan de Deu and Pompeu Fabra University, Barcelona, Spain. DNA was extracted from peripheral blood leukocytes using standard techniques. We measured DNA concentration with the Quant-iT Picogreen (Invitrogen) and then normalized all concentrations to 25-50 ng/µl before proceeding with the Access Array amplification.

Primer design

48 pairs of primers were designed using the Access Array Amplicon Tagging Assay design service from Fluidigm to cover all the 26 exons of the *JMJD1C* gene (NM_032776.1), including exon-intron boundaries. These primers generated 48 fragments varying in size from 360-489 bp. Amplicons were designed of approximately the same length to obtain an optimal sequencing result on the Junior 454. We also used the 96 Access Array Barcode Library (Fluidigm) to identify all the sequences in a pool of samples. This also contained the sequencing adaptors necessary for subsequent sequencing in the 454 GS Junior Sequencer.

Fluidigm access array

The Fluidigm Access Array is a microfluidic chip on which 48 patient samples and 48 primer pairs can be loaded. The outcome is a pool of 48 fragments per patient sample. By incorporating a unique identifier or barcode for each sample and the necessary sequencing adaptors, it is possible to pool the samples on the sequencing platform. Three access arrays were used to amplify the DNA samples in this study. For each experiment, 25-50 ng DNA per sample was used as input for the system. Experiments were performed according to the manufacturer's 4-Primer Amplicon Tagging protocol. Briefly, the target-specific primers were injected into the primer inlets and the sample-specific primers with their unique MID were loaded into the sample inlets along with the DNA samples and the PCR reagents. The primers and DNA mixture were then combined in the reaction chambers in the chip. After PCR, 10 µl of the samples were collected from their original wells, now containing a pool of 48 amplicons.

Verification and quantitation of harvested pcr products

Before running the samples on the GS Junior 454, we verified the amplification of the fragments using an Agilent 2100 BioAnalyzer with DNA 1000 chips, following the manufacturer's instructions. We ran 1 µl from all the amplified samples to ensure that the amplicon size and distribution were within the expected range. We also checked that primer dimer contamination was less than 25%. In addition, we obtained a concentration value used to ensure equimolar pools of amplified samples. 11 pools were obtained, which were sequenced on a 454 Titanium PicoTiterPlate device before purifying the pooled samples with Agencourt AMPure XP system (Beckman Coulter Genomics), following the manufacturer's instructions. This consists of magnetic beads that allow a high level of recovery of amplicons, efficient removal of unincorporated dNTPs, primers, primer dimers and salts.

Multiplex ligation-dependent probe amplification (mlpa)

All samples had been screened for large rearrangements in the *JMJD1C*. We designed 9 MLPA probes specific to the *JMJD1C* gene and 6 control probes according to the instructions provided by MRC-Holland (www.mrc-holland.com/pages/support_desing_synthetic_probespag.html). Probes are

available from the authors on request. Unique sequences were identified using the BLAT program from UCSC (www.genome.ucsc.edu), and care was taken to avoid the presence of known sequence variants in the probe annealing site. Probes were designed to produce PCR products differing by 3 bp to allow correct separation by size. Oligonucleotides were obtained from Sigma–Aldrich (Haverhill, UK). The signal of each probe was adjusted after visual examination of preliminary results by raising or lowering the concentration in the probe mix. PCR products were analyzed on an ABI 3100 capillary sequencer using Gene Mapper software (Applied Biosystems, Foster City, CA, USA). The proportion of each peak relative to the height of all peaks was calculated for each sample and then compared with the proportions of the corresponding peak averaged over a set of at least ten normal DNA samples. Ratios between 0.8-1.2 were considered to have a normal copy number (2n).

Sanger sequencing

The variants were validated by Sanger sequencing using a *BigDye*® Terminator v3.1 Cycle Sequencing Kit in an Applied Biosystems 3730/DNA Analyzer. The raw data were analyzed with Codon Code Aligner Software.

Exome sequencing

The patient and healthy parents were analyzed by whole exome sequencing with TruSeq Sample Preparation Kit (Illumina). Exomes were captured with TruSeq Exome Enrichment Kit (Illumina) and paired-end 100x2 sequenced with the equip HiScan SQ. The raw data were analyzed in Centre Nacional d'Anàlisi Genòmica (CNAG), in Barcelona, Catalonia, Spain. FASTQ files were analyzed as is it follows:

1.Alignment and variant calling. Sequence reads were aligned to the human reference genome build GRCh37 (hg19) by using the Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009). Properly mapped reads were filtered with SAMtools [Li, 2009], which was also used for sorting and indexing mapping files. GATK [McKenna, 2012] was used to realign the reads around known indels and for base quality score recalibration. Once a satisfactory alignment was achieved, identification of single nucleotide variants and indels was performed using GATK standard multisample variant calling protocol, including variant

recalibration (DePristo et al 2011). For the final exome sequencing analysis report we used the ANNOVAR [Wang et al 2010] annotation tool to provide additional variant information to ease the final selection of candidates. In particular, minor allele frequency (MAF), obtained from dbSNP (Sherry 2001) and 1000 Genomes project was provided to help to select previously undescribed variants in healthy population.

2.SNV. To identified de novo single nucleotide variations, the patient´s variant were filtered first for the parental variants and then for the variants of a pool of controls made up by all healthy parents included in the study. Also SIFT (Kumar 2009) and Polyphen (Adzhubei 2010) damage scores were computed to predict putative impact over protein structures. The successive application of quality control filters and the prioritization by the parameters with potential functional impact was used to construct a list of candidate genes (and variants) ranked by its uniqueness in the cases (or very low frequency in the control population, as derived from the MAFs) and the putative potential impact. The variants were validated by sanger sequencing using BigDye® Terminator v3.1 Cycle Sequencing Kit in a Applied Biosystems 3730/DNA Analyzer

3.CNV. To identify Copy Number Variation, we used the C++ software XHMM (eXome-Hidden Markov Model) (Poultney 2013). The CNV events were filtered by DGV Data Base to remove common CNV and validated by Quantitative PCR using for normalization Type-it CNV KIT primers from QIAGEN and 2 multicopy amplicons with similar results.

Cell culture and vectors

JMJD1C coding sequence in pCMV6-AC-GFP vector was purchased from Origene (RG214878). The mutants were generated with Mutant QuikChange™ Site-Directed Mutagenesis Kit. WT and Pro163Leu, His2336Ala mutant was subcloned in pCMV6-Entry vector to introduce Myc-DDK-tag. shRNAs against the coding sequence of mouse Jmjd1c gene were cloned in pLVX-shRNA2 vector between the BamHI and EcoRI restriction sites (shRNA24 target: CAGAGACTGCTTGAGGAAT). Hek293 cells were cultivated in DMEM 10% FBS. To generate stable WT or mutant clones, Hek293 cells were transfected with Lipofectamine 2000 (Invitrogen), selected with G418 antibiotic, and individual clones were isolated 2 weeks later. For transient expression, 6 mg of

vector were transfected in 35 mm 6-well plates with jetPRIME™ transfection reagent following the manufacturer's instructions. Primary cultures of hippocampal neurons were prepared from neonate mice (P0). Brains were dissected out with forceps on ice and placed in petri dishes containing ice-cold Hibernate® media, meninges were removed under a dissecting microscope and hippocampuses were removed and saved in Hibernate. The tissues were transferred to 15-ml tubes containing trypsin-EDTA (0.025% in PBS). The tubes were incubated in a 37°C chamber for 15 minutes and agitated every 5 minutes. After stopping trypsinization with 5 ml 20% FBS, these tissues were triturated 15 times with a Pasteur pipette. The suspension was filtered through a 75 µm pore-sized filter and centrifuged in a tube for 2 minutes at 1000 *g*. The cells were resuspended in DMEM 10% FSB, Glutamax-Pyruvate and seeded at $2x10^6$ cell/cm^2. Three hours later, the medium was replaced with Neurobasal medium (GIBCO) containing 2% B27, 0.5 mM glutamine, and 50 U/ml penicillin/streptomycin, AraC 5 □M and cultured at 37°C, 5% CO_2. Cultures were infected at 3DIV with lentiviral vectors to express scramble or shRNAs against JMJD1C together with a GFP tracer (pLVX-shRNA2 system). Coverslips were fixed and protein was extracted at 15DIV.

Cellular fractionation

Cell were harvested by scratching, washed in PBS buffer, and incubated 5 minutes in RBS buffer (10mM Tris HCl pH7.6, 10mM NaCl, 1.5mM MgCl2, 0.1% NP40) and centrifuge 1 minute at max speed. Supernatant was considerate cytoplasm fraction. The pellet was washed in once in RBS buffer and then incubated 5 minutes with RIPA (50 mM Tris HCl pH 8.0, 150mM NaCl, 1.0% NP-40, 0.5% Sodium Deoxycholate), the supernatant was considerate nuclear fraction, the pellet containing the chromatin fraction was resuspended en RIPA and sonicated to fragment the DNA.

Immunoprecipitation

750µg of chromatin fraction was diluted 10 fold in IP buffer (5mM Tris-HCl pH 7.6, 15mM HEPES pH 8.0, 1mM EDTA, 1mM EGTA, 0.1% SDS, 1% Triton X-100), incubated with 2µg of antibodies anti-Me-Lysine (Abcam ab23366) overnight at 4°C and 2 hours with PureProteome Protein A/G Magnetic Beads.

Beads were washed twice with IP buffer, twice in RBS NP-40 and eluted in laemli buffer in reduction condition at 70°C by 10 minutes. For the MeCP2 immunoprecipitation procedure, anti-JMJD1C and anti-MeCP2 antibodies were coupled to Dynabeads Protein G (Invitrogen). JMJD1C transfected HEK293F cells were transiently transfected with MeCP2-Flag tagged plasmid and the nuclear fraction was obtained by RIPA buffer (10 mM TRIS-Cl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 1% TRITON x-100, 0.1 % Sodium deoxycholate, 0.1% SDS and 140 mM NaCl) supplemented with protease inhibitors (complete, ROCHE) and hybridized with the antibodies at 4ºC for 2h. 150 mM NaCl RBS buffer was used for washing. Human IgG was used as negative control. Anti-Flag HRP (M2-SIGMA) antibody was used to visualize the binding.

Western blot

Protein extract of Hek293 cells and primary neuronal culture was obtained in RIPA buffer supplemented with cOmplete Protease Inhibitor Cocktails tablet (Roche) and sonicated. Protein concentrations were determined using a DC Protein Assay kit from Bio-RAD. 50 µg of each protein sample were denatured in Laemli buffer 4% β-mercaptoethanol for 10 minutes at 95ºC and separated on a 7.5% or 15% SDS-polyacrylamide gel, then transferred onto a PVDF membrane (Immobilon-P, Millipore) by liquid electroblotting for 90 minutes at 100 V. The membrane was blocked in 5% nonfat dry milk in TBS-0.05% Tween 20. The antibodies and dilutions used are as follows: rabbit anti-JMJD1C 1:2000 (Millipore 09-817), mouse anti-nucleolin 1:1000 (Santa-Cruz SC-8031), rabbit anti-MeCP2 1:5000 (Sigma M9317), rabbit anti-H3 1:10000 (Abcam AB1791); mouse anti-H3 1:4000 (Abcam AB10799), rabbit anti-H3K9Me2 1:4000 (Abcam AB32521), rabbit anti-MDC1 1:5000 (Abcam AB11171). The blots were developed with Luminata™ Crescendo Western HRP Substrate or with the LiCor Odyssey System.

Immunofluorescence

Cells were fixed in 4% PFA-PBS, quenched in 100 mM glycine-PBS, permeabilized with 0.25% Triton X-100, 1% BSA, PBS. The cells were blocked with 0.2% gelatin, 0.25% triton X-100. Antibody dilutions were prepared in 0.25% Triton X-100, 1% BSA, PBS. The dilutions used were: rabbit anti-

JMJD1C 1:200, chicken anti-Map2 1:5000, anti β-tubulin 1:1500 (Abcam AB21058). Nuclei were stained with 2 mg/ml Hoechst 33342. Coverslips were mounted in ProLong® Gold antifade reagent.

Microscopy

Confocal images were captured with a Leica SP5 confocal microscope. For FRAP analysis the cells were maintained at 37ºC in an atmosphere of 5% CO_2. We captured images every 70 μs at 63x, 128x128 resolution, at 1400Hz with bidirectional acquisition. We captured 25 control images at 3% laser transmission before bleaching, then bleached the ROI inside a nucleus 25 times at nominal level of 100% laser transmission. For this experiment, 150 images were captured after bleaching. The raw data were analyzed with FrapAnalyzer Software (http://actinsim.uni.lu/eng/Downloads/FRAPAnalyser).

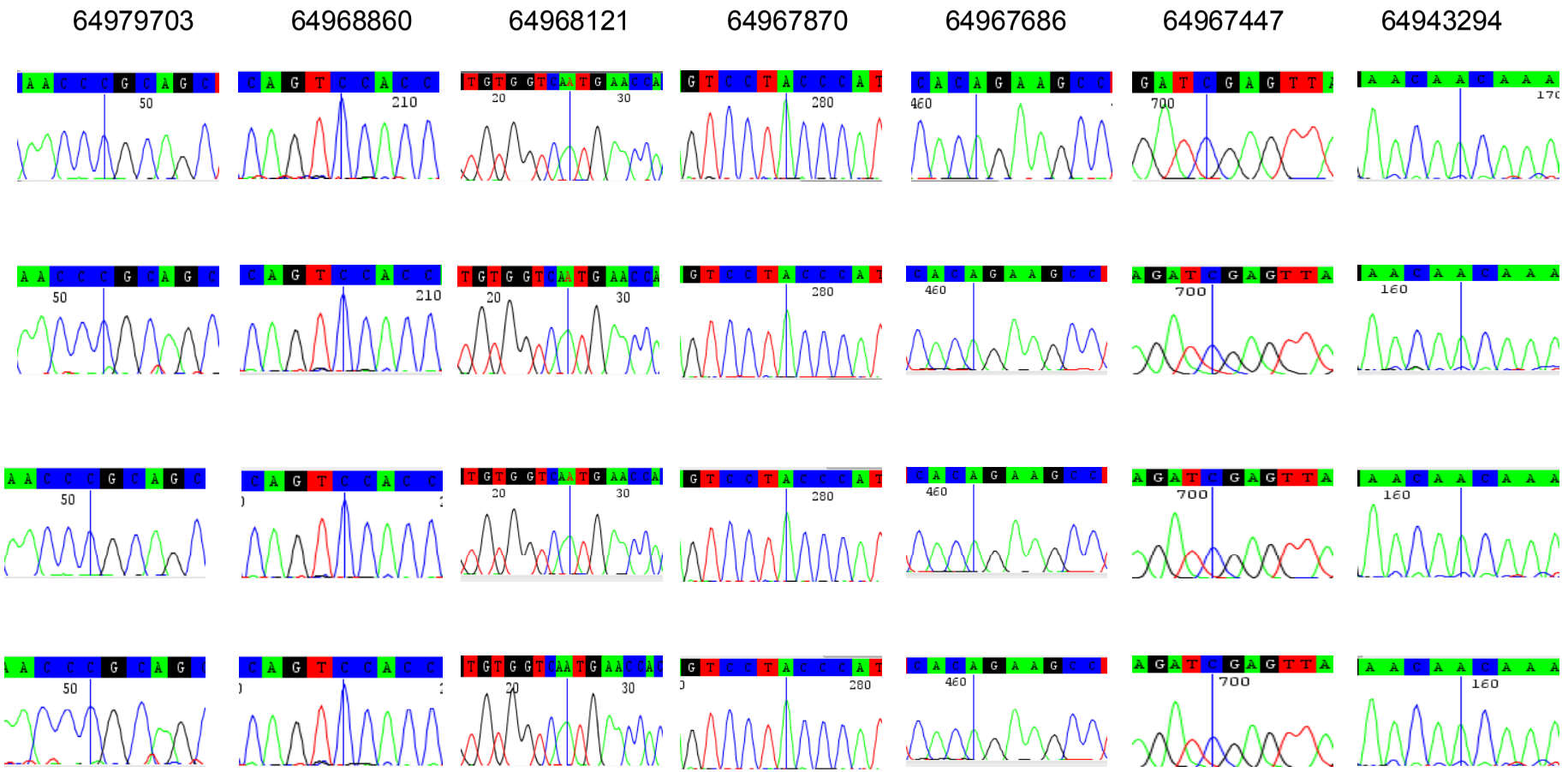
RNA extraction and real-time PCR

Total RNA was extracted from cell lysates using TRIzol Reagent (Invitrogen), purified using the RNeasy Kit (Qiagen) and 2 μg were retrotranscribed using the ThermoScriptTM RT-PCR System (Invitrogen). Real-time PCR reactions were performed in triplicate on an Applied Biosystems 7,900HT Fast Real-Time PCR system using 20 ng cDNA, 5 μl SYBR Green PCR Master Mix (Applied Biosystem) and 150 nM specific primers (sequences are available upon request) in a final volume of 10 μl.

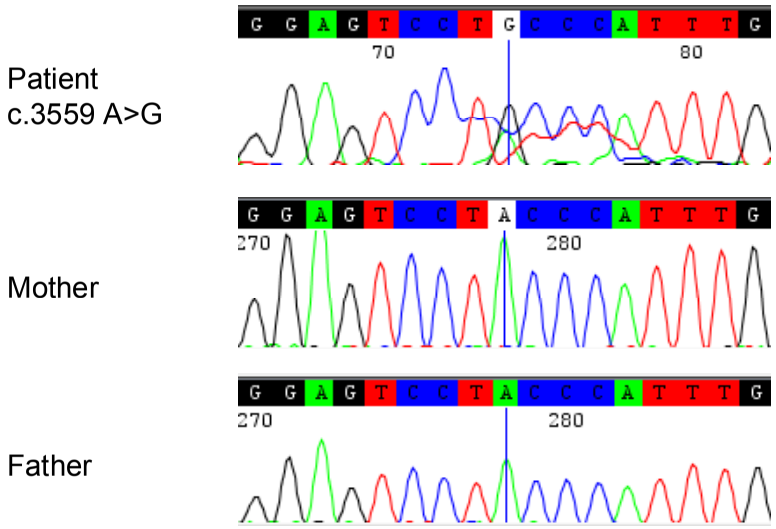
Supplementary Table S1. JMJD1C synonymous variants found in the studied patients					
	Chr 10 position	**Nucleotide Change**	**Exon**	**Amino acid**	**Protein Position**
1	64974871	c. 1056A>G	8	R	352
2	64973860	c. 2067T>A	8	I	689
3	64973890	c.2037T>C	8	H	679
4	64968828	c. 2862T>C	9	H	954
5	64967478	c. 3951T>C	10	S	1317
6	64967445	c. 3984T>A	10	R	1146
7	64966764	c. 4665C>A	10	L	1555
8	64966860	c. 4569C>A	10	I	1523
9	64952834	c.5940G>A	16	P	1980
10	64945364	c. 6789C>T	20	D	2263

Supplementary Table S2. Previously informed JMJD1C variants in studied patients							
	Chr10 position	**Nucleotide Change**	**Exon/ Intron**	**Amino acid change**	**Protein Position**	**Exome Variant Server**	**Patient**
1	64975327	c. 808G>A	Exon 8	V/I	270	1/6019	Intellectual Disability
2	64974224	c. 1703A>T	Exon 8	D/V	568	15/5953	1 Intellectual Disability 1 Autism Spectrum Disorder
3	64973978	c.1949C>T	Exon 8	T/I	650	114/5981	1 Autism Spectrum Disorder 1 Intellectual Disability
4	64967249	c. 4180A>T	Exon 10	T/S	1394	8/6058	1 Intellectual Disability 1 Autism Spectrum Disorder
5	64966621	c. 4808T>C	Exon 10	I/T	1603	2/5954	Intellectual Disability
6	64945336	c. 6817A>G	Exon 20	M/V	2273	1/5918	Autism Spectrum Disorder
7	64948925	c.6570+3G>A	Intron 18	-	-	107/5896	Autism Spectrum Disorder

Supplementary Table S3. JMJD1C mutational status in studied patients									
	Chr 10 position	**Nucleotide Change**	**Exon**	**Amino acid change**	**Protein Position**	**Exome Variant Server**	**Patient**	**Gender**	**Parental Status**
1	64979703	c. 488C>T	4	P/L	163	0/5953	Rett syndrome	Female	Wild Type
2	64968860	c. 2830C>T	9	P/S	944	0/5943	Autism Spectrum Disorder	Male	Not Available
3	64968121	c. 3308A>G	10	N/S	1103	0/5989	Intellectual Disability	Female	Not Available
4	64967870	c. 3559A>G	10	T/A	1187	0/6086	Intellectual Disability	Male	Wild Type
5	64967686	c. 3743A>G	10	Q/R	1248	0/5946	Intellectual Disability	Female	Not Available
6	64967447	c. 3982C>G	10	R/G	1328	0/6000	Autism Spectrum Disorder	Male	Not Available
7	64943294	c. 6997A>G	22	T/A	2333	0/5880	Autism Spectrum Disorder	Male	Not Available



Supplementary Figure S1. Illustrative chromatograms of Sanger sequencing from control population (n=500). Numbers represent chromosome position and nucleotides of interest are marked with anvertical line .



Supplementary Figure S2. Chromatograms of Sanger sequencing showing the *de novo* status of the c.3559 A>G JMJD1C mutation. Parents are wild type for the amino acid.

ANNEX II

Hum Genet (2016) 135:1343–1354
DOI 10.1007/s00439-016-1721-3

ORIGINAL INVESTIGATION

Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype

Mario Lucariello[1] · Enrique Vidal[1] · Silvia Vidal[2] · Mauricio Saez[1] · Laura Roa[1] · Dori Huertas[1] · Mercè Pineda[3] · Esther Dalfó[4] · Joaquin Dopazo[5,6,7] · Paola Jurado[1] · Judith Armstrong[2,8,11] · Manel Esteller[1,9,10]

Received: 15 May 2016 / Accepted: 31 July 2016 / Published online: 19 August 2016

Abstract Classical Rett syndrome (RTT) is a neurodevelopmental disorder where most of cases carry *MECP2* mutations. Atypical RTT variants involve mutations in *CDKL5* and *FOXG1*. However, a subset of RTT patients remains that do not carry any mutation in the described genes. Whole exome sequencing was carried out in a cohort of 21 female probands with clinical features overlapping with those of RTT, but without mutations in the customarily studied genes. Candidates were functionally validated by assessing the appearance of a neurological phenotype in *Caenorhabditis elegans* upon disruption of the corresponding ortholog gene. We detected pathogenic variants that accounted for the RTT-like phenotype in 14 (66.6 %) patients. Five patients were carriers of mutations in genes already known to be associated with other syndromic neurodevelopmental disorders. We determined that the other patients harbored mutations in genes that have not previously been linked to RTT or other neurodevelopmental syndromes, such as the ankyrin repeat containing protein *ANKRD31* or the neuronal acetylcholine receptor subunit alpha-5 (CHRNA5). Furthermore, worm assays demonstrated that mutations in the studied candidate genes caused locomotion defects. Our findings indicate that mutations in a variety of genes contribute to the development of RTT-like phenotypes.

Introduction

Rett syndrome (RTT, MIM 312750) is a postnatal progressive neurodevelopmental disorder (NDD), originally

Electronic supplementary material The online version of this article (doi:10.1007/s00439-016-1721-3) contains supplementary material, which is available to authorized users.

✉ Paola Jurado
pjurado@idibell.cat

✉ Judith Armstrong
jarmstrong@hsjdbcn.org

✉ Manel Esteller
mesteller@idibell.cat

[1] Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet, 08908 Barcelona, Catalonia, Spain

[2] Servei de Medicina Genètica i Molecular, Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Esplugues De Llobregat, Catalonia, Spain

[3] Fundació Hospital Sant Joan de Déu (HSJD), Barcelona, Catalonia, Spain

[4] Genetics Department, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Catalonia, Spain

[5] Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), 46012 Valencia, Spain

[6] Bioinformatics of Rare Diseases (BIER), CIBER de Enfermedades Raras (CIBERER), Valencia, Spain

[7] Functional Genomics Node (INB) at CIPF, 46012 Valencia, Spain

[8] CIBER Enfermedades Raras, Barcelona, Catalonia, Spain

[9] Department of Physiological Sciences, School of Medicine and Health Sciences, University of Barcelona, Barcelona, Catalonia, Spain

[10] Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

[11] Department of Neurology, Hospital Sant Joan de Déu (HSJD), Barcelona, Catalonia, Spain

described in the 1960s by Andreas Rett (Rett 1966), that most frequently manifests itself in girls during early childhood, with an incidence of approximately 1 in 10,000 live births (Chahrour and Zoghbi 2007). RTT patients are asymptomatic during the first 6–18 months of life, but gradually develop severe motor, cognitive, and behavioral abnormalities that persist for life. It is the second most common cause of intellectual disability in females after Down's syndrome (Chahrour and Zoghbi 2007). Around 90 % of the cases are explained by more than 800 reported mutations in the methyl CpG-binding protein 2 gene (*MECP2*) (RettBASE: MECP2 Variation Database) (Christodoulou et al. 2003), which is located in the X chromosome and which causes most of the classical or typical forms of RTT (Chahrour and Zoghbi 2007), and it was originally identified as encoding a protein that binds to methylated DNA (Lewis et al. 1992). Individuals affected by atypical or variant RTT present with many of the clinical features of RTT, but do not necessarily have all of the classic characteristics of the disorder (Neul et al. 2010). Approximately 8 % of classic RTT and 42 % of variant RTT patients are *MECP2* mutation-negative (Monros et al. 2001; Percy 2008). Some of the latter group have mutations in other genes, such as that of the cyclin-dependent kinase-like 5 (*CDKL5*), which is described in individuals with an early seizure onset variant of RTT (Kalscheuer et al. 2003) or the forkhead box G1 (*FOXG1*), which is responsible for the congenital variant of RTT (Ariani et al. 2008). However, there remains a subset of patients with a clinical diagnosis of RTT who are mutation-negative for all the aforementioned genes. Next generation sequencing (NGS) has emerged as a potentially powerful tool for the study of such genetic diseases (Zhu et al. 2015).

Herein, we report the use of a family based exome sequencing approach in a cohort of 20 families with clinical features of RTT, but without mutations in the usually studied genes. We establish the neurological relevance of the newly identified candidate genes by assessing them in *Caenorhabditis elegans* model.

Materials and methods

Patient samples

A cohort of 19 Spanish parent–child trios and one family with two affected daughters who exhibited clinical features associated with RTT were recruited at Sant Joan de Deu Hospital in Barcelona, Catalonia, Spain. These patients had been diagnosed on the basis of the usual clinical parameters (Monros et al. 2001), and according to the recently revised RettSearch International Consortium criteria and nomenclature (Neul et al. 2010), but were found to be mutation-negative for *MECP2*, *CDKL5* and *FOXG1* in the original single-gene screening. The parents were clinically evaluated and it was not observed any evidence of intellectual disability. Genomic DNA from these patients was extracted from peripheral blood leukocytes using standard techniques, and analyzed by exome sequencing at the Cancer Epigenetics and Biology Program (PEBC) in Barcelona, Catalonia, Spain. Ethical approval for the molecular genetic studies was obtained from each institutional review board.

Whole exome sequencing and Sanger validation

Coding regions were captured using the TruSeq DNA Sample Preparation and Exome Enrichment Kit (Illumina, San Diego, California). Paired-end 100 × 2 sequences were sequenced with the Illumina HiScan SQ system at the National Center for Genomic Analysis in Barcelona. We also included the exome sequencing data of an *MECP2*, a *CDKL5* and a *JMJD1C* (Sáez et al. 2016) RTT-associated family for data processing to improve the de novo single nucleotide variant calling. The complete exome sequencing data of all the studied samples are available from the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) with the ID: SRP073424 (private link for the reviewer until publication: http://www.ncbi.nlm.nih.gov/sra/SRP073424). The overall coverage statistics for each individual of the families, considering the regions captures using Exome Enrichment Kit, and number of reads in the position of the variation is shown in Supplementary Table 1. The identified variants were validated by Sanger sequencing using a BigDye® Terminator v3.1 Cycle Sequencing Kit in an Applied Biosystems 3730/DNA Analyzer. The raw data were analyzed with Codon Code Aligner Software. The primers used for Sanger sequencing are shown in Supplementary Table 2.

Caenorhabditis elegans handling

The techniques used for the culture of *Caenorhabditis elegans* were essentially as described (Brenner 1974). The worms were backcrossed at least three times to avoid background mutations. The behavior of three sets of ten animals was independently assessed in locomotion assays without food that were performed at 20 °C, as previously described (Sawin et al. 2000).

Results

Clinical criteria for selecting RTT trios

The 21 patients (derived from the 20 families studied) included in this study fulfilled the recently revised clinical criteria for the diagnosis of RTT following the usual clinical parameters (Monros et al. 2001), and the RettSearch International Consortium criteria and nomenclature (Neul et al. 2010). Specifically, all patients presented stereotypic hand movements, 90.5 % of them (19/21) showed microcephaly and also presented onset of the first signs of the disease before the age of 12 months. 66.7 % of patients (14/21) acquired motor skills, while a further seven (33.3 %), who had a more severe phenotype, never walked. Language skills were progressively lost in 28.6 % of the patients and 71.4 % of them (15/21) never acquired them. Additionally, important episodes of epilepsy were experienced by 81.0 % of the patients (17/21), and 57.1 % of them (12/21) manifested apneas and/or hyperventilation.

Bioinformatic process for filtering and selecting pathogenic variants

Before their inclusion in this study, patients underwent an extensive clinical and genetic work-up to detect genetic alterations in *MECP2*, *CDKL5*, and *FOXG1*. However, no molecular diagnosis could be established. We performed whole exome sequencing (WES) on the 61 individuals (20 pairs of healthy parents and 21 affected daughters) separately by subjecting whole blood derived genomic DNA to exome enrichment and sequencing. We focused our analysis on de novo single nucleotide variants (SNVs) due to their known relevance in autism and mental retardation-related diseases (Vissers et al. 2010). On average, WES gave rise to 419,045 variants, including SNVs and indels, of which 19,951 non-synonymous variants per family (4.7 %) were predicted to have a functional impact on the genomic sequence. To select variants that had not previously been described in the healthy population, we filtered out the variants with an allele frequency of 1 % or higher (the classic definition of a polymorphism) formerly observed in the Single Nucleotide Polymorphism database (dbSNP) and the 1000 Genomes Pilot Project data. Afterwards, to focus on de novo inheritance, patients' variants were filtered first against variants found in their own parents and then against a pool of controls comprising all the healthy parents included in the study. Following this process, we achieved an average of 106 SNVs per family, which corresponded to 81 mutated genes per family. De novo candidate variants were selected on the basis of the quality of the alignments, damage score predictors and the conservation level of each of the genes during evolution. The complete exome sequencing data of all the studied samples are available from the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra).

The global yield of genomic analysis following the bioinformatic process described herein enabled 22 coding de novo mutations to be identified in 66.7 % (14 of 21) of Rett-like patients: 20 SNVs and 2 indels. The identified variants and their de novo status were confirmed by conventional Sanger sequencing. Illustrative samples are shown in Fig. 1. Interestingly, in seven (33.3 %) of the studied RTT probands, exome sequencing did not detect any genetic change relative to their respective parents. The clinical characteristics of these seven patients without obvious pathogenic variants are summarized in Table 1. In one of the families, there were two affected children, and an analysis of potentially relevant recessive variants was performed. For the recessive analysis, and following the same criteria to define a variant as deleterious, we selected the variants with homozygous recessive genotype, and then at the gene level, we also selected the genes presenting more than one heterozygotic variant in the same gene (compound heterozygosity). We did not find any candidate gene consistent with the phenotype of the family with the two affected sisters.

Variants in genes previously associated with neurodevelopmental disorders

Of the 22 identified coding de novo mutations in the assessed RTT-like patients, five (22.7 %) occurred in genes previously associated with neurodevelopmental disorders that presented a clinicopathological phenotype with features coinciding with those of Rett syndrome (Table 2). In particular, we identified four mutations in genes such as *HCN1* (Nava et al. 2014) and *GRIN2B* (Endele et al. 2010; Lemke et al. 2014), which are associated with early infantile epileptic encephalopathy; *SLC6A1*, which is associated with epilepsy and myoclonic-atonic seizures (Carvill et al. 2015); *TCF4*, which is associated with Pitt–Hopkins syndrome (Sweatt 2013); and *SCN1A*, which is associated with Dravet syndrome (Brunklaus and Zuberi 2014) (Table 2). The clinical characteristics of these five patients with variants in genes previously associated with neurodevelopmental phenotypes are summarized in Table 3. A comparison of the clinical features of our RTT-like patients, where we have identified mutations in candidate genes previously associated with other neurodevelopmental disorders, with those observed for these diseases is summarized in Table 4.

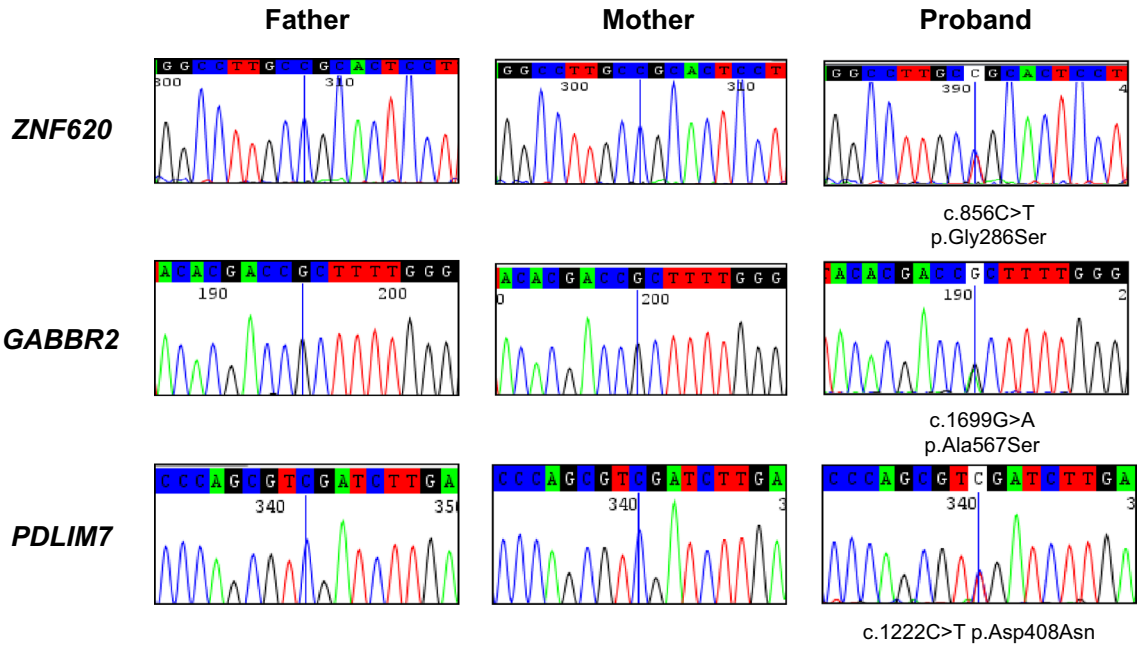


Fig. 1 Sanger sequencing validation of the de novo variants identified by exome sequencing. Illustrative examples for ZNF620 (c.856C > T p.Gly286Ser), GABBR2 (c.1699G > A p.Ala567Ser) and PDLIM7 (c.1222C > T p.Asp408Asn) are shown

Table 1 Clinical summary of patients without exome candidates

Proband	Age (years)	Onset of signs	Micro-cephaly	Sitting alone	Ambula-tion	Respira-tory func-tion	Epilepsy	Hand use	Stereotyp-ies	Language	Total score
1	15	1	1	1	1	0	1	1	1	1	8
2	28	3	1	1	0	1	2	2	3	2	15
5.1	7	3	1	3	4	1	0	3	2	2	19
5.2	5	3	1	1	2	0	0	2	3	2	14
7	16	2	1	0	0	0	2	1	2	1	9
15	8	3	0	1	2	0	2	1	1	1	10
18	3.5	3	1	1	1	0	0	1	3	0	10

Clinical scores of our series of patients according to Pineda scale. Severity classification ranges from 0 to 4 as follows: age of onset of first signs (1: >24 months; 2: 12–24 months; 3: 0–12 months), microcephaly (0: absent; 1: present), sitting alone (1: acquired < 8 months; 2: seat and mantains; 3: seat and lost), ambulation (0: acquired < 18 months, 1: acquired < 30 months; 2: acquired > 30 months; 3: lost acquisition; 4: never acquired), respiratory function (0: no dysfunction; 1: hyperventilation and/or apnea), epilepsy (0: absent; 1: present and controlled; 2: uncontrolled or early epilepsy), hands use (0: acquired and conserved; 1: acquired and partially conserved; 2: acquired and lost; 3: never acquired), onset of stereotypies (1: > 10 years, 2: > 36 months; 3: 18-36 months) and languages (0: preserved and propositive; 1: lost; 2: never acquired). The total score is the sum of the scores of each clinical feature

Variants in genes previously not associated with neurodevelopmental disorders

Of the 22 identified coding de novo variants in the RTT-like patients assessed here, 17 (77.3 %) occurred in genes that had not previously been associated with neurodevelopmental disorders (Table 5). However, two of these variants were associated with non-neurodevelopmental disorders: a *BTBD9* variant linked to restless leg syndrome (Kemlink et al. 2009), and an *ATP8B1* SNV associated with familial cholestasis (Klomp et al. 2004), respectively. Interestingly, the *BTBD9* variant was detected in the same patient that carried the *SCN1A* variant associated with Dravet syndrome (Table 2). The other 15 potentially pathogenic variants identified occurred in genes that had not been linked to any genetic disorder of any type. However, there was an enrichment of genes with a potential role in neuronal biology and functionality, such as the gamma-aminobutyric type

Table 2 List of patients with variants found in genes previously associated with neurodevelopmental phenotypes

Proband	Gene	Protein	NM number	Variant: genomic coordinates	cDNA change	Protein change	Gene-disease association
4	*HCN1*	Hyperpolarization Activated Cyclic Nucleotide Gated Potassium Channel 1	NM_021072.3	5:45396665	c.1159G > T	p.Ala387Ser	Early infantile epileptic encephalopathy 24
8	*SCN1A*	Sodium Channel Protein Type I Subunit Alpha	NM_001165963.1	2:166866266	c.3965C > G	p.Arg1322Thr	Dravet syndrome
10	*TCF4*	Transcription Factor 4	NM_001243236.1	18:52901827	c.958delC	p.Gln320Ser_fs8X	Pitt–Hopkins syndrome
11	*GRIN2B*	Glutamate receptor ionotropic, NMDA 2B	NM_000834.3	12:13764782	c.1657C > A	p.Pro553Thr	Autosomal Dominant Mental Retardation 6; Early infantile epileptic encephalopathy 27
17	*SLC6A1*	Solute Carrier Family 6 Member 1	NM_003042.3	3:11067528	c.919G > A	p.Gly307Arg	Myoclonic-atonic epilepsy and schizophrenia

B receptor subunit 2 (*GABBR2*), the neuronal acetylcholine receptor subunit alpha-5 (*CHRNA5*), the Huntington-associated protein 1 (*HAP1*), the axon guider semaphorin 6B, the ankyrin repeat containing proteins *ANKRD31* and *AGAP6*, and the neuronal voltage-gated calcium channel *CACNA1* (Table 5). Proband 14 was a particularly interesting case in which four potential pathogenic variants were present, affecting zinc finger (*ZNF620*), a nucleolar complex (*NOC3L*), G patch domain (*GPATCH2*) and GRAM domain (*GRAMD1A*)-related proteins (Table 5). The clinical characteristics of these patients with variants in genes previously not associated with neurodevelopmental disorders are summarized in Table 6.

Neurological phenotype of candidate genes in *C. elegans*

To demonstrate a neurological effect for a loss of function of the detected genes that had not previously been associated with neurodevelopmental disorders (Table 5), we used the model organism *C. elegans* to confirm the genotype-phenotype correlation. We obtained all the available *C. elegans* mutants that carry deleterious mutations in the orthologous genes to those human genes with potentially pathogenic mutations in the patients. In this model, backcrossing is a commonly used procedure to obtain a specific mutant strain without any secondary mutations from its genetic composition. Under these conditions, we were able to test six available mutant strains that were backcrossed at least three times to prove that any observed phenotype was really associated to specific mutations in the orthologous genes. To this end, we studied the *C. elegans* mutants carrying deleterious mutations in the gene orthologs of the human genes *PDLIM7*, *ANKRD31*, *ZNF620*, *CHRNA5*, *MGRN1* and *GABBR2* described in Table 7. Considering that the loss of normal movement and coordination is one of the clearest signs shown by Rett patients, we performed a locomotion assay of the nematodes as previously described (Sawin et al. 2000), using the wild-type N2 strain as a control (Supplementary Video 1). We observed that in 83.3 % (5 of 6) of the cases the mutation of the ortholog of the human exome sequencing identified genes in *C. elegans* exhibited a locomotion defective phenotype (Fig. 2). The most severe phenotypes were represented by *alp-1*, *unc-44* and *pag-3*, with mutations in the orthologs of PDZ and LIM domain protein 7 (*PDLIM7*), ankyrin repeat containing protein *ANKRD31* and the zinc protein *ZNF620*, respectively (Fig. 2 and Supplementary Videos 2, 3 and 4). The case of *alp-1* was particularly interesting, because mutant worms were not only thinner than usual and completely locomotion defective, but they exhibited transitory spasms. Significant defects, such as slower locomotion and uncoordinated movement, were also observed in the mutants of *unc-63* and *C11H1.3*, the *C. elegans* orthologs

Table 3 Clinical summary of the patients with variants in genes previously associated with neurodevelopmental phenotypes

Proband	Gene variant	Age (years)	Onset of the signs	Microcephaly	Sitting alone	Ambulation	Respiratory function	Epilepsy	Hand use	Stereotypies	Language	Total score
4	*HCN1*	24	3	1	1	3	1	2	3	1	2	16
8	*SCN1A, MGRN1, BTBD9*	7	3	1	1	4	1	2	1	3	2	18
10	*TCF4*	16	3	1	1	2	1	2	1	1	2	14
11	*GRIN2B, SEMA6B*	3	2	0	1	3	0	1	1	1	2	10
17	*SLC6A1*	36	3	0	1	1	0	1	1	1	0	7

Clinical scores of our series of patients according to Pineda scale. Severity classification ranges from 0 to 4 as follows: age of onset of first signs (1: > 24 months; 2: 12–24 months; 3: 0–12 months), microcephaly (0: absent; 1: present), sitting alone (1: acquired < 8 months; 2: seat and mantains; 3: seat and lost), ambulation (0: acquired < 18 months, 1: acquired < 30 months; 2: acquired > 30 months; 3: lost acquisition; 4: never acquired), respiratory function (0: no dysfunction; 1: hyperventilation and/or apnea), epilepsy (0: absent; 1: present and controlled; 2: uncontrolled or early epilepsy), hands use (0: acquired and conserved; 1: acquired and partially conserved; 2: acquired and lost; 3: never acquired), onset of stereotypies (1: > 10 years, 2: > 36 months; 3: 18-36 months) and languages (0: preserved and propositive; 1: lost; 2: never acquired). The total score is the sum of the scores of each clinical feature

of the genes coding for the neuronal acetylcholine receptor subunit alpha-5 (*CHRNA5*) and mahogunin RING finger protein 1 (*MGRN1*), respectively. Although we did not find a clear locomotion defect in the *gbb-2* mutant (the ortholog of *GABBR2*) (Fig. 2), it occurs in the *gbb-1*;*gbb-2* double mutant (Dittman and Kaplan 2008), *gbb-1* being the *C. elegans* ortholog of *GABBR1* (gamma-aminobutyric acid type B receptor subunit 1). The clinical picture of the particular RTT cases with mutations in the genes studied in *C. elegans* is shown in Table 6.

Discussion

Our results indicate that the existence of de novo variants in genes with potential neurological functionalities, such as neuronal receptors (*GABBR2* and *CHRNA5*), axon guiders (*SEMA6B*), synaptic ionic channels (*CACNA1I*) and others, contribute to the development of RTT-like clinical phenotypes in the context of wild-type sequences for standard Rett genes such as *MECP2* and *FOXG1*. These patients share most of the clinicopathological features of classic RTT syndrome, such as stereotypic hand movements, relative microcephaly, and onset of the disease after the age of 12 months. Thus, exome sequencing is a powerful tool for genetically characterizing these enigmatic cases. In this regard, once a new candidate gene has been identified, it is now possible to design specific sequencing strategies for the molecular screening of this particular target in larger populations of patients with intellectual disability. The strategy based on exome sequencing patients who have RTT features, but no known mutations in the usual genes, has recently been used in other smaller series of patients (Grillo et al. 2013; Okamoto et al. 2015; Hara et al. 2015; Olson et al. 2015; Lopes et al. 2016). Most importantly, our study and the aforementioned previous reports strengthen the concept that a mutational heterogeneous profile hitting shared neurological signaling pathways contributes to RTT-like syndromes. Examples of confluence in the same molecular crossroads include the gamma-aminobutyric type B receptor subunit 2 (*GABBR2*) de novo variant, described here, and the formerly identified variant in the gamma-aminobutyric acid receptor delta gene (*GABRD*) (Hara et al. 2015). Interestingly, a second RTT-like patient has been identified as being a carrier of a de novo *GABBR2* variant (Lopes et al. 2016), highlighting the likelihood that this gene and pathway contribute to the clinical entity. Another example of similarly targeted genes in RTT-like patients is that of the proteins containing ankyrin-repeats that are involved in postsynaptic density (Durand et al. 2007). This study has revealed de novo variants in the ankyrin repeat containing proteins *AGAP6* and *ANKRD31* in RTT-like patients, and the presence of de novo variant of the SH3 and multiple ankyrin repeat domain3 protein (*SHANK3*) (Hara et al. 2015) and ankyrin-3 (*ANK3*) (Grillo et al. 2013) has been reported in two RTT-like patients. A final example of the convergence of cellular pathways to provide a common RTT-like phenotype is represented by the disruption of the ionic channels. We found the existence of a voltage-gated calcium channel subunit alpha 11 (*CANA1I*) de novo variant in an RTT-like patient. Additionally, the presence of de novo variants in the calcium release channel *RYR1* (Grillo et al. 2013) and the sodium voltage-gated channel alpha subunit 2 (*SCN2A*) (Baasch et al. 2014) in two other RTT-like

Table 4 Comparison of the clinical features of RTT-like patients from who we have identified mutations in candidate genes previously associated with other neurodevelopmental disorders with those observed for these diseases

Disease	Rett		Atypical Rett		Pitt–Hopkins		Dravet		EEIE27		MAE		EEIE24	
GENE	*MECP2*		*CDKL5*		*TCF4*		*SCN1A*		*GRIN2B*		*SLC6A1*		*HCN1*	
OMIM/Patient	312,750	6	308,350	16	602,272	10	182,390	8	616,139	11	616,421	17	615,871	4
Onset age	6–18 m	12–24 m	1–3 m	0–12 m	6–12 m	0–12 m	2 days–7 m	0–12 m	0–24 m	12–24 m	0–24 m	0–12 m	0–24 m	0–12 m
Microcephaly	Yes	Yes	Yes	Yes	Yes	Yes	±	Yes	±	No	±	No	NA	Yes
Hypotonia	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	±	Yes	±	Yes	NA	Yes
Epilepsy	80 %	Yes	Yes	Yes	Yes	Yes	Yes	Yes	±	Yes	Yes	Yes	Yes	Yes
Respiratory dysfunction	80 %	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	NA	Yes
Expressive language dysfunction	Yes	Yes	Yes	Yes	Yes	Yes	±	Yes	Yes	Yes	No	No	NA	Yes
Preserved use of hands	No	No	±	No	No	No	Yes	No	±	No	±	No	NA	No
Stereotypies	Yes	Yes	Yes	Yes	Yes	Yes	±	Yes	±	Yes	±	Yes	NA	Yes
Inheritance	XL		XL		AD		AD		AD		AD		AD	

EEIE epileptic encephalopathy, early infantile, *MAE*, myoclonic-atonic epilepsy, *m* months, *XL* X linkage, *AD* autosomal dominant, *NA* unavailable

probands have been reported. It is also intriguing that in our study a variant in HAP was found, whereas in similar series heterozygous variants in huntingtin (HTT) have been described (Lopes et al. 2016; Rodan et al. 2016), further reinforcing the links between Huntington's disease and Rett syndrome (Roux et al. 2012). Another interesting case is provided by TCF4, which is associated with Pitt–Hopkins syndrome (Sweatt 2013), where in addition to our study, others have found mutations in RTT-like patients (Lopes et al. 2016). This observation could be of interest for clinicians due to phenotypic similitudes such as intellectual disability, stereotypic movement, apneas and seizures (Marangi et al. 2012).

Our findings also suggest that a substantial degree of clinical overlap can exist between the features associated with RTT and those of other neurodevelopmental disorders. Our exome sequencing effort indicated that probands originally diagnosed as RTT-like patients were, in fact, carriers of well-known pathogenic de novo mutations linked to Dravet Syndrome (*SCN1A*), myoclonic-atonic epilepsy (*SCLC6A1*), or early infantile epileptic encephalopathies 24 (*HCN1*) and 27 (*GRIN2B*). The purely clinical classification of these patients, without a thorough genetic study, can be difficult because some of these patients are composites that carry at least two pathogenic variants. For example, in our cases, the Dravet syndrome patient also had a de novo variant in *BTBD9* associated with the development of restless leg syndrome. In addition, among the newly identified candidate genes associated with RTT-like features, a few of these patients simultaneously carried two de novo variants (e.g., probands 8, 19 and 21), further complicating the tasks of correctly diagnosing and managing these individuals.

Finally, the studies performed in *C. elegans* validate the functional relevance for nervous system function of the newly proposed candidate genes. Future studies would be necessary to assess the role of the specific variants identified, such as rescuing the defects with the expression of normal cDNAs versus cDNAs containing the mutation, ideally using cDNAs of human origin to prove similar function of the gene in the two species. It is also relevant to mention that for some of the newly reported mutated genes in our RTT-like patients, there are mice models targeting the described loci that show neurological phenotypes such as BTBD9 (motor restlessness and sleep disturbances) (DeAndrade et al. 2012), MGRN1 (spongiform neurodegeneration) (He et al. 2003), SEMA6B (aberrant mossy fibers) (Tawarayama et al. 2010), CHRNA5 (alterations in the habenulo-interpeduncular pathway) (Fowler et al. 2011), GABBR2 (anxiety and depression-related behavior) (Mombereau et al. 2005) and HAP1 (depressive-like behavior and reduced hippocampal neurogenesis) (Chan et al. 2002; Xiang et al. 2015).

Table 5 List of patients with variants in new candidate disease genes

Proband	Gene	Protein	Function	NM number	Variant: genomic coordinates	cDNA change	Protein change	ExAC	SIFT	Polyphen2	PROVEAN	Mutation Taster2	Conservation
3	*AGAP6*	ArfGAP with GTPase domain, ankyrin repeat and PH domain 6	Putative GTPase-activating protein	NM_001077665.2	10:51748528	c.53insC	p.Asp18Ala_fs10X	Not present	NA	NA	B	P	405
8	*MGRN1*	Mahogunin RING Finger Protein 1	E3 ubiquitin-protein ligase	NM_001142290.2	16:4723583	c.880C > T	p.Arg294Cys	0.000077	P	P	P	P	573
8	*BTBD9*	BTB (POZ) Domain-Containing 9	Putative protein–protein interactor	NM_001099272.1	6:38256093	c.1409C > T	p.Ala470Val	Not present	B	P	P	P	512
11	*SEMA6B*	Semaphorin-6B	Role in axon guidance	NM_032108.3	19:4555540	c.508G > A	p.Gly170Ser	Not present	P	P	P	P	510
12	*VASH2*	Vasohibin 2	Angiogenesis inhibitor	NM_001301056.1	1:213161902	c.1044A > C	p.Glu348Asp	Not present	B	B	B	B	473
13	*CHRNA5*	Neuronal acetylcholine receptor subunit alpha-5	Excitator of neuronal activity	NM_000745.3	15:78882481	c.748C > A	p.Pro250Thr	Not present	B	P	P	P	519
14	*ZNF620*	Zinc Finger Protein 620	Transcriptional regulator	NM_175888.3	3:40557941	c.856G > A	p.Gly286Ser	Not present	P	P	P	P	317
14	*GRAMD1A*	GRAM Domain-Containing 1A	Not described	NM_020895.3	19:35506764	c.1106G > A	p.Arg369His	Not present	P	P	P	P	358
14	*NOC3L*	Nucleolar complex protein 3 homolog	Regulator of adipogenesis	NM_022451.10	10:96097586	c.2137G > A	p.Ala713Thr	Not present	B	B	B	B	0
14	*GPATCH2*	G patch domain-containing protein 2	Regulator of cell proliferation	NM_018040.3	1:217784371	c.878G > A	p.Gly293Asp	Not present	B	P	P	P	304

Table 5 continued

Proband	Gene	Protein	Function	NM number	Variant: genomic coordinates	cDNA change	Protein change	ExAC	SIFT	Polyphen2	PROVEAN	Mutation Taster2	Conservation
19	*GABBR2*	Gamma-aminobutyric acid type B receptor subunit 2	Inhibitor of neuronal activity	NM_005458.7	9:101133817	c.1699G > A	p.Ala567Thr	Not present	P	P	P	P	412
19	*ATP8B1*	Phospholipid-transporting ATPase IC	Aminophospholipid translocator	NM_005603.4	18:55328507	c.2606C > T	p.Thr869Ile	Not present	P	P	P	P	361
20	*HAP1*	Huntingtin-Associated Protein 1	Vesicular transporter	NM_177977.2	17:39890655	c.232G > A	p.Ala78Thre	Not present	P	B	B	B	0
21	*PDLIM7*	PDZ and LIM domain protein 7	Scaffold protein	NM_005451.4	5:176910933	c.1222G > A	p.Asp408Asn	Not present	P	P	B	P	515
21	*SRRM3*	Serine/Arginine Repetitive Matrix 3	Splicing activator	NM_001291831.1	7:75890878	c.655C > G	p.Ser218Cys	Not present	P	P	P	P	491
22	*ANKRD31*	Ankyrin Repeat Domain 31	Not described	NM_001164443.1	5:74518166	c.196A > T	p.Ile66Phe	Not present	P	P	B	B	401
23	*CACNA1I*	Voltage-Gated Calcium Channel Subunit Alpha 1I	Calcium signaling in neurons	NM_021096.3	22:40066855	c.4435C > T	p.Leu1479Phe	Not present	B	P	B	P	695

ExAC, frequency of the identified variants in the exome aggregation consortium. Four in silico prediction tools of functional mutation impact were used: 'Sorting Tolerant From Intolerant' (*SIFT*), 'Polymorphism Phenotyping v2' (*Polyphen2*); 'Protein Variation Effect Analyzer' (*PROVEAN*) and Mutation Taster2. The output results were classified as: likely pathogenic (*P*), likely benign (*B*) and not available (*NA*). Conservation scores refer to the conservation level of the nucleotide at the position of the identified variant between 46 species of vertebrates based on PhastCons. It ranges from 0 to 1000: the highest, the more conserved during evolution

Table 6 Clinical summary of patients with variants in new candidate disease genes

Proband	Gene variant	Age (years)	Onset of the signs	Micro-cephaly	Sitting alone	Ambu-lation	Res-piratory function	Epilepsy	Hand use	Stereo-typies	Lan-guage	Total score
3	*AGAP6*	14	3	1	2	4	1	1	3	2	2	19
8	*SCN1A, MGRN1, BTBD9*	7	3	1	1	4	1	2	1	3	2	18
11	GRIN2B, SEMA6B	3	2	0	1	3	0	1	1	1	2	10
12	*VASH2*	11	3	1	1	4	1	0	2	2	1	15
13	*CHRNA5*	10	3	1	2	4	1	1	2	3	2	19
14	ZNF620, GRAMD1A, NOC3L, GPATCH2	2	3	1	1	2	0	1	3	2	2	15
19	GABBR2, ATP8B1	2	3	1	1	4	0	0	3	2	2	16
20	*HAP1*	24	3	1	1	1	1	0	2	2	1	12
21	*PDLIM7, SRRM3*	5	3	1	3	4	0	1	2	1	2	17
22	*ANKRD31*	17	3	0	1	2	0	1	2	3	2	14
23	*CACNA1I*	1/8	3	1	1	2	1	0	3	3	2	16

Clinical scores of our series of patients according to Pineda scale. Severity classification ranges from 0 to 4 as follows: age of onset of first signs (1: >24 months; 2: 12–24 months; 3: 0–12 months), microcephaly (0: absent; 1: present), sitting alone (1: acquired < 8 months; 2: seat and mantains; 3: seat and lost), ambulation (0: acquired < 18 months, 1: acquired < 30 months; 2: acquired > 30 months; 3: lost acquisition; 4: never acquired), respiratory function (0: no dysfunction; 1: hyperventilation and/or apnea), epilepsy (0: absent; 1: present and controlled; 2: uncontrolled or early epilepsy), hands use (0: acquired and conserved; 1: acquired and partially conserved; 2: acquired and lost; 3: never acquired), onset of stereotypies (1: > 10 years, 2: > 36 months; 3: 18–36 months) and languages (0: preserved and propositive; 1: lost; 2: never acquired). The total score is the sum of the scores of each clinical feature

Table 7 Phenotype in *Caenorhabditis elegans*

Human gene	Ortholog in C.elegans	Similarity (%)	Identity (%)	Mutation in C.elegans	Locomotion phenotype	Neurological phenotypes	Other phenotypes
GABBR2	*gbb-2*	53	34	Deletion	normal	Hypersensitivity to aldicarb	–
MGRN1	*C11H1.3*	58	41	Deletion	locomotion defective	–	–
CHRNA5	*unc-63*	58	40	Deletion	locomotion defective	Uncoordinated locomotion with strong levamisole resistance	–
ZNF620	*pag-3*	65	47	Deletion	locomotion defective	Altered neurosecretion and up-regulation of DCV (Dense Core Vesicles) components	–
ANKRD31	*unc-44*	59	39	Deletion	locomotion defective	Asymmetric dynamics of axonal and dendritic microtubules defects	–
PDLIM7	*alp-1*	65	47	Deletion	locomotion defective	–	Defects in actin filament organization in muscle cells

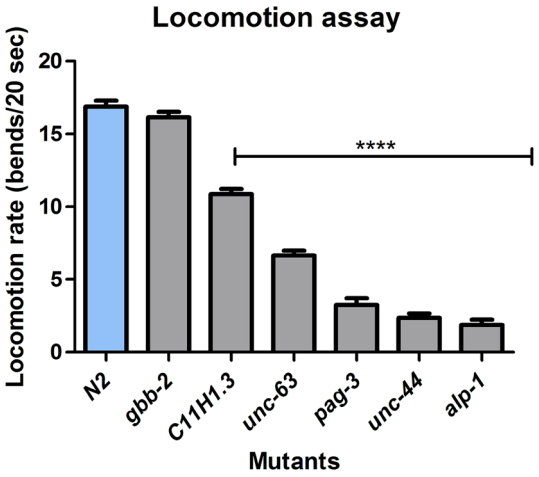


Fig. 2 Locomotion assay in *Caenorhabditis elegans*. Functional validation of mutations was performed by measuring the locomotion rate, expressed in average of measuring, *in C. elegans*. Each mutant strain was compared to a wild-type N2 control strain by measuring worm body bends during 20 s in three independent sets of experiments. Locomotion rates of mutants, represented by *C11H1.3* (*MGRN1*), *unc-63* (*CHRNA5*), *pag-3* (*ZNF620*), *unc-44* (*ANKRD31*) and *alp-1* (*PDLIM7*) are significantly lower compared to that of the N2 control strain ($p < 0.0001$), on the contrary *gbb-2* (*GABBR2*) mutant move similarly. Standard error of the mean (SEM) values is shown. *p* values obtained according to Student's *t* test. ***$p < 0.0001$

Conclusions

Overall, this study demonstrates the genetic mutational diversity that underlies the clinical diagnosis of patients with clinical features that resemble RTT cases. Once the recognized *MECP2*, *CDKL5* and *FOXG1* mutations have been discarded, exome sequencing emerges as a very useful strategy for the more accurate classification of these patients. The de novo variants identified by this approach can modify the first diagnostic orientation towards another neurodevelopmental disorder, or pinpoint new genes involved in the onset of RTT-like features. Interestingly, most of these new targets are involved in the same functional networks associated with correct neuronal functionality. Further research is required to understand the role of these proteins in the occurrence of neurodevelopmental diseases. Additional functional experiments, such as the *C. elegans* assays used in this study, would be extremely helpful for this purpose.

Acknowledgments The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2012 under REA Grant agreement PITN-GA-2012-316758 of the EPITRAIN project and PITN-GA-2009–238242 of DISCHROM; the E-RARE EuroRETT network (Carlos III Health Institute project PI071327); the Foundation Lejeune (France); the Cellex Foundation; the Botín Foundation; the Finestrelles Foundation; the Catalan Association for Rett Syndrome; and the Health and Science Departments of the Catalan government (Generalitat de Catalunya). M.E. is an ICREA Research Professor.

Compliance with ethical statement

Conflict of interest The authors declare that there is no conflict of interest associated with this manuscript.

References

Ariani F et al (2008) FOXG1 is responsible for the congenital variant of Rett syndrome. Am J Hum Genet 83:89–93

Baasch AL et al (2014) Exome sequencing identifies a de novo SCN2A mutation in a patient with intractable seizures, severe intellectual disability, optic atrophy, muscular hypotonia, and brain abnormalities. Epilepsia 55:e25–e29

Brenner S (1974) The genetics of Caenorhabditis elegans. Genetics 77:71–94

Brunklaus A, Zuberi SM (2014) Dravet syndrome–from epileptic encephalopathy to channelopathy. Epilepsia 55:979–984

Carvill GL et al (2015) Mutations in the GABA transporter SLC6A1 cause epilepsy with myoclonic-atonic seizures. Am J Hum Genet 96:808–815

Chahrour M, Zoghbi HY (2007) The story of Rett syndrome: from clinic to neurobiology. Neuron 56:422–437

Chan EY et al (2002) Targeted disruption of Huntingtin-associated protein-1 (Hap1) results in postnatal death due to depressed feeding behavior. Hum Mol Genet 11:945–959

Christodoulou J, Grimm A, Maher T, Bennetts B (2003) RettBASE: the IRSA MECP2 Variation Database—a new mutation database in evolution. Hum Mutat 21:466–472

DeAndrade MP et al (2012) Motor restlessness, sleep disturbances, thermal sensory alterations and elevated serum iron levels in Btbd9 mutant mice. Hum Mol Genet 21:3984–3992

Dittman JS, Kaplan JM (2008) Behavioral impact of neurotransmitter-activated G-protein-coupled receptors: muscarinic and GABAB receptors regulate Caenorhabditis elegans locomotion. J Neurosci 28:7104–7112

Durand CM et al (2007) Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet 39:25–27

Endele S et al (2010) Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat Genet 42:1021–1026

Fowler CD, Lu Q, Johnson PM, Marks MJ, Kenny PJ (2011) Habenular α5 nicotinic receptor subunit signalling controls nicotine intake. Nature 471:597–601

Grillo E et al (2013) Revealing the complexity of a monogenic disease: Rett syndrome exome sequencing. PLoS One 8:e56599

Hara M, Ohba C, Yamashita Y, Saitsu H, Matsumoto N, Matsuishi T (2015) De novo SHANK3 mutation causes Rett syndrome-like phenotype in a female patient. Am J Med Genet A 167:1593–1596

He L et al (2003) Spongiform degeneration in mahoganoid mutant mice. Science 299:710–712
Kalscheuer VM et al (2003) Disruption of the serine/threonine kinase 9 gene causes severe X-linked infantile spasms and mental retardation. Am J Hum Genet 72:1401–1411
Kemlink D et al (2009) Replication of restless legs syndrome loci in three European populations. J Med Genet 46:315–318
Klomp LW et al (2004) Characterization of mutations in ATP8B1 associated with hereditary cholestasis. Hepatology 40:27–38
Lemke JR et al (2014) GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. Ann Neurol 75:147–154
Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 69:905–914
Lopes F et al (2016) Identification of novel genetic causes of Rett syndrome-like phenotypes. J Med Genet 53:190–199
Marangi G et al (2012) Proposal of a clinical score for the molecular test for Pitt-Hopkins syndrome. Am J Med Genet A 158A:1604–1611
Mombereau C, Kaupmann K, Gassmann M, Bettler B, van der Putten H, Cryan JF (2005) Altered anxiety and depression-related behaviour in mice lacking GABAB(2) receptor subunits. NeuroReport 16:307–310
Monrós E, Armstrong J, Aibar E, Poo P, Canós I, Pineda M (2001) Rett syndrome in Spain: mutation analysis and clinical correlations. Brain Dev 23:S251–S253
Nava C et al (2014) De novo mutations in HCN1 cause early infantile epileptic encephalopathy. Nat Genet 46:640–645
Neul JL et al (2010) Rett syndrome: revised diagnostic criteria and nomenclature. Ann Neurol 68:944–950
Okamoto N et al (2015) Targeted next-generation sequencing in the diagnosis of neurodevelopmental disorders. Clin Genet 88:288–292
Olson HE et al (2015) Mutations in epilepsy and intellectual disability genes in patients with features of Rett syndrome. Am J Med Genet A 167A:2017–2025
Percy AK (2008) Rett syndrome: recent research progress. J Child Neurol 23:543–549
Rett A (1966) On an unusual brain atrophy syndrome in hyperammonemia in childhood. Wien Med Wochenschr 116:723–772
Rodan LH et al (2016) A novel neurodevelopmental disorder associated with compound heterozygous variants in the huntingtin gene. Eur J Hum Genet. doi:10.1038/ejhg.2016.74
Roux JC, Zala D, Panayotis N, Borges-Correia A, Saudou F, Villard L (2012) Modification of Mecp2 dosage alters axonal transport through the Huntingtin/Hap1 pathway. Neurobiol Dis 45:786–795
Sáez MA et al (2016) Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. Genet Med 18:378–385
Sawin ER, Ranganathan R, Horvitz HR (2000) C. elegans locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. Neuron 26:619–631
Sweatt JD (2013) Pitt-Hopkins syndrome: intellectual disability due to loss of TCF4-regulated gene transcription. Exp Mol Med 45:e21
Tawarayama H, Yoshida Y, Suto F, Mitchell KJ, Fujisawa H (2010) Roles of semaphorin-6B and plexin-A2 in lamina-restricted projection of hippocampal mossy fibers. J Neurosci 30:7049–7060
Vissers LE et al (2010) A de novo paradigm for mental retardation. Nat Genet 42:1109–1112
Xiang J, Yan S, Li SH, Li XJ (2015) Postnatal loss of hap1 reduces hippocampal neurogenesis and causes adult depressive-like behavior in mice. PLoS Genet 11:e1005175
Zhu X et al (2015) Whole-exome sequencing in undiagnosed genetic disorders: interpreting 119 trios. Genet Med 17:774–781

Supplementary Table 1 Technical statistics of WES. Overall coverage statistics for each individual of the families, considering the regions captures using Exome Enrichment Kit, and number of reads in the position of the variation.

Gene	IDs	Average of coverage	Numer of reads	Variant: genomic coordinates
AGAP6	**Patient 3**	51,7222	30	10:51748528
	Mother	19,1904	59	
	Father	19,8947	49	
HCN1	**Patient 4**	37,9988	33	5:45396665
	Mother	21,2688	11	
	Father	53,6885	22	
SCN1A	**Patient 8**	63,9124	98	2:166866266
	Mother	55,4485	98	
	Father	66,9241	91	
MGRN1	**Patient 8**	63,9124	4	16:4723583
	Mother	55,4485	1	
	Father	66,9241	3	
BTBD9	**Patient 8**	63,9124	100	6:38256093
	Mother	55,4485	101	
	Father	66,9241	105	
TCF4	**Patient 10**	57,353	44	18:52901827
	Mother	45,2247	87	
	Father	58,4836	77	
SEMA6B	**Patient 11**	38,1451	21	19:4555540
	Mother	43,3847	16	
	Father	31,635	31	
GRIN2B	**Patient 11**	38,1451	45	12:13764782
	Mother	43,3847	44	
	Father	31,635	41	
VASH2	**Patient 12**	56,4905	84	1:213161902
	Mother	68,9685	84	
	Father	63,1402	109	
CHRNA5	**Patient 13**	63,6187	53	15:78882481
	Mother	65,3851	64	
	Father	58,9521	67	
ZNF620	**Patient 14**	55,267	27	3:40557941
	Mother	57,7491	34	
	Father	60,002	28	
GRAMD1A	**Patient 14**	55,267	56	19:35506764
	Mother	57,7491	42	
	Father	60,002	43	
NOC3L	**Patient 14**	55,267	26	10:96097586
	Mother	57,7491	52	
	Father	60,002	36	

GPATCH2	**Patient 14**	55,267	96	1:217784371
	Mother	57,7491	111	
	Father	60,002	104	
SLC6A1	**Patient 17**	62,2434	43	3:11067528
	Mother	56,8694	53	
	Father	61,0401	66	
GABBR2	**Patient 19**	67,6492	46	9:101133817
	Mother	35,6295	30	
	Father	65,7731	65	
ATP8B1	**Patient 19**	67,6492	61	18:55328507
	Mother	35,6295	38	
	Father	65,7731	70	
HAP1	**Patient 20**	57,3956	11	17:39890655
	Mother	63,086	19	
	Father	52,3069	15	
PDLIM7	**Patient 21**	63,5277	56	5:176910933
	Mother	59,2951	28	
	Father	58,6697	20	
SRRM3	**Patient 21**	63,5277	52	7:75890878
	Mother	59,2951	59	
	Father	58,6697	53	
ANKRD31	**Patient 22**	58,6697	75	5:74518166
	Mother	56,1542	66	
	Father	57,2364	65	
CACNA1I	**Patient 23**	20,1538	32	22:40066855
	Mother	9,48446	9	
	Father	56,6495	75	

Supplementary Table 2 Primers used for Sanger sequencing

Gene	Primers	Sequence	Genomic position (GRCh37/hg19)
AGAP6	F	AGCGGGAAGACCATCTCTG	chr10:51748364-51748795
	R	GAAAGGAGCTCGAAGTGTGG	
HCN1	F	CAGCAGACTGTTTCCACTTCA	chr5:45396198-45397021
	R	CATGCACAATAGCTGCCTGT	
SCN1A	F	TTTTGTGTGTGCAGGTTTCATT	chr2:166866053-166866365
	R	AGGCCTATTTCTCTTGCATATCA	
MGRN1	F	CTGGATTTGAGCCTGGTGAT	chr16:4723322+4723803
	R	CCCACGTTCAGCACAGACTA	
BTBD9	F	TCCTGATGCCAAATCTTGTT	chr6:38255873-38256270
	R	GCACGCTATATCTCGTTGTTG	
TCF4	F	TCAGCGCCCTCTAGTGAAAC	chr18:52901609-52902050
	R	TTAGCGGGCGAAGTTCTAAA	
SEMA6B	F	TGGCCAAGGTCACACAGTAA	chr19:4555127-4555699
	R	AAGGCAGGCAAGAGATGAG	
GRIN2B	F	TACAATCTAACCTAGGCCCTGG	chr12:13764558-13764935
	R	TGGATATGCTAGGGAAAATGCAG	
VASH2	F	GCAAGGTTCAAGAGTACTGGGT	chr1:213161536+213162249
	R	TGGTGAGGCATAATGTTCAAAGC	
CHRNA5	F	GAGCAGGGTCCCTATGTAGC	chr15:78881983+78882780
	R	CGCCATGGCATTATGTGTTGA	
ZNF620	F	TAGCGTCAGCACACAACTCA	chr3:40557616+40558235
	R	GCTGGTGCTGAATCAGGGT	
GRAMD1A	F	CTCACCCCTGAACCAATTGC	chr19:35506400+35507085
	R	AGAAGGAGAAACTGAGGCACA	
NOC3L	F	TGTAGAAAATAGAAGTGGCAGGT	chr10:96097185-96097950
	R	CACATGAAGCACCTATAGCCA	
GPATCH2	F	TGCTGGCAGTTCTTAGAGTCT	chr1:217784024-217784774
	R	TCAATGAGCCTAGCAAGAAAGC	
SLC6A1	F	CTGTCTGACTCCGAGGTGAG	chr3:11067255+11067936
	R	GACGATGATGGAGTCCCTGA	
GABBR2	F	GTGACCTGGGTCTGGTAAGTG	chr9:101133405-101134056
	R	TTCCTTTGACAAGGTCCCCAG	
ATP8B1	F	AATCTTGGGAATGGTACTCCTGG	chr18:55328163-55328763
	R	ACCTTATTTTCCTCTTCGCATCC	
HAP1	F	CATCCGGAACTTGCACTCG	chr17:39890418-39891088
	R	CTTCCTCCAGCTCCCGAATA	
PDLIM7	F	GTGTTCCCGTGACCCAGG	chr5:176910565-176911259
	R	AGCCCTACCCAGAAATGCAG	
SRRM3	F	GTCACTCTGTACAAGGGACCT	chr7:75890525+75891203
	R	CTGCTTGTCTAACTGGCACC	
ANKRD31	F	GGACGCATCAATAGGTGCAG	chr5:74517744-74518408
	R	GCTTCCAGTCAACAGTAGGC	
CACNA1I	F	CCACTGCCAACCTGAGTGA	chr22:40066439+40067235
	R	CACAGTCATTGCCACCCATG	

F, Forward; R, Reverse.