Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Skewed X-chromosome Inactivation in Unsolved Neurodevelopmental Disease Cases Can Guide Re-evaluation for X-linked Genes

Alfredo Brusco (alfredo.brusco@unito.it)
University of Turin https://orcid.org/0000-0002-8318-7231
Chiara Giovenino
University of Turin https://orcid.org/0000-0003-2041-4019
Slavica Trajkova
Lisa Pavinato
University of Torino https://orcid.org/0000-0002-7630-8365
Simona Cardaropoli
Verdiana Pullano
https://orcid.org/0000-0002-0409-9832
Elena Sukarova-Angelovska
University Sv. Kiril i Metodij https://orcid.org/0000-0001-9702-3994
Silvia Carestiato
Paola Salmin
Antonina Rinninella
Anthony Battaglia
Luca Bertoli
Antonio Fadda
Flavia Palermo
Diana Carli
University of Torino https://orcid.org/0000-0001-5690-6504
Alessandro Mussa
Paola Dimartino
Alessandro Bruselles
Tawfiq froukh
Philadelphia University
Giorgia Mandrile
Barbara Pasini
Università degli Studi Di Torino https://orcid.org/0000-0002-4373-1212
Silvia De Rubeis
Icahn School of Medicine at Mount Sinai https://orcid.org/0000-0001-9383-6883
Joseph Buxbaum
Mount Sinai School of Medicine https://orcid.org/0000-0001-8898-8313
Tommaso Pippucci
University of Bologna https://orcid.org/0000-0001-7737-7963
Marco Tartaglia
Ospedale Pediatrico Bambino Gesù https://orcid.org/0000-0001-7736-9672
Marzia Rossato
University of Verona
Massimo Delledonne
University of Verona https://orcid.org/0000-0002-7100-4581
Giovanni Battista Ferrero

Article

Keywords: chromosome X inactivation, X-linked gene, neurodevelopmental disorders, exome sequencing reanalysis, breakpoint definition, Xdrop technology

Posted Date: November 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2179710/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at European Journal of Human Genetics on March 6th, 2023. See the published version at https://doi.org/10.1038/s41431-023-01324-w.

Abstract

Despite major technical and genetic advances, more than half of the neurodevelopmental disorders (NDDs) cases remain undiagnosed.

We explored the frequency of non-random XCI in the mothers of male patients and in affected females from a clinically heterogeneous cohort of unsolved NDD cases, negative at FRAXA, chromosomal microarray analysis and Trio Exome Sequencing. We hypothesize that an unbalanced XCI could unmask previously discarded genetic variants on the X chromosome connected both to XCI and NDD.

A multiplex fluorescent-PCR-based assay was used to screen the XCI pattern after methylation sensitive *HhaI* digestion. Trio-based ES re-analysis was performed in families with skewed XCI occurrence. Linkage analysis and RT-PCR were used to further study the X-chromosome inactive allele. X-drop was used to define the chromosome deletion boundaries.

We found a skewed XCI (>90%) in 16/186 mothers of affected NDD males (8.6%) and 12/90 female patients (13.3%), far beyond the expected XCI in normal population (3.6%, OR=4.10; OR=2.51). Reanalyzing ES and clinical data, we solved 7/28 cases (25%). These included variants in the *KDM5C, PDZD4, PHF6, TAF1, OTUD5*, and *ZMYM3,* and a genomic deletion spanning exons 3-4 of the *ATRX* gene.

The identification of a skewed XCI is an easy assay that can help selecting a subgroup of patients for the re-evaluation of X-linked variants, improving the diagnostic yield in NDD patients, and allowing the identification of new X-linked disorders.

Introduction

Major advances in exome sequencing (ES) technologies and data analysis, along with the continue identification of novel disease genes, have greatly increased the diagnostic rate of neurodevelopmental disorders (NDDs). Yet, 50–70% of the cases remain unsolved.(1–3) Several reasons make the molecular diagnosis still challenging, including (i) unspecific phenotypes, (ii) difficult interpretation of variants of uncertain significance (VUS)(4), (iii) newly defined disease with few patients described, not able to draw clear conclusions about phenotypic expansion(4), (iv) technical limitations of the used diagnostic tool.(5) In these unsolved NDD cases, several complementary approaches can be attempted to increase diagnostic yield, such as transcriptome analysis and whole genome sequencing. Re-analysis of ES data has proven to be the most effective increasing the diagnostic yield by 10–15%.(6)

Approximately 6% of NDDs are estimated to be X-linked. The causative event is often though to lead to a non-random X-chromosome inactivation (XCI) proportion (skewing).

XCI process and its fine-tuning, although known for many years, is still to be defined and further explored.(7) Physiologically, XCI is random and results in an approximate equal ratio of cells expressing maternal or paternal X chromosome genes.(8) XCI skewing can be considered preferential (80:20) or extreme (90:10), and is a frequent indication of an X-linked pathogenetic variant, in both affected females(9) or mother of male patients.(10)

In healthy females, heterozygote for an X-linked pathogenetic variant, skewed XCI favors the expression of the wild type allele, thus protecting from the effect of the deleterious variant.(10, 11) More recently, skewed XCI has been associated also with affected females.(9) This latter case could be explained by skewing towards a deleterious allele that reaches a pathogenicity threshold. In these patients, females result susceptible to X-linked recessive conditions with a phenotype similar to male patients.(9) For X-linked dominant conditions, which are usually lethal in males, a skewing against the deleterious allele that decreases expression below lethality has also been observed.(12)

We exploited the analysis of XCI skewing on 276 deeply phenotyped NDD undiagnosed patients (90 females and 186 mothers of unsolved males). We show that this test combined with targeted re-evaluation of ES data and functional analyses can increase the diagnostic yield and identify novel X-linked disease genes.

Material And Methods

Study cohort

Within a large international collaborative study aimed at identifying the genetic bases of NDD, we selected 91 affected females, and 189 mothers of affected males negative at trio-ES, CMA (50K Agilent), and FRAXA (see Supplemental materials and methods).

X-chromosome inactivation analysis

XCI was tested in blood extracted DNA using an in-house developed protocol.

The XCI pattern was calculated using three independent microsatellite polymorphic markers on the X chromosome : (i) the CA-repeat in the promoter region of the SLIT and NTRK Like Family Member 4 (*SLITRK4*) gene; (ii) the CAG-repeat located in exon 1 of androgen receptor (*AR*) gene(13); (iii) the CA and AG tandem repeats in the first intron of Proprotein Convertase Subtilisin/Kexin Type 1 Inhibitor (*PCSK1N*) gene (Supplemental Fig. 1; supplemental materials and methods).

Characterization of ATRX breakpoints by indirect sequence capture coupled with Illumina sequencing

The Xdrop-based enrichment and subsequent amplification of enriched DNA was conducted at Samplix's facility as previously described(14) and subsequently sequenced in 150PE on a NovaSeq6000 (Illumina)(Supplemental materials and methods).

RNA extraction and RT-PCR

To analyze the effect of the variants on cDNA, we generated and amplified cDNA from total RNA extracted from patients' fresh blood (Supplemental materials and methods).

X-chromosome inactivation phasing by linkage analysis

Families with variants inherited from the mother and segregated in different subjects were analyzed by a set of markers to phase the identified variant with the active or inactive X chromosome and exclude recombination events. The following genetic markers: DXS993, DXS991, DXS986, DXS1068, DXS990, were amplified using AmpliTaq Gold DNA Polymerase (Thermo Fisher Scientific) (see Supplemental Material and Methods for details), separated by capillary electrophoresis on an ABI 3130xl DNA analyzer (Thermo Fisher Scientific) with the GeneScan 500 LYZ size standard (Thermo Fisher Scientific) and analyzed using the GeneMapper software v.4 (Thermo Fisher Scientific).

Results

X chromosome inactivation assay

HUMARA test, based on the analysis of a CAG repeat at the *AR* locus, is currently the most used method for XCI evaluation.(13) HUMARA is uninformative in ~ 21% of females because homozygotes, or with alleles of difficult interpretation.(15) For this reason, we set up a method based on a fluorescent multiplex methylation-sensitive PCR that contemporarily amplifies *AR* and two additional independent polymorphic microsatellites within the *SLITRK4* and *PCSK1N* genes (Supplemental Fig. 1). Firstly, we evaluated our assay in a female patient with a balanced Xq25;8q24 translocation and complete XCI-skewing (100:0), previously assessed by HUMARA.(16) Complete XCI skewing was confirmed using the two informative loci (Supplemental table 2), showing the validity of the test.

To further test the assay, we evaluated the XCI pattern in four females affected by NDD and four mothers of NDD males who previously received clinical and molecular diagnosis of an X-linked condition with a potential skewed XCI (Supplemental Table 3). XCI skewing (> 80%) was documented in three affected and three healthy mothers. Our findings were consistent with the literature for *NAA10*, *PQBP1*, *MECP2*, and *ACSL4* (occurrence of XCI skewing).(17) Similarly, we found a random XCI in the healthy mother heterozygous for a pathogenic *IDS* variant, which is in line with previous observation indicating occurrence of skewing in one affected female only.(17) Finally, our female patient with *DDX3X* had a random XCI as reported for half of the patients affected by MRXSSB (MIM #300958).(17) No or limited information was available in the literature for *HNRNPH2*, *RBM10*.

XCI screening in unsolved NDD cases

We tested 92 NDD females and 189 healthy mothers of NDD male patients. Probands had been deeply studied by CMA/trio-based ES without finding a genetic cause. Patients' phenotype was highly variable: 35% (98/281) presented intellectual disability (ID), 41% (115/281) autism spectrum disorder (ASD) and 24% (68/281) were complex syndromic cases with facial dysmorphism (Supplemental Tables 4–5).

XCI assay was informative in 276 out of 281 cases (98.2%) (90 undiagnosed female NDD patients and 186 healthy mothers) for at least one marker (Supplemental Tables 6–7). We noted an enrichment of cases with extreme skewing (> 90%) in both groups: 12 out of 90 affected females (13.3% vs. 3.6%; OR = 4.10; IC 95% 1.85–9.10) and 16 out of 186 healthy mothers (8.6% vs. 3.6%; OR = 2.51; IC 95% 1.21–5.19) (Table 1; Supplemental Fig. 2–3, Supplemental Table 2).

Genetic analyses in XCI skewed cases

To exclude a possible genetic cause of skewed X inactivation, we sequenced the *XIST* minimal promoter in the identified 28 cases. We aimed at finding possible rare variants which may underlie epigenetic and functional differences between X chromosome in females as described by Plenge *et al.*(18) No such variant was found. Thus, we reasoned to re-evaluate the ES data focusing on X-chromosome variants, assessing their relevance also in the light of newly available clinical information. We identified an X-linked variant consistent with the phenotype in three females and five males (8/28; 28.6%) (Table 2), as detailed below.

Table 2
Variants found in the eight families with X-skewed females

Family_ID	Sex	Diagnosis	OMIM	Inheritance	Gene	Variant c.DNA	Variant protein	SIFT DANN LRT MutationTaster	A C
113	F	MRXS33	300966	maternal	*TAF1*	NM_004606.5: c.745G > A	p.Gly249Arg	D 0.99 D D	C F F F
NWM24	F	BFLS	301900	de novo	*PHF6*	NM_001015877.2: c.890G > T	p.Cys297Phe	D 0.99 D D	C F F F
237	F	MRXSJ	300534	maternal	*KDM5C*	NM_004187.5: c.1204G > A	p. Asp402Asn	D 0.99 D D	C F F
234	M	MCAND	301056	maternal	*OTUD5*	NM_017602.4: c.1526C > T	p.Pro509Leu	D 0.99 N D	C F F
236	M	MRXFH1	309580	maternal	*ATRX*	NC_000023.10:g.76953033_76959004del	p.?	/ / / /	C
TF110	F		* 300634	de novo	*PDZD4*	NM_032512.5: c.10_16del	p.(Asn4Alafs*12)	/ / / /	C F F
NWM25	M		* 300634	maternal	*PDZD4*	NM_032512.5: c.2190G > C	p.Lys736Asn	D 0.99 D D	C F
NWM127	M		* 300061	maternal	*ZMYM3*	NM_201599.3: c.1322G > A	p.Arg441Gln	T 0.99 / D	C F

Notes. *Data from(17); **Data from(38); D = Deleterious; N = Neutral. Variants submitted to ClinVar (SUB12130121)

A novel TAF1 variant in multiple affected members of family 113

Proband was a 2-year-old girl affected by global developmental delay and delayed psychomotor development with an almost completely skewed XCI (95:5; Fig. 1A, Supplemental Table 2). She had a 14-year-old brother affected by ID, delayed speech and language development, food restriction, and behavioral abnormalities (II.1); he had a healthy 11-year-old brother (II.2). Their mother (I.2) reported she had teaching support at school.

We found the c.745G > A p.(Gly249Arg) variant in Transcription initiation factor TFIID subunit 1 (*TAF1*), a gene associated with X-linked syndromic intellectual developmental disorder-33 (MRX33; MIM# 300966). MRX33 is characterized by delayed psychomotor development, ID, and typical facial dismorphisms (Supplemental Table 8).(19)

The variant was inherited from the mother, who also had a skewed XCI (90:10), segregated to her affected brother, and was absent in the healthy brother (Fig. 1A, B). The p.(Gly249Arg) variant was absent in the GnomAD database (ver.2.1.1), and hits a highly conserved nucleotide (PhyloP = 9.37; PhastCons = 1) and amino acid residue, which is maintained from vertebrates to *Drosophila melanogaster* (Fig. 1C). The change was predicted intolerant by MetaDome(20) (Fig. 1D) and deleterious by CADD (Phred: 24.7)(21). Most reported likely pathogenic/pathogenic *TAF1* variants are missense clustering between exon 16 and 30, whereas p.(Gly249Arg) is located in exon 6 (Fig. 1E). Nevertheless, using MutScore, which takes into consideration positional clustering of variants already detected in disease-associated genes and variants found in the population, we noted the variant reaches a predicted pathogenicity score of 0.96 (maximum 1).(22) Pathogenicity of p.(Gly249Arg) was also supported by the predicted structural damage triggered by disallowed phi/psi alert(23) (Supplemental Material and Methods, Supplemental Fig. 4A).

Using a series of microsatellite markers on the X chromosome, we segregated the haplotype containing the c.745G > A *TAF1* variant in the family. Because the haplotype also spanned the AR microsatellite, we could determine the c.745G > A was located on the inactive X chromosome (X_i) in both the mother (I.2) and her daughter (II.3)(Fig. 1A).

A rare de novo PHF6 variant affecting the female proband of family NWM24

Proband of family NWM24 was a 7-year-old girl, second child from healthy parents. She was born small for gestational age (SGA), and presented global developmental delay, autism disorder, several dysmorphic features, divergent strabismus and brachy/syndactyly. Her XCI was completely skewed (100:0) (Fig. 1F, Supplemental Table 2). We identified a *de novo* c.890G > T p.(Cys297Phe) variant in PHD finger protein 6 (*PHF6;* Fig. 1G), a gene associated with the

X-linked recessive Borjeson-Forssman-Lehmann syndrome (BFLS; MIM# 301900). This previously missed variant was reconsidered since *de novo* heterozygous variants have recently been described in affected females with an overlapping but distinct phenotype including characteristic facial dysmorphism, dental, finger and toe anomalies, and linear skin pigmentation (Supplemental Table 9).(24, 25) These features are shared with our patient. The variant is absent in GnomAD (ver 2.1.1), and hits a very conserved nucleotide (PhyloP = 9.36; PhastCons = 1) and amino acid (Fig. 1H). Cys297 is located within the PHD-like zinc-binding domain and showed is considered intolerant to changes by MetaDome (Fig. 1I; PF13771; a.a. 239–330; UniProt: Q8IWS0), where most *PHF6* pathogenic and likely pathogenic variants reported in ClinVar map (MutScore = 0.949). Bioinformatic predictions showed the change was deleterious (CADD Phred = 29.5; REVEL = 0.97; Table 2). Pathogenicity of the p.(Cys297Phe) was also supported by the predicted structural damage, the amino acid substitution trigger a clash alert(23) (local clash score: wild type = 10.47; mutant = 35.67; Supplemental Fig. 4B).

A KDM5C variant with variable expressivity in family 237

In family 237, we identified a 10- year-old female presenting with moderate ID with a skewed XCI (Fig. 1J). She was the second child of four siblings: an affected male (III.1) and two healthy sisters (III.3 and III.4). Parents were healthy, but several maternal male relatives were reported with ID. We found a maternally inherited c.1204G > A p.(Asp402Asn) missense variant in Lysine-specific demethylase 5C (*KDM5C*), associated with the intellectual developmental disorder, X-linked, syndromic, Claes-Jensen type (MRXSCJ; MIM# 300534; Fig. 1K). The variant was shared by the proband's affected brother (III.1), and one of her healthy sisters (III.3). The variant was predicted as deleterious by bioinformatic analysis (CADD Phred: 29.7; REVEL: 0.866), and the affected residues mapped in a region that was considered intolerant to variation by MetaDome (Fig. 1L), and conserved from vertebrates to fly (Fig. 1M). A variant at the same amino acid c.1204G > T p.(Asp402Tyr) had previously been demonstrated to compromise KDM5C stability and enzymatic activity.(26)

MRXSCJ is an X-linked recessive disorder, characterized by DD/ID with clinical heterogeneity in affected males.(27) Recurrent features include short stature, microcephaly, hyperreflexia, and aggressive behavior which were shared by the male sibling III.1. In females, as case III.2, variants in *KDM5C* have only recently been associated with an incomplete penetrance and variable phenotype ranging from mild to severe ID (Supplemental Table 10).(27) The presence of both a male and female in this family initially led us to discard X-linked genes.

By determining the phase of AR alleles and the *KDM5C* alleles by linkage analysis, we demonstrated that the affected sister (III.2) had a preferentially active mutant allele (90%); conversely, the unaffected sister (III.3) and her mother had a preferentially inactive mutant allele (Fig. 1J).

A genomic ATRX deletion characterized by Xdrop method in family 236

The proband of family 236 was 13 years old boy with a long diagnostic odyssey (Fig. 2A). At 3 years, he presented with hypotonia, DD/ID and dysmorphisms. The phenotype was suggestive for mental retardation-hypotonic facies syndrome (MRXFH1, MIM# 309580); however, both Sanger sequencing and ES resulted negative for an intragenic *ATRX* pathogenic variant (MIM* 300032). We identified a complete XCI skewing (100:0) in his mother, which prompted us to re-evaluate genetic data. Visually inspecting ES reads using IGV(28), we noticed that exons 3 and 4 of *ATRX* were not covered (Fig. 2B upper panel), suggesting the presence of an intragenic deletion. According to linkage analysis, the X_i chromosome in the mother carried the haplotype with the deletion (Fig. 2A).

For the in-depth characterization of the deletion, we used the Indirect Sequence Capture (Xdrop technology)(29, 30), a powerful method to characterize specific genomic regions. We enriched a tract of ~ 100 kb within the *ATRX* gene, spanning the deletion, that was subsequently sequenced at high coverage with an Illumina NGS platform. The analysis allowed identifying the breakpoints of the deletion (Supplemental Fig. 5), with an uncertainty of 3 bp identical on the two sides of the interrupted region (hg38; chrX:77,697,545 – 77,703,516; chrX:77,697,542 – 77,703,513) (Fig. 2B). Remapping of Illumina reads on the reconstructed sequence demonstrated perfect alignment, without mismatches, thus confirming the correctness of the breakpoints (Fig. 2B, lower panel). The deletion of 5,971 bp was confirmed by Sanger sequencing using flanking PCR primers and shown to be inherited from the mother (Fig. 2C, D).

PDZD4: a possible novel NDD gene in family NWM25

In family NWM25, we identified a mother of an affected boy with a 90:10 XCI (Fig. 2E, Supplemental Table 2). Since 2 years of age, her son presented symptoms of developmental delay, later evolving in kyphoscoliosis with pectus excavatum hyperelastic skin and joints, persistent hand tremors, facial dysmorphisms and polymicrogyria at brain MRI. Two of his maternal uncles were reported to be affected by undefined ID.

Reanalysis of the X-chromosome variants allowed to identify a c.2190G > C p.(Lys736Asn) missense variant in PDZ domain-containing 4 gene (*PDZD4*; MIM* 300634) (Fig. 2F), which was inherited from the healthy mother. Lysine736 is conserved in vertebrates (Fig. 2G). Using linkage analysis, we showed the haplotype with the p.(Lys736Asn) was on the inactive X-chromosome (Fig. 2E).

Exploiting GeneMatcher (https://genematcher.org/), we identified a second affected 12 year-old female (II.1; family TF110, Fig. 2H) with a *de novo* frameshift c.10_16del p.(Asn4Alafs*12) variant in *PDZD4*. She presented with an overlapping phenotype, including DD, microcephaly, ID and dysmorphisms. Also in this family, an almost complete XCI (95:5) in the unaffected mother was observed, although we could not determine if it the variant laid on the inactive X chromosome.

OTUD5 a novel recently identified gene in family 234

The probands of family 234 were two male siblings with mild ID, aged 16 and 26 years respectively. Their healthy mother showed a complete XCI skewing (100:0) (Fig. 2I). We identified a missense c.1526C > T p.(Pro509Leu) variant in *OTUD5*, recently associated with the Multiple Congenital Anomalies-Neurodevelopmental syndrome, (MCAND; MIM# 301056).(31) MCAND is an X-linked recessive congenital multisystemic disorder characterized by poor growth, global developmental delay with impaired intellectual development, and variable abnormalities of the cardiac, skeletal, and genitourinary systems. The disease severity is highly variable, ranging from death in early infancy to survival into the second or third decade, suggesting the variant found is hypomorphic.(31)

We confirmed that the c.1526C > T allele is expressed in patient blood (II.1, Fig. 2J). Then, we compared sequenced *OTUD5* cDNA from blood with genomic DNA (gDNA) in the mother and showed that the c.1526C > T allele was not detectable, suggesting that the skewed X inactivation silenced preferentially the chromosome with the variant (I.2, Fig. 2J). Bioinformatics analyses supported the variant as likely pathogenic (Table 2). The substitution of a leucine with a proline triggers a structural damage with a local clash score of 33.58 versus a score of 15.21 calculated for the wild type protein (Supplemental Fig. 4C).(23)

ZMYM3: a possible novel NDD gene in family NWM127

Subject II.1 (Fig. 2K) was a 13-year-old male with DD, moderate ID, cryptorchidism, porosis of bones and dysmorphic features. He was the fourth child in a family of European ancestry. He had an affected sister (II.2) presenting severe ID due *de novo* tetrasomy for 15q11.2-q13.1 (MIM *608636). He was severely hypotonic in early infancy and showed relevant delay in his gross motor milestones (head control at 1 year and sitting position at 5 years). He never developed fine motor skills nor acquired toilet training. Dysmorphic features included long face, tall forehead, thick eyebrows, deeply set eyes, broad nasal tip, and low-set flashy ears with cupped formed ear lobes. Re-analyzing ES data, we found a *ZMYM3* c.1322G > A p.(Arg441Gln) variant. His mother showed a completely skewed XCI (100:0), and a similar XCI skewing was found in both the affected and unaffected sisters (II.4; 90:10). Furthermore, p.(Arg441Trp) variant was described by Philips et al. 2014 in three male probands with ID and several dysmorphic features shared with proband II.1, and recently confirmed as a recurrent variant in a novel *ZMYM3*-associated NDD.(32, 33) Other potentially causative ES-variants were excluded by functional analysis (*de novo OSBPL8*: c.1535T > C; p.(Val512Ala) that did not showed alteration of the protein activity (Prof. T. Balla, Bethesda NY, personal communication).

Discussion

We reasoned that unbalanced XCI could be a guide for re-evaluating clinical and molecular data in undiagnosed NDD cases. To test XCI, we set up a multiplex fluorescent PCR simultaneously analyzing the methylation status of three independent polymorphic markers on the X-chromosome. This assay allowed us to increase informativeness to > 98%, compared to 80% using the standard HUMARA.(13)

The analysis of 91 females NDD patients and 186 mothers of male NDD patients, previously undiagnosed by CMA and trio-ES, showed a significant enrichment of subjects with extremely unbalanced XCI, defined as a > 90:10 XCI ratio (28/277, 10%) in line with a similar study.(9) This suggested that several of our undiagnosed cases could be attributed to a gene on the X chromosome. A proof of principle came from the re-evaluation of available trio-ES data: focusing on X-linked coding regions, we identified variants likely causative of a disease in 7/28 cases, solving 25% of skewed XCI cases. In our original survey of 575 NDD cases, we had 9 patients with X-linked variants classified as class 4 or 5, and 12 with class 3 variants. Considering the 28 cases with skewed XCI, we may estimate that X-linked genes account from 6.4 to 8.5% (9 + 28/575; 21 + 28/757) in our survey. These figures agree with a recent evaluation of the burden of X-linked coding variation based on 11,044 Developmental Disorder patients, who estimated X-linked causes in 6.0% of males and 6.9% of females.(34)

We previously missed the seven variants on the X chromosome because of three reasons: the gene was unknown at the time of the analysis (*OTUD5, PDZD4*, and *ZMYM3*), the variant was a structural rearrangement missed by ES (*ATRX*) or overlooked because apparently inconsistent with X-linked segregation, since both males and females were affected (*TAF1, PHF6*, and *KDM5C*).

In the former category, LINKage-specific-deubiquitylation-deficiency-induced embryonic defects (LINKED) syndrome, has been associated to pathogenic *OTUD5* variants only in 2021,(31) whereas *PDZD4* and *ZMYM3* have been presently only proposed as a disease gene.(33, 35)

The microdeletion identified in *ATRX* (family 236) highlights the importance of searching for genomic rearrangements, exploiting exome data, or performing genome sequencing. In this case the deletion was missed by CMA due to lack of array probes in the deleted tract. The strong clinical suspicion of ID-hypotonic facies syndrome (MIM# 309580) prompted us to analyze the coverage of all *ATRX* exons on ES data and to finally identify the deletion of exon 3–4.

We also chose to locate the precise breakpoints using a novel method based on the enrichment for targeted resequencing by the Xdrop technology, which combines high-resolution droplet PCR (dPCR) with droplet sorting and Multiple Displacement Amplification in droplets (dMDA).

This approach proved successful to fine-mapping the deletion breakpoints, narrowing it down from a large putative region of ≈ 20 kb between exon 2 and 5. Given the flexibility of this technology, we expect that it might be useful in the future in other similar cases where the large size of the involved region hampers the efficient use of traditional assays for the characterization of structural variations at the single-base resolution. Alternatively, to achieve the same results, we would have required either a genome sequencing (with higher costs) or a very large sets of PCR-based assays and a laborious work to map the whole 20 kb region, also because the deletion maps within a region rich in repeated sequences. The availability of the deletion boundaries allowed to set up a simple PCR test to segregate the variant in the family.

The X chromosome is often underestimated in the diagnosis of female NDD patients, because of the common perception that females are less susceptible to X-linked conditions.(17) Although many X-linked conditions show a profound sex-linked bias, given the specific mechanism of inheritance, an increasing number of X-linked disease has been described to occur similarly in both female and male patients.(17)

In families 113 and NWM24, we identified a missense variant in a female for *TAF1* and *PHF6*; we overlooked/ignored these variants at first ES reading because inconsistent with an X-linked recessive disease. However, literature data reported females with phenotypes consistent with variants in those genes. In *TAF1*, the completely skewed XCI is consistent with other recently described cases where the phenotype, different between females and males, is uniform within each sex.(36) XCI unbalance is favoring the wild type allele in both the mother (mild phenotype) and her affected daughter, leaving unclear the pathogenic mechanism. We can speculate that (i) 5% of the pathogenic allele is enough to cause the phenotype or (ii) that the XCI pattern is different in affected tissues such as brain, where the pathogenic allele is for some reason more expressed than in blood.

In family NWM24, the phenotype associated with *PHF6* is consistent with literature data reporting two females carrying the *de novo* p.Cys305Phe, a few amino acids distant from our proband's variant.(24, 37)

Family 237 is another example of X-linked gene complexity: three females carrying a missense variant in *KDM5C*, but we detected skewed XCI towards the deleterious allele only in the one with the disease phenotype. Segregation analysis showed skewing towards the deleterious allele. *KDM5C* is known to escape XCI and thus the role of skewing in the phenotype is not clear.(38)

Among the various causes of female susceptibility to X-linked conditions, XCI certainly plays a key role at the penetrance level. Although XCI mechanism has been known for a very long time now, evaluation of its influence on the phenotype remains challenging.

In some cases, the presence of skewed XCI in more easily explainable by the selection of cells that inactivate the mutated allele, expressing only the wild type allele and gaining a selective advantage at the first stages of development.(10) Examples are given by mothers heterozygous for *OTUD5*, *ATRX*, *ZMYM3*, *PDZD4* variants, who are protected against the deleterious effect of an X-linked pathogenetic variant by skewed XCI. In females affected by X-linked conditions, XCI can modulate the expression of phenotype;(39) likely, there are several mechanisms that may underlie disease and skewing that currently escape our understanding and are not always easily identifiable.

In 20 XCI skewed cases, we did not identify any potentially causative variant. We hypothesize that the phenotype could be explained by variants in coding regions not covered by exome or noncoding variation such as deep intronic variant affecting splicing or regulatory regions.

Overall, our data suggest that XCI testing may be a simple and inexpensive analysis to help focusing the reanalysis of exomic data on the X chromosome.

Declarations

DATA AVAILABILITY

All variants have been deposited into ClinVar: SCV002583290, SCV002583291, SCV002583292, SCV002583293, SCV002583294, SCV002583295, SCV002583296, SCV002583297, SCV002583298, SCV002583299, SCV002583300, SCV002583301.

ACKNOWLEDGMENTS

We are grateful to patients and their family members for their participation in this study. We acknowledge Samplix for the technical support.

AUTHOR CONTRIBUTIONS

Conceptualization: C.G., S.T, A.B.; Data curation: C.G., S.T, A.B., E.S., F.P., D.C., A.M., T.F., G.M., B.P., G.B.F; Formal analysis: C.G., S.T, L.P., S.C., V.P., S.C., A.R., A.B., P.D., A.B., T.P., M.T.; Investigation: C.G., S.T, L.P., S.C., V.P., S.C., L.B., A.F., P.S., S.D.R., J.B.; Methodology: C.G., P.S., L.B., A.F., M.R., M.D., Visualization: Writing-original draft: C.G., S.T, A.B; Writing-review & editing: C.G., S.T, A.B, E.S., F.P., D.C., A.M., T.F., G.M., B.P., G.B.F, M.R., M.D.

FUNDING SUPPORT

This research received funding from the Italian Ministry for Education, University and Research (Ministero dell'Istruzione, dell'Università e della Ricerca - MIUR) PRIN2020 code 20203P8C3X to AB, "Associazione E.E. Rulfo per la genetica Medica" to AB, Fondazione Cassa di Risparmio di Torino to AB, and Fondazione Bambino Gesù (Vite Coraggiose) to M.T. Sample collection was supported by the NIMH (U01MH111661 to JDB).

ETHICS DECLARATION

All individuals and families from the different institutions agreed to participate in this study and signed appropriate consent forms. The Ethics Committee of Città della Salute e della Scienza University Hospital (n. 0060884) and University of Skopje (n. 03-6116/7) approved this study.

DECLARATION OF INTERESTS

The authors declare no competing interests.

References

1. Miles JH. Autism spectrum disorders--a genetics review. Genet Med. 2011;13(4):278-94.
2. Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res. 2011;1380:42-77.
3. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell. 2020;180(3):568-84.e23.
4. Hartley T, Lemire G, Kernohan KD, Howley HE, Adams DR, Boycott KM. New Diagnostic Approaches for Undiagnosed Rare Genetic Diseases. Annu Rev Genomics Hum Genet. 2020;21:351-72.
5. Boycott KM, Hartley T, Biesecker LG, Gibbs RA, Innes AM, Riess O, et al. A Diagnosis for All Rare Genetic Diseases: The Horizon and the Next Frontiers. Cell. 2019;177(1):32-7.

6. Basel-Salmon L, Orenstein N, Markus-Bustani K, Ruhrman-Shahar N, Kilim Y, Magal N, et al. Improved diagnostics by exome sequencing following raw data reevaluation by clinical geneticists involved in the medical care of the individuals tested. Genet Med. 2019;21(6):1443-51.
7. Gjaltema RAF, Schwämmle T, Kautz P, Robson M, Schöpflin R, Ravid Lustig L, et al. Distal and proximal cis-regulatory elements sense X chromosome dosage and developmental state at the Xist locus. Mol Cell. 2022;82(1):190-208.e17.
8. Harper PS. Mary Lyon and the hypothesis of random X chromosome inactivation. Hum Genet. 2011;130(2):169-74.
9. Fieremans N, Van Esch H, Holvoet M, Van Goethem G, Devriendt K, Rosello M, et al. Identification of Intellectual Disability Genes in Female Patients with a Skewed X-Inactivation Pattern. Hum Mutat. 2016;37(8):804-11.
10. Giorgio E, Brussino A, Biamino E, Belligni EF, Bruselles A, Ciolfi A, et al. Exome sequencing in children of women with skewed X-inactivation identifies atypical cases and complex phenotypes. Eur J Paediatr Neurol. 2017;21(3):475-84.
11. 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(1):168-73.
12. Li D, Strong A, Shen KM, Cassiman D, Van Dyck M, Linhares ND, et al. De novo loss-of-function variants in X-linked MED12 are associated with Hardikar syndrome in females. Genet Med. 2021;23(4):637-44.
13. Amos-Landgraf JM, Cottle A, Plenge RM, Friez M, Schwartz CE, Longshore J, et al. X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. Am J Hum Genet. 2006;79(3):493-9.
14. Blondal T, Gamba C, Møller Jagd L, Su L, Demirov D, Guo S, et al. Verification of CRISPR editing and finding transgenic inserts by Xdrop indirect sequence capture followed by short- and long-read sequencing. Methods. 2021;191:68-77.
15. 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(6):1229-39.
16. Di Gregorio E, Bianchi FT, Schiavi A, Chiotto AM, Rolando M, Verdun di Cantogno L, et al. A de novo X;8 translocation creates a PTK2-THOC2 gene fusion with THOC2 expression knockdown in a patient with psychomotor retardation and congenital cerebellar hypoplasia. J Med Genet. 2013;50(8):543-51.
17. Migeon BR. X-linked diseases: susceptible females. Genet Med. 2020;22(7):1156-74.
18. Plenge RM, Hendrich BD, Schwartz C, Arena JF, Naumova A, Sapienza C, et al. A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. Nat Genet. 1997;17(3):353-6.
19. O'Rawe JA, Wu Y, Dörfel MJ, Rope AF, Au PY, Parboosingh JS, et al. TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations. Am J Hum Genet. 2015;97(6):922-32.
20. Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. Hum Mutat. 2019;40(8):1030-8.
21. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-5.
22. Quinodoz M, Peter VG, Cisarova K, Royer-Bertrand B, Stenson PD, Cooper DN, et al. Analysis of missense variants in the human genome reveals widespread gene-specific clustering and improves prediction of pathogenicity. Am J Hum Genet. 2022;109(3):457-70.
23. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? J Mol Biol. 2019;431(11):2197-212.
24. Zweier C, Kraus C, Brueton L, Cole T, Degenhardt F, Engels H, et al. A new face of Borjeson-Forssman-Lehmann syndrome? De novo mutations in PHF6 in seven females with a distinct phenotype. J Med Genet. 2013;50(12):838-47.
25. Gerber CB, Fliedner A, Bartsch O, Berland S, Dewenter M, Haug M, et al. Further characterization of Borjeson-Forssman-Lehmann syndrome in females due to de novo variants in PHF6. Clin Genet. 2022;102(3):182-90.
26. Brookes E, Laurent B, Õunap K, Carroll R, Moeschler JB, Field M, et al. Mutations in the intellectual disability gene KDM5C reduce protein stability and demethylase activity. Hum Mol Genet. 2015;24(10):2861-72.
27. Carmignac V, Nambot S, Lehalle D, Callier P, Moortgat S, Benoit V, et al. Further delineation of the female phenotype with KDM5C disease causing variants: 19 new individuals and review of the literature. Clin Genet. 2020;98(1):43-55.
28. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
29. Grosso V, Marcolungo L, Maestri S, Alfano M, Lavezzari D, Iadarola B, et al. Characterization of. Front Genet. 2021;12:743230.
30. Madsen EB, Höijer I, Kvist T, Ameur A, Mikkelsen MJ. Xdrop: Targeted sequencing of long DNA molecules from low input samples using droplet sorting. Hum Mutat. 2020;41(9):1671-9.
31. Beck DB, Basar MA, Asmar AJ, Thompson JJ, Oda H, Uehara DT, et al. Linkage-specific deubiquitylation by OTUD5 defines an embryonic pathway intolerant to genomic variation. Sci Adv. 2021;7(4).
32. Philips AK, Sirén A, Avela K, Somer M, Peippo M, Ahvenainen M, et al. X-exome sequencing in Finnish families with intellectual disability--four novel mutations and two novel syndromic phenotypes. Orphanet J Rare Dis. 2014;9:49.
33. Hiatt SM. Deleterious, protein-altering variants in the X-linked transcriptional coregulator ZMYM3 in 22 individuals with a neurodevelopmental delay phenotype. https://www.medrxiv.org/content/10.1101/2022.09.29.22279724v12022.
34. Martin HC, Gardner EJ, Samocha KE, Kaplanis J, Akawi N, Sifrim A, et al. The contribution of X-linked coding variation to severe developmental disorders. Nat Commun. 2021;12(1):627.

35. Leitão E. Systematic analysis and prediction of genes associated with disorders on chromosome X. https://www.medrxiv.org/content/10.1101/2022.02.16.22270779v1.
36. Cheng H, Capponi S, Wakeling E, Marchi E, Li Q, Zhao M, et al. Missense variants in TAF1 and developmental phenotypes: challenges of determining pathogenicity. Hum Mutat. 2019.
37. Wieczorek D, Bögershausen N, Beleggia F, Steiner-Haldenstätt S, Pohl E, Li Y, et al. A comprehensive molecular study on Coffin-Siris and Nicolaides-Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. Hum Mol Genet. 2013;22(25):5121-35.
38. Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, et al. Landscape of X chromosome inactivation across human tissues. Nature. 2017;550(7675):244-8.
39. Franco B, Ballabio A. X-inactivation and human disease: X-linked dominant male-lethal disorders. Curr Opin Genet Dev. 2006;16(3):254-9.

Table 1

Table 1 is available in the Supplementary Files section.

Figures

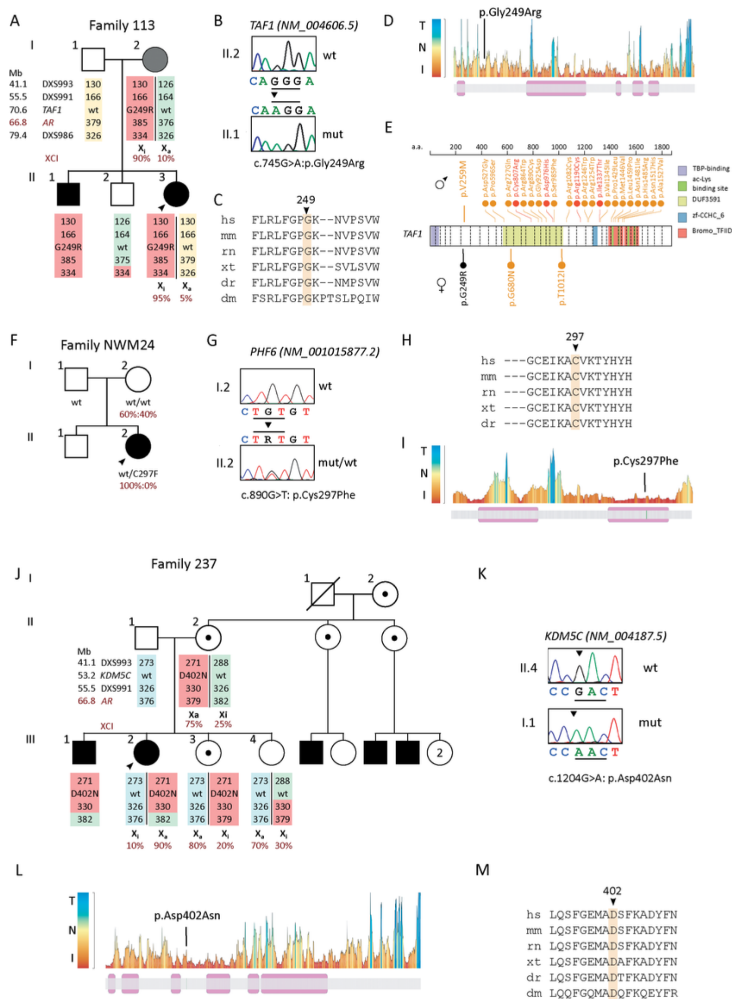

Figure 1

Pedigree and variants analysis in the three families with X-skewed female cases.

A, F, J. Genealogic trees of families 113, NWM24, and 237. We used X-chromosome polymorphic microsatellites to reconstruct the haplotypes and to phase the pathogenetic variant on the inactive/active X chromosome (percentage indicated below the symbol of tested females). Haplotypes are colored to illustrate the segregation and the presence of recombinants. For each marker, and the gene involved, the physical position in Mb is reported (GRCh37/hg19 reference genome). **B, G, K.** Sanger sequencing used to confirm the variants in *TAF1* (NM_004606.5), *PHF6* (NM_01015877.2) and *KDM5C* (NM_004187.5). Example electropherograms are shown: wild type (wt); mutant hemizygous (mut); mutant heterozygous (mut/wt). **C, H, M.** Multiple sequence alignment of the protein amino-acid sequences in different species obtained using Marvel software for the relevant changed aminoacids (highlighted in yellow; http://marvel.org/)

(hs: *Homo sapiens*; mm: *Mus musculus*; rn: *Rattus norvegicus*; xt: *Xenopus tropicalis*; dr: *Danio rerio*; dm: *Drosophila melanogaster*). **D, I, L**. Tolerance Landscape obtained using MetaDome Web Server visualizes regional tolerance to normal genetic variation (https://stuart.radboudumc.nl/metadome/). The position of the missense change is indicated for each gene. The Tolerance Landscape Y-axis is reported as a color scale from blue (position tolerant to variation, T), to yellow (position neutral to variation, N), to red (position intolerant to variation, I). Below the X- axis, a schematic representation of the protein known domains (pink). **E**. Localization of the pathogenic (red) and likely pathogenic (orange) variants reported in the literature for *TAF1* gene in male (upper panel) and female cases (lower panel). Our patient's variant is shown in black.

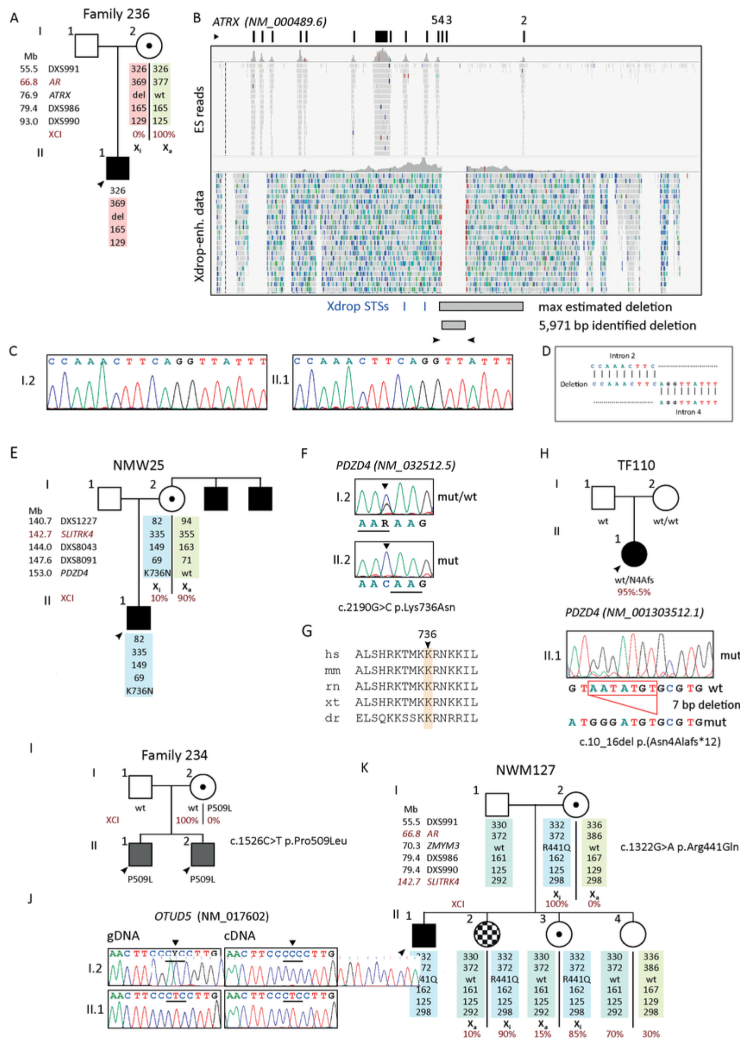

Figure 2

Pedigrees and variants analysis in the three families with XCI skewed mothers of affected males.

A, E, H, I, K. Genealogic trees of families 236, NMW25, TF110, 234 and NWM127. See legend in figure 2A, F, J. **B**. NGS Coverage of *ATRX* exons (schematized above) in ES data (upper panel) and with Xdrop enrichment (lower panel) in the II.1 proband from family 236. Xdrop enrichment primers (blue bars below) were designed 5' of the maximum estimated deletion. After the region enrichment and the subsequent Illumina Sequencing, we were able to precisely identify a 5,971 bp deletion spanning exons 3 and 4. **C**. Sanger sequencing validation of the *ATRX* deletion in II.1 and his mother (I.2) using primers flanking the deleted segment (arrows). The deletion breakpoint is shown in panel **D**. **F, H.** Sanger sequencing validation of the identified variants. **J.** In family 234, we sequenced the genomic region (gDNA) and the corresponding transcript (cDNA) in one of the probands (II.1) and their mother (I.2). The wild type allele only was detected in both cases on cDNA, showing that the pathogenic variant was not expressed and thus located on the inactive X-chromosome. **G.** Multiple alignment of the protein amino-acid sequences in different species as described in legend at figure 2 **C, H, M**.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementalall.pdf
- VariantValidator.pdf
- Table1.xlsx