Article

Understanding rare variant contributions to autism: lessons from dystrophin-deficient model

Check for updates

Claudia Ismania Samogy Costa, Luciana Madanelo, Jaqueline Yu Ting Wang,
Gabriele da Silva Campos , Ana Cristina De Sanctis Girardi , Marília Scliar, Frederico Monfardini ,
Rita de Cássia Mingroni Pavanello, Vivian Romanholi Cória, Maria Dulcetti Vibranovski,
Ana Cristina Krepischi, Naila Cristina Vilaça Lourenço, Mayana Zatz, Guilherme Lopes Yamamoto,
Elaine Cristina Zachi & Maria Rita Passos-Bueno ✉

Duchenne and Becker Muscular Dystrophy are dystrophinopathies with a prevalence of 1:5000–6000 males, caused by pathogenic variants in *DMD*. These conditions are often accompanied by neurodevelopmental disorders (NDDs) like autism (ASD; ~20%) and intellectual disability (ID; ~30%). However, their low penetrance in dystrophinopathies suggests additional contributing factors. In our study, 83 individuals with dystrophinopathies were clinically evaluated and categorized based on ASD (36 individuals), ID risk (12 individuals), or controls (35 individuals). Exome sequencing analysis revealed an enrichment of risk de novo variants (DNVs) in ASD-DMD individuals (adjusted p value = 0.0356), with the number of DNVs correlating with paternal age (p value = 0.0133). Additionally, DMD-ASD individuals showed a higher average of rare risk variants (RRVs) compared to DMD-Controls (adjusted p value = 0.0285). Gene ontology analysis revealed an enrichment of extracellular matrix-related genes, especially collagens, and Ehlers-Danlos syndrome genes in ASD-DMD and DMD-ID groups. These findings support an oligogenic model for ASD in dystrophinopathies, highlighting the importance of investigating homogenized samples to elucidate ASD's genetic architecture.

The dystrophinopathies Duchenne (DMD) and Becker Muscular Dystrophy (BMD) are allelic genetic disorders characterized by progressive muscle weakness and degeneration, affecting about 1:5000–6000 males[1]. They are caused by pathogenic variants in *DMD*, which encode dystrophin, a protein that plays a crucial role in maintaining fiber stability in skeletal muscles[1]. Neurodevelopmental changes in dystrophinopathies are well-established, encompassing delays in developmental milestones and the co-occurrence of neurodevelopmental disorders (NDDs), including intellectual disability (ID; ~30%) and autism spectrum disorders (ASD; ~20%)[2,3]. The low penetrance of these conditions in dystrophinopathies suggests the presence of additional factors to their manifestation. The role of *DMD* shorter isoforms Dp140 and Dp71 in increasing the risk of NDD is well recognized. However, it is important to note that their absence alone cannot account for all cases of NDDs observed in individuals with dystrophinopathies[4,5].

Genomic studies in individuals with dystrophinopathies have primarily focused on identifying modifier genes influencing the muscle phenotype[6,7]. To date, despite a few case reports in the literature of individuals with dystrophinopathy carrying additional genetic variants[8–10], there is no systematic genomic analysis of additional genetic contributions to the manifestations of NDDs in this group of individuals. This gap exists despite the recognized relevance of the genetic background to the manifestation of complex neurodevelopmental phenotypes[11,12].

ASD encompasses a heterogeneous group of neurodevelopmental alterations characterized by persistent deficiencies in social interaction and communication, as well as repetitive behaviors and restricted interests (DSM-V)[13]. ASD occurs in ~1% of the global population, with a prevalence four times higher in boys than in girls. Despite the known contribution of environmental factors to its etiology, ASD presents one of the highest heritability among NDDs (38–90%[14,15]). It is remarkable that, although

Departamento de Genética e Biologia Evolutiva, Centro de Estudos do Genoma Humano e Células-tronco, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brasil. ✉e-mail: passos@ib.usp.br

1

hundreds of genes have been associated with ASD, only up to 30% of ASD cases carry a known pathogenic, highly penetrant variant. The majority of these cases represent monogenic forms of ASD. The remaining cases are believed to be associated with more complex patterns of inheritance, such as the oligogenic and polygenic/multifactorial models. In these models, both inherited or de novo rare and common variants act as risk variants, and can contribute to the phenotypic manifestation.

In general, rare variants confer higher risk compared to common variants because they can lead to more significant functional damage to the protein. Despite the advances in understanding the contribution of both common and rare variants to ASD[16,17], several questions remain largely unanswered, such as the minimum required number of variants and which combinations of such variants are sufficient to cause the phenotype at the individual level. In this sense, the analysis of the genetic background of ASD individuals carrying a known risk variant may be helpful in understanding these more complex patterns of inheritance.

Through the analysis of the genetic background of 83 individuals with dystrophinopathies, this study main aims were: to characterize ASD in males with dystrophinopathies; to determine the relevance of additional de novo (DNV) and inherited rare risk variants (RRVs) for their neurodevelopmental phenotype; and to clinically and functionally characterize the identified risk variants.

Results

Participants, clinical characterization, and CNV analysis

In total, 83 individuals (36 DMD-ASD, 12 DMD-ID, 35 DMD-Control), from 77 families, were kept for further analysis. No significant difference in motor development was observed among DMD-subgroups (Kruskal–Wallis test, $p = 0.2320$). Speech impairment was common (62.5%) and significantly higher in the DMD-ASD group, compared to DMD-Control (Fisher's exact test, adjusted p value < 0.0002). Additional details on *DMD* variants and clinical data are provided in Table 1 and Supplementary Table 1.

Four individuals (P15, P21, P61, P91) were diagnosed with BMD, while the remaining individuals exhibited typical DMD progression. DMD-ASD (61.1%) and DMD-ID (75.0%) had a large proportion of variants affecting shorter isoforms (Dp140 and Dp71), as compared to those affecting only the largest isoform, though not statistically significant (Fisher's exact test, adjusted p value > 0.999 and 0.5382, respectively; Supplementary Fig. 2). No differences occurred in DMD-Control (51.4%). PCA analysis showed no ancestry differences between groups (Supplementary Fig. 3).

Finally, one pathogenic CNV (copy number variation) was identified in the individual P41: a 1.5 Mb microdeletion at 1q21.1q21.2 (minimum deletion region chr1:146877659–148483252; Supplementary Fig. 4), overlapping the 1q21.1 recurrent region (BP3-BP4, distal) (includes GJA5 from ClinGen. This 1q21.1 microdeletion confers an increased risk of global developmental delay, ID, and other neurodevelopmental conditions such as ASD, among other clinical signs, with variable expressivity and reduced penetrance (OMIM #612474 CHROMOSOME 1q21.1 DELETION SYNDROME, 1.35-MB).

De novo variant analysis in DMD-ASD and DMD-control groups

In total, 39 DNVs were identified (Supplementary Table 3, Supplementary Fig. 6), leading to a DNV rate of 1.1 variant/exome/individual. DMD-ASD group presented a large average number of DNVs than the DMD-Control group (26/18 [1.44] vs. 13/17 [0.76] variants/individuals; one-tailed Fisher's exact test, adjusted p value = 0.2732, alpha = 0.025), although not statistically significant. However, when considering risk DNVs, DMD-ASD group was significantly enriched (7/18 vs. 0/17 variants/individuals; one-tailed Fisher's exact test, adjusted p value = 0.0356, alpha = 0.025; Fig. 1a).

Paternal age was positively associated with the number of DNVs (mean (SD) = 32.51 (±8.02) years; Poisson regression analysis, beta = 0.05708, p value = 0.0133, Supplementary Fig. 5), while maternal age showed no significant association (mean (SD) = 29.66 (±5.42) years; beta = 0.04463, p value = 0.133). Moreover, DMD-ASD fathers were, on average, older than

DMD-Control fathers (DMD-ASD: 34.83 (±9.43) years, DMD-Control: 29.94 (±5.31) years; one-tailed Welch's t test, p value = 0.0347; Fig. 1b). Parental origin of three out of 39 variants (7.7%) was determined, being all of them paternal, Supplementary Table 3.

Enrichment of RRVs analysis

A total of 221 RRVs (in 209 genes), and 379 synonymous variants (in 338 genes), were selected. DMD-ASD individuals showed a higher average of RRVs compared to DMD-Controls (3.25 [117/36] vs. 2.23 [78/35] variants/individual; Mann–Whitney test, adjusted p value = 0.0285, alpha = 0.025). No difference was observed for synonymous variants (DMD-ASD: 4.69 [169/36] vs. 4.69 [164/35] variants/individual; Mann–Whitney test, adjusted p value > 0.999, alpha = 0.025), Fig. 1c.

Comparing DMD-ID with DMD-Control, no significant differences were found for RRVs or synonymous variants (RRVs: 2.17 [26/12] variants/individual, Mann–Whitney test, adjusted p value > 0.999, alpha = 0.025; synonymous: 3.83 [46/12] variants/individual, Mann–Whitney test, adjusted p value = 0.7348, alpha = 0.025).

Considering the parental WES data for 38 individuals from our sample (18 DMD-ASD, 17 DMD-Control, and 3 DMD-ID individuals), we were able to infer the origin of 60 out of 117 RRVs in the DMD-ASD group, 8/26 in the DMD-ID group, and 45/78 in the DMD-Control group. DMD-ASD presented 27 (45.0%) maternally inherited and 29 (48.33%) paternally inherited variants (Fishers' exact test, p value = 0.8549); DMD-ID presented four (50.0%) maternally inherited and three (37.5%) paternally inherited variants (Fishers' exact test, p value > 0.9999); and DMD-Control presented 21 (46.7%) maternally inherited and 26 (53.3%) paternally inherited variants (Fishers' exact test, p value = 0.6735). In addition, five variants were classified as de novo, and all de novo risk variants were found in the DMD-ASD (four variants) and DMD-ID (one variant) groups.

Frequency of dual diagnosis in the DMD-ASD and DMD-ID groups

Two pathogenic and one likely pathogenic variants associated with autosomal dominant neurodevelopmental conditions (*EBF3*:NM_001005463.1:c.616 C > T [p.Arg206Ter], #MIM617330; *SCAF4*: NM_020706.2:c.1276 C > T [p.Arg426Ter], #MIM620511; and *UBE3A*:NM_000462.3:c.29delA [p.Lys10fs], #MIM105830) were identified in DMD-ASD individuals (P89, P86 and P85, respectively). The pathogenic variants in *EBF3* and *SCAF4* were DNVs, while the likely pathogenic variant in *UBE3A* was an RRV, maternally inherited. The rate of dual diagnosis in the DMD-ASD group was 8.33% (3/36 individuals) and 0.0% (0/12 individuals) in the DMD-ID group. Of note, a likely pathogenic DNV was identified in individual P69 (*HEPACAM*: NM_152722.5:c.382 G > A [p.Asp128Asn], DMD-ID group), which is associated with ID, although with low penetrance (around 40%, #MIM613926).

Gene enrichment analysis

Joint analysis of RRVs from DMD-ASD and DMD-ID groups ($n = 138$ genes) revealed enrichment for molecular functions related to extracellular matrix (ECM) constituents and ECM structural constituent conferring tensile strength (Supplementary Table 4). It was also enriched for the following cellular components: collagen trimmer fibrillar, collagen trimmer, and banded collagen fibril. Similarly, pathway analysis showed significant pathway enrichment associated with integrin and non-integrin cell surface interactions, collagen trimerization, collagens, and syndecan interactions. Finally, genes were enriched for Ehlers-Danlos Syndrome (EDS), Supplementary Table 4. No enrichment terms or pathways were observed for the DMD-Control group ($n = 78$ genes).

Discussion

DMD pathogenic variants are established risk factors for NDDs[18]. In this study, we explored the impact of RRVs on the neurodevelopmental phenotype of 83 individuals with dystrophinopathies, particularly ASD. The relevance of the oligogenic model for ASD has been recognized[12]. However,

Table 1 | Clinical and molecular data from the 83 probands included in this study

ID	Family_ID	Age – evaluation (years old)	CARS classification	RAVEN	Mother-carrier status	Pregnancy and childbirth	Speech impairment	Motor development (first steps)	Seizures	Learning difficulties	Variant type	Exon(s)
P1	1	17	No ASD	ID	No	Cesarean section, uneventfully	NA	30 months	NA	Yes	Del	45
P2	2	13	No ASD	No ID	Yes	NA	No	10 months	NA	Yes	Del	10-17
P3	3	15	No ASD	ID	Yes	Cesarean section, uneventfully	Yes	13 months	NA	No	Dup	10-17
P4	4	13	No ASD	ID	Yes	NA	NA	15 months	NA	Yes	Del	51-55
P6	3	9	No ASD	No ID	Yes	NA	NA	14 months	NA	No	Dup	10-17
P7	5	15	No ASD	ID	Yes	Cesarean section, hypoxia	NA	24 months	Yes, 2 episodes	Yes	Stopgain	NM_004006.3:c.9568 C > T (exon 66)
P8	6	16	No ASD	ID	Yes	Cesarean section, uneventfully	No	14 months	NA	No	Del	5-7
P9	7	15	Previous diagnosis	NA	No	Cesarean section, uneventfully	No	18 months	NA	No	Del	3-43
P10	8	14	No ASD	No ID	Yes	Uneventfully	No	19 months	NA	No	Del	46-49
P12	9	11	No ASD	No ID	Yes	Cesarean section, uneventfully	No	18 months	NA	NA	Stopgain	NM_004006.3:c.5159del (exon 37)
P13	11	12	No ASD	No ID	Yes	Cesarean section, uneventfully	Yes	14 months	No	Yes	Del	51
P14	12	14	ASD	No ID*	Yes	Cesarean section, suspected meningitis (not confirmed)	No	16 months	NA	No	Splicing	NM_004006.3:c.2803+1 G > C (exon 21)
P15	13	12	No ASD	No ID	No	Uneventfully	No	17 months	NA	Yes	Del	45-46
P16	14	12	ASD	NA	No	Cesarean section, uneventfully	Yes	19 months	NA	Yes	Del	52
P17	15	15	ASD	NA	Yes	Cesarean section, uneventfully	Yes	18 months	NA	Yes	Frameshift	NM_004006.3:c.7567del (exon 52)
P19	17	14	No ASD	No ID	Yes	Uneventfully	No	12 months	NA	No	Stopgain	NM_004006.3:c.G5974T (exon 42)
P20	18	11	No ASD	No ID	Yes	Cesarean section, uneventfully	No	24 months	NA	NA	Del	3-45
P21	19	15	No ASD	No ID	NA	Uneventfully	Yes	15 months	Na	Yes	ntronic (VUS)	50 (IVS50)
P23	21	13	No ASD	No ID	Yes	Uneventfully	Yes	24 months	Suspected	Yes	Stopgain	NM_004006.3:c.4271 T > A (exon 31)
P24	16	7	ASD	ID	Yes	Cesarean section, uneventfully	Yes	NA	NA	Yes	Stopgain	NM_004006.3:c.2436 G >A (exon 20)
P25	22	11	No ASD	No ID	Yes***	Uneventfully	No	13 months	NA	No	Stopgain	NM_004006.3:c.2605 C > T (exon 20)
P26	23	8	No ASD	No ID	Yes	Uneventfully	Yes	13 months	NA	Yes	Del	45-52
P27	24	11	No ASD	No ID	NA	Cesarean section (mother with Hepatitis B)	No	24 months	NA	Yes	Del	46-47
P28	25	7	ASD	ID	Yes	IVF, premature, cesarean section, uneventfully	Yes	18 months	NA	Yes	Indel	NM_004006.3:c.88_93+17del (exon 2)

Table 1 (continued) | Clinical and molecular data from the 83 probands included in this study

| ID	Family_ID	Age – evaluation (years old)	CARS classification	RAVEN	Mother - carrier status	Pregnancy and childbirth	Speech impairment	Motor development (first steps)	Seizures	Learning difficulties	Variant type	Exon(s)
P30	26	8	ASD	No ID	Yes	Uneventfully	NA	12 months	NA	NA	Dup	63-67
P31	27	12	ASD	NA	No	Cesarean section, uneventfully	Yes	14 months	NA	Yes	Del	48-52
P32	28	13	No ASD	No ID	Yes	Uneventfully	NA	12 months	NA	Yes	Del	3-7
P33	20	6	No ASD	No ID	Yes	NA	Yes	15 months	NA	No	Stopgain	NM_004006.3:c.572 C > A (exon 7)
P34	29	13	ASD	ID	No	Cesarean section, uneventfully	Yes	24 months	NA	Yes	Stopgain	NM_004006.3:c.10651 C > T (exon 75)
P35	30	17	No ASD	No ID	Yes	Uneventfully	No	NA	NA	No	Del	46-50
P36	31	12	No ASD	No ID	No	Cesarean section, uneventfully	No	12 months	NA	No	Del	12-29
P37	32	11	No ASD	No ID	Yes	Cesarean section, hypoxia	Yes	16 months	Suspected	Yes	Dup	2
P38	33	15	No ASD	NA	No	Cesarean section, jaundice	NA	11 months	NA	No	Del	3-7
P39	34	7	No ASD	ID	No	Cesarean section, jaundice	Yes	24 months	NA	No	Del	45-52
P40	35	14	No ASD	ID	Yes	Cesarean section, uneventfully	Yes	18 months	No	Yes	Del	48-52
P41	2	6	ASD	No ID	Yes	Cesarean section, uneventfully	No	19 months	NA	No	Del	10-17
P42	36	3	ASD	NA	Yes	NA	Yes	12 months	Yes, 1 episode	NA	Stopgain	NM_004006.3:c.2665 C > T (exon 21)
P43	36	9	No ASD	NA	Yes	NA	No	12 months	No	No	Stopgain	NM_004006.3:c.2665 C > T (exon 21)
P44	37	10	No ASD	No ID	NA	Vaginal delivery, uneventfully	No	12 months	NA	No	Del	49-50
P45	38	10	No ASD	No ID	No	Cesarean section, uneventfully	No	15 months	NA	No	Del	8-9
P46	39	6	ASD	No ID	Yes	Vaginal delivery, uneventfully	Yes	12 months	NA	Yes	Del	45-52
P47	40	17	ASD	ID	Yes	Cesarean section, uneventfully	Yes	14 months	No	Yes	Del	45-49
P48	41	11	No ASD	NA	Yes	Cesarean section, uneventfully	Yes	12 months	NA	No	Del	10-43
P49	23	5	ASD	ID	Yes	Vaginal delivery, uneventfully	Yes	15 months	NA	Yes	Del	45-52
P50	42	15	Previous diagnosis	NA	NA	Vaginal delivery, uneventfully	Yes	14 months	No*	Yes	Del	51-55
P51	43	14	No ASD	ID	Yes	Vaginal delivery, corticosteroid for pulmonar maturarion	Yes	21 months	NA	Yes	Splicing	NM_004006.3:c.10086+3_10086+4insT (exon 69)
P52	44	14	No ASD	ID	No	Cesarean section, uneventfully	No	13 months	NA	Yes	Del	44

Table 1 (continued) | Clinical and molecular data from the 83 probands included in this study

ID	Family_ID	Age – evaluation (years old)	CARS classification	RAVEN	Mother - carrier status	Pregnancy and childbirth	Speech impairment	Motor development (first steps)	Seizures	Learning difficulties	Variant type	Exon(s)
P54	46	10	No ASD	No ID	No	Vaginal delivery, uneventfully	NA	NA	NA	Yes	Del	1–2 + promoter
P55	47	16	No ASD	ID	NA	Cesarean section, uneventfully	NA	15 months	NA	Yes	Del	52-54
P56	48	9	No ASD	No ID	No	Cesarean section, uneventfully	No	24 months	NA	No	Del	44
P57	49	11	No ASD	No ID	Yes	Cesarean section, uneventfully	Yes	12 months	NA	Yes	Dup	20-25
P58	50	10	No ASD	ID**	Yes	NA	NA	12 months	NA	Yes	Del	45-48
P59	51	5	ASD	No ID	No	Vaginal delivery, hypoglycemia	Yes	17 months	No	Yes	Del	21
P60	52	7	No ASD	No ID	NA	Cesarean section, uneventfully	NA	21 months	NA	NA	Del	49-52
P61	53	6	No ASD	No ID	NA	Cesarean section, uneventfully	Yes	NA	NA	Yes	Del	45-57
P62	54	12	ASD	NA	Yes	Cesarean section, uneventfully	Yes	17 months	NA	Yes	Del	53
P64	56	7	No ASD	No ID	Yes	Vaginal delivery, hypoxia	No	16 months	NA	No	Intronic	10 (IVS10)
P65	57	13	No ASD	No ID	No	Vaginal delivery, jaundice	Yes	15 months	NA	Yes	Del	45
P66	43	10	No ASD	No ID	Yes	Vaginal delivery, premature, uneventfully	No	24 months	NA	Yes	Splicing	NM_004006.3:c.10086+3_10086+4insT (exon 69)
P67	58	10	ASD	NA	Yes	Cesarean section, uneventfully	Yes	24 months	NA	Yes	Del	18-44
P68	59	6	ASD	NA	No	Vaginal delivery, uneventfully	Yes	18 months	NA	Yes	Del	65
P69	60	6	No ASD	ID	Yes	Cesarean section, uneventfully	Yes	20 months	NA	Yes	Dup	53-54
P70	61	7	Previous diagnosis	NA	Yes	Cesarean section, uneventfully	Yes	14 months	NA	NA	Del	49-50
P71	62	5	ASD	NA	No	Cesarean section, uneventfully	Yes	15 months	NA	Yes	Del	46-51
P72	63	4	ASD	NA	No	Threatened miscarriage	Yes	16 months	NA	Yes	Del	51-59
P74	65	10	ASD	NA	Yes	NA	Yes	14 months	NA	Yes	Del	49-50
P75	66	13	ASD	NA	No	Cesarean section, corticosteroid for pulmonar maturarion	Yes	16 months	No	Yes	Del	48-52
P76	67	10	No ASD	NA	No	Cesarean section, premature, weak heartbeat	No	NA	No	No	Del	45

Table 1 (continued) | Clinical and molecular data from the 83 probands included in this study

ID	Family_ID	Age – evaluation (years old)	CARS classification	RAVEN	Mother - carrier status	Pregnancy and childbirth	Speech impairment	Motor development (first steps)	Seizures	Learning difficulties	Variant type	Exon(s)
P77	68	3	No ASD	NA	No	Cesarean section, uneventfully	No	16 months	No	No	Del	46–52
P78	69	7	No ASD	NA	Yes	NA	No	15 months	NA	No	Del	51–55
P79	70	8	ASD	NA	Yes	Cesarean section, uneventfully	No	12 months	NA	NA	Frameshift	NM_004006.3:c.3639del (exon 27)
P80	71	3	ASD	NA	No	Uneventfully	Yes	26 months	No	Yes	Del	56–67
P81	72	8	Previous diagnosis	NA	Yes	Cesarean section, uneventfully	Regression	17 months	No	Yes	Del	46–49
P83	73	9	Previous diagnosis	NA	Yes	NA	Regression	16 months	No	NA	Dup	52
P84	74	6	No ASD	NA	Yes	Vaginal delivery, uneventfully	No	12 months	No	No	Del	52
P85	75	3	ASD	NA	Yes	Cesarean section, uneventfully	Regression	NA	No	Yes	Stopgain	NM_004006.3:c.3742 C > T (exon 27)
P86	76	7	Previous diagnosis	NA	Yes	Cesarean section, threatened miscarriage	Yes	NA	No	Yes	Stopgain	NM_004006.3:c.10033 C > T (exon 69)
P87	77	5	ASD	NA	No	Cesarean section, uneventfully	Yes	NA	No	No	Del	45–50
P88	78	6	Previous diagnosis	NA	NA	Corticosteroid for pulmonar maturarion, jaundice	Yes	17 months	NA	No	Del	44
P89	10	14	ASD	NA	Yes	Vaginal delivery, twin pregnancy	Yes	18 months	No	Yes	Stopgain	NM_004006.3:c.1615 C > T (exon 14)
P90	79	11	ASD	No ID	Yes	Cesarean section, uneventfully	No	18 months	No	Yes	Del	1–79 + promoter
P91	80	16	ASD	NA	Yes	Cesarean section, uneventfully	Yes	12 months	No	Yes	Del	2–9
P92	81	3	Previous diagnosis	NA	No	Cesarean section, premature, pre-eclampsia	Yes	NA	NA	No	Frameshift	NM_004006.3:c.5602_5605del (exon 40)

Data were collected from probands' parents interviews and/or retrieved from clinical records. Probands age refers to the age at ASD and/or ID evaluation. Carrier status provided whenever available/ possible. *CARS* Childhood Autism Rating Scale (scores are not shown). *RAVEN* The Raven's Progressive Matrices test classification, *NA* not available, *ID* cognitive impairment, *ASD* autism spectrum disroder. *Del* deletion. *Dup* duplication, *VUS* variant of uncertain significance. *WISC-III (vocabulary subset). **WISC-IV, Quantitative Intelligence Test = 63. ***based on pedigree analysis and CK levels.
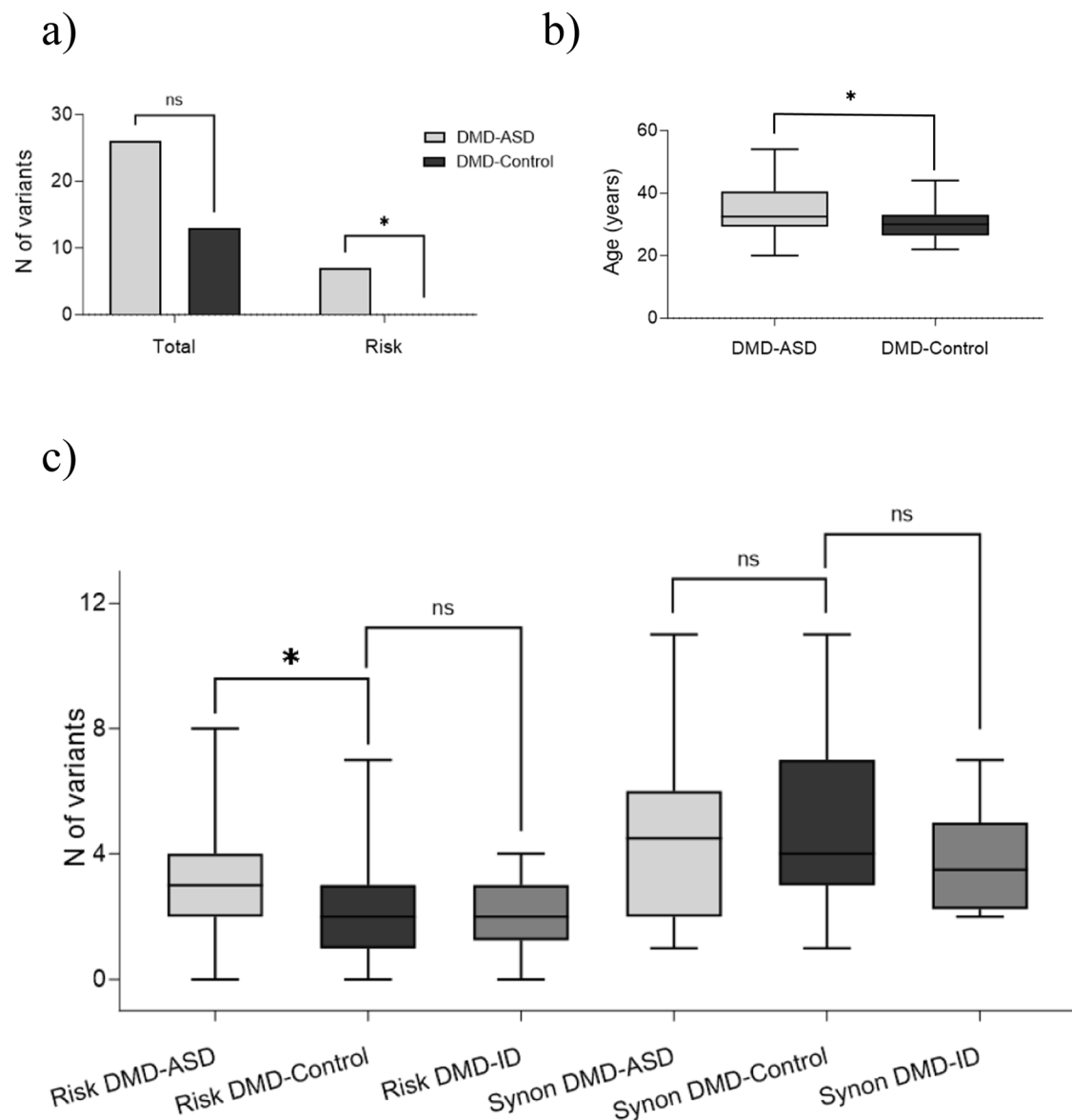
Fig. 1 | Distribution of total and risk DNVs and differences in paternal age between groups. **a** Bar plot showing the distribution of total and risk DNVs in DMD-ASD and DMD-Control groups. Fishers' exact test was used to test whether there was a difference between groups regarding total and risk DNVs. **b** Boxplot showing the distribution of paternal age between DMD-ASD and DMD-Control groups. Welch's *t* test was applied to determine whether there was a difference in the mean paternal age. **c** Boxplot showing the distribution of rare variants for each group. Mann–Whitney test was performed to verify whether there was a difference in the average number of RRVs and synonymous variants between cases (DMD-ASD and DMD-ID) and the DMD-Control group. The DMD-ASD group presented, on average, a higher number of RRVs than DMD-Controls, after Bonferroni correction. ns = non-significant; *p value < 0.05.

due to the low populational frequency of RRVs, detecting causative combinations with statistical significance often requires larger sample sizes than those available in well-established autism cohorts[19]. In this scenario, homogenizing the sample for the known primary risk variant allowed the identification of genetic patterns and genotype-phenotype associations with a significantly reduced sample size.

In this study, we have proposed the analysis of a cohort with a known primary ASD risk variant (*DMD* gene) to verify if it allows the identification of genetic patterns and genotype-phenotype associations with a significantly reduced sample size. We have taken advantage of the current knowledge of the genetic architecture in ASD for rare variants, which include de novo and inherited variants together with one of the best-known environmental risk factors for ASD, parental age[20].

All the included individuals had a confirmed molecular diagnosis of dystrophinopathy. Speech impairment was prevalent (62.5%) and significantly higher in DMD-ASD individuals than in DMD-Controls,

suggesting an association between speech development and ASD in dystrophinopathies. In contrast, motor development delay was observed across all DMD-subgroups, as commonly described in boys with dystrophinopathies regardless of their neurodevelopment[21]. No differences were observed in the proportion of individuals carrying variants affecting the *DMD* shorter isoforms Dp140 and Dp71 for any DMD-subgroup. While a higher frequency of cognitive impairment is associated with pathogenic variants affecting these isoforms[3,22], their role in other neurodevelopmental alterations is unclear. Indeed, the highest frequency of variants affecting these isoforms was observed in our DMD-ID group. The lack of statistical difference may be due to a limited statistical power due to the sample size.

The analysis of DNVs revealed that DMD-ASD exhibited an enrichment of risk DNVs compared to DMD-Controls. The findings on the higher DNV rate for risk genes and increased paternal age in the DMD-ASD group suggest that ASD in dystrophinopathies is being partly driven by paternal age, similarly to the results observed in ASD large cohorts[23].

DMD-ASD cases also showed a higher average of RRVs than DMD-Controls, supporting an oligogenic model in ASD among dystrophinopathies. No bias in the parental origin of RRVs was observed in the three analyzed groups. In addition, the rate of dual diagnosis for the DMD-ASD group was 8.33%. The three affected genes are associated with monogenic forms of ASD and/or ID, with a high (> 90%) penetrance (*EBF3*, *UBE3A*, and *SCAF4*). Notably, in all cases, the primary clinical concern reported in these individuals was their behavioral alterations and/or delayed development. While these variants are causative of NDDs, how they interact with the *DMD* variant to determine the final clinical outcome is unclear. Nevertheless, given its relevance for genetic counseling, exome testing should be considered in dystrophinopathies cases with a severe neurodevelopmental phenotype.

Moreover, one individual from the DMD-ASD group (P41) and one from the DMD-ID group (P69) presented an additional clinically relevant variant (a pathogenic 1q21.1 microdeletion and a likely pathogenic missense variant in *HEPACAM*, respectively), both associated with neurodevelopmental alterations with reduced penetrance[24,25]. The presence of these additional known RRVs further supports an oligogenic model in the manifestation of NDDs among individuals with dystrophinopathies. Additionally, although our analysis mainly focused on rare single nucleotide variants (SNVs), we cannot rule out contributions from other types of variants, including common and regulatory ones[26].

Enrichment analysis based on the RRVs identified in DMD-ASD and DMD-ID groups converged to ECM components, most particularly collagen genes. Moreover, disease analysis revealed enrichment for EDS, a group of genetic disorders that affect connective tissues caused by pathogenic variants in collagens or collagen processing proteins[27]. ECM genes have been shown to contribute to ASD[28], and individuals with EDS and hypermobility syndrome are at increased risk for NDDs, including ASD[29,30], although the molecular mechanisms are still unclear. In addition, dystrophin is essential in the DGC (dystrophin-glycoprotein complex), which links the ECM with intracellular components in different cell types[1,31]. Our findings thus suggest that insults in the DGC, through dystrophin pathogenic variants, in conjunction with insults in ECM components, elevate the risk of ASD in dystrophinopathies. However, despite their functional proximity, it remains unclear how alterations in one affect the other, and how they jointly contribute to neurodevelopmental changes.

In summary, the observation of enrichment of rare DNVs and RRVs in ASD-DMD-affected individuals suggests that ASD in DMD may be dependent on several variants in addition to dystrophin pathogenic variants, favoring an oligogenic model of inheritance. Paternal age seems also to contribute to ASD risk in this DMD cohort. Finally, this study shed light on the biological functions and pathways altered in this subgroup of ASD-affected individuals. We thus support that the study of genetically homogenized cohorts represents a strategy to explore complex genetic models.

Despite our significant results, this study presents some limitations, such as the relatively low sample size, particularly for the DMD-ID and DMD-ASD-ID groups, and the unavailability of clinical data for all individuals, limiting a more comprehensive genotype-phenotype correlation. We expect that in the near future, with the analysis of larger DMD-ASD, DMD-ID, and DMD-ASD-nonID cohorts with more in-depth phenotyping, we will be able to investigate if the genetic architecture differs among these subgroups, which represent a relevant challenge to be addressed. Furthermore, it is significant to evaluate regulatory or common variants besides rare variants in their contribution to the final outcome.

Methods
Subjects, DNA samples, and clinical characterization
We collected DNA samples from 92 individuals (from 81 families; Supplementary Fig. 1). Families were recruited at Centro de Estudos do Genoma Humano e Células-Tronco (CEGH-CEL), Universidade de São Paulo, or through the foundation Aliança Distrofia Brasil (ADB). DNA was extracted from whole blood or saliva at CEGH-CEL. The main criteria of inclusion were male individuals confirmed to carry a known pathogenic *DMD*

variant. Exclusion criteria, applied after neuropsychological characterization (described below), are detailed in the "DMD-subgroups definition" section. Clinical data was provided by parents or retrieved from medical records. This study was performed in accordance with the ethical standards of the Declaration of Helsinki. The Ethics Committee from Instituto de Biociências, Universidade de São Paulo, CEP/USP, fully approved the study (CAEE 56459522.0.0000.5464). Written informed consent was obtained from parents or legal guardians, whenever possible. As this study involved the use of stored DNA of families previously seen by our group, our Ethics Committee allowed us to include individuals whom we were not able to contact after at least five attempts. All these families had previously signed a consent term allowing the use of biological and clinical data for research purposes, and all clinical data at the individual level was de-identified.

Neuropsychological characterization was performed using CARS (Childhood Autism Rating Scale, *n* = 83) or previous ASD diagnosis (*n* = 9). Individuals with CARS score ≥30 were classified as at risk for autism. Moreover, 55 individuals were also evaluated for cognitive impairment using the Raven's Progressive Matrices test or the WISC-IV (Wechsler Intelligence Scale for Children, 3rd or 4th edition). Three groups were defined: DMD-ASD: individuals with CARS ≥ 30 or previous ASD diagnosis; DMD-ID: individuals with CARS < 30 and RAVEN/WISC indicating cognitive impairment; DMD-Control: individuals with CARS < 30 and Raven/WISC (when available) negative for cognitive impairment, and individuals with CARS < 30 without cognitive evaluation, but without a history of speech delay nor reported learning difficulties.

DMD-subgroups definition
For the ten families with more than one individual with dystrophinopathy in our sample, at most two individuals were kept in the sample and only if they were ASD/ID discordant. In cases in which more than one individual from the same family was in the same DMD group, the following criteria were adopted for individual's selection: (1) clinical evaluation–the presence of previous ASD or ID diagnosis over only CARS/cognitive evaluation (for DMD-ASD and DMD-ID groups) or best performance on cognitive evaluation (for DMD-Control group), (2) in cases for which both individuals presented the same clinical performance, the younger individuals was kept in the sample, (3) in case of dizygotic twins, WES quality was considered, keeping the individual with higher mean coverage. Detailed information on clinical data, DNA extraction, sequencing metrics, and group definition is provided in Supplementary Tables 1 and 2.

Based on cognitive and ASD evaluations, the 92 individuals were firstly distributed into three groups as following: DMD-ASD: 38 individuals (29 with CARS ≥ 30 and nine individuals with previous ASD diagnosis); DMD-ID: 12 individuals (11 based on RAVEN and one based on WISC-subtest results); DMD-Control: 42 individuals (including 28 individuals with normal cognitive evaluation). Two individuals from the DMD-ASD group (P29 and P82) were excluded since they have another affected brother in the same group. In addition, seven individuals were excluded from the DMD-Control group: four individuals (P18, P53, P63, and P73) without cognitive evaluation that presented both speech impairment and learning difficulties and three individuals with an affected relative individual in the same group (P5, P11, and P22).

Parents' WES was available for 38/83 individuals (DMD-ASD group: 18 individuals, DMD-ID group: three individuals, and DMD-Control group: 17 individuals; Supplementary Table 2).

DMD pathogenic variants
All *DMD* variants were identified through either MLPA (Multiplex ligation-dependent probe amplification, kits P034 and P035 - MRC-Holland BV, Amsterdam, the Netherlands) or NGS Next Generation Sequencing (NGS) analysis: target sequencing of the *DMD* gene, a panel for myopathy and dystrophy associated genes, containing 266 genes as target, or WES, and were confirmed through Sanger (for NGS results), or exome sequencing (WES; for MLPA results). Individuals were categorized based on variant type and affected isoforms, distinguishing between those impacting the

Dp427 isoforms (exons 1–44) and those affecting the brain isoforms Dp140 and Dp71 (exons 45–79), *DMD* accession number NM_004006.3.

Exome sequencing, variant selection, validation, and copy number variation (CNV) analysis

DNA samples were prepared using one of the following kits: SureSelect QXT Target Enrichment for Illumina Multiplexed Sequencing - V6 (Agilent Technologies, Inc., Santa Clara, CA - USA), IDT – xGen Exome Research Panel V1.0 or IDT – xGen Exome Research Panel V2.0 (Integrated DNA Technologies [IDT], Inc., Iowa, USA), following the manufacturer's recommendations. The libraries were sequenced with Illumina HiSeq 2500 sequencer or Illumina NovaSeq 6000 sequencer, in paired-end reads of ~100 bp (Supplementary Tables 1 and 2). Sequence alignments to the human genome reference (UCSC hg38) were performed with Burrows-Wheeler Aligner, data processing and variant calling was performed with Picard (version 2.18.7[32]) and Genome Analysis Toolkit package (GATK, version 4.0.9.0[33]). SNVs and small indels located in exons or within 50 bp of the splicing junction were called in individual samples using GATK v.4.0.9.0 HaplotypeCaller, and were jointly genotyped using GATK GenotypeGVCFs. The mean coverage of exome data from the 92 individuals was 126.3, with 97% of the exome being covered by at least 10 reads, and 96% by at least 20 reads, on average.

After variant quality score recalibration, called variants were annotated using ANNOVAR v.2016Feb01[34] and Ensembl Variant Effect Predictor (VEP[35]), including allele frequency in the general population (GnomAD, ABraOM and 1000 G databases[36–38]), predictive tools for deleteriousness of the alternate allele (CADD score, SpliceAI[39,40]), gene probability of being loss-of-function intolerant (pLI) or missense intolerant (Z score) scores (GnomAD database) and gene-disease association databases (OMIM, ClinVar, SFARI[41–43]). For all analyses performed, only variants that reached the minimum quality criteria defined as variants with at least 10 reads (and alternate allele number of reads ≥5) and allele balance between 0.3 and 0.7 (heterozygous) or ≥0.9 (homozygous), were selected. Next, variants were visually inspected using the Integrative Genomic Viewer (IGV[44]) software and, for individuals for which parents were sequenced, variant segregation was also inspected using IGV. All variants were verified using the variant description validation software VariantValidator[45].

For each analysis performed, a set of variants with poor quality on IGV's visual inspection were selected for Sanger sequencing, which was performed using BigDye Terminator v3.1 Cycle Sequencing Kit, in the ABI 3730 DNA Analyzer (Thermo Fisher Scientific, Massachusetts, USA).

The aim of CNV analysis in this dataset was to identify additional pathogenic or likely pathogenic CNVs mapped to 48 chromosomal regions with recurrent CNVs[46]. eXome Hidden Markov Model (XHMM)[47] was used in conjunction with at least one additional software (CNVkit[48], panelcn.MOPS[49], or NextGene (https://softgenetics.com/products/nextgene/cnv-analysis/), following the recommended parameters, to investigate CNVs from exome sequencing data in individuals from the DMD-ASD and DMD-ID groups (detailed information on software used for each individual is available in Supplementary Table 1). After the CNV call, the variants were annotated with AnnotSV[50].

The following prioritization criteria were used to select relevant CNVs of high confidence: only CNVs identified by XHMM and one additional software were considered (considering an overlap of, at least, 10% with the minimum chromosomal region detected by XHMM). Rare (frequency ≤1% according to DGV (Database of Genomic Variants[51]) deletions and duplications with at least 200 kb in size and SQ ≥ 30, unique in our sample (or in a family, in cases where more than one individual was evaluated) and mapped to each of the 48 target chromosomal regions from CLinGen recurrent regions. CNVs automatically classified as ACMG (American College of Medical Genetics and Genomics[52]) score 3 (Variant of Uncertain Significance, VUS), 4 (Likely pathogenic) or 5 (Pathogenic) by AnnotSV were further evaluated. Additional CNV classification criteria included (a) the exclusion of CNVs highly recurrent on DGV, after manual curation; characterization according to (b) size, (c) type (deletion or duplication), (d)

gene content, (e) literature and clinical information about disease-associated genes, and (f) inheritance (whenever possible), (g) Franklin by Genoox (https://franklin.genoox.com/) automated CNV classification (based on the ACMG guidelines). For all prioritized CNVs for manual inspection, the size was considered as the minimum CNV size considering the overlap between the software calls.

Kinship and ancestry analysis

All trios, quartets, and related individuals were tested for kinship. In summary, in search for rare variants that are most likely inherited by descent, we determine the frequency of shared rare (frequency <1% in public databases) variants in all loci with high quality (properly genotyped in at least 80% of the cohort). Individuals with at least 2% of shared rare variants were considered related to some degree (Supplementary Table 1). We also performed ancestry characterization based on PCA analysis aiming to determine whether our groups potentially present differences in their genetic background due to ancestry (Supplementary Fig. 3).

De novo and RRvs analysis

High-confidence de novo variants (DNVs) were annotated using DeNovoGear[53] and PossibleDeNovo[54]. *DMD* variants were excluded from analysis due to the study's inclusion criteria. Selected DNVs, identified by at least one software, included rare (frequency ≤1%), missense, splicing, synonymous, and loss-of-function variants (LoF; frameshift insertion, frameshift deletion, stopgain, and stop loss variants). The DMD-ASD and DMD-Control groups were compared for total and risk DNVs (defined as LoF, missense (CADD ≥ 25), and splicing (spliceAI ≥ 0.8) variants in brain-expressed genes). Correlations between the number of DNVs in the offspring and parental age were also tested. Finally, to obtain parental phasing information for the DNVs, DeNovoGear phaser[53] analysis was performed.

To test whether DMD-ASD and the DMD-ID individuals present a higher average of RRVs in their genetic background, compared to DMD-Controls, we selected rare (frequency ≤1%) variants in a subset of genes relevant to neurodevelopment, here pre-defined as a merge of the following gene sets: (a) genes expressed in brain (HPA list) with pLI ≥ 0.5 or Z score ≥3.0 (4292 genes); (b) genes associated with neurodevelopmental conditions with autosomal or X-linked recessive conditions (based on the Decipher[55] list of genes; 545 genes) and (3) genes with pLI < 0.5 previously associated with ASD (based on SFARI database genes score 1 and S, https://gene.sfari.org/, retrieved in March, 23th 2023; 393 genes), in addition to a set of 154 genes retrieved from different articles on ASD candidate genes identification[56–58], totalizing 4934 unique genes, here referred as risk genes (Supplementary Data 1). Regarding the variant type, only LoF, missense (CADD ≥ 30), and splicing (spliceAI ≥ 0.8), variants were analyzed. For the Z-score set of brain-expressed genes (with pLI < 0.5 and not listed in the other set of genes), the variant was kept in the analysis only if it was classified as missense with CADD ≥ 30. The subset of such variants identified in our DMD groups are defined as "RRVs". As a form of control for the false discovery rate, we also tested if the groups differed regarding the synonymous ultra-rare (absent in the populational databases) variants in the same set of 4934 risk genes.

Frequency of dual diagnosis

To determine the number of individuals carrying additional pathogenic or likely pathogenic variants associated with monogenic NDDs, rare risk DNVs and RRVs from DMD-ASD and DMD-ID groups were clinically characterized. Genes were characterized according to disease association (OMIM, Decipher, and SFARI databases) and disease inheritance patterns. Variants in genes associated with monogenic neurodevelopmental conditions with autosomal dominant or X-linked inheritance, in addition to variants affecting known ASD risk genes, were classified according to the ACMG guidelines[59], using the Franklin by Genoox (https://franklin.genoox.com/) automatized classification, followed by manual curation. The frequency of individuals with dual diagnosis (DMD diagnosis plus a monogenic, highly penetrant NDD) was then determined for each group.

Gene enrichment and statistical analysis

Toppgene portal (http://toppgene.cchmc.org) was used to perform gene ontology (GO) and pathway enrichment analysis with the RRVs identified in the DMD-subgroups (biological processes, cellular components, molecular functions, biological pathways or disease). All terms enriched after Bonferroni correction for multiple tests were returned.

Statistical analysis

All statistical analyses were performed using either GraphPad Prism 9.0 (version 9.5.1) or RStudio (version 4.2.1). Statistical descriptive analyses were performed for clinical data, total number and frequency of individuals in different groups. Shapiro–Wilk and Kolmogorov–Smirnov tests were used to test data normality, and then nonparametric and parametric tests were used to compare clinical and molecular data, such as motor development among the three groups. Fishers' exact test was used to determine whether there was a difference in frequencies of speech impairment, learning difficulties, and *DMD*-affected isoforms among groups. It was also used to determine the differences between the number of total and risk DNVs between DMD-ASD and DMD-Control groups, and to determine the differences between the frequency of paternal and maternal risk variants in all groups. Poisson regression analysis was used to determine the association between parental age (number of variants ~ parental age) and de novo mutations, and Welch's t test was applied to determine whether there was a difference in the mean paternal age. Mann–Whitney Test (Wilcoxon Rank Sum Test) was used to compare the average of RRVs and synonymous variants between DMD-ASD/ID group and DMD-Control. It was also used to compare the average correlation coefficient between genes from the DMD-ASD/ID group and the total number of genes affected by RRVs. Bonferroni (BFC) or Benjamini-Hochberg False Discovery Rate (BH-FDR) correction was applied whenever multiple tests were performed.

Data availability

Genomic data from whole exome sequencing is available through EGA (European Genome-Phenome Archive; accession number EGA50000000706). The remaining data generated as part of this study is available from the corresponding author upon reasonable request.

References

1. Duan, D., Goemans, N., Takeda, S., Mercuri, E. & Aartsma-Rus, A. Duchenne muscular dystrophy. *Nat. Rev. Dis. Prim.* **7**, 13 (2021).
2. Zhang, X.-F., Luo, Y.-Y., Jiang, L. & Hong, S.-Q. Clinical study on cognitive impairment in Duchenne muscular dystrophy. *Neuromuscul. Disord.* **33**, 596–604 (2023).
3. Ricotti, V. et al. Neurodevelopmental, emotional, and behavioural problems in Duchenne muscular dystrophy in relation to underlying dystrophin gene mutations. *Dev. Med. Child Neurol.* **58**, 77–84 (2016).
4. Fujino, H. et al. Autism spectrum disorders are prevalent among patients with dystrophinopathies. *Neurol. Sci.* **39**, 1279–1282 (2018).
5. Colombo, P. et al. Assessing mental health in boys with Duchenne muscular dystrophy: Emotional, behavioural and neurodevelopmental profile in an Italian clinical sample. *Eur. J. Paediatr. Neurol.* **21**, 639–647 (2017).
6. Vo, A. H. & McNally, E. M. Modifier genes and their effect on Duchenne muscular dystrophy. *Curr. Opin. Neurol.* **28**, 528–534 (2015).
7. Hightower, R. M. & Alexander, M. S. Genetic modifiers of Duchenne and facioscapulohumeral muscular dystrophies. *Muscle Nerve* **57**, 6–15 (2018).
8. Pagnamenta, A. T. et al. A family with autism and rare copy number variants disrupting the Duchenne/Becker muscular dystrophy gene DMD and TRPM3. *J. Neurodev. Disord.* **3**, 124–131 (2011).
9. Chen, J.-J. et al. A Chinese girl with Turner syndrome and Duchenne muscular dystrophy: diagnosis and management of this "dual diagnosis". *Chin. Med. J.* **134**, 743–745 (2020).
10. Chaubey, A. et al. Low-pass genome sequencing: validation and diagnostic utility from 409 clinical cases of low-pass genome sequencing for the detection of copy number variants to replace constitutional microarray. *J. Mol. Diagn.* **22**, 823–840 (2020).
11. Pizzo, L. et al. Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants. *Genet. Med.* **21**, 816–825 (2019).
12. Wang, T., Zhao, P. A. & Eichler, E. E. Rare variants and the oligogenic architecture of autism. *Trends Genet.* **38**, 895–903 (2022).
13. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th edn. Arlington, VA: American Psychiatric Publishing; (2013).
14. Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 551–563 (2015).
15. Sandin, S. et al. The heritability of autism spectrum disorder. *JAMA* **318**, 1182–1184 (2017).
16. Schaaf, C. P. et al. A framework for an evidence-based gene list relevant to autism spectrum disorder. *Nat. Rev. Genet.* **21**, 367–376 (2020).
17. Leblond, C. S. et al. Both rare and common genetic variants contribute to autism in the Faroe Islands. *NPJ Genom. Med.* **4**, 1 (2019).
18. Darmahkasih, A. J. et al. Neurodevelopmental, behavioral, and emotional symptoms common in Duchenne muscular dystrophy. *Muscle Nerve* **61**, 466–474 (2020).
19. Pounraja, V. K. & Girirajan, S. A general framework for identifying oligogenic combinations of rare variants in complex disorders. *Genome Res.* **32**, 904–915 (2022).
20. Lord, C., Elsabbagh, M., Baird, G. & Veenstra-Vanderweele, J. Autism spectrum disorder. *Lancet* **392**, 508–520 (2018).
21. van Dommelen, P., van Dijk, O., de Wilde, J. A. & Verkerk, P. H. Early developmental milestones in Duchenne muscular dystrophy. *Dev. Med. Child Neurol.* **62**, 1198–1204 (2020).
22. Naidoo, M. & Anthony, K. Dystrophin dp71 and the neuropathophysiology of duchenne muscular dystrophy. *Mol. Neurobiol.* **57**, 1748–1767 (2020).
23. Sandin, S. et al. Autism risk associated with parental age and with increasing difference in age between the parents. *Mol. Psychiatry* **21**, 693–700 (2016).
24. Edwards, S. D. et al. Clinical characterization of individuals with the distal 1q21.1 microdeletion. *Am. J. Med. Genet. A* **185**, 1388–1398 (2021).
25. López-Hernández, T. et al. Mutant GlialCAM causes megalencephalic leukoencephalopathy with subcortical cysts, benign familial macrocephaly, and macrocephaly with retardation and autism. *Am. J. Hum. Genet.* **88**, 422–432 (2011).
26. Dominguez-Alonso, S., Carracedo, A. & Rodriguez-Fontenla, C. The non-coding genome in autism spectrum disorders. *Eur. J. Med. Genet.* **66**, 104752 (2023).
27. Malfait, F. et al. The Ehlers-Danlos syndromes. *Nat. Rev. Dis. Prim.* **6**, 64 (2020).
28. Long, K. R. & Huttner, W. B. How the extracellular matrix shapes neural development. *Open Biol.* **9**, 180216 (2019).
29. Casanova, E. L., Baeza-Velasco, C., Buchanan, C. B. & Casanova, M. F. The relationship between autism and ehlers-danlos syndromes/hypermobility spectrum disorders. *J. Pers. Med.* **10**, 260 (2020).
30. Baeza-Velasco, C. Neurodevelopmental atypisms in the context of joint hypermobility, hypermobility spectrum disorders, and Ehlers-Danlos syndromes. *Am. J. Med. Genet. C. Semin. Med. Genet.* **187**, 491–499 (2021).
31. Waite, A., Brown, S. C. & Blake, D. J. The dystrophin-glycoprotein complex in brain development and disease. *Trends Neurosci.* **35**, 487–496 (2012).

32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. Genomics in the Cloud [Book]. https://www.oreilly.com/library/view/genomics-in-the/9781491975183/.
34. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
35. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
36. Naslavsky, M. S. et al. Whole-genome sequencing of 1,171 elderly admixed individuals from São Paulo, Brazil. *Nat. Commun.* **13**, 1004 (2022).
37. Fairley, S., Lowy-Gallego, E., Perry, E. & Flicek, P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* **48**, D941–D947 (2020).
38. Chen, S. et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024).
39. Schubach, M., Maass, T., Nazaretyan, L., Röner, S. & Kircher, M. CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. *Nucleic Acids Res.* **52**, D1143–D1154 (2024).
40. Jaganathan, K. et al. Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548.e24 (2019).
41. Abrahams, B. S. et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).
42. Amberger, J. S., Bocchini, C. A., Scott, A. F. & Hamosh, A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* **47**, D1038–D1043 (2019).
43. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
44. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
45. Freeman, P. J., Hart, R. K., Gretton, L. J., Brookes, A. J. & Dalgleish, R. VariantValidator: accurate validation, mapping, and formatting of sequence variation descriptions. *Hum. Mutat.* **39**, 61–68 (2018).
46. Rehm, H. L. et al. ClinGen–the clinical genome resource. *N. Engl. J. Med.* **372**, 2235–2242 (2015).
47. Fromer, M. & Purcell, S. M. Using XHMM software to detect copy number variation in whole-exome sequencing data. *Curr. Protoc. Hum. Genet.* **81**, 7.23.1–7.23.21 (2014).
48. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
49. Povysil, G. et al. panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. *Hum. Mutat.* **38**, 889–897 (2017).
50. Geoffroy, V. et al. AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics* **34**, 3572–3574 (2018).
51. MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* **42**, D986–D992 (2014).
52. Riggs, E. R. et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet. Med.* **22**, 245–257 (2020).
53. Ramu, A. et al. DeNovoGear: de novo indel and point mutation discovery and phasing. *Nat. Methods* **10**, 985–987 (2013).
54. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **11**, 11.10.1–11.10.33 (2013).
55. Firth, H. V. et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
56. Zhou, X. et al. Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nat. Genet.* **54**, 1305–1319 (2022).
57. Fu, J. M. et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* **54**, 1320–1331 (2022).
58. Trost, B. et al. Genomic architecture of autism from comprehensive whole-genome sequence annotation. *Cell* **185**, 4409–4427.e18 (2022).
59. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).

Author contributions
C.I.S.C. and M.R.P.B. conceptualized and designed the study. G.L.Y. helped in study conceptualization, writing review, and manuscript editing. M.Z. provided access to most of participants' clinical data and to biological samples. C.I.S.C. wrote the manuscript. L.M., A.C.S.G., and E.Z. performed the CARS, RAVEN, and/or WISC tests. R.C.M.P. and V.R.C. performed DMD-affected individuals' clinical evaluation. J.Y.T.W. performed all exome data processing prior analysis. C.I.S.C., J.Y.T.W., and G.S.C. performed data collection and formal analysis. G.S.C. performed all Sanger sequencing analysis. F.M. and M.D.V. helped with all statistical analysis. M.S. and A.C.K. helped with CNV analysis. N.C.V.L. helped with data collection and additional genetic tests of the participants. All authors contributed to data discussion and interpretation of the results and revised the manuscript.

Competing interests
Dr. Maria Rita Passos-Bueno serves as associate editor of this journal and had no role in the peer-review or decision to publish this manuscript. The remaining authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-025-00469-5.

Correspondence and requests for materials should be addressed to Maria Rita Passos-Bueno.

Reprints and permissions information is available at http://www.nature.com/reprints