Genomic analyses in a large clinical cohort reveal high prevalence of *MECP2* variants associated with neuropsychiatric phenotypes in adulthood

Claudia Gonzaga-Jauregui[1,ç], Alina Kurolap[2,3], Jeffrey Staples[1], Cristopher Van Hout[1], Daren Liu[1], Regeneron Genetics Center, Geisinger-Regeneron DiscovEHR Collaboration, Rodney Samaco[4,5], Karin Weiss[2], Hagit Baris Feldman[2,3,6]

[1] Regeneron Genetics Center, Tarrytown, NY, USA

[2] The Genetics Institute, Rambam Health Care Campus, Haifa, Israel.

[3] The Genetics Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel (current affiliation)

[4] Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, TX, USA

[5] Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA

[6] Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[ç] To whom correspondence should be addressed:

Claudia Gonzaga-Jauregui, PhD.

Regeneron Genetics Center, 777 Old Saw Mill River Road, Tarrytown, NY, USA, 10591.

E-mail: cgonzagaj@gmail.com

Abstract

Purpose: To evaluate the phenotypes of individuals with pathogenic and likely pathogenic variants in the *MECP2* gene.

Methods: We surveyed exome sequencing data from a large clinical care cohort for deleterious variation in the *MECP2* gene. We reviewed de-identified clinical information for these individuals to interrogate for neurodevelopmental and neuropsychiatric phenotypes.

Results: We identified pathogenic and likely pathogenic variants in *MECP2* in individuals with typical and atypical Rett syndrome, and neuropsychiatric phenotypes, and estimate a prevalence of MECP2-associated disorders of 1 in 2,645 individuals. We observed a 7.45x increased relative risk of neuropsychiatric phenotypes, especially major depression, in adult individuals with deleterious variants in *MECP2* without a diagnosis of Rett syndrome. Male individuals with missense pathogenic variants in *MECP2* appear to have more severe neuropsychiatric phenotypes.

Conclusions: We identified and report individuals with heterozygous pathogenic variants in *MECP2* and their phenotypes in a large clinical cohort. The observed prevalence of MECP2-associated disorders in our cohort is higher than estimated in the literature. Individuals with pathogenic variants in *MECP2* can survive into adulthood but are at increased risk of developing neuropsychiatric disorders, mainly major depression. Pathogenic variation in *MECP2* is a likely important contributor to neuropsychiatric disorders in the general population.

Keywords: *MECP2, Depression, atypical Rett syndrome, neuropsychiatric disorders*

Introduction

Rett syndrome (RTT, MIM #312750) is a rare neurodevelopmental disorder characterized by normal development in early infancy with subsequent developmental arrest and regression of psychomotor acquired skills between 6 and 18 months of age, accompanied by additional features, including acquired progressive microcephaly, seizures, stereotypical hand movements, autistic behaviors, loss of acquired speech, and severe intellectual disability[1,2]. Deleterious variants, generally *de novo*, in the methyl-CpG-binding protein-2 gene, *MECP2*, are the cause of typical RTT[3]. *MECP2* is located on the X chromosome and consequently Rett syndrome is observed as an X-linked dominant disorder in females, although RTT phenotypes have been observed in a few males with Klinefelter syndrome or somatic mosaicism[4]. The presence of two X chromosomes in females and random chromosome X inactivation during development contributes to variable expressivity of the phenotype in affected females with typical RTT. Conversely, RTT-causing variants in *MECP2* in males with a normal single X chromosome present with a severe phenotype of neonatal encephalopathy with microcephaly, intellectual disability, failure to thrive, and respiratory insufficiency with early death before 2 years of age[5,6]. In other cases, MECP2 pathogenic variants that do not cause classic RTT have been observed in males with X-linked syndromic intellectual disability [MIM # 300055], where the mutant allele is typically transmitted from asymptomatic or mildly affected mothers. Some of these "carrier" females and other females with generally hypomorphic missense, rather than complete loss of function, variants in *MECP2* can present with a milder phenotype, generally called atypical Rett syndrome. Atypical RTT is characterized by mild to moderate intellectual disability and autistic features without regression. Additionally, non-random X-inactivation in females can contribute to milder and atypical RTT phenotypes in females[7].

The estimated prevalence of RTT is about 1 in 10,000 to 1 in 20,000 live births[5]. However, there are no estimates for the prevalence of the milder presentations of the disorder in the form of atypical RTT. The availability of large clinically unascertained population cohorts with available genomic data presents an opportunity to evaluate the prevalence of genetic disorders based on unbiased molecular data, and to understand the phenotypic spectrum of MECP2 variation at a deeper level[8]. Furthermore, a better assessment and understanding of the phenotype of individuals with genetic disorders in the general population can serve to inform prognosis for newly diagnosed pediatric cases.

Materials and Methods

Participants. Proband 1 was evaluated for developmental delay and facial dysmorphisms. The proband and her parents consented for research genomic studies through an institutional research study approved by the Rambam Helsinki Committee, study number 0038-14-RMB.

Separately, DiscovEHR participants are a subset of the Geisinger MyCode® Community Health Initiative. The MyCode® Community Health Initiative is a Geisinger-wide repository of blood, serum, and DNA samples from individuals that have been consented to participate in research and donate samples for broad research use, including genomic analysis that can be linked to de-identified electronic health record (EHR) information. Participants were consented in accordance with the Geisinger Health System (GHS) Institutional Review Board approved protocol, study number 2006-0258.

Exome sequencing and data analyses. Sample preparation, exome sequencing, and sequence data production were performed at the Regeneron Genetics Center (RGC) as previously described[8,9]. For downstream genetic analyses, variants were further annotated and analyzed using an in-house implemented annotation and analysis pipeline, and additional customized Perl bioinformatics scripts

for data processing. For ascertainment and survey of pathogenic variants, a union of NCBI's ClinVar Pathogenic/Likely_Pathogenic and HGMD (Human Gene Mutation Database) high-confidence disease_causing mutations (DM-High) reported variants were considered. In addition to standard bioinformatic QC filters, all candidate variants in *MECP2* were visually confirmed using the Integrative Genomics Viewer (IGV)[10] to ensure they were of good quality and to exclude false positives.

We performed burden analysis of rare deleterious variants identified in *MECP2* versus non-carrier individuals to test for association with phenotypes of interest using a Fisher's exact test (FET).

Data availability. All relevant genetic variants and deidentified phenotypic information are included in the manuscript and supplementary information. *MECP2* reported variants will be deposited in ClinVar for public data access.

Results

Initially, we evaluated a pediatric (<10y) female proband (Proband 1) referred for global developmental delay without history of regression, mild intellectual disability, and facial dysmorphisms, including posterior helix pit and macroglossia. Trio-based exome sequencing and analysis identified a *de novo* missense mutation [hg38.chrX:154030903(G>A); c.925C>T; p.Arg309Trp] in the *MECP2* gene (Figure 1A). Based on the observed phenotype of the proband, this variant was a strong candidate as the molecular cause, although the clinical presentation was not of typical RTT. Further variant evaluation and curation revealed that this p.Arg309Trp variant had been previously associated with atypical Rett syndrome and X-linked intellectual disability in both females and males[11]. The p.Arg309 residue is a CpG site and likely a hotspot for recurrent *de novo* mutation, as in our case. Affected individuals with this variant have some overlapping features with RTT but do not fulfill criteria for typical RTT. For

example, while all patients have intellectual disability and some present with hand stereotypies and seizures, the majority did not suffer regression, and whereas microcephaly is a common feature observed in RTT patients, these atypical p.Arg309Trp heterozygous patients have macrocephaly instead, among other differences[11]. Interestingly, when evaluating this variant for our proband, we identified two additional heterozygous individuals in our internal database. These two individuals are adult female participants of the Geisinger MyCode® Community Health Initiative. Manual evaluation of their EHR information showed that they have diagnostic codes for mild intellectual disability and learning difficulties, attention deficit disorder, depressive disorder, and psychotic disorder. These observations prompted us to further investigate the genetic prevalence of *MECP2*-associated disorders in the DiscovEHR clinical cohort and their contribution to neurodevelopmental and neuropsychiatric disorders.

We performed a gene-wide survey of pathogenic and likely pathogenic variants in *MECP2* to evaluate the prevalence of *MECP2*-associated neurodevelopmental disorders in 92,455 participants of the Geisinger-Regeneron DiscovEHR cohort that links exome sequencing genomic data with de-identified EHR data. We identified 43 heterozygous individuals for 17 distinct pathogenic or likely pathogenic variants in *MECP2* (Table 1, Supplementary Table 1). Of these, the majority (N=35, 81.39%) were females and 8 were males (18.6%). Eight of the variants identified were loss of function variants (7 frameshift and 1 nonsense), of which only 3 had been previously reported. Additionally, we identified 36 heterozygous individuals for 9 pathogenic missense variants (Supplementary Table 1).

Of these 43 individuals with *MECP2* pathogenic and expected pathogenic loss of function variants, 4 individuals were minors (2 males and 2 females) and had received a clinical diagnosis of RTT and/or been tested for mutations in *MECP2*, and consequently had a diagnosis code of Rett Syndrome (ICD-10

F84.2) in their EHR (Patients 1-4, Figure 1B and 1C, Table 1 and Supplementary Table 1). The remaining 39 individuals were adults and had no diagnosis code or a clear indication of suspected RTT in their EHR. However, manual investigation of their EHR and clinical information revealed an enrichment for diagnosis codes associated with neuropsychiatric and mood disorder phenotypes (Supplementary Table 1). We observed a significant increased relative risk of 7.45 (95% CI [3.63-15.30], P=1.14 x10^{-10}) for any mood disorder including major depressive disorder, bipolar disorder, and anxiety disorder (Table 1). The major enrichment among these phenotypes was observed for major depressive disorder with 24 of the 39 individuals (61.53%) having a diagnosis versus 28% of the DiscovEHR cohort, resulting in a 4.11 increased risk (95% CI [2.15-7.84], P=3.11 x10^{-6}) for carriers of *MECP2* pathogenic variants (Table 1). Of note, some of these individuals were related and we were able to observe co-segregation of the *MECP2* pathogenic variant with affectation for neuropsychiatric phenotypes (Figure 1D). Overall, we calculate a prevalence of approximately 1 in 2,645 individuals with MECP2-associated disorders in our clinical cohort (35 unrelated individuals in 92,455 sequenced participants) for previously reported pathogenic and expected pathogenic loss of function variants.

In addition to the previously reported pathogenic variants and expected pathogenic loss of function variants in *MECP2*, our genetic analyses identified six additional candidate pathogenic missense variants. Two of these (c.286C>T; p.Arg96Trp and c.598C>T; p.Arg200Trp) had been previously reported in ClinVar as variants of unknown significance (VUSs) for severe neonatal-onset encephalopathy with microcephaly. Manual exploration of the EHR data for the individuals carrying these VUSs showed support for c.598C>T; p.Arg200Trp to be deleterious with a similar phenotypic spectrum in these two adult females as the one demonstrated above for carriers of pathogenic *MECP2* variants, whereas the information on the c.286C>T; p.Arg96Trp heterozygous female individual was

inconclusive. One other variant (c.784C>T; p.Arg262Cys) is a reported VUS with conflicting interpretations of pathogenicity. In our DiscovEHR cohort, we identified five heterozygous individuals for this variant, including two males. Detailed review of their clinical information showed no support for pathogenicity of this particular variant (Supplementary Table 2). The remaining 3 candidate missense variants (c.307C>T; p.Arg103Trp, c.658C>A; p.Gln220Lys, and c.677A>G; p.Glu226Gly) represent novel candidate pathogenic alleles in *MECP2*. Exploration of the EHR data of these adult heterozygous individuals showed a similar clinical presentation as described above for carriers of known and expected loss of function pathogenic variants in *MECP2* with diagnoses of anxiety, major depressive disorder, and other neuropsychiatric disorders; however the data are inconclusive to classify these variants (Supplementary Table 2).

Discussion

The prevalence of rare genetic disorders has historically been based on literature reports of patients seeking medical care and diagnoses. The availability of large population cohorts[9,12] linked to de-identified medical information provides excellent resources to evaluate the true prevalence of monogenic disorders in an unbiased way, and to assess their phenotypic spectrum through a genotype-driven approach. Some studies performed in these cohorts have already shown that monogenic conditions are often mis- or underdiagnosed, and that their prevalence may be higher than reported in the literature[8].

In this report, our work was prompted by the analysis of a child with an unspecific neurodevelopmental disorder, the identification of a *de novo* deleterious variant in *MECP2*, and the

observation that carriers of this particular recurrent variant are present in the general population. The availability of de-identified clinical information for these individuals further revealed that while they do not have major cognitive impairments, they are more likely to develop depressive and other mood disorders later in adulthood. Furthermore, this was observed to be the case for the majority of carriers of previously reported pathogenic variants in *MECP2*, as well as carriers of novel loss of function expected pathogenic variants, showing a 7.45 and 4.11 increased risk of any mood disorder and major depressive disorder, respectively. Although we observed a strong association with depression, anxiety and mood disorders in these participants, these diagnoses were not present in the totality of our cohort. This could be due to a few possible reasons, including that these individuals do not seek regular healthcare at Geisinger and consequently there is not enough information in their EHR. It is also possible that they have not self-reported incidents of depression or anxiety, or that they have not been clinically evaluated for mood or neuropsychiatric disorders. Finally, in the case of X-linked disorders, penetrance and phenotype expressivity can be further modified by X chromosome inactivation skewing in females. Indeed, this has been the case observed for MECP2-associated disorders[7]. It is possible that in some of these female carriers strong X chromosome inactivation skewing has not only prevented them from presenting with RTT in infancy, but also from developing the neuropsychiatric phenotypes observed in other carriers or pathogenic variation in the gene. Additionally, given the presence of male adult carriers of some pathogenic missense variants in *MECP2*, it is possible that there are additional genetic modifiers of MECP2-associated disorders that have allowed these individuals to develop normally and reach adulthood. Further clinical and genetic evaluation of these individuals could potentially reveal these modifiers and possibly suggest therapeutic interventions to ameliorate MECP2-deficiency phenotypes.

Most patients reported with neurodevelopmental disorders are pediatric, and in most cases follow-up and natural history of these disorders have not been documented, including additional comorbidities that may develop later in adulthood. Recent studies looking for genetic associations with neuropsychiatric disorders have identified variants in known neurodevelopmental disorder genes contributing significantly to so called complex disorders like schizophrenia[13] and bipolar disorder in adults. However, it is also possible that known rare pathogenic variants responsible for neurodevelopmental disorders with variable expressivity contribute significantly to neuropsychiatric disorders in adults with mild phenotypes that were never assessed and diagnosed with a disorder of neurodevelopment in childhood. Furthermore, agnostic genotype-driven studies in large populations are relevant to evaluate prognosis of newly diagnosed patients with rare genetic disorders as the information can help families plan for the future and seek appropriate medical and support resources to aid in the care of children with neurodevelopmental disorders.

Disclosures

JS, DL, are full-time employees of the Regeneron Genetics Center and receive salary and stock options as part of compensation. All other authors have no conflict of interest to disclose.

Acknowledgements

The authors wish to thank the family of Proband 1 and the Geisinger participants that consented to participate in research as part of the Geisinger MyCode Health Initiative. The authors thank Dr. Huda Y. Zoghbi for insightful conversations on this project and comments on the manuscript.

Author contributions

Conceptualization: CGJ; Data curation: CGJ, AK, Formal analysis: CGJ, CVH, DL; Investigation: CGJ, AK, CVH, KW; Methodology: CGJ; Resources: HBF; Software: JS; Supervision: HBF; Validation: CGJ; Visualization: CGJ; Writing - original draft: CGJ; Writing - review and editing of manuscript: CGJ, AK, CVH, RS, HBF.

References

1. Rett, A. Ueber ein eigenartiges hirnatrophisches Syndrom bei Hyperammoniamie in Kindesalter. Wien. Med. Wschr. 1966; 116: 723-738.
2. Chahrour M, Zoghbi HY. The story of Rett syndrome: from clinic to neurobiology. Neuron. 2007 Nov 8;56(3):422-37.
3. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet. 1999; 23(2):185-8.
4. Jan MM, Dooley JM, Gordon KE. Male Rett syndrome variant: application of diagnostic criteria. Pediatr Neurol. 1999; 20(3):238-40.
5. Moog U, Smeets EEJ, van Roozendaal KEP, Schoenmakers S, Herbergs J, Schoonbrood-Lenssen AMJ, Schrander-Stumpel CTRM. Neurodevelopmental disorders in males related to the gene causing Rett syndrome in females (MECP2). Europ. J. Paediat. Neurol. 2003; 7: 5-12.
6. Villard L. MECP2 mutations in males. J Med Genet. 2007; 44(7):417-23.
7. Hoffbuhr KC, Moses LM, Jerdonek MA, Naidu S, Hoffman EP. Associations between MeCP2 mutations, X-chromosome inactivation, and phenotype. Ment Retard Dev Disabil Res Rev. 2002; 8(2):99-105.

8. Gonzaga-Jauregui C, Ge W, Staples J, Van Hout C, Yadav A, Colonie R, Leader JB, Kirchner HL, Murray MF, Reid JG, Carey DJ, Overton JD, Shuldiner AR, Gottesman O, Gao S, Gromada J, Baras A, Altarejos J; Geisinger-Regeneron DiscovEHR collaboration. Clinical and Molecular Prevalence of Lipodystrophy in an Unascertained Large Clinical Care Cohort. Diabetes. 2020; 69(2):249-258.
9. Dewey FE, Murray MF, Overton JD, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science. 2016; 354(6319).
10. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011; 29(1):24-6.
11. Schönewolf-Greulich B, Tejada MI, Stephens K, Hadzsiev K, Gauthier J, Brøndum-Nielsen K, Pfundt R, Ravn K, Maortua H, Gener B, Martínez-Bouzas C, Piton A, Rouleau G, Clayton-Smith J, Kleefstra T, Bisgaard AM, Tümer Z. The MECP2 variant c.925C>T (p.Arg309Trp) causes intellectual disability in both males and females without classic features of Rett syndrome. Clin Genet. 2016; 89(6):733-8.
12. Van Hout C, Tachmazidou I, Backman JD, Hoffman JX, Ye B, Pandey AK, Gonzaga-Jauregui C, Khalid S, Liu D, et al. Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. Nature 2020.
13. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Stephan Ripke, James TR Walters, Michael C O'Donovan. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. medRxiv 2020.09.12.20192922; doi: https://doi.org/10.1101/2020.09.12.20192922

Figure Legends:

Figure 1. Example pedigrees of individuals identified with pathogenic and likely pathogenic variants in *MECP2.* Family pedigrees and segregation of pathogenic and likely pathogenic variants in *MECP2*. The upper panel shows three different trios, including the index case Proband 1 (A) with an identified *de novo* recurrent missense variant (c.925C>T; p.R309W) associated with atypical Rett syndrome (RTT); and two trios where the probands G1 (B) and G4 (C) have a clinical diagnosis of Rett syndrome and were found to have *de novo* truncating mutations in *MECP2*. The lower panel (D) shows an extended pedigree with three females diagnosed with anxiety and depressive disorder found to have a novel *MECP2* missense (c.658C>A; p.Gln220Lys) variant of unknown significance (VUS).

Figure 1. Example pedigrees of individuals identified with pathogenic and likely pathogenic variants in *MECP2*.

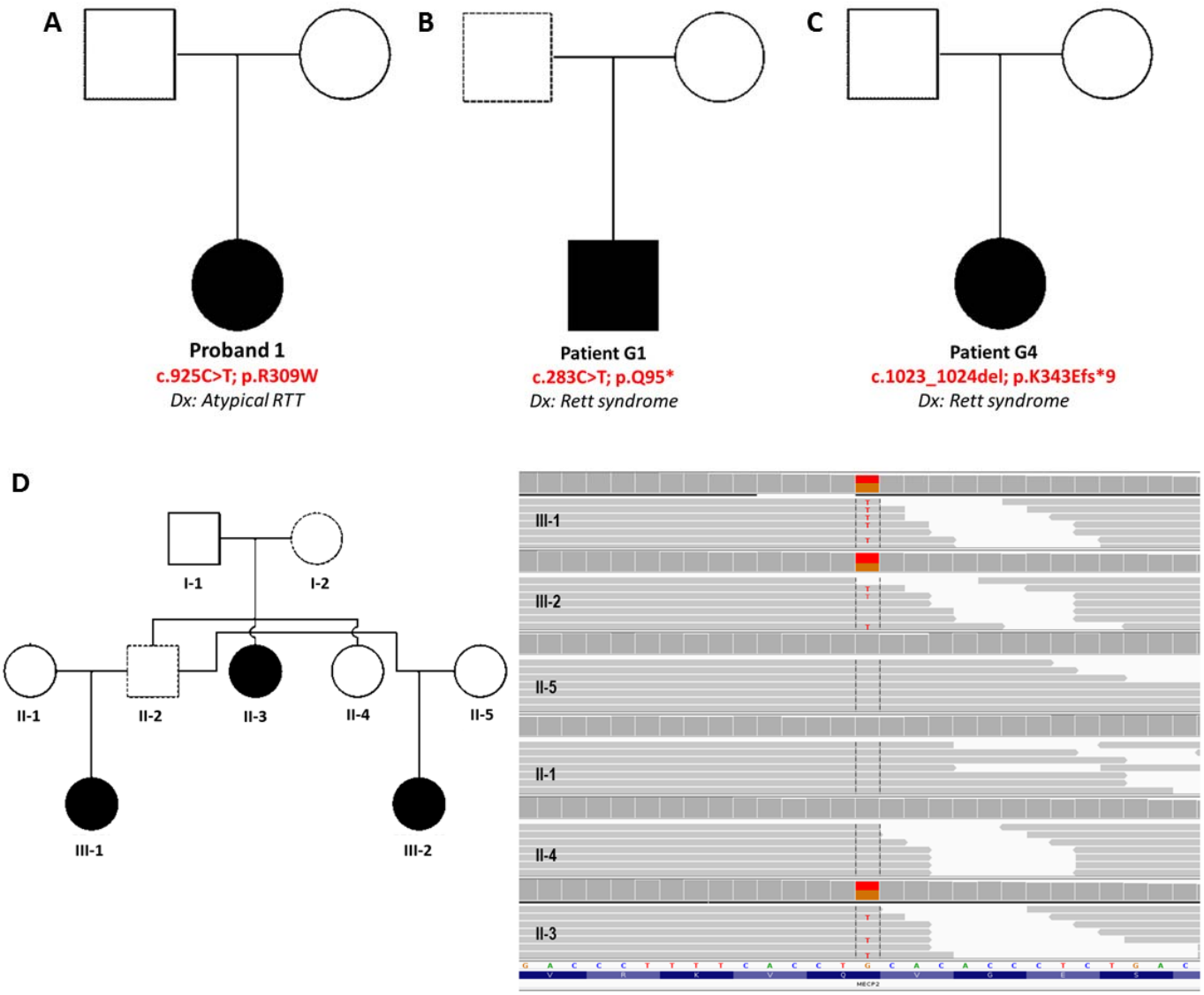

Table 1. Clinical characteristics of individuals identified carrying pathogenic (P) and likely pathogenic (LP) variants in *MECP2*.

Patient	NT_Variant	AA_Variant	Age	Gender	Clinical Significance	RTT	ID/DD	Epilepsy	Anxiety	Depression	Other mood disorder(s)	Neuropathy	Migraine	Sleep disorder
G1	c.283C>T	p.Gln95*	10's	Male	P	?	?	?						
G2	c.455C>T	p.Ala152Val	10's	Male	P	?	?	?						?
G3	c.952C>T	p.Arg318Cys	<10	Female	P	?	?	?			?			
G4	c.1023_1024del	p.Lys343Glufs*9	10's	Female	P	?	?	?						
G5	c.1199_1200ins	p.Pro402fs	30's	Female	P		?	?	?				?	
G6	c.1403delG	p.Gly468fs	40's	Female	LP					?		?		
G7	c.1403delG	p.Gly468fs	50's	Female	LP				?	?		?		
G8	c.1244dupC	p.Glu416fs	60's	Female	LP								?	
G9	c.1244dupC	p.Glu416fs	40's	Female	LP					?				
G10	c.1187_1205del	p.Pro396fs	30's	Female	LP					?				
G11	c.1174_1199del	p.Val392fs	30's	Female	LP					?				
G12	c.1122delC	p.Lys375fs	70's	Female	LP				?			?	?	?
G13	c.604C>T	p.Arg202Cys	30's	Female	P;VUS				?	?	?			
G14	c.604C>T	p.Arg202Cys	30's	Female	P;VUS				?	?	?		?	?
G15	c.961C>T	p.Arg321Trp	40's	Female	P; LP		?			?	?			
G16	c.961C>T	p.Arg321Trp	30's	Female	P; LP				?	?	?			
G17	c.433C>T	p.Arg145Cys	10's	Female	P; LP		?	?	?		?			
G18	c.535C>T	p.Arg179Trp	60's	Female	P; LP				?	?	?	?		?
G19	c.535C>T	p.Arg179Trp	30's	Male	P; LP					?	?			?
G20	c.535C>T	p.Arg179Trp	60's	Female	P; LP			?						
G21	c.838C>T	p.Arg280Trp	30's	Female	LP; VUS					?	?			
G22	c.838C>T	p.Arg280Trp	40's	Female	LP; VUS				?		?	?		
G23	c.838C>T	p.Arg280Trp	30's	Male	LP; VUS				?		?			?
G24	c.379C>T	p.Arg127Cys	50's	Female	LP; VUS					?	?			
G25	c.490C>G	p.Pro164Ala	50's	Male	P; LP								?	
G26	c.490C>G	p.Pro164Ala	40's	Female	P; LP									
G27	c.490C>G	p.Pro164Ala	40's	Female	P; LP				?	?				?
G28	c.490C>G	p.Pro164Ala	80's	Female	P; LP					?		?	?	?
G29	c.490C>G	p.Pro164Ala	60's	Male	P; LP					?			?	?
G30	c.490C>G	p.Pro164Ala	50's	Female	P; LP									
G31	c.490C>G	p.Pro164Ala	70's	Male	P; LP									
G32	c.490C>G	p.Pro164Ala	20's	Female	P; LP				?	?	?		?	
G33	c.490C>G	p.Pro164Ala	40's	Female	P; LP					?	?		?	
G34	c.490C>G	p.Pro164Ala	60's	Male	P; LP									
G35	c.490C>G	p.Pro164Ala	40's	Female	P; LP									
G36	c.490C>G	p.Pro164Ala	20's	Female	P; LP					?	?		?	
G37	c.490C>G	p.Pro164Ala	70's	Female	P; LP				?	?				
G38	c.490C>G	p.Pro164Ala	30's	Female	P; LP									
G39	c.490C>G	p.Pro164Ala	40's	Female	P; LP								?	
G40	c.490C>G	p.Pro164Ala	90's	Female	P; LP				?					
G41	c.490C>G	p.Pro164Ala	20's	Female	P; LP			?	?	?	?		?	
G42	c.490C>G	p.Pro164Ala	60's	Female	P; LP				?	?				
G43	c.490C>G	p.Pro164Ala	80's	Female	P; LP									

P: Pathogenic; LP: Likely Pathogenic; VUS: Variant of Unknown Significance.