



AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Exome sequencing in children of women with skewed X-inactivation identifies atypical cases and complex phenotypes**This is the author's manuscript**

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1621163> since 2017-04-28T13:57:36Z

Published version:

DOI:10.1016/j.ejpn.2016.12.005

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)



This Accepted Author Manuscript (AAM) is copyrighted and published by Elsevier. It is posted here by agreement between Elsevier and the University of Turin. Changes resulting from the publishing process - such as editing, corrections, structural formatting, and other quality control mechanisms - may not be reflected in this version of the text. The definitive version of the text was subsequently published in EUROPEAN JOURNAL OF PAEDIATRIC NEUROLOGY, 21, 2017, 10.1016/j.ejpn.2016.12.005.

You may download, copy and otherwise use the AAM for non-commercial purposes provided that your license is limited by the following restrictions:

- (1) You may use this AAM for non-commercial purposes only under the terms of the CC-BY-NC-ND license.
- (2) The integrity of the work and identification of the author, copyright owner, and publisher must be preserved in any copy.
- (3) You must attribute this AAM in the following format: Creative Commons BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>), 10.1016/j.ejpn.2016.12.005

The publisher's version is available at:

<http://linkinghub.elsevier.com/retrieve/pii/S1090379816302823>

When citing, please refer to the published version.

Link to this full text:

<http://hdl.handle.net/2318/1621163>

Exome sequencing in children of women with skewed X-inactivation identifies atypical cases and complex phenotypes

Elisa Giorgio^a, Alessandro Brussino^a, Elisa Biamino^b, Elga Fabia Belligni^b, Alessandro Bruselles^c, Andrea Ciolfi^{d,e}, Viviana Caputo^f, Simone Pizzi^d, Alessandro Calcia^a, Eleonora Di Gregorio^{a,g}, Simona Cavalieri^a, Cecilia Mancini^a, Elisa Pozzi^a, Marta Ferrero^a, Evelise Riberi^b, Iolanda Borelli^a, Antonio Amoroso^a, Giovanni Battista Ferrero^b, Marco Tartaglia^d, Alfredo Brusco^{a,g}

^a University of Torino, Department of Medical Sciences, 10126, Turin, Italy

^b University of Torino, Department of Public Health and Pediatrics, 10126, Turin, Italy

^c Department of Hematology, Oncology and Molecular Medicine, Istituto Superiore di Sanità, Rome, Italy.

^d Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù IRCCS, Rome, Italy.

^e Food and Nutrition Research Center, CREA, Rome, Italy ^f Department of Experimental Medicine, Sapienza University of Rome, Rome, Italy.

^g Città della Salute e della Scienza University Hospital, Medical Genetics Unit, Turin, Italy.

Corresponding author: Alfredo Brusco, University of Torino, Department of Medical Sciences, via Santena 19, 10126, Torino, Italy. Fax +390116706582; e-mail: alfredo.brusco@unito.it

Conflicts of interest: none

Keywords: Whole exome sequencing, WES, skewed X-inactivation, ATRX, DMD, RPS6KA3, MECP2

ABSTRACT

More than 100 X-linked intellectual disability (X-LID) genes have been identified to be involved in 10-15% of intellectual disability (ID). To identify novel possible candidates, we selected 18 families with a male proband affected by isolated or syndromic ID. Pedigree and/or clinical presentation suggested an X-LID disorder. After exclusion of known genetic diseases, we identified seven cases whose mother showed a skewed X-inactivation (>80%). Whole exome sequencing (WES, 50X average depth) allowed to solve the genetic basis in four cases, two of which (Coffin-Lowry syndrome, *RPS6K3* gene; ATRX syndrome, *ATRX* gene) had been missed by previous clinical/genetics tests. One further *ATRX* case showed a complex phenotype including pontocerebellar atrophy (PCA), possibly associated to an unidentified PCA gene mutation. In a case with suspected Lujan-Fryns syndrome, a c.649 C>T (p.Pro217Ser) *MECP2* missense change was identified, likely explaining the neurological impairment, but not the marfanoid features, which were possibly associated to the p.Thr1020Ala variant in fibrillin 1. Finally, a c.707T>G variant (p.Phe236Cys) in the *DMD* gene was identified in a patient retrospectively recognized to be affected by Becker muscular dystrophy (BMD, OMIM 300376).

Overall, our data show that WES may give hints to solve complex ID phenotypes with a likely X-linked transmission, and that a significant proportion of these orphan conditions might result from concomitant mutations affecting different clinically associated genes.

1. INTRODUCTION

The Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) defines intellectual disability (ID) as an impairment “of general mental abilities that impact adaptive functioning” involving the conceptual, social and practical domains, usually associated with an IQ score of 70 or below (<http://www.dsm5.org>). From the end of the 19th century, it was noted that males with ID outnumbered females with a ratio of 1.3-1.4 to 1. This is due to the high proportion of genes on the X chromosome causing ID ^{1,2}. Overall, mutations in genes causing X-LID are responsible of 10-15% of males ID cases, affecting ~1:600 individuals ³⁻⁵, and more than 200 forms of syndromic and non-syndromic X-linked ID (X-LID) have been recognized to date due to mutations in over than 100 different genes.

In mothers of males with X-LID, carriers of the genetics/cytogenetics causative anomaly, it is a common feature finding an unbalanced X-chromosome inactivation (XCI) ^{6,7}. XCI is the process through which females balance the quantity of genes expressed by the X chromosome with respect to males ^{8,9}. The process is usually random, and the final proportion of maternal vs. paternal active chromosomes can range from 50:50 to the complete inactivation of one of the two X-chromosomes ¹⁰. Nevertheless, it is estimated that at least 15% of the genes on the X chromosome escape XCI ¹¹. An extremely unbalanced inactivation pattern is particularly rare in the normal population, with a ratio >90:10 being present in ~1.8% of the whole female population and ~3.6% of adult females, and a ratio >80:20 being present in ~8.8% of the whole female population and ~14.2% of adult females ¹⁰.

In this paper, we selected 18 males affected by syndromic or non-syndromic X-LID, and screened their mothers for the X-inactivation pattern. Patients, whose mothers carried a skewed X-inactivation, underwent a whole exome sequencing (WES) to identify the causative mutation.

2. PATIENTS AND METHODS.

2.1 Patients

Eighteen male subjects aged 6-14 yrs. affected by syndromic or non-syndromic ID, ranging from mild to severe, were included in the study (Table A.1). Two cases presented isolated ID; in two, ID was associated with autism spectrum disorders (ASD), while the remaining cases showed syndromic forms of ID. Their family history was suggestive of an X-linked ID (i.e., the presence of one or more affected brothers and/or other affected males in the maternal lineage) in 15, and their clinical presentation was compatible with an X-linked disease in three (Table A.1).

2.2 Karyotyping and array-CGH analyses

Karyotyping was performed on circulating leukocytes by GTG-banding. Array-CGH was performed at 60K resolution following the manufacturer's protocol (Agilent Technologies, Santa Clara, California, USA): slides were read with a G2565BA microarray scanner, and analyzed with the Agilent CGH Analytics software v. 4.0.81 (Agilent Technologies) using GRCh37/hg19 as the reference human DNA. We used the statistical algorithm ADM-2 and a sensitivity threshold of 6.0. Three consecutive aberrant probes were considered as the minimum to identify a copy-number variant. Positive array-CGH results were confirmed by real-time PCR.

2.3 X-inactivation analysis

The X-inactivation pattern was evaluated exploiting the CAG triplet-repeat in the first exon of the androgen receptor gene (AR, Xq12), using the protocol and calculations described by ¹². Genomic DNA was amplified with and without prior *HaeII* / *HhaI* methylation sensitive restriction enzyme digestion ¹³. The forward primer was fluorescently labeled and PCR products were run on an ABI Prism 3730xl automatic sequencer (Applied Biosystems, Foster City, CA, USA). The X-inactivation pattern was considered skewed if the proportion of the two alleles after digestion was at least 20:80, in two independent experiments.

2.4 WES analysis

WES was outsourced at BGI-Tech (Hong-Kong, PRC) using genomic DNA extracted from circulating leukocytes. Targeted enrichment was performed using Nimblegen SeqCap EZ Library v.3.0 (64 M) (Roche, Mannheim, Germany), and captured libraries were loaded onto an Illumina HiSeq 2000 platform (Illumina, San Diego, CA, USA). WES data analysis was performed using an in-house implemented pipeline ¹⁴⁻¹⁶. In brief, paired-end reads were aligned to human genome (UCSC GRCh37/hg19) with the Burrows–Wheeler Aligner (BWA V. 0.7.5a-r405) ¹⁷, and presumed PCR duplicates were discarded using the Picard's MarkDuplicates utility (<http://picard.sourceforge.net>). The alignment process was refined by local realignment and base-quality-score recalibration steps by means of Genome Analysis Toolkit (GATK 3.4) ¹⁸. GATK HaplotypeCaller was used to identify single nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs) ¹⁹. Variants with quality score < 50 and quality-by-depth < 1.5 or resulting from 4 or more reads having ambiguous mapping (this number being greater than 10% of all aligned reads) were discarded. ExAC 0.3, dbSNP146, retaining only variants with MAF < 0.001 or with a known clinical association), and in-house databases (retaining variants with frequency < 1%). SnpEff toolbox v4.2 ²⁰ was used to predict the functional impact of variants, and retain missense/nonsense/frameshift changes, coding indels, and intronic variants at exon-intron junctions. Functional annotation of variants was performed by using snpEff v4.2 and dbNSFP2.9 ^{20, 21}. WES data analysis was originally focused on functionally relevant X-linked variants, but both autosomal recessive and autosomal dominant models of inheritance were also taken into consideration. Sequence validation and segregation analyses were performed by Sanger sequencing using an ABI 3130XL and the ABI BigDye Terminator Sequencing Kit V.3.1 (Life Technologies), and using the SeqScape v2.6 Software (Life Technologies), for the following genes: *RPS6KA3* (NM_004586), *MECP2* (NM_004992), *DMD* (NM_004006), *ATRX* (NM_000489), *FBN1* (NM_000138), and *RPS4X* (NM_001007). Pathogenicity of each variant was predicted by PolyPhen-2, SIFT, Mutation T@sting, PMUT, and PHD.

Copy Number Variants (CNVs) analysis was carried out using EXCAVATOR2²², a read count based tool that permits to extend the analysis of copy numbers and breakpoints to the genomic level making use of WES In- and Off-Target reads.

3. RESULTS

3.1 Genetic analysis

Karyotype and *FMR1* analysis were normal in all probands. Array-CGH revealed two CNVs in cases XEX-08 and XEX-15, both transmitted by an unaffected parent: a paternally inherited duplication [arr 9q31.1(107,062,733x2,107,156,421-107,696,328x3, 107,809,963x2] and a maternally inherited duplication [arr 6q15(88,136,278x2,88,170,758-88,366,616 x3,88,366,616 x2]. In both cases, the rearrangements were not shared by the affected brothers, and consequently were considered as benign variants.

In seven families, an extremely skewed X-inactivation (>80%) was noted, and WES was performed exclusively on the proband of each family. A summary of the sequencing data output is provided in Table A.2.

Detected Variants were kept for further evaluation if one of the following conditions was satisfied:

(i) the mutation was already reported in literature; (ii) the variant affected a gene known to be associated with a disorder clinically fitting or overlapping the observed phenotype and was predicted to be damaging *in silico*. Using these parameters, putative causative variants were identified in five subjects (Table 1). In two, the gene mutated could fully explain the phenotype. One carried a mutation associated with ID but that could not explain the whole phenotype alone. Two carried a complex phenotype that could be explained by the presence of variants in two different genes (see below). Finally, two cases could not be solved analyzing sequencing data. WES data were also analyzed taking into consideration possible occurrence of small CNVs that could have been missed by array-CGH, which however was ruled out in all cases.

3.2 Case summary

3.2.1 Case XEX-03-TO

The proband was born at 30 weeks by urgent caesarean section from unrelated parents. Birth weight was 721 gr (<3rd centile) and the APGAR score was 6/9. He showed hypotonia, gastroesophageal reflux and numerous apparent life threatening events (ALTE). Brain ultrasonography revealed a first-degree ventricular hemorrhage. Brain NMR showed pontocerebellar and cerebellar hemisphere atrophy. An MRI at 5 yrs. confirmed the cerebellar hemispheres atrophy and showed cerebral trunk hypoplasia, moderate leukomalacia, and a thin corpus callosum.

At nine years of age, the clinical evaluation reported microcephaly, open mouth posture with protruding tongue, severe neuropsychomotor delay, absence of language, behavioral disturbances with self-aggressive and hetero-aggressive outbursts, and insomnia. Its metabolic panel and sialotransferrin isoelectric focusing showed no alterations. Methylation test for Prader-Willi and Angelman syndromes was normal. Exome sequencing revealed a c.7366_7367insA (p.Met2456Asnfs*42) variant in the *ATRX* gene (OMIM 300032) that was predicted to be damaging (Fig. 1, Table 2). A similar frameshift variant affecting the same nucleotide was reported as likely pathogenic in a patient with *ATRX* syndrome (<http://www.ncbi.nlm.nih.gov/clinvar/variation/196938/>).

3.2.2 Case XEX-11-TO

The subject showed ID associated with microcephaly, short stature, and facial dimorphism, which was suggestive of an X-linked mental retardation-hypotonic facies spectrum disorder (OMIM 309580). X-exome analysis revealed a c.6253C>T (p.Arg2085Cys) variant in the *ATRX* gene. Bioinformatics analysis confirmed the pathogenicity of the mutation by 4 out of 5 software (Fig. 1, Table 2).

3.2.3 Case XEX-16-TO

The subject was one of two dizygotic twins, born after an uneventful pregnancy from healthy non-consanguineous parents. At 11 years of age, he showed severe ID, facial dysmorphism (triangular-shaped and coarse face, prominent ears, widely spaced eyes, downslanted palpebral fissures, short nose with broad columella, thick alae nasi and septum, thick and everted lips vermillion), short and flashy hands with tapering fingers, microcephaly (OFC = 50 cm, < 3rd centile), short stature (height = 127 cm, < 3rd centile), small teeth with open bite and hypodontia, mild mitral valve insufficiency, and scoliosis.

His brother presented an overlapping phenotype and exhibited stimulus-induced drop attacks (SIDAs) triggered by auditory stimuli. Overall, the clinical presentation was suggestive of the Coffin-Lowry syndrome (OMIM 303600). Exome data identified the c.2188A>G variant (p.Arg730Gly) in the Ribosomal Protein S6 Kinase, 90 kDa, Polypeptide 3 gene (*RPS6KA3*, OMIM 300075) on chromosome Xp22.1. *RPS6KA3* is known to be mutated in patients with Coffin-Lowry syndrome ²³. The c.2188A>G variant was novel, segregated with the diseases, occurring in the two affected siblings, and was predicted to be damaging using five different *in silico* software (Fig.1, Table 2).

3.2.4 Case XEX-10-TO

The subject was born from healthy consanguineous parents, although the degree of kindship could not be determined. He showed learning disabilities and gained independent walking at 11 months of age. At 14 years, the clinical evaluation showed a marfanoid habitus with a height of 180 cm (>97th centile) a weight of 50 kg (50th centile) an OFC of 56 cm (75th centile) and an arm-span of 182 cm. Cardiac ultrasound evaluation was normal. He developed psychotic mania. Sanger sequencing excluded mutations in *MED12* (OMIM 300188), associated with the Lujan-Fryns syndrome (OMIM 309520).

His brother, examined at the age of 12 years, showed a similar phenotype, with ID, lumbar scoliosis, joint laxity, and myopia. His height was 163 cm (>97th centile), his weight was 40 kg (50th centile) and his OFC was 53 cm (25th centile). ECG showed a right bundle branch block. Cardiac ultrasound and brain MRI were normal.

A third brother, aged 11 months, was healthy and did not show any specific clinical feature. Exome sequencing revealed a c.649 C>T (p.Pro217Ser) in the methyl-CpG-binding protein 2 gene (*MECP2*; OMIM 300005) and a c.3058 A>G (p.Thr1020Ala) in the Fibrillin 1 gene (*FBN1*; OMIM 134797). Both variants were present also in the affected brother. The *MECP2* variant was inherited from the mother and the maternal grandmother, who both showed a skewed X-inactivation. *In silico* prediction tools supported the causative role for the *MECP2* variant, whereas the apparently unaffected father transmitted the *FBN1* variant (Fig. 2, Table 2).

3.2.4 Case XEX-12-TO

The subject, first child of unrelated parents, was born after an uneventful pregnancy by spontaneous deliver with vacuum extraction. The mother reported a three-day antibiotic therapy before the delivery because of a group B streptococcus infection at the vaginal swab. Birth weight was 3,320 gr (50th centile), length was 50.6 cm (50th centile), occipital frontal circumference (OFC) was 33 cm (10th centile), and APGAR score was 7/9. Mild hypotonia and hyporeflexia were present. The proband showed severe developmental delay (no formal IQ evaluation was possible); crawling and independent walking were acquired at 2.5 and 3.5 years.

At 6 yrs, his weight was 22 kg (50th centile), height was 112 cm (25th centile), and OFC was 50 cm (10th centile). At the last evaluation (7 yrs.), he presented a severe impairment of intellectual abilities, sphincter control was not yet acquired, and language was completely absent. The proband used only syllables and non-verbal communication. A training for adaptive/augmentative communication was ongoing. Motor skills were moderately compromised as well as social and affective behavior. X-exome analysis showed the c.138 T>G (p.Ile46Met) variant in the ribosomal

protein S4 gene (*RPS4X*). The mutation was predicted to be pathogenic and it was present in both the mother and maternal grandmother, and absent in a healthy maternal uncle (Fig. 2 and Table 2). This gene is predicted to undergo non-random X-inactivation, presenting a paralogous copy on the Y chromosome (93% identity)^{11,24}.

Remarkably, a c.707T>G variant (p.Phe236Cys) in the *DMD* gene was also identified. A clinical re-evaluation showed moderate proximal weakness and a creatine kinase (CK) level of 235 UI/L were compatible with the diagnosis of Becker muscular dystrophy (BMD, OMIM 300376) (Fig. 2 and Table 2).

3.2.5 Case XEX-06-TO

The subject was born after an uneventful pregnancy from non-consanguineous parents (Fig. 1). Birth weight was 4,210 gr (>90th centile) and the APGAR score was 9/9. He showed a globally delayed development with first words at three years. At eight years, he presented motor and vocal tics and an obsessive-compulsive behavior. The IQ scored 70 in the Wechsler Intelligence Scale for Children (WISC) test. The NMR revealed a pineal gland cyst. His younger brother was diagnosed with autism. WES did not reveal any potentially deleterious variant both on the X chromosome and on autosomes. Based on siblings' phenotype, we deeply analyzed CNVs and point mutations in the *NLGN3* and *NLGN4* genes; no functionally relevant mutation was identified.

3.2.6 Case XEX-18-TO

The subject was born by spontaneous delivery at 40 weeks. Birth weight was 3,250 gr (50th centile), length was 48 cm (25th centile) and the APGAR score was 9/9. At 13 days, he suffered an apneic episode. He was able to walk autonomously at 18 months. At 2 yrs., he suffered an episode of focal seizures: brain CT scan and metabolic evaluation were normal. At 7 yrs., a neuromotor delay was noted: clinical examination revealed an OFC at the 90-97th centile. Audiometry was normal. He attended secondary school with education assistance. At the last clinical evaluation, at 20 yrs.,

height was 168 cm (5-10th centile); weight was 71 kg (50th centile) and OFC was 58.5 cm (75-90th centile). He showed severe dyspraxia, fine motor skills impairment, sparse hair, long eyelashes, hoarse voice, fine hairs on the legs and overlapping 2nd-3rd toes. Growth hormone levels were normal. Exome sequencing did not reveal any potentially deleterious variant both on the X chromosome and on autosomes.

4. DISCUSSION

The aim of our work was to identify novel ID genes using WES on families enriched for X-LID. We selected 18 males with both syndromic and non-syndromic ID, the majority of whom with a family history compatible with an X-linked transmission. We analyzed their mothers to determine the X-inactivation pattern, and found seven cases with an extremely skewed X-inactivation.

Although this pattern can be incidental ¹⁰, it has frequently been associated with a mutant gene on the X chromosome causing positive selection for the cells expressing the wild type allele, and it is a common finding in mothers of X-LID patients ^{6,7}. Nevertheless, our strategy missed genes which do not affect X-inactivation, estimated to account for one third of males with X-LID ²⁵.

In five sons, we found pathogenic variants associated with ID (71%), in agreement with literature data, reporting a detection rate of 67% in cases with a family history compatible with X-LID and a mother with skewed X-inactivation ²⁵. The gene mutated could fully explain the phenotype in only one case: subject XEX-16-TO, clinically reminding a Coffin-Lowry syndrome, had a novel missense mutation in the *RPS6KA3* gene, potentially deleterious as inferred by bioinformatics analysis and segregating with the disease in the affected brother. The mutation was missed in a previous Sanger sequencing screening performed by another laboratory.

Two cases were found with *ATRX* mutations. Specifically, in subject XEX-11-TO, clinical features were only partially indicative of *ATRX*. However, the mutation had already been described in at least four patients with this disorder, suggesting phenotypic variability ^{26,27}. In subject XEX-03-TO, we unexpectedly found a likely pathogenic variant in the *ATRX* gene. Bioinformatics predictions

and its absence from ExAC/SNP databases strongly supported this assumption. However, we could not causally relate it with the complex clinical phenotype, which included pontocerebellar atrophy. We deep searched for further mutated genes in exome data and, in particular, we carefully inspected genes involved in pontocerebellar atrophy (*AMPD2*, *CHMP1A*, *CLP1*, *EXOSC3*, *RARS2*, *SEPSECS*, *TSEN2*, *TSEN34*, *TSEN54*, and *VRK1*) without finding any potentially remarkable variant.

In three further cases, molecular data pointed to a combined phenotype arising from two Mendelian defects. Indeed, the occurrence of concomitant disorders does represent a relatively common event recognized by WES, estimating to occur in approximately 5% of the analyzed cases, and may be mistaken for new disorders or variant phenotypes of a known disorder²⁸⁻³⁰.

Subject XEX-12-TO carried a missense mutation in the Duchenne Muscular Dystrophy (DMD) gene: this was considered an incidental finding, and prompted a clinical re-evaluation of the patient, which indeed documented signs and symptoms compatible with the diagnosis of Becker Muscular Dystrophy (BMD). Although missense changes in *DMD* are reported in ID, they are located in the carboxy-terminal end of the protein³¹. Thus, it is unlikely DMD mutation causes ID, also considering its severity.

Exome data also showed a missense change in the *RPS4X* gene, encoding for the 40S ribosomal protein S4. This protein is part of the postsynaptic proteome (PSP), the complement of proteins localized within the postsynaptic terminal. Genes encoding for these proteins have been found enriched in mutations in X-LID, and although no specific human disease has specifically been linked to *RPS4X*³². *RPS4X* lies within the candidate region of Abidi type X-linked mental retardation syndrome (MRXSAB, OMIM 300262), which is characterized by short stature and small head circumference. Less frequent features are sloping forehead (5 of 8), small testes (4 of 6), hearing loss (3 of 8) and cupped ears (2 of 8)³³. Short stature and small head circumference are also present in patient XEX-12-TO. Of note, zebrafish studies revealed that knocking-down *rsp4* (the orthologue of *RPS4X*) results in smaller telencephalon and aplastic midbrain-hindbrain boundary

with smaller head size³⁴. These data overall support a role for *RPS4X* in ID. However, both *RPS4X* and *DMD* escape X-inactivation^{35,36}, suggesting their mutation cannot explain the complete skewed X-inactivation in the proband's mother and grandmother. We need to hypothesize skewed X-inactivation is a random event or it is associated to a third genetic variant undetected by our analysis. Overall, the role of *RPS4X* in X-LID is controversial and needs to be further explored.

Subject XEX-10-TO and his brother showed ID associated with a marfanoid phenotype, which led to a clinical diagnosis of Lujan-Fryns syndrome. Exome sequencing excluded mutations in *MED12*, and revealed a c.649 C>T (p.Pro217Ser) variant in the *MECP2* gene. Mutations in this gene cause Rett syndrome (MIM: 312750), a severe neurodevelopmental disorder that affects 1:10,000 females³⁷. *MECP2* mutations have been initially supposed to be lethal in males, with the only viable exceptions being mosaicism or the presence of an extra X chromosome in subjects with Klinefelter syndrome^{38,39}. Later, a broad range of *MECP2*-related clinical presentations have been described in males, ranging from severe phenotypes (congenital encephalopathy or Rett syndrome) to mild ID or ASD. The absence of the c.649C>T variant in ExAc and dbSNP databases and *in silico* analysis suggested its pathogenicity. Moreover, the mutation was located in the transcriptional repression domain (TRD), where mutations associated with ID in males seem to cluster^{39,40} (Fig. 2c). While the neuropsychiatric phenotype was related to *MECP2*, we suspected marfanoid features could represent an independent event, possibly related to the *FBNI* p.Thr1020Ala variant. This change was shared by the affected brother, having an identical phenotype, and absent from the healthy sibling. The p.Thr1020Ala has been found in patients with marfanoid phenotypes^{41,42} and in a patient with the Lujan-Fryns phenotype⁴³. According to the UMD-FBN1 mutation database, it is probably pathogenic^{44,45}. In our family, however, the p.Thr1020Ala was inherited from the healthy father, suggesting it has an incomplete penetrance or it is unrelated to the phenotype.

Mutation analysis of the two remaining cases (XEX-06-TO and XEX-18-TO) did not yield any significant result. We hypothesize that WES failed detecting the X-linked mutation due to technical

limitations. Beside a known proportion of exons uncaptured by WES (Table A.2), this strategy can also miss mutations in well-covered regions. In a recent work, ~3% of variants in coding regions were missed by WES and identified by whole genome sequencing (WGS)⁴⁶. Alternatively, the mutation may be located in intronic or regulatory regions uncovered by WES.

As further speculation, the skewed inactivation in the two probands' mothers may also be unrelated with the pathology: in fact, skewed X chromosomes are relatively common in adult females, with ~14% of individuals having greater-than-20% skewed XCI^{10,47}. Thus, we also excluded autosomal recessive diseases. Sequencing the *trios* may also have enabled the identification of rare *de novo* dominant variants, likely missed by sequencing one proband only.

Our work demonstrates that a careful clinical selection coupled with high throughput exome sequencing is useful to narrow the field in which searching for potentially harmful mutations. However, strategies unbiased by coverage/depth such as WGS are necessary to solve complex cases.

ACKNOWLEDGEMENTS

We are grateful to the participating families. This work was supported by the following grants: MURST60% (to A.B.), Fondazione Bambino Gesù (Vite Coraggiose to M.T.), Ministero della Salute (RC2016, to M.T.), and CINECA (computational resources, to M.T.).

LEGENDS TO FIGURES

Figure 1. Pedigrees of XEX-03-TO, XEX-06-TO, XEX-11-TO, XEX-16-TO, and XEX-18-TO.

Black symbols indicate affected cases. A black dot in female symbols indicates a carrier status. A line above symbols indicate DNA availability. Electropherograms are shown for *ATRX* and *RPS6KA3* mutations. For details on clinical phenotypes, see Table A.1.

Figure 2. Pedigrees of XEX-10-TO and XEX-12-TO, mutation analysis and *MECP2* mutations in males.

In panel a, pedigrees with mutation segregation are shown. Black symbols indicate affected cases. A black dot in female symbols indicates a carrier status. A line above symbols indicate DNA availability. For details on clinical phenotypes, see Table A.1. In panel b, electropherograms for *MECP2*, *FBN1*, *DMD*, and *RPS4X* mutations are reported. In panel c, the genomic and domain structure of *MECP2* gene and its encoded protein are illustrated. Coding (black boxes) and non-coding (grey boxes) *MECP2* exons are shown, with the corresponding functional domains at protein level (methyl-CpG-binding domain, MBD; transcriptional repression domain, TRD; three AT-hooks, H1–H3; C-terminal domain, C-ter). Known *MECP2* mutations identified in male patients associated with Autism Spectrum Disorders (ASD, red) or intellectual disability (green) are reported (HGMD professional, june 2016). Nonsense changes cluster towards the C-terminal region of the protein, suggesting in both cases a residual protein function. The aminoacid change (p.Pro217Ser, boxed) here described is located at the beginning of the TRD domain.

Tables

Table 1. Cases whose mother had a skewed X-inactivation.

<i>Family</i>	<i>X-inactivation (mother)^a</i>	<i>Mutated gene</i>	<i>Mutation</i>	<i>Protein change</i>	<i>Clinical diagnosis</i>
XEX-03-TO	100%	<i>ATRX</i>	c.7366_7367insA	p.Met2456Asnfs*42	Pontocerebellar hypoplasia and ID.
XEX-06-TO	84%	-	-	-	ASD with ID (IQ 70);OCS
XEX-10-TO	87%	<i>MECP2</i>	c.649C>T	p.Pro217Ser	Lujan-Fryns phenotype
		<i>FBNI</i>	c.3058A>G	p.Thr1020Ala	
XEX-11-TO	100%	<i>ATRX</i>	c.6253C>T	p.Arg2085Cys	ATRX
XEX-12-TO	100%	<i>DMD</i>	c.707T>G	p.Phe236Cys	Syndromic severe ID/BMD
XEX-16-TO	97%	<i>RPS6KA3</i>	c.2188A>G	p.Arg730Gly	Coffin-Lowry syndrome
XEX-18-TO	100%	-	-	-	Non-syndromic ID

Note: a) X-inactivation was calculated as described in materials and methods. ASD: Autism Spectrum Disorder. BMD: Becker Muscular Dystrophy. ID: Intellectual disability. OCS: Obsessive compulsive disorder.

Table 2. *In silico* pathogenicity prediction and population frequency of the putative disease-causing variants identified by WES.

Software	XEX-03-TO		XEX-11-TO		XEX-16-TO		XEX-10-TO		XEX-12-TO					
	<i>ATRX</i> p.M2456Nfs*42		<i>ATRX</i> p.R2085C		<i>RPS6KA3</i> p.R730G		<i>MECP2</i> p.P217S		<i>FBN1</i> p.T1020A					
	Description	Output value	Description	Output value	Description	Output value	Description	Output value	Description	Output value				
PolyPhen-2 ⁴⁸	Probably damaging	1	Probably damaging	1	Probably damaging	1	Benign	0.26	Benign	0	Probably damaging	1	Probably damaging	0.91
SIFT ⁴⁹	Deleterious	3.7	Deleterious	3.2	Deleterious	4.3	Deleterious	3.8	Tolerated	0.12	Deleterious	3.6	Deleterious	3.8
Mutation Tasting ⁵⁰	Disease causing	0.99	Disease causing	1.0	Disease causing	1	Disease causing	1	Polymorphism	0.99	Disease causing	1	Disease causing	1.0
PMUT ⁵¹	Pathogenic	8	Pathogenic	7	Pathogenic	7	Pathogenic	5	Neutral	7	Pathogenic	9	Neutral	8
PHD-SNP ⁵²	Disease	5	Disease	n.a.	Disease	8	Neutral	6	Neutral	6	Disease	5	Neutral	2
ExAC	0	0	0	0	0	0		0.00046	0	0	0	0	0	
dbSNP147	0	0	0	0	0	0		0.0006	0	0	0	0	0	

Table A.1. Clinical features of the 18 patients included in this study.

Family	Key clinical features	Other affected relatives
XEX-01-TO	ID, epilepsy, dysmorphisms	Brother affected (ID, dysmorphisms)
XEX-02-TO	Severe hypotonia, pontocerebellar hypoplasia – diagnosis of myotubular myopathy (mut MTM1)	Brother affected
XEX-03-TO	ID, microcephaly, cerebellar hypoplasia, pontocerebellar atrophy	Maternal uncle affected (ID)
XEX-05-TO	Severe ID, ASD and epilepsy	Brother affected
XEX-06-TO	ID, ASD, OCS	Brother affected (ASD)
XEX-07-TO	ID	Brother affected
XEX-08-TO	ID, overweight	Brother affected
XEX-09-TO	ID, dysmorphisms	Brother affected
XEX-10-TO	ID, marfanoid habitus (Lujan-Fryns)	Brother affected
XEX-11-TO	ID, microcephaly, short stature, dysmorphisms	No siblings affected; suspected ATRX
XEX-12-TO	Hypotonia, ID, ASD, dysmorphisms	No siblings affected; suspected ATRX
XEX-15-TO	ID, microcephaly	Brother affected
XEX-16-TO	ID, Coffin-Lowry phenotype	Brother affected
XEX-17-TO	ID, epilepsy	Sister, three sons of maternal aunts; maternal grandfather
XEX-18-TO	Moderate ID (Wisc/III IQ<45)	Brother affected, mild ID (Wisc/III IQ 55)
XEX-19-TO	Mild ID, short stature, dysmorphisms	No siblings affected; Aarskog like (FGD1 negative)
XEX-20-TO	ID, tall stature, dysmorphisms	Brother affected
XEX-21-TO	ASD	Brother affected (ID)

Notes: ID: Intellectual Disability; ASD: Autism Spectrum Disorder; OCS: Obsessive Compulsive Disorder.

Table A.2. WES coverage and depth

Family	# paired-end reads	High Quality variants	WES			Chromosome X		
			Coverage 2x (%)	Coverage 20x (%)	Median depth (x)	Coverage 2x (%)	Coverage 20x (%)	Median depth (x)
XEX-03-TO	132,082,938	60,723	99.0	92.6	62.0	99.0	83.9	44.7
XEX-06-TO	137,998,334	61,612	98.9	91.9	59.6	98.9	81.5	42.7
XEX-10-TO	71,573,574	57,994	98.3	89.3	55.1	98.5	77.5	40.1
XEX-11-TO	112,924,964	58,528	98.6	90.0	54.9	98.7	77.9	40.1
XEX-12-TO	148,117,414	61,680	99.0	93.2	64.7	98.9	85.2	46.7
XEX-16-TO	71,259,282	59,338	98.7	91.2	56.8	98.6	78.2	40.6
XEX-18-TO	116,926,392	58,758	98.8	88.9	51.9	98.8	74.8	38.8

REFERENCES

1. Stevenson RE, Schwartz CE. X-linked intellectual disability: unique vulnerability of the male genome. *Dev Disabil Res Rev.* 2009; **15**: 361-368.
2. Lubs HA, Stevenson RE, Schwartz CE. Fragile X and X-linked intellectual disability: four decades of discovery. *Am J Hum Genet.* 2012; **90**: 579-590.
3. Raymond FL. X linked mental retardation: a clinical guide. *J Med Genet.* 2006; **43**: 193-200.
4. de Brouwer AP, Yntema HG, Kleefstra T, et al. Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium. *Hum Mutat.* 2007; **28**: 207-208.
5. Crow YJ, Tolmie JL. Recurrence risks in mental retardation. *J Med Genet.* 1998; **35**: 177-182.
6. Van den Veyver IB. Skewed X inactivation in X-linked disorders. *Semin Reprod Med.* 2001; **19**: 183-191.
7. Plenge RM, Stevenson RA, Lubs HA, Schwartz CE, Willard HF. Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. *Am J Hum Genet.* 2002; **71**: 168-173.
8. Lyon MF. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature.* 1961; **190**: 372-373.
9. Lyon MF. Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet.* 1962; **14**: 135-148.
10. Amos-Landgraf JM, Cottle A, Plenge RM, et al. X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. *Am J Hum Genet.* 2006; **79**: 493-499.
11. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature.* 2005; **434**: 400-404.
12. Uchida T, Ohashi H, Aoki E, et al. Clonality analysis by methylation-specific PCR for the human androgen-receptor gene (HUMARA-MSP). *Leukemia.* 2000; **14**: 207-212.
13. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet.* 1992; **51**: 1229-1239.
14. Cordeddu V, Redeker B, Stellacci E, et al. Mutations in ZBTB20 cause Primrose syndrome. *Nat Genet.* 2014; **46**: 815-817.
15. Kortum F, Caputo V, Bauer CK, et al. Mutations in KCNH1 and ATP6V1B2 cause Zimmermann-Laband syndrome. *Nat Genet.* 2015; **47**: 661-667.
16. Niceta M, Stellacci E, Gripp KW, et al. Mutations Impairing GSK3-Mediated MAF Phosphorylation Cause Cataract, Deafness, Intellectual Disability, Seizures, and a Down Syndrome-like Facies. *Am J Hum Genet.* 2015; **96**: 816-825.
17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; **25**: 1754-1760.
18. McKenna N, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010; **20**: 1297-1303.
19. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011; **43**: 491-498.
20. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly.* 2012; **6**: 80-92.
21. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat.* 2013; **34**: E2393-2402.
22. D'Aurizio R, Pippucci T, Tattini L, Giusti B, Pellegrini M, Magi A. Enhanced copy number variants detection from whole-exome sequencing data using EXCAVATOR2. *Nucleic Acids Res.* 2016.
23. Trivier E, De Cesare D, Jacquot S, et al. Mutations in the kinase Rsk-2 associated with Coffin-Lowry syndrome. *Nature.* 1996; **384**: 567-570.
24. Omoe K, Endo A. Relationship between the monosomy X phenotype and Y-linked ribosomal protein S4 (Rps4) in several species of mammals: a molecular evolutionary analysis of Rps4 homologs. *Genomics.* 1996; **31**: 44-50.
25. Tzschach A, Grasshoff U, Beck-Woedl S, et al. Next-generation sequencing in X-linked intellectual disability. *Eur J Hum Genet.* 2015; **23**: 1513-1518.

26. Gibbons RJ, Wada T, Fisher CA, et al. Mutations in the chromatin-associated protein ATRX. *Hum Mutat*. 2008; **29**: 796-802.
27. Lacoste C, Leheup B, Agouti I, Mowat D, Giuliano F, Badens C. Mutations of codon 2085 in the helicase domain of ATRX are recurrent and cause ATRX syndrome. *Clin Genet*. 2014; **86**: 502-503.
28. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013; **369**: 1502-1511.
29. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014; **312**: 1870-1879.
30. Tarailo-Graovac M, Shyr C, Ross CJ, et al. Exome Sequencing and the Management of Neurometabolic Disorders. *N Engl J Med*. 2016; **374**: 2246-2255.
31. Lenk U, Hanke R, Thiele H, Speer A. Point mutations at the carboxy terminus of the human dystrophin gene: implications for an association with mental retardation in DMD patients. *Hum Mol Genet*. 1993; **2**: 1877-1881.
32. Laumonnier F, Cuthbert PC, Grant SG. The role of neuronal complexes in human X-linked brain diseases. *Am J Hum Genet*. 2007; **80**: 205-220.
33. Abidi F, Hall BD, Cadle RG, et al. X-linked mental retardation with variable stature, head circumference, and testicular volume linked to Xq12-q21. *Am J Med Genet*. 1999; **85**: 223-229.
34. Uechi T, Nakajima Y, Nakao A, et al. Ribosomal protein gene knockdown causes developmental defects in zebrafish. *PLoS One*. 2006; **1**: e37.
35. Fisher EM, Beer-Romero P, Brown LG, et al. Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell*. 1990; **63**: 1205-1218.
36. Zhang Y, Castillo-Morales A, Jiang M, et al. Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. *Mol Biol Evol*. 2013; **30**: 2588-2601.
37. Lyst MJ, Bird A. Rett syndrome: a complex disorder with simple roots. *Nat Rev Genet*. 2015; **16**: 261-275.
38. Villard L. MECP2 mutations in males. *J Med Genet*. 2007; **44**: 417-423.
39. Bianciardi L, Fichera M, Failla P, et al. MECP2 missense mutations outside the canonical MBD and TRD domains in males with intellectual disability. *J Hum Genet*. 2016; **61**: 95-101.
40. Schonewolf-Greulich B, Tejada MI, Stephens K, et al. 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**: 733-738.
41. Attanasio M, Lapini I, Evangelisti L, et al. FBN1 mutation screening of patients with Marfan syndrome and related disorders: detection of 46 novel FBN1 mutations. *Clin Genet*. 2008; **74**: 39-46.
42. Tiecke F, Katzke S, Booms P, et al. Classic, atypically severe and neonatal Marfan syndrome: twelve mutations and genotype-phenotype correlations in FBN1 exons 24-40. *Eur J Hum Genet*. 2001; **9**: 13-21.
43. Stheneur C, Collod-Beroud G, Faivre L, et al. Identification of the minimal combination of clinical features in probands for efficient mutation detection in the FBN1 gene. *Eur J Hum Genet*. 2009; **17**: 1121-1128.
44. Collod-Beroud G, Le Bourdelles S, Ades L, et al. Update of the UMD-FBN1 mutation database and creation of an FBN1 polymorphism database. *Hum Mutat*. 2003; **22**: 199-208.
45. Frederic MY, Lalande M, Boileau C, et al. UMD-predictor, a new prediction tool for nucleotide substitution pathogenicity -- application to four genes: FBN1, FBN2, TGFBR1, and TGFBR2. *Hum Mutat*. 2009; **30**: 952-959.
46. Belkadi A, Bolze A, Itan Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A*. 2015; **112**: 5473-5478.
47. Wong CC, Caspi A, Williams B, Houts R, Craig IW, Mill J. A longitudinal twin study of skewed X chromosome-inactivation. *PLoS One*. 2011; **6**: e17873.
48. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature methods*. 2010; **7**: 248-249.
49. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res*. 2001; **11**: 863-874.
50. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature methods*. 2014; **11**: 361-362.

51. Ferrer-Costa C, Gelpi 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.
52. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. 2006; **22**: 2729-2734.

Figure 1.

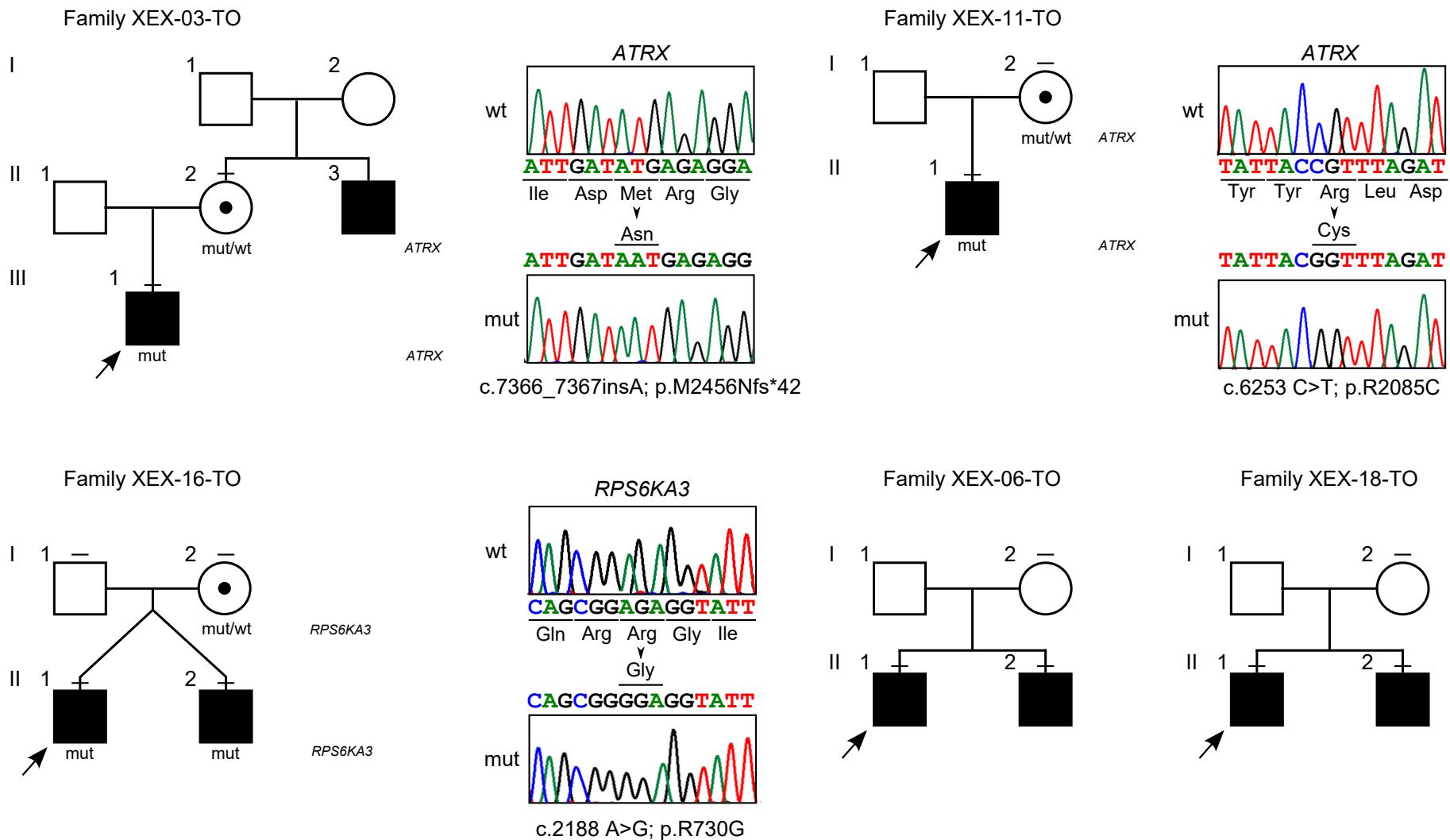
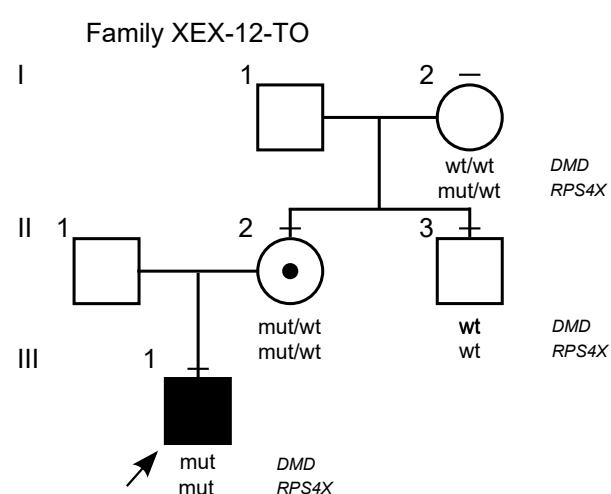
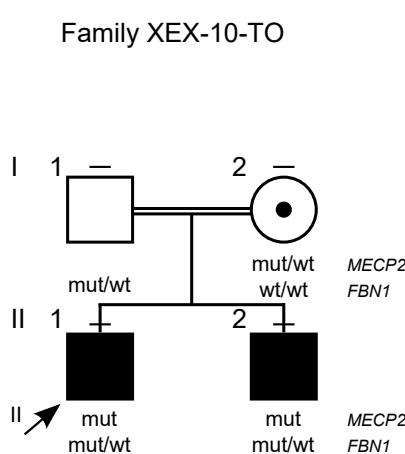
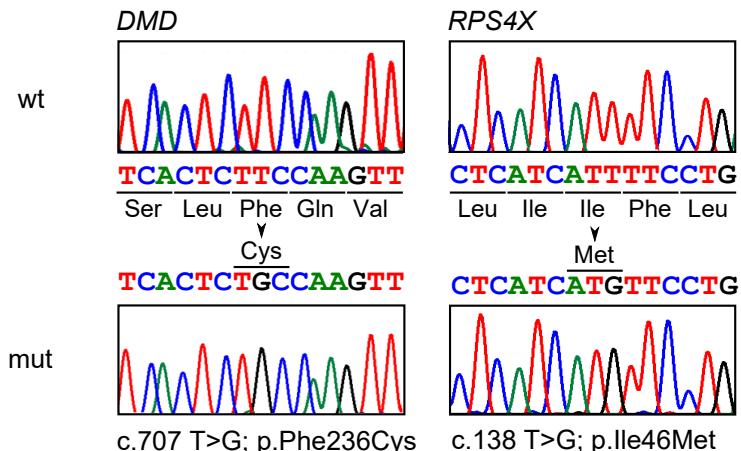
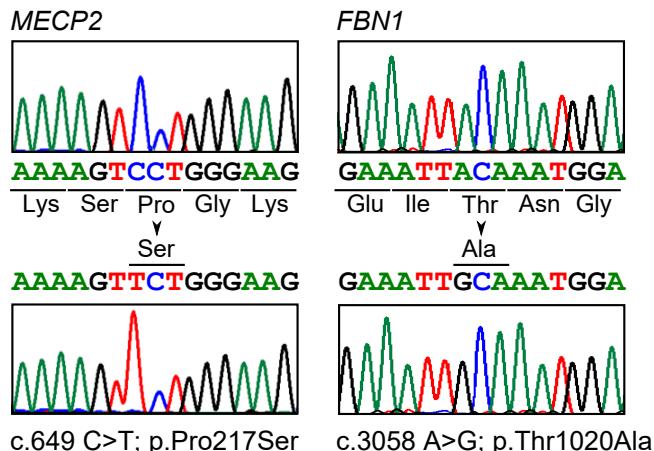


Figure 2

a)



b)



c)

