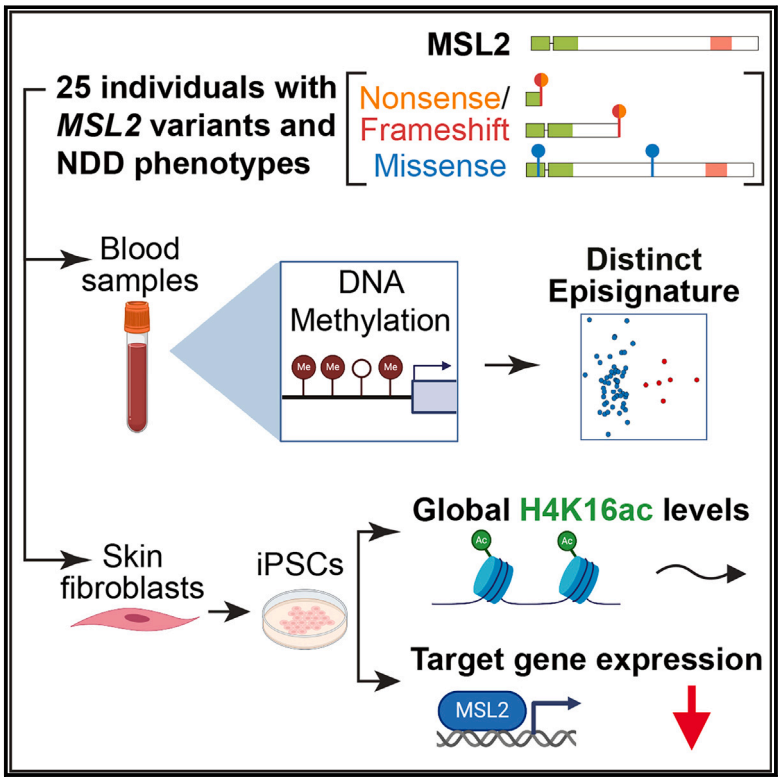
MSL2 variants lead to a neurodevelopmental syndrome with lack of coordination, epilepsy, specific dysmorphisms, and a distinct episignature

Graphical abstract



Authors

Remzi Karayol, Maria Carla Borroto, Sadegheh Haghshenas, ..., Bekim Sadikovic, Asifa Akhtar, Philippe M. Campeau

Correspondence

bekim.sadikovic@lhsc.on.ca (B.S.), akhtar@ie-freiburg.mpg.de (A.A.)

***MSL2* encodes a member of the MSL complex, an epigenetic regulator acetylating histone H4. We identify *MSL2* variants leading to a neurodevelopmental disorder with intellectual disability, developmental delay, motor issues, seizures, dysmorphisms, and a specific blood methylation episignature. Patient-derived reprogrammed cells reveal developmental gene dysregulation without altered global H4 acetylation.**

Karayol et al., 2024, The American Journal of Human Genetics *111*, 1330–1351
July 11, 2024 © 2024
https://doi.org/10.1016/j.ajhg.2024.05.001

ARTICLE

MSL2 variants lead to a neurodevelopmental syndrome with lack of coordination, epilepsy, specific dysmorphisms, and a distinct episignature

Remzi Karayol,[1,54] Maria Carla Borroto,[2,54] Sadegheh Haghshenas,[3,54] Anoja Namasivayam,[1] Jack Reilly,[4] Michael A. Levy,[3] Raissa Relator,[3] Jennifer Kerkhof,[3] Haley McConkey,[3,5] Maria Shvedunova,[1] Andrea K. Petersen,[6] Kari Magnussen,[6] Christiane Zweier,[7,8] Georgia Vasileiou,[7] André Reis,[7] Juliann M. Savatt,[9] Meghan R. Mulligan,[10] Louise S. Bicknell,[10] Gemma Poke,[11] Aya Abu-El-Haija,[12] Jessica Duis,[13] Vickie Hannig,[14] Siddharth Srivastava,[15] Elizabeth Barkoudah,[16] Natalie S. Hauser,[17] Myrthe van den Born,[18] Uri Hamiel,[19] Noa Henig,[20] Hagit Baris Feldman,[19] Shane McKee,[21] Ingrid P.C. Krapels,[22] Yunping Lei,[23] Albena Todorova,[24,25] Ralitsa Yordanova,[26,27]

(Author list continued on next page)

Summary

Epigenetic dysregulation has emerged as an important etiological mechanism of neurodevelopmental disorders (NDDs). Pathogenic variation in epigenetic regulators can impair deposition of histone post-translational modifications leading to aberrant spatiotemporal gene expression during neurodevelopment. The male-specific lethal (MSL) complex is a prominent multi-subunit epigenetic regulator of gene expression and is responsible for histone 4 lysine 16 acetylation (H4K16ac). Using exome sequencing, here we identify a cohort of 25 individuals with heterozygous *de novo* variants in MSL complex member *MSL2*. *MSL2* variants were associated with NDD phenotypes including global developmental delay, intellectual disability, hypotonia, and motor issues such as coordination problems, feeding difficulties, and gait disturbance. Dysmorphisms and behavioral and/or psychiatric conditions, including autism spectrum disorder, and to a lesser extent, seizures, connective tissue disease signs, sleep disturbance, vision problems, and other organ anomalies, were observed in affected individuals. As a molecular biomarker, a sensitive and specific DNA methylation episignature has been established. Induced pluripotent stem cells (iPSCs) derived from three members of our cohort exhibited reduced MSL2 levels. Remarkably, while NDD-associated variants in two other members of the MSL complex (*MOF* and *MSL3*) result in reduced H4K16ac, global H4K16ac levels are unchanged in iPSCs with *MSL2* variants. Regardless, *MSL2* variants altered the expression of MSL2 targets in iPSCs and upon their differentiation to early germ layers. Our study defines an *MSL2*-related disorder as an NDD with distinguishable clinical features, a specific blood DNA episignature, and a distinct, MSL2-specific molecular etiology compared to other MSL complex-related disorders.

Introduction

Epigenetic regulators are a diverse class of proteins responsible for modulating chromatin accessibility and gene transcription in response to intrinsic or extrinsic cell signals. Epigenetic regulators are key mediators of lineage-specific developmental gene programs, particularly in the nervous system.[1,2] Unsurprisingly, variants in epigenetic regulators

[1]Max-Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany; [2]Centre de recherche Azrieli du CHU Sainte-Justine, Montreal, QC H3T 1C5, Canada; [3]Verspeeten Clinical Genome Centre, London Health Sciences Centre, London, ON N6A 5W9, Canada; [4]Department of Pediatrics, Clinical Neurological Sciences and Epidemiology, Western University, London, ON N6A 3K7, Canada; [5]Department of Pathology and Laboratory Medicine, Western University, London, ON N6A 3K7, Canada; [6]Department of Genetics and Metabolism, Randall Children's and Legacy Emanuel Hospitals, Portland, OR 97227, USA; [7]Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany; [8]Department of Human Genetics, Inselspital, Bern University Hospital, University of Bern, 3010 Bern, Switzerland; [9]Autism & Developmental Medicine Institute, Geisinger, Danville, PA, USA; [10]Department of Biochemistry, University of Otago, Dunedin, New Zealand; [11]Genetic Health Service New Zealand, Wellington, New Zealand; [12]Division of Genetics, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA; [13]Section of Genetics & Metabolism, Department of Pediatrics, University of Colorado, Children's Hospital Colorado, Aurora, CO, USA; [14]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, USA; [15]Translational Neuroscience Center, Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA; [16]Department of Neurology, Boston Children's Hospital, Boston, MA, USA; [17]Medical Genetics, Inova Fairfax Hospital, Falls Church, VA 22042, USA; [18]Department of Clinical Genetics, Erasmus MC, University Medical Center, Rotterdam, the Netherlands; [19]Genetics Institute and Genomics Center, Tel Aviv Sourasky Medical Center & Faculty of Medicine, Tel Aviv University, Tel Aviv 6423906, Israel; [20]Genetics Institute and Genomics Center, Tel Aviv Sourasky Medical Center, Tel Aviv 6423906, Israel; [21]Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast Health & Social Care Trust, Belfast BT9 7AB, UK; [22]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, the Netherlands; [23]Center for Precision Environmental Health, Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX, USA; [24]Department of Medical Chemistry and Biochemistry, Medical University Sofia, Sofia, Bulgaria; [25]Genetic Medico-Diagnostic Laboratory "Genica", Sofia, Bulgaria; [26]Department of pediatrics "Prof. Ivan Andreev", Medical university - Plovdiv, Plovdiv, Bulgaria; [27]Department of Pediatrics, University Hospital "St. George", Plovdiv, Bulgaria; [28]Department of Neurogenetics, Kennedy Krieger Institute, Baltimore, MD, USA; [29]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [30]Department of Pediatrics, Division of Medical Genetics, Duke University School of Medicine, Durham, NC 27710, USA; [31]Department of Medical Genetics, ASST Papa Giovanni XXIII, Bergamo, Italy; [32]Epilepsy Center – Sleep Medicine Center, Childhood and Adolescence Neuropsychiatry Unit, ASST

(Affiliations continued on next page)

Slavena Atemin,[25] Mihael Rogac,[29] Vivienne McConnell,[21] Anna Chassevent,[28] Kristin W. Barañano,[29] Vandana Shashi,[30] Jennifer A. Sullivan,[30] Angela Peron,[52,53] Maria Iascone,[31] Maria P. Canevini,[32,33] Jennifer Friedman,[34,35] Iris A. Reyes,[35] Janell Kierstein,[13] Joseph J. Shen,[36] Faria N. Ahmed,[37] Xiao Mao,[38,39] Berta Almoguera,[40,41] Fiona Blanco-Kelly,[40,41] Konrad Platzer,[42] Ariana-Berenike Treu,[43] Juliette Quilichini,[44] Alexia Bourgois,[45] Nicolas Chatron,[46,47] Louis Januel,[46] Christelle Rougeot,[48] Deanna Alexis Carere,[49] Kristin G. Monaghan,[49] Justine Rousseau,[2] Kenneth A. Myers,[50] Bekim Sadikovic,[3,5,55,*] Asifa Akhtar,[1,55,*] and Philippe M. Campeau[2,51,55]

and chromatin-related genes are frequently linked to neurodevelopmental disorders. Among the chromatin regulators, lysine acetyltransferases were found to be the subfamily with the 3rd highest frequency of *de novo* missense and protein-truncating variants in autism and developmental disorders,[3] making it imperative to gain a deeper understanding of the involvement of this category of epigenetic factor in the neurodevelopmental context. Within the lysine acetyltransferase class, MOF (KAT8) is particularly interesting for its ability to change substrate specificity according to the protein complex with which it associates. As part of the MSL complex, MOF predominantly acetylates histone H4 lysine 16 (H4K16).[4–6] On the other hand, as a member of the NSL complex, MOF has the capacity to acetylate multiple H4 tail residues (H4K5, H4K8, H4K12, H4K16) as well as several non-histone substrates.[7–9] MOF's direct or indirect interactions with MSL complex components are key determinants for guiding its activity toward H4K16ac. The mammalian MSL complex is formed by the homodimerization of the heterotetramer sub-complex consisting of MSL1, MSL2, MSL3, and MOF. MOF is the acetyltransferase and MSL3 is required to stimulate MOF's catalytic activity. MSL2 confers DNA binding, whereas MSL1 provides the dimerization interface required for complex integrity.[10–12]

In recent years, H4K16ac has been linked to neurodevelopmental disorders (NDDs) via the discovery of *de novo* pathogenic variants in the MSL complex members. Pathogenic variants in *MOF* (*KAT8* [MIM: 609912]) are linked to Li-Ghorbani-Weisz-Hubshman syndrome (LIGOWS [MIM: 618974]), an NDD with global developmental delay (DD) and intellectual disability (ID). Patient-mimicking mutations in MOF result in reduced *in vitro* H4K16 acetylation activity of the protein.[13] *De novo* variants in another member of the MSL complex, *MSL3* (MIM: 300609), lead to the Basilicata-Akhtar syndrome (MRXSBA [MIM: 301032]), an X-linked condition affecting both sexes and characterized by ID, global DD, hypotonia, gait disturbance, and spasticity.[14,15] Primary human dermal fibroblasts taken from individuals with MRXSBA exhibit reduced bulk H4K16ac.[14]

Most recently, *MSL2* (MIM: 614802) was identified as a candidate gene with recurrent *de novo* pathogenic variants in autism spectrum disorder (ASD).[16] In addition, two *MSL2* protein-truncating variants were detected in a large cohort of individuals with neurodevelopmental disorders (NDDs).[17] Accordingly, three individuals with *MSL2 de novo* pathogenic variants (two frameshift, one missense) have been reported as a case study revealing strikingly similar phenotypes: ASD, obsessive compulsive disorder (OCD), mild ID, and joint hypermobility.[18] However, the limited number of probands described to date prevents a detailed description of the disorder. Furthermore, the molecular basis of the phenotypic effects of likely pathogenic variants of *MSL2* has not been analyzed in previous studies.[17,18]

Genome-wide DNA methylation profiles in individuals with a number of rare diseases have been used to derive highly sensitive and specific biomarkers called episignatures[19,20] and provide functional insights related to the genome-wide DNA methylation profiles in genetic neurodevelopmental disorders.[21] This technology has recently been applied in clinical diagnostics for reclassification of ambiguous genomic findings and for screening of individuals with clinical features overlapping this growing list of episignature-positive rare diseases.[22,23]

Santi Paolo e Carlo, San Paolo Hospital, Milan, Italy; [33]Department of Health Sciences, University of Milan, Milan, Italy; [34]Departments of Neurosciences and Pediatrics, University of California, San Diego, La Jolla, CA, USA; [35]Rady Children's Institute for Genomic Medicine and Rady Children's Hospital, San Diego, CA, USA; [36]Division of Genomic Medicine, Department of Pediatrics, MIND Institute, UC Davis, Sacramento, CA 95817, USA; [37]Division of Genomic Medicine, Department of Pediatrics, UC Davis, Sacramento, CA 95817, USA; [38]National Health Commission Key Laboratory of Birth Defects Research and Prevention, Hunan Provincial Maternal and Child Health Care Hospital, Hunan, China; [39]Nanhua University, Chiayi County, Taiwan; [40]Department of Genetics and Genomics, Fundacion Jimenez Diaz University Hospital, Health Research Institute-Fundacion Jimenez Diaz, Universidad Autonoma de Madrid (IIS-FJD, UAM), Madrid, Spain; [41]Center for Biomedical Network Research on Rare Diseases (CIBERER), Madrid, Spain; [42]Institute of Human Genetics, University of Leipzig Medical Center, Leipzig, German; [43]Epilepsy Center Kleinwachau, Radeberg, Germany; [44]Service de Médecine Génomique des maladies de système et d'organe, APHP, Centre Université Paris Cité, Paris, France; [45]Normandy University, UNICAEN, Caen University Hospital, Department of Genetics, UR 7450 BioTARGen, FHU G4 Genomics, Caen, France; [46]Department of Genetics, Lyon University Hospital, Lyon, France; [47]Pathophysiology and Genetics of Neuron and Muscle (PGNM, UCBL - CNRS UMR5261 - INSERM U1315), Université Claude Bernard Lyon 1, Lyon, France; [48]Department of Neuropediatrics, Lyon University Hospital, Lyon, France; [49]GeneDx, Gaithersburg, MD 20877, USA; [50]Child Health and Human Development, Research Institute of the McGill University Health Centre, Montreal, QC, Canada; [51]Department of Pediatrics, University of Montreal, Montreal, QC H3T 1C5, Canada; [52]SOC Genetica Medica, Meyer Children's Hospital IRCCS, Florence, Italy; [53]Department of Biomedical, Experimental and Clinical Sciences "Mario Serio", Università degli Studi di Firenze, Florence, Italy
[54]These authors contributed equally
[55]Senior author
*Correspondence: bekim.sadikovic@lhsc.on.ca (B.S.), akhtar@ie-freiburg.mpg.de (A.A.)
https://doi.org/10.1016/j.ajhg.2024.05.001.

Here we describe 25 individuals with heterozygous, missense, or protein-truncating variant alleles in *MSL2*, delineating genotypes and phenotypes associated these variants with global DD, ID, hypotonia, and motor issues such as coordination problems, feeding difficulties, and gait disturbance. Dysmorphisms and behavioral and/or psychiatric conditions including ASD were also observed in affected individuals. Epilepsy, connective tissue disease (CTD) signs, sleep disturbance (i.e., sleep apnea and/or fragmentation), and dysfunction of other organs were also present. Using blood samples from six cohort members, we have established a sensitive and specific DNA methylation episignature as a molecular biomarker. Furthermore, we have derived induced pluripotent stem cells (iPSCs) from three cohort members, allowing molecular characterization of the *MSL2* variants. Patient-derived iPSCs with truncating variants showed reduced intact MSL2 levels without an effect on bulk H4K16ac levels. On the other hand, the expression levels of putative MSL2 target genes were dysregulated both in iPSCs and upon their differentiation into early germ layers. Overall, heterozygous *MSL2* variants alter the expression of MSL2 targets leading to a neurodevelopmental disorder with specific phenotypic features and a distinct episignature. We tentatively suggest a name for this phenotype: the *MSL2*-associated neurodevelopmental syndrome (MANDS).

Material and methods

Clinical and genetic data

We established a cohort with 25 individuals in whom an *MSL2* variant had been identified. Their clinical information was obtained through their respective physicians. An international collaboration was facilitated by the online platform GeneMatcher,[24] which allows clinicians and researchers interested in the same gene to find and communicate with each other. Clinical information is shared according to the rules in place by Institutional Review Boards of each institution. Affected individuals or their guardians provided informed consent to participate in research studies. Signed consent to publish photographs was obtained where applicable.

Phenotypic quantitative analysis

As previously described,[25,26] phenotypes were annotated with Human Phenotype Ontology (HPO) terms corresponding to their clinical features, then used to generate symmetric Lin scores with the OntologyX suite of packages[27] in R (https://www.R-project.org/). The resulting matrix of phenotype similarity scores was used to generate a distance matrix, then used to generate a hierarchical clustering with the Ward method. A heatmap was generated with the ComplexHeatmap package in R,[28] and clusters were traced after determining the optimal number of clusters ($k = 3$) by applying the Elbow method to a gap statistic curve.

DNA isolation and methylation analysis

DNA extracted from peripheral blood was processed using Illumina Infinium methylation EPIC bead chip arrays, following bisulfite conversion, according to the manufacturer's protocol. These arrays cover >860,000 human genome CpGs. The process of DNA methylation analysis was described in detail previously.[22,29,30] In summary, analysis was performed in R (v.4.1.2), where methylated and unmethylated signal intensities were normalized by background correction using minfi package (v.1.40.0).[31] Probes meeting certain criteria (such as detection p value above 0.01, located on X and Y chromosomes, containing SNPs at or near CpG interrogation or single nucleotide extension, or known to cross-react with chromosomal locations other than their target regions) were excluded from the analysis. Principal component analysis (PCA) was used to observe batch structure and detect possible outliers. The samples used for the study were from six affected individuals (three individuals with the c.694_697delTCTG [p.Ser232Thrfs*10] variant and one each with c.767delT [p.Phe256Serfs*5], c.796_797delCT [p.Leu266Valfs*5], and c.949A>G [p.Met317Val] variants).

Selection of matched control subjects and DNA methylation profiling

Fifty-four control individuals were selected from the EpiSign Knowledge Database (EKD, https://episign.lhsc.on.ca/index.html) using MatchIt package (v.4.3.3),[32] matched to six *MSL2* subjects by age and sex. PCA was performed after every round of matching until no outliers were detected in the first two components of PCA. Probe selection procedure was described in detail previously.[33] In summary, methylated signal intensity was divided by the sum of methylated and unmethylated signal intensities to obtain the methylation level (β-value), which was then converted into M-values using logit transformation. The limma package (v.3.50.0)[34] was used to perform a linear model to obtain the mean methylation difference and the corresponding p value for each probe, taking into account the blood cell type compositions estimated by Houseman's algorithm[35] as confounding variables. The resulting p values were moderated using eBayes function in the limma package and corrected for multiple testing using the Benjamini and Hochberg (BH) algorithm.

Probes most differentiating the samples with MSL2 variants from controls were selected using a three-step process. First, 900 probes with the highest products of mean methylation differences and p values were chosen. Second, 450 probes with the highest areas under the receiver operating characteristic curve (AUROC) were retained. Thirdly, probes with a correlation higher than 0.7 were excluded, resulting in 239 probes forming the MSL2-related disorder episignature. To verify the robustness of the selected probes, hierarchical clustering was performed using Ward's method on Euclidean distance, and multidimensional scaling (MDS) was conducted by computing the pairwise Euclidean distances between samples.

Datasets used in this study that are available publicly were previously described.[20] Some of the datasets used in this study are publicly available and may be obtained from Gene Expression Omnibus (GEO) using the following accession numbers: GEO: GSE116992, GSE66552, GSE74432, GSE97362, GSE116300, GSE95040, GSE104451, GSE125367, GSE55491, GSE108423, GSE116300, GSE89353, GSE52588, GSE42861, GSE85210, GSE87571, GSE87648, GSE99863, and GSE35069.

Binary classifier construction

Using 239 selected probes, a support vector machine (SVM) classifier was constructed to classify *MSL2* subjects with greater

accuracy. This process has been previously described in detail.[22,29,30] The e1071 package (v.1.7.9) was utilized for constructing the SVM model. The model produces a methylation variant pathogenicity (MVP) score for each individual, ranging from 0 to 1, with a higher score indicating greater similarity to the MSL2-related disorder episignature. The construction of the model involved training six samples with *MSL2* variants against 54 matched control individuals, 75% ($N = 587$) of other control subjects, and 75% ($N = 1315$) of subjects from other rare genetic disorders with known episignatures from the EKD. The remaining 25% ($N = 635$) of control samples and individuals with other rare genetic disorders were used as testing samples.

Differentially methylated regions and gene set enrichment analysis

To investigate differentially methylated regions (DMRs), the DMRcate package (v.2.8.3) was utilized.[36] A region was considered differentially methylated if the mean methylation difference between case and control groups was >0.05 for at least 3 CpG sites within 1 kb. Following this, missMethyl package (v.1.28.0) was used for gene ontology (GO) enrichment analysis of the identified DMRs.[37]

Isolation of dermal fibroblasts and iPSC reprogramming

Skin fibroblasts were biopsied and cultured using standard methods.[38] These were reprogrammed to iPSCs using integration-free Sendai virus expressing the OSKM Yamanaka factors, then iPSC colonies were stained with antibodies for SSEA-4, Sox2, OCT4, and TRA1, as described previously.[39]

iPSC maintenance

The control iPSC line used in this study, HMGU1, is a well-characterized male iPSCs line (ISFi001-A, hpscreg.eu/cell-line/ISFi001-A), provided by Drs. Mishca Drukker and Ejona Rusha from Helmholtz Zentrum München (HMGU). All iPSCs lines were adapted, maintained, and passaged in mTeSR Plus medium (Stemcell Technologies) on Matrigel (Corning)-coated cell culture plates according to manufacturer's guidelines. hPSC Genetic Analysis Kit (Stemcell Technologies) was used to test all iPSC lines for the majority of karyotypic abnormalities reported in human iPSCs.

Protein extraction and western blotting

iPSCs were harvested at the day of passaging using ReLeSR Human PSC Selection & Passaging Reagent (Stemcell Technologies). Subcellular protein fractions were prepared using Nuclear NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific), and protein concentrations were measured using Pierce Rapid Gold BCA Protein Assay Kit (Thermo Scientific). Chromatin-bound proteins were extracted from the remaining insoluble fraction using reducing buffer ROTILoad 1 (Carl Roth) and sonication. For Western blot loading, ROTILoad 1 was added to protein samples, followed by boiling for 5 min. Proteins were separated by SDS-PAGE using NuPAGE Bis-Tris 4-12% gels (Invitrogen) and 1X MOPS buffer (Invitrogen), or NuPAGE Bis-Tris 12% gels (Invitrogen) and 1X MES buffer (Invitrogen) for proper separation of histones. Proteins were transferred onto 0.2 μm polyvinylidene difluoride (PVDF) membranes using Trans-Blot Turbo Transfer System (Bio-rad) according to manufacturer's instructions. Membranes were blocked in 5% milk in PBS with 0.3% Tween 20 for 1 h at room temperature, then incubated with primary antibodies overnight at 4°C. Primary antibodies used are rabbit anti-MSL2 (Atlas Antibodies/Sigma-Aldrich, Cat.# HPA003413), rabbit anti-DHX9 (Abcam, Cat.# ab26271), rabbit anti-MOF (Abcam, Cat.# ab200660), rabbit anti-H4K16ac (Sigma-Aldrich, Cat.# 07–329), and mouse anti-H4-HRP (Abcam, Cat.# ab197517); all were 1:1,000 diluted in the blocking solution. HRP-conjugated secondary antibodies were used and bands were detected using Lumilight Western blotting substrate (Roche) or SuperSignal West Atto Ultimate Sensitivity Substrate (Thermo Scientific) and visualized using a Bio-Rad Imager.

RNA isolation and reverse transcription quantitative PCR (RT-qPCR)

Total RNA was extracted using the Direct-zol RNA Miniprep Plus Kit (Zymo Research) according to the manufacturer's instructions. GoScript Reverse Transcription System (Promega) was used to synthesize cDNA from total RNA according to the manufacturers' instructions. Quantitative reverse transcription PCR (RT-qPCR) was carried out on Roche LightCycler II using the Faststart SYBR Green Master (Rox) mix (Roche) at a final volume of 10 μL. Primer sequences are provided in Table S1.

Trilineage differentiation

iPSCs were differentiated to all three germ layers using StemDiff Trilineage Differentiation Kit (Stemcell Technologies) according to manufacturer's instructions. Differentiation success was confirmed based on the relative expression of lineage specific markers measured by RT-qPCR.

Statistical analysis

Methods used in HPO and DNA methylation analyses are described above. For qPCR and RT-qPCR data, Graphpad Prism 9 was used to generate bar graphs and perform the statistical analyses, one-way and two-way ANOVA, which are detailed in figure legends.

Results

Clinical findings

We identified 25 individuals aged from 4 months to 37 years (average of 10 years, median of 6) with start-loss, missense, nonsense, or frameshift variants in *MSL2* (see Table 1). The complete clinical information can be found in Table S2.

Impaired neurodevelopment

All individuals had some form of DD and/or ID. Please note that the term DD is usually reserved for children younger than 5 years and refers to a delay in the achievement of milestones in the motor domain, speech domain, or both (global DD). Most of the individuals had motor (21/24, 88%) and/or speech (18/23, 78%) delay, as well as hypotonia (18/24, 75%), either neonatal or permanent. One individual (13, c.659dup [p.Cys221Metfs*2]) presented with the specific neuromuscular condition of spastic diplegic cerebral palsy (see Table S2 for detailed clinical data). Coordination (11/19, 58%) and feeding (8/22, 36%) difficulties were present in approximately half of the individuals. Other signs of neurodevelopmental impairment in our cohort included dyspraxia, ataxic gait, heel-toe gait, and

Table 1. Clinical summary

Individual	1	2	3	4	5	6	7	8	9	10	11	12	13
Sex	M	M	M	M	M	F	M	M	M	M	F	M	F
Age	10 y 11 mo	3 y	5 y	8 y	5 y	4 mo	5 y	37 y	3 y 5 mo	34 y	18 y	2 y	5 y
Variant	c.1A>G p.Met1?	c.44G>T p.Arg15 Leu	c.949A>G p.Met317 Val	c.511C>T p.Gln171*	c.534dupT p.Glu179*	c.778delA p.Ile260*	c.1057 C>T p.Gln353*	c.1231_ 1232del p.His412*	c.105dup p.Pro36 Alafs*36	c.112_ 115dup p.Arg39 Leufs*34	c.396_399 delGCTT p.Leu133 Metfs*22	c.521dupT p.Leu174 Phefs*6	c.659dup p.Cys221 Metfs*2
Coding impact	start-loss	missense	missense	nonsense	nonsense	nonsense	nonsense	nonsense	frameshift	frameshift	frameshift	frameshift	frameshift
CADD score	–	31	17.15	–	–	–	–	–	–	–	–	–	–
Inheritance	N/A	dn	dn	dn	dn	dn	dn	het	dn	dn	dn	dn	dn
Clinical information													
Perinatal complications	+	–	+	+	+	+	–	–	+	–	–	+	–
OFC anomaly	+	–	+	–	N/A	+	–	N/A	–	–	–	–	N/A
ID/DD	+	+	+	+	+	+	+	+	+	+	+	+	+
Speech delay	+	+	+	+	+	N/A	–	–	–	+	+	+	N/A
Motor delay	+	–	+	+	+	+	+	–	+	+	+	+	N/A
Hypotonia	+	+	+	+	+	+	+	–	+	N/A	–	–	+
Feeding difficulties	–	–	+	–	+	+	+	–	+	N/A	–	N/A	N/A
Coordination issues	+	–	N/A	+	+	N/A	N/A	–	–	+	N/A	N/A	+
Breathing assistance	–	–	–	–	–	+	–	–	–	–	–	–	–
Behavioral abnormalities	+	+	N/A	+	+	N/A	–	+	+	N/A	+	+	+
Autistic features	+	+	N/A	–	–	N/A	–	–	+	N/A	+	+	–
Attention deficit	+	N/A	N/A	–	–	N/A	–	+	–	N/A	+	–	+
Psychiatric concerns	+	–	N/A	–	–	N/A	–	+	–	N/A	+	–	–
Sleep issues	+	+	+	–	–	–	+	+	–	–	+	–	–
Seizures	–	–	+	–	–	+	+	+	+	+	+	–	–
MRI abnormalities	–	+	+	+	–	N/A	–	–	–	+	–	–	+
EEG abnormalities	–	–	–	N/A	–	+	+	+	+	N/A	–	–	–
Dysmorphisms	+	+	+	+	+	+	+	–	+	+	–	+	–
Deep-set eyes	+	+	+	–	+	–	+	–	–	+	–	–	–
CTD signs	+	–	–	+	+	–	+	–	+	–	–	–	–
Visual anomalies	+	–	+	+	–	–	+	–	–	–	–	–	–
Gastrointestinal anomalies	–	+	+	+	–	–	+	N/A	+	–	–	–	–
Renal anomalies	–	–	–	–	–	+	–	N/A	–	–	–	–	–
Cardiac anomalies	–	–	+	–	–	–	–	N/A	–	–	–	–	–
Genitourinary anomalies	–	–	+	–	–	–	–	N/A	–	–	+	–	–

M, male; F, female; y, years; mo, months; OFC, occipital frontal circumference; ID, intellectual disability; DD, developmental impairment; CTD, connective tissue disease; dn, *de novo*; +, yes; −, no; N/A, not available.
[a]g.[136152145_136152149delinsC;136152802_136152981inv] (GenBank: NC_000003.12)

14	15	16	17	18	19	20	21	22	23	24	25
M	F	M	F	F	F	M	F	M	M	M	F
13 y 6 mo	2 y	2 y	6 y	33 y	8 y 6 mo	2 y	8 y 7 mo	9 y 3 mo	5 y	6 y 1 mo	13 y
c.684_685del p.Glu229 Glyfs*4	c.694_697 delTCTG p.Ser232 Thrfs*10	c.694_697 delTCTG p.Ser232 Thrfs*10	c.694_697 delTCTG p.Ser232 Thrfs*10	c.694_697 delTCTG p.Ser232 Thrfs*10	c.707_708 insTGCC p.Pro237 Alafs*5	c.767delT p.Phe256 Serfs*5	c.796_797delCT p.Leu266 Valfs*5	c.796_797 delCT p.Leu266 Valfs*5	c.823del p.Arg275 Alafs*3	c.1102delG p.Ala368 Hisfs*5	complex[a]
frameshift	frameshift	frameshift	frameshift	frameshift	frameshift	frameshift	frameshift	frameshift	frameshift	frameshift	
–	–	–	–	–	–	–	–	–	–	–	–
maternal	dn	dn	dn	dn	dn	dn	dn	dn	dn	dn	dn
Clinical information											
–	–	–	+	–	+	+	–	+	–	–	–
–	N/A	N/A	–	+	+	N/A	–	–	N/A	–	–
+	+	+	+	+	+	+	+	+	+	+	+
+	–	+	+	+	+	+	+	+	–	+	+
+	+	+	+	+	+	+	+	+	–	+	+
+	–	+	+	+	+	+	+	+	–	+	–
–	+	+	–	–	–	–	+	–	–	–	–
+	N/A	–	+	+	+	–	+	–	–	–	+
–	–	–	–	–	–	–	–	–	–	–	–
+	+	–	–	–	+	–	+	+	–	+	+
+	+	–	–	–	+	N/A	+	+	–	+	–
–	–	–	–	+	–	N/A	+	+	–	+	+
–	+	–	–	–	–	N/A	+	–	–	+	–
–	–	–	–	–	–	–	–	–	–	+	–
–	–	–	–	–	+	–	–	–	+	–	–
–	N/A	–	–	–	+	+	N/A	+	–	+	N/A
–	–	–	–	–	+	–	N/A	–	–	–	–
+	+	+	+	–	+	+	–	+	–	+	+
+	+	+	–	–	+	–	–	N/A	–	+	–
+	–	+	+	–	+	+	–	+	–	+	+
+	–	–	+	+	–	–	–	+	–	+	–
–	–	–	–	–	–	+	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	+	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	+	–

excessive drooling. The youngest individual (6, c.778delA [p.Ile260*]) was severely affected and required respiratory support (1/25, 4%). These observations indicate that neuromuscular impairment is a recurrent element in our cohort.

Over a third of the individuals had seizures (9/25, 36%), while around half of them had MRI abnormalities (9/21, 43%). These included delayed myelination, white matter loss, small corpus callosum, enlarged extracerebral spaces, dilated ventricles, hydrocephaly, periventricular nodular heterotopia, cavum septum pellucidum, polymicrogyria, and schizencephaly. EEG abnormalities were observed in around a quarter of the assessed individuals (5/22, 23%) and they included hypsarrhythmia, slowing of background rhythms with spike and polyspike wave complexes, and burst suppression patterns (the latter seen on individual 6, p.Ile260*, along with an epilepsy onset from the first 24 hr after birth). Occipital frontal circumference (OFC) anomalies (5/18, 28%) must also be underlined as one individual had absolute macrocephaly (OFC $>$ 2 SD), three had absolute microcephaly (<-2 SD), and others had relative macrocephaly (close to OFC $>$ 2 SD) or microcephaly (close to <-2SD) (see Table S2 for detailed clinical data). These observations suggest that cerebral malformations and neurophysiological anomalies are key in the MSL2-related disorder, which may help explain the DD, neuromotor conditions, epileptic activity, and even behavioral abnormalities commonly seen in our cohort.[40]

Behavioral abnormalities

Behavioral abnormality (16/22, 73%) was predominant in our cohort, including anxiety, depression, echolalia, repetitive movements, OCD, aggressivity, hyperactivity, self-injurious behavior, mixed receptive-expressive language delay, sensory processing disorder, oppositional defiant disorder, food obsession, strong mimicking, and poor regulation. Over half of the assessed individuals presented with autistic features (11/21, 52%). Attention deficit (9/20, 45%) and psychiatric disorders (6/21, 29%) were relatively common. Difficulty with sleep was noted in close to a third of the individuals (7/25, 28%), while three individuals (1, c.1A>G [p.(Met1?)]; 3, p.Met317Val; and 7, c.1057C>T [p.Gln353*]) were diagnosed with sleep apnea. These observations suggest that individuals with pathogenic or potentially pathogenic variants in MSL2 may require thorough medical monitoring in regard to autistic features and psychiatric conditions.

Perinatal complications

Around half of the individuals overcame perinatal complications (11/25, 44%). These included prematurity, intrauterine growth restriction, reanimation due to fetal distress, dystocia, and maternal substance abuse. Prenatal stress is linked to cognitive, behavioral, and developmental problems later in life,[41] so there might be some etiological overlap between these and other clinical features previously described, but excluding some conditions representative of our cohort such as cerebral anomalies, seizures, and neuromotor impairment.

Dysmorphisms and appearance

Most of the individuals (19/25, 76%) had dysmorphic features (see Table S3), including deep-set eyes, abnormal calvaria, toe anomalies, strabismus, downslanted palpebral fissures, large ear/earlobe, low-set/posteriorly rotated ears, short/upturned/flat nose, retrognathia, dermal translucency, narrow/tall forehead, abnormal philtrum, and thin vermillion (see Table S2 for detailed clinical data). Deep-set eyes (11/24, 46%) and CTD signs (13/25, 52%) were present in around half of the individuals. CTD signs encompass abnormalities of any connective tissue: tendons, ligaments, skin, cartilage, bone, and/or blood vessels. In our cohort, CTD signs included joint subluxations, mild pectus excavatum, translucent skin, epistaxis, and inguinal hernias. One individual (individual 13, p.Cys221Metfs*2) had a femoral anteversion, and two of them (individuals 11, c.396_399delGCTT [p.Leu133Metfs*22], and 18, p.Ser232Thrfs*10) had atopic dermatitis, but neither of them had CTD signs.

We also obtained facial photos of individuals 4 (c.511C>T [p.Gln171*]), 7 (c.1057C>T [p.Gln353*]), 10 (c.112_115dup [p.Arg39Leufs*34]), 21 (p.Leu266Valfs*5), and 25 (complex variant) (Figure 1). Individual 4 had large ears and visible veins on the forehead. Individual 7 had a higher forehead, mild metopic ridge, mild malar hypoplasia, slightly smoother philtrum, thin vermillion border of the upper lip, and mild retrognathia. Individual 10 had a high forehead, mild bitemporal narrowing, downslanting palpebral fissures, and anteverted ears. Individual 20 had slight micrognathia and telecanthus. Individual 25 had downslanting palpebral fissures. Photos showing hand anomalies of individual 1 (p.Met1?) were also obtained.

Involvement of other systems

Visual problems were relatively common (9/25, 36%) and included strabismus, high myopia, and congenital cataracts. Other anomalies were diagnosed in the gastrointestinal (6/24, 25%), genitourinary (3/24, 13%), and renal (2/24, 8%) systems. These included dysmotility, bilateral inguinal hernias, low levels of creatinine in urine, elevated sulfites in urine, polycystic kidney, and hypospadias. Two individuals also had cardiac anomalies (2/24, 8%), including a ventricular septal defect (VSD) that closed spontaneously, a smaller muscular VSD in the apical trabecular septum, a persistent foramen ovale which also closed spontaneously, left aortic arch with common origin of right brachiocephalic and left common carotid arteries (left bovine arch), as well as mild dilatation of aorta with normal tricuspid aortic valve, also a sign of CTD. These observations indicate that the MSL2-related disorder may be linked to a variety of multisystemic conditions besides the syndrome-defining neurodevelopmental impairment.

Variant description

A total of 21 different *MSL2* variants were identified among our 25 individuals (Figure 2A). There were 12 different frameshift variants and 5 nonsense variants in 17 individuals. 15 of these variants are predicted

Figure 1. Photos of individuals with the MSL2-related disorder
(A) Individual 4 (p.Gln171*). Note prognathism as well as broad and high forehead.
(B) Individual 7 (p.Gln353*). Note high forehead and retrognathia.
(C) Individual 10 (p.Arg39Leufs*34). Note high forehead and prominent ridge of the nose.
(D) Individual 21 (p.Leu266Valfs*5). Note mild telecanthus and retrognathia.
(E) Individual 25 (complex variant). Note downslanting palpebral fissures.
(F) Individual 1 (p.Met1?) fingers. Note thumb and 5th finger leading to a Beighton score of 4. Note mild malar hypoplasia in all facial photos.

to lead to truncated proteins by escaping from the nonsense-mediated decay mechanism, as they are located in the last exon, or within the last 50 nucleotides of the penultimate exon.[42] c.105dup (p.Pro36Alafs*36) (individual 9) and p.Arg39Leufs*34 (individual 10) are located in exon 1 causing immediate-early stop codons and predicted to undergo nonsense-mediated decay or lead to non-functional peptides lacking any intact annotated domains and thus targeted for degradation.

All variants in our cohort were *de novo*, with the exception of p.Glu229Glyfs*4 (individual 14), inherited from the mother, who had ID, as did her other two children, and p.Met1? (individual 1), whose inheritance was unknown. The latter was analyzed as a start-loss variant, given that most missense variants affecting the first methionine usually lead to a significant loss of protein expression.[45] However, the effect of this variant is more complicated to predict since additional transcript variants such as MSL2 transcript variant 2 (GenBank: NM_001145417.2) that lacks the sequence translating to the 74th amino acid (aa) from the N-terminal of the main *MSL2* transcript variant 1 (GenBank: NM_018133.4) or those carrying an alternative start codon (AUG) could still be produced. Notably, there are

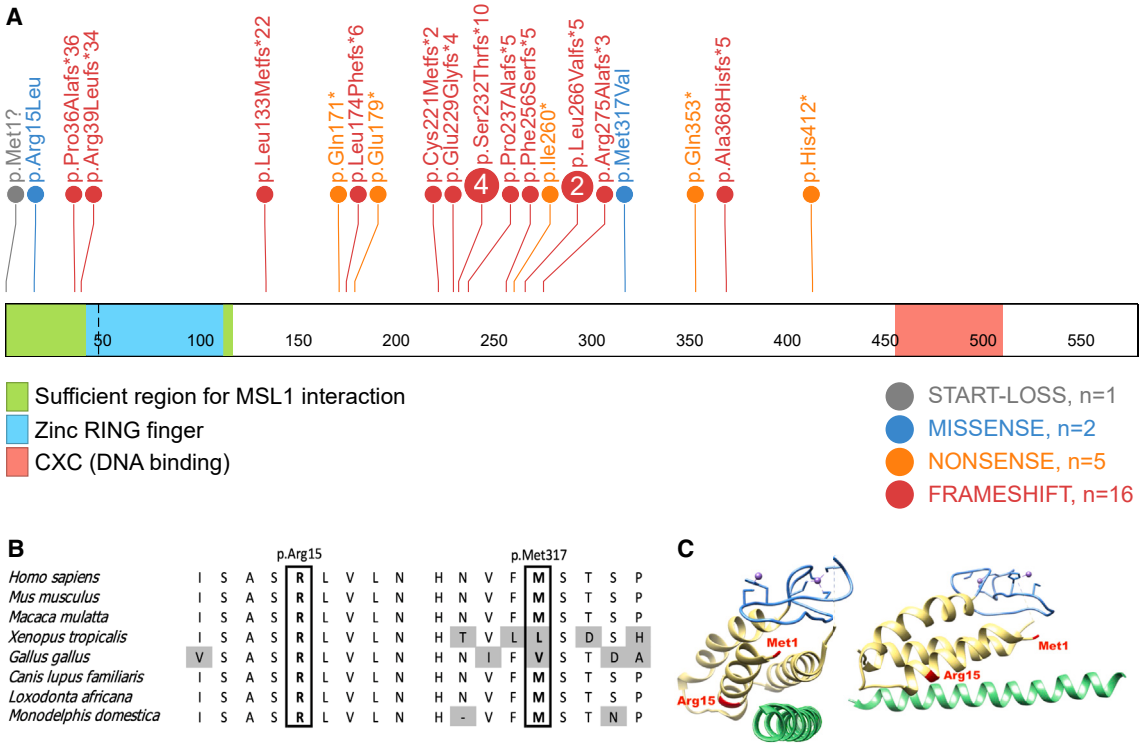


Figure 2. Variant location along MSL2

(A) The variants studied in our cohort are represented on the MSL2 protein, excluding the complex variant of individual 25. The active regions are named and highlighted in different colors according to the legend. Variant types are also indicated, and if the same one was shared by multiple individuals, their total number is in a circle. Illustration adapted from ProteinPaint[43] and domain sequences obtained from Pfam.[44]

(B) Amino acid conservation of the two missense variants (individuals 2 and 3).

(C) Homology model of human MSL2 as experimentally crystalized (PDB: 4B86),[10] represented by UCSF Chimera, from two angles. MSL2 is in yellow, the missense variants of our cohort in red, the zinc finger domain in blue, the two zinc atoms in fuchsia, and the MSL1 protein in green. MSL2 is illustrated from the 1st to the 115th residues, so Met317 is not represented.

no loss-of-function variants in the gnomAD v3.1.2 non-neuro cohort, which suggests human neurodevelopment is likely intolerant to losing one allele of *MSL2*. Additionally, the pLI and LOEUF scores of 1 and 0.17, respectively (gnomAD v.4.0.0), further suggest MSL2 loss-of-function intolerance.

One complex variant in our cohort (individual 25), affecting the last exon of *MSL2* at aa Thr244, involved a 4-bp deletion where a cytosine and a 180 nt sequence from the neighboring intron are inserted. The variant can be written g.[136152145_136152149delinsC; 136152802_136152981inv] (GenBank: NC_000003.12). The phenotype was not only similar but also typical of the rest of the cohort: neurodevelopmental impairment, behavioral anomalies, dysmorphisms, and CTD signs.

The two missense variants in our cohort, c.44G>T (p.Arg15Leu) (individual 2) and p.Met317Val (individual 3), had CADD scores of 31 and 17.15, respectively. This means that p.Met317Val is among the 10% most deleterious possible substitutions in the human genome ($10 < \text{CADD} < 20$), while p.Arg15Leu is among the 1% ($\text{CADD} \geq 20$).[46] p.Arg15Leu is in a gene region highly conserved across eukaryotes (see Figure 2B). Some evolutionary variation was noted in and around p.Met317Val, with valine being the normal 317th residue encoded by the chicken *msl2*. In theory, since this variant leads to a functional protein in chickens, p.Met317Val would be predicted to be less likely to have a significant impact on MSL2 function, which was not necessarily corroborated by the phenotype. Individual 3 presented with a severe global DD, complex epilepsy, MRI anomalies, and several dysmorphisms and multi-systemic anomalies (cardiac, gastrointestinal, genitourinary, and visual). These observations suggest that residue conservation in other species is not necessarily proportional to the clinical impact of *MSL2* missense variants. Alternatively, it is possible that the variant is benign and that some of the features of the individual and the abnormal functional results are due to another cause, such as stochasticity or a *de novo COL4A1* variant in this individual (MIM: 120130) (c.2008G>A [GenBank: NM_001845.6] [p.Gly670Arg]).

Both missense variants were associated with MRI abnormalities and multi-systemic involvement, which might indicate a correlation. The p.Met1? phenotype was restricted to neurodevelopmental and behavioral problems, dysmorphisms including CTD signs, sleep issues, and visual anomalies; whereas the variants predicted to undergo nonsense-mediated decay and therefore likely to exhibit reduced MSL2 levels (p.Pro36Alafs*36 and p.Arg39Leufs*34)

led to a more severe phenotype as the individuals presented with early-onset epilepsy. These observations may suggest that the start-loss variant has a different impact than do the variants predicted to lead to reduced MSL2 levels, the latter being associated with seizures.

Comparisons based on sex

The majority of our individuals were males; only 9/25 (36%) were females. Despite this asymmetry and keeping in mind that sex-based tendencies in our cohort would not necessarily be representative of the MSL2-related disorder, we searched for sex-specific phenotypic differences (Table 2). The vast majority of males presented with dysmorphisms, while only around half of the females did (88% vs. 56%). CTD signs were twice as common in males (63% vs. 33%). Vision problems were also twice as common (44% vs. 22%), while gastrointestinal ones (40% vs. 0%) were seen only in males. Sleep issues were also more common in males (38% vs. 11%).

Females were more than twice as likely to have an OFC abnormality (43% vs. 18%), and speech delay was slightly more pronounced (86% vs. 75%). Behavioral anomalies were equally seen in both sexes (~73%), with the particularity that psychiatric concerns (38% vs. 23%) and attention deficit (63% vs. 33%) were more common in females. The most relevant observation is that females seemed to be more affected in the neuromotor aspect since they all presented with motor delay (100% vs. 81%) and, especially, coordination issues (100% vs. 38%).

Comparisons based on variant position along MSL2

One missense variant, p.Arg15Leu (individual 2), was located in the characterized region interacting with MSL1 (see Figure 2C). Interaction between MSL2 and scaffolding protein MSL1 is critical for its incorporation into the MSL complex.[10] The p.Arg15Leu missense variant would potentially disrupt the protein domain conformation, hindering MSL1-MSL2 binding. Interestingly, individual 2 did not necessarily have a more severe nor diverse phenotype compared to the rest of the cohort (ID, speech delay, mild hypotonia, behavioral problems, dysmorphisms, one MRI anomaly, constipation). The frameshift variants p.Pro36Alafs*36 (individual 9) and p.Arg39Leufs*34 (individual 10), located in the same characterized region, would be predicted to lead not only to the premature truncation of this MSL1-interacting domain, but also to the loss of the zinc-binding site of the RING finger domain, further diminishing the possibility of an efficient interaction with MSL1.[47] They were both associated with seizures. Overall, no generalization could be observed with respect to the variants located directly in the N-terminal MSL1-binding portions of the MSL2.

When comparing variants producing a loss and/or disruption of the MSL1-interacting region (p.Met1?, p.Arg15Leu, p.Pro36Alafs*36, and p.Arg39Leufs*34) to the rest (see Figure 2A), which are predicted to lead to an isolated loss of the DNA-binding CXC domain, a few observations could be highlighted (see Table 2). Behavioral abnormalities (100% vs. 67%), including autistic features (100% vs. 47%) and sleep issues (50% vs. 25%), were around twice more common in individuals with variants affecting the MSL1-interacting region. Hypotonia (100% vs. 75%), seizures (50% vs. 35%), gastrointestinal anomalies (50% vs. 21%), and dysmorphisms (100% vs. 68%) although not CTD signs (both 50%) were also more common. However, renal (0% vs. 11%), cardiac (0% vs. 11%), and genitourinary (0% vs. 16%) anomalies were seen only with an isolated loss of the CXC domain, and vision problems were more common in this subgroup too (25% vs. 40%).

Same-variant comparison

Four individuals (15–18) shared the p.Ser232Thrfs*10 variant. The phenotypes were very similar, associated with the absence of seizures, brain imaging anomalies, or multisystemic involvement, with the exception of cardiac anomalies seen in individual 16. When compared in terms of sex-specific differences, one does not notice any of the particularities previously suggested in that section, which could indicate that variant-specific features are more pronounced than sex-specific ones. It also reiterates the importance of corroborating our cohort-limited observations. Finally, there were two individuals (21 and 22) who shared the p.Leu266Valfs*5 variant. Their phenotypes were also similar: ID/DD, hypotonia, behavioral problems, the absence of seizures, and limited multi-systemic involvement.

Phenotypic clusters

Human Phenotype Ontology (HPO) analysis was performed using the OntologyX package in R[27] (see Table S4 for HPO annotations assigned to each individual). No appropriate clustering was identifiable after gap statistic curve analysis (Figure 3). This suggests that, overall, there are no specific clinical subgroups of the MSL2-related disorder and that individuals cannot be clustered based on their genotypes.

Establishment of a specific episignature for the MSL2-related disorder

DNA methylation episignatures have been used as a reliable biomarker for the molecular diagnosis of rare genetic disorders in a clinical setting and for the reclassification of variants of unknown clinical significance (VUSs).[19] Global DNA methylation profiles were analyzed in peripheral blood samples of 6 individuals with *MSL2* variants (see Table S6 for individual identification, age, sex, and ethnicity), revealing a robust, sensitive, and specific episignature. Hierarchical clustering and MDS models were performed using the 239 probes selected in the three-step process outlined in the material and methods section (for detailed information about the selected probes, refer to Table S5). These models demonstrated a clear separation between the individuals with *MSL2* variants and control groups, further indicating the robustness of the episignature (Figures 4A and 4B). Methylation levels (β values) at these probes for individuals with *MSL2* variants and control individuals can be found in Table S6.

Table 2. Clinical features frequency according to variant coding impact, variant position along MSL2 protein, and individual sex, as well as comparison with MRXSBA

	Start-loss	Missense	Frameshift/nonsense		Complex	MSL1-interacting region loss/disruption		Only DNA-interacting domain loss		M		F		Total cohort		MRXSBA cohorts[14,15]	
	= 1	*n* = 2	*n* = 21		*n* = 1	*n* = 4		*n* = 20		*n* = 16		*n* = 9		*n* affected/ *n* assessed	%	*n* affected/ *n* assessed	%
Perinatal complications	+	1/2	43	(9/21)	–	50	(2/4)	45	(9/20)	50	(8/16)	33	(3/9)	11/25	44	N/A	N/A
OFC anomaly	+	1/2	21	(3/14)	–	25	(1/4)	31	(4/13)	18	(2/11)	43	(3/7)	5/18	28	≥16/33	≥48
ID/DD	+	2/2	100	(21/21)	+	100	(4/4)	100	(20/20)	100	(16/16)	100	(9/9)	25/25	100	37/37	100
Speech delay	+	2/2	74	(14/19)	+	75	(3/4)	78	(14/18)	75	(12/16)	86	(6/7)	18/23	78	37/38	100
Motor delay	+	1/2	90	(18/20)	+	75	(3/4)	89	(17/19)	81	(13/16)	100	(8/8)	21/24	88	36/37	97
Hypotonia	+	2/2	75	(15/20)	–	100	(3/3)	75	(15/20)	80	(12/15)	67	(6/9)	18/24	75	33/37	89
Feeding difficulties	–	1/2	39	(7/18)	–	33	(1/3)	39	(7/18)	36	(5/14)	38	(3/8)	8/22	36	≥17/37	≥46
Coordination issues	+	0/1	56	(9/16)	+	50	(2/4)	57	(8/14)	38	(5/13)	100	(6/6)	11/19	58	N/A	N/A
Breathing assistance	–	0/2	5	(1/21)	–	0	(0/4)	5	(1/20)	0	(0/16)	11	(1/9)	1/25	4	1/37	3
Behavioral abnormalities	+	1/1	68	(13/19)	+	100	(3/3)	67	(12/18)	71	(10/14)	75	(6/8)	16/22	73	20/31	65
Autistic features	+	1/1	50	(9/18)	–	100	(3/3)	47	(8/17)	54	(7/13)	50	(4/8)	11/21	52	10/20	50
Attention deficit	+	0/0	39	(7/18)	+	50	(1/2)	41	(7/17)	33	(4/12)	63	(5/8)	9/20	45	4/19	21
Psychiatric concerns	+	0/1	28	(5/18)	–	33	(1/3)	29	(5/17)	23	(3/13)	38	(3/8)	6/21	29	N/A	N/A
Sleep issues	+	2/2	19	(4/21)	–	50	(2/4)	25	(5/20)	38	(6/16)	11	(1/9)	7/25	28	N/A	N/A
Seizures	–	1/2	38	(8/21)	–	50	(2/4)	35	(7/20)	38	(6/16)	33	(3/9)	9/25	36	5/35	14
MRI abnormalities	–	2/2	39	(7/18)	NA	50	(2/4)	41	(7/17)	44	(7/16)	40	(2/5)	9/21	43	20/35	57
EEG abnormalities	–	0/2	28	(5/18)	–	33	(1/3)	22	(4/18)	21	(3/14)	25	(2/8)	5/22	23	N/A	N/A
Dysmorphisms	+	2/2	71	(15/21)	+	100	(4/4)	70	(14/20)	88	(14/16)	56	(5/9)	19/25	76	23/24	96
Deep-set eyes	+	2/2	40	(8/20)	–	75	(3/4)	42	(8/19)	60	(9/15)	22	(2/9)	11/24	46	≥2/37	≥5
CTD signs	+	0/2	52	(11/21)	+	50	(2/4)	50	(10/20)	63	(10/16)	33	(3/9)	13/25	52	N/A	N/A

(*Continued on next page*)

Table 2. *Continued*

	Start-loss	Missense	Frameshift/nonsense		Complex	MSL1-interacting region loss/disruption		Only DNA-interacting domain loss		M		F		Total cohort		MRXSBA cohorts[14,15]	
	= 1	*n* = 2	*n* = 21		*n* = 1	*n* = 4		*n* = 20		*n* = 16		*n* = 9		*n* affected/ *n* assessed	%	*n* affected/ *n* assessed	%
Visual anomalies	+	1/2	33	(7/21)	–	25	(1/4)	40	(8/20)	44	(7/16)	22	(2/9)	9/25	36	12/22	55
Gastrointestinal anomalies	–	2/2	20	(4/20)	–	50	(2/4)	21	(4/19)	40	(6/15)	0	(0/9)	6/24	25	28/35	80
Renal anomalies	–	0/2	5	(1/20)	–	0	(0/4)	11	(2/19)	7	(1/15)	11	(1/9)	1/24	4	7/35	20
Cardiac anomalies	–	1/2	5	(1/20)	–	0	(0/4)	11	(2/19)	13	(2/15)	0	(0/9)	2/24	8	7/34	21
Genitourinary anomalies	–	1/2	10	(2/20)	–	0	(0/4)	16	(3/19)	13	(2/15)	11	(1/9)	3/24	13	N/A	N/A

M, male; F, female; OFC, occipital frontal circumference; ID, intellectual disability; DD, developmental impairment; CTD, connective tissue disease; MRXSBA, MSL3-related X-linked Syndrome Basilicata-Akhtar, or Basilicata-Akhtar syndrome; N/A, not available.

Classification model construction

To ensure accurate classification of individuals with NDD-associated *MSL2* variants with high specificity, an SVM was constructed using the 239 selected probes. This was achieved by training the SVM model on 6 samples with *MSL2* variants against matched control individuals, as well as 75% of other controls and samples from other disorders with established episignatures from the EpiSign v3 clinical classifier within the EKD, as described in the materials and methods section. The remaining 25% of samples were used for testing the model. With the exception of 2 samples, all testing samples received MVP scores close to 0, indicating the high specificity of the classifier (Figure 4C).

Differentially methylated regions and gene set enrichment analysis

Differentially methylated regions (DMRs) were identified as those containing a minimum of 3 CpGs with a mean methylation difference >0.05 between case and control groups and a Fisher's multiple comparison p value < 0.01 within 1 kb. Based on these criteria, 8 DMRs were identified and are listed in Table S7. GO enrichment analysis was performed for these DMRs; however, no enriched terms were detected in those regions with an adjusted p value < 0.01.

Modeling the effect of *MSL2* variants *in vitro*

Generation of patient-derived iPSCs

To examine the molecular phenotypes of *MSL2* variants, we have derived *MSL2* patient-specific iPSCs from three of our individuals with distinct traits representative of the cohort and complementary for comparative analyses: (1) individual 16, a male carrying the most frequent truncating variant p.Ser232Thrfs*10 (ST), (2) individual 6, a female with a truncating variant p.Ile260* (IX), and (3) individual 3, a male with a missense variant p.Met317Val (MV) (Figure 5A). Two distinct clones were picked for each proband, and the generation of iPSCs confirmed by the expression of pluripotency markers TRA-1-60, SOX2, OCT4, and SSEA4 (Figure 5B). The genetic stability of these clones was also confirmed using a qPCR-based assay for recurrent karyotypic abnormalities in iPSCs (Figure S1A).

Characterization of MSL2 expression and linked epigenetic marks in patient-derived iPSCs

In the patient-derived iPSC clones, we sought to determine the effect of variants on MSL2 levels and stability (Figure 5C). Full-length MSL2 was detectable in all iPSC lines, including those with truncating variants (ST and IX) (Figures 5C and 5D, nuclear soluble fraction), indicating that full-length MSL2 is still produced from the intact allele. However, there is an appreciable decrease in the overall levels of MSL2 in ST and IX iPSCs relative to controls (C), indicating that the intact allele does not produce more MSL2 to compensate for the variant. We could not detect the truncated variants with the MSL2 antibody since the antibody is generated against an antigen with the amino acid (aa) sequence 211–353 (Figure 5A). The level of full-length MSL2 in iPSCs with the missense variant (MV)

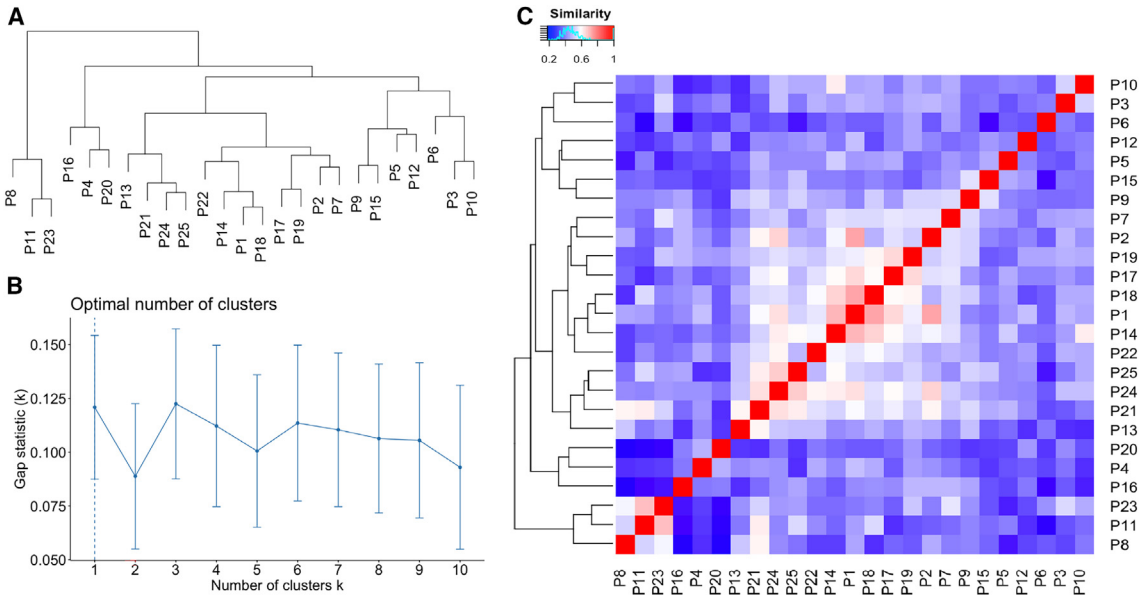


Figure 3. HPO analysis of the MSL2-related disorder cohort
(A) Hierarchical clustering with Ward method of the phenotypes of our 25 probands.
(B) Gap statistic curve serving to determine the optimal number of clusters (vertical dashed line); error bars represent the uncertainty in the estimation of the Gap statistic parameter (k).
(C) Quantitative phenotypic similarity among individuals based on HAC analysis.

was similar to the control (Figure 5D), suggesting that the missense variant, which is expected to be the same size as the intact MSL2, is stable and detected by the MSL2 antibody.

Interestingly, global levels of H4K16ac, the main epigenetic mark deposited by the MSL complex, were not altered in any of the patient-derived iPSCs compared to the control (Figure 5D, nuclear insoluble fraction). We also saw that the levels of MOF, the catalytic subunit of the MSL complex responsible for H4K16ac, were the same across our iPSC samples (Figure 5D, nuclear soluble fraction), indicating that MSL2 variants did not alter the protein stability of another member of MSL complex. Together, our results suggest a distinct epigenetic mechanism underlying the MSL2-related disorder compared to the MRXSBA.

MSL2 target gene expression profiles in patient-derived iPSCs and differentiated germ layers

MSL2 regulates the expression of X-linked (e.g., *Zfp185*, *Bex2*, *Tsix*) as well as autosomal (*Bscl2*, *Gng3*) genes in mouse embryonic stem cells.[48] Furthermore, MSL2-mediated regulation of gene expression has been shown to be dynamic in more differentiated cellular states compared to naive pluripotency.[48] To investigate the effect of MSL2 variants on MSL2-dependent gene regulation during early lineage differentiation, we differentiated patient-derived iPSCs into the three germ layers: ectoderm, mesoderm, and endoderm (Figure 6A, top). All clones from patient-derived iPSCs successfully differentiated into the three germ layers assessed by the expression of germ layer-specific marker genes (*PAX6* and *NES* for ectoderm, *TBXT* and *CXCR4* for mesoderm, and *SOX17* and *FOXA2* for endoderm) comparable to the control line (Figure 6A).

We next tested whether human orthologs of the putative mouse MSL2 target genes (*ZNF185* [MIM: 300381], *BEX2* [MIM: 300691], *TSIX* [MIM: 300181], *BSCL2* [MIM: 606158], *GNG3* [MIM: 608941]) are dysregulated in the MSL2 patient-derived iPSCs or differentiated early germ layers. We observed significant downregulation in the expression of these target genes in patient-derived iPSCs compared to the control (Figure 6B), suggesting that both copies of intact MSL2 are necessary for the expression of these genes in humans. The expression of *MSL2* and *MSL1* was also downregulated in patient-derived iPSCs (Figure S2A) which may have contributed to the downregulation of target genes.

When differentiated into three germ layers, significant differences in target gene expression persisted in patient-derived cells compared to control (Figure 6B), while gene-, individual-, and lineage-specific dysregulations emerged, suggesting varying vulnerabilities to *MSL2* variants. For example, truncating variants (ST and IX) showed more significant downregulation in target gene expression, especially after differentiation. The expression of each target gene was also different among the three germ layers (Figure S2B), indicating that dynamic regulation of gene expression is necessary during differentiation. The expression of *MSL2* and other complex members *MOF* (*KAT8*) and *MSL1* was similar across the iPSC-derived germ layers compared to control, except for increased *MSL2* expression in IX and MV mesoderm (Figure S2B). This may explain the stable, and occasionally increased, target gene expression in IX and MV mesoderm lineage, highlighting potential sex-, variant-, and lineage-specific effects.

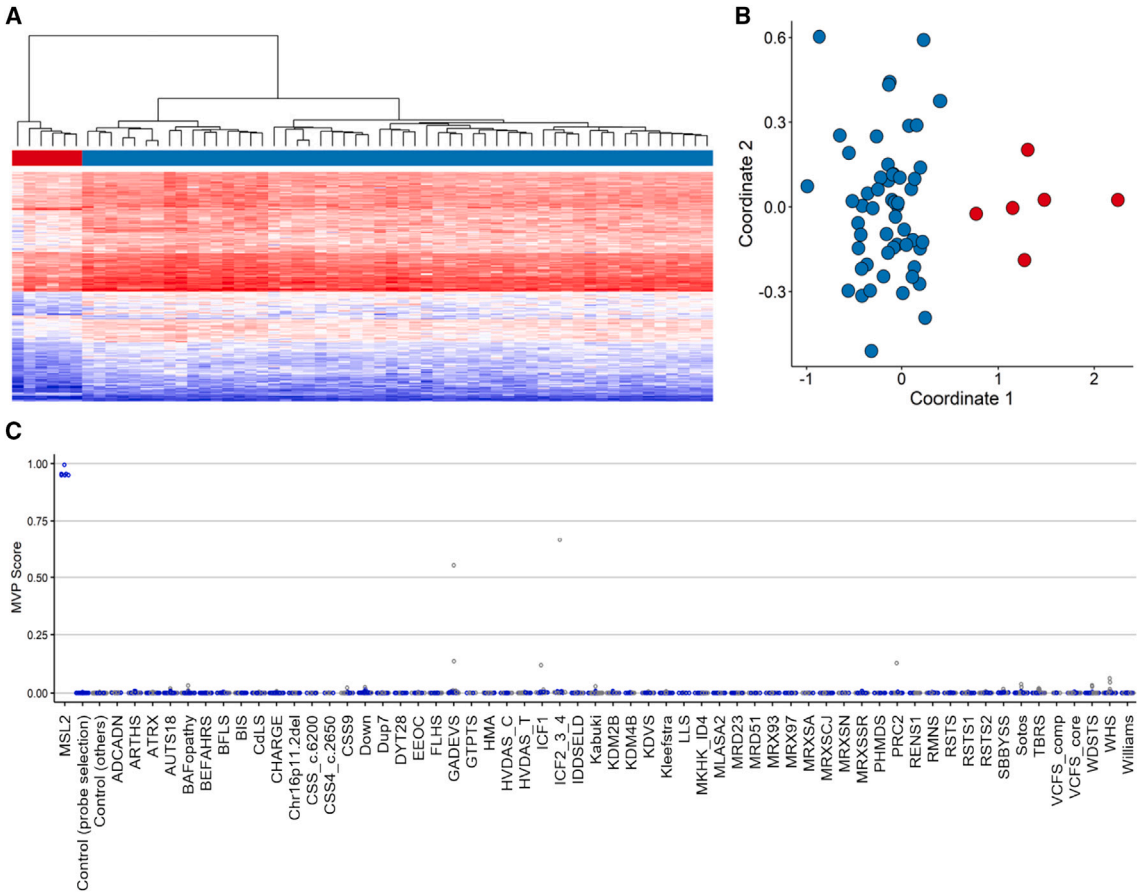


Figure 4. Verification of the identified episignature and MVP scores generated by the SVM classifier

(A) The heatmap depicts hierarchical clustering, where the rows represent the 239 selected probes and the columns represent the affected and control individuals. The case and control samples are indicated by the colors red and blue, respectively. The color gradient ranges from blue (representing no methylation or 0) to red (representing full methylation or 1).

(B) MDS plot, where the *MSL2* case samples are represented by red circles and the control individuals by blue circles.

Both plots (A and B) clearly demonstrate a distinct separation between the case and control groups.

(C) The SVM-generated MVP scores ranging from 0 to 1, indicating similarity to the *MSL2* episignature, with high scores indicating greater similarity. Blue circles represent training samples, while gray circles represent testing samples. The low MVP score of testing control samples and testing samples from other rare genetic disorders, except for two individuals, suggests high specificity of the classifier. ADCADN, cerebellar ataxia, deafness, and narcolepsy, autosomal dominant; ARTHS, Arboleda-Tham syndrome; ATRX, X-linked alpha-thalassemia/impaired intellectual development syndrome; AUTS18, autism, susceptibility to, 18; BAFopathy, Coffin-Siris syndrome-1,2,3,4 (CSS1,2,3,4) and Nicolaides-Baraitser syndrome (NCBRS); BEFAHRS, Beck-Fahrner syndrome; BFLS, Borjeson-Forssman-Lehmann syndrome; BIS, blepharophimosis intellectual disability SMARCA2 syndrome; CdLS, Cornelia de Lange syndromes 1,2,3,4 (CdLS1,2,3,4); CHARGE, CHARGE syndrome; Chr16p11.2del, chr16p11.2 deletion syndrome, 593-KB; CSS_c.6200, Coffin-Siris syndrome-1,2 (CSS1,2); CSS4_c.2656, Coffin-Siris syndrome-4 (CSS4); Down, Down syndrome; Dup7, Williams-Beuren duplication syndrome (chr7q11.23 duplication syndrome); DYT28, dystonia 28, childhood-onset; EEOC, epileptic encephalopathy, childhood-onset; FLHS, Floating Harbor syndrome; GADEVS, Gabriele-de Vries syndrome; GTPTS, genitopatellar syndrome; HMA, Hunter McAlpine craniosynostosis syndrome; HVDAS_C and HVDAS_T, Helsmoortel-van der Aa syndrome; ICF_1, immunodeficiency-centromeric instability-facial anomalies syndrome 1 (ICF1); ICF_2_3_4, immunodeficiency-centromeric instability-facial anomalies syndromes 2,3,4 (ICF2,3,4); IDDSELD, intellectual developmental disorder with seizures and language delay; Kabuki, Kabuki syndrome 1,2 (KABUK1,2); KDVS, Koolen de Vreis syndrome; Kleefstra, Kleefstra syndrome 1 (KLEFS1); LLS, Luscan-Lumish syndrome; MKHK_ID4, Menke-Hennekam syndromes 1,2 (MKHK1,2); MLASA2, myopathy, lactic acidosis, and sideroblastic anemia 2; MRD23, intellectual developmental disorder, autosomal dominant 23; MRD51, intellectual developmental disorder, autosomal dominant 51; MRX93, intellectual developmental disorder, X-linked 93; MRX97, intellectual developmental disorder, X-linked 97; MRXSA, intellectual developmental disorder, X-linked, syndromic, Armfield type; MRXSCJ, intellectual developmental disorder, X-linked, syndromic, Claes-Jensen type; MRXSN, intellectual developmental disorder, X-linked syndromic, Nascimento-type; MRXSSR, intellectual developmental disorder, X-linked, Snyder-Robinson type; PHMDS, Phelan-McDermid syndrome; PRC2, Cohen-Gibson syndrome (COGIS) and Weaver syndrome (WVS); RENS1, Renpenning syndrome; RMNS, Rahman syndrome; RSTS, Rubinstein-Taybi syndrome; SBBYSS, Ohdo syndrome, SBBYSS variant; Sotos, Sotos syndrome 1; TBRS, Tatton-Brown-Rahman syndrome; VCFS, velocardiofacial syndrome; WDSTS, Wiedemann-Steiner syndrome; WHS, Wolf-Hirschhorn syndrome; Williams, Williams-Beuren deletion syndrome.

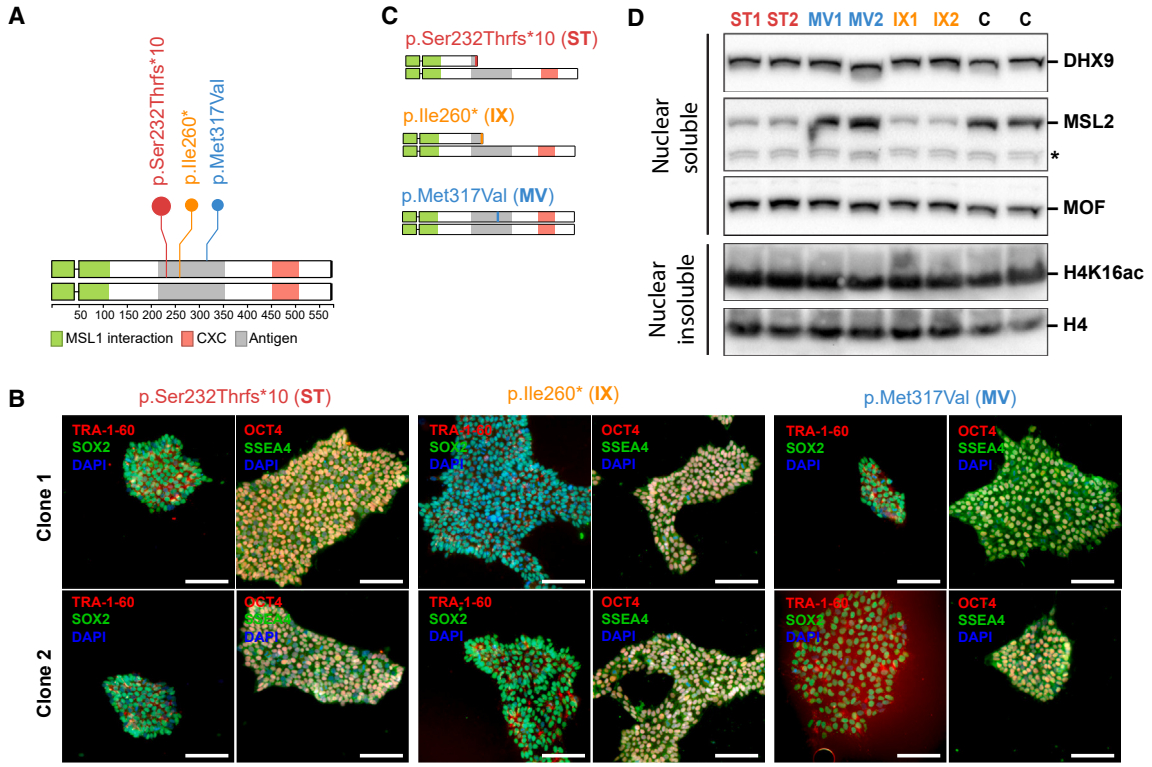


Figure 5. Generation and molecular characterization of induced pluripotent stem cell (iPSC) lines from individuals with the MSL2-related disorder

(A) Schematic highlighting the *MSL2* variants of the three probands whose dermal fibroblasts have been reprogrammed into iPSCs.

(B) Immunofluorescence images showing the expression of pluripotency markers (TRA-1060, SOX2, OCT4, and SSEA4) in patient-derived iPSC clones, two clones per individual. Scale bars, 100 μm.

(C) Schematic depicting the predicted outcome of the variants on MSL2. Heterozygous *de novo* frameshift (ST) or nonsense (IX) variants lead to the generation of early stop codons leading to truncation of MSL2 whereas the intact allele would produce the full-length MSL2. The missense variant (MV) would not alter the size of the translated MSL2 with the variant.

(D) Immunoblotting on nuclear soluble extracts from patient-derived iPSC clones showed lower levels of full-length MSL2 in ST and IX compared to a control (C) iPSC line. Note the comparable levels of detection of MSL2 in MV to the C replicates. The asterisk shows unspecific bands on the MSL2 blot. MOF levels are similar across patient-derived clones compared to Cs. DHX9 is used as the loading control. Immunoblotting on nuclear insoluble fractions shows similar levels of H4K16ac in all affected-individual-derived clones compared to the control. H4 is used as the loading control for the insoluble fractions.

Discussion

In the current study, we identified 25 individuals with variants in *MSL2*. Genotype-phenotype analysis linked these variants to ID, global DD, hypotonia, and other motor limitations such as coordination problems, feeding difficulties, and gait disturbance, defining a distinctive syndrome. Dysmorphisms as well as behavioral and psychiatric conditions including ASD were frequently observed. To a lesser extent, seizures, CTD signs, sleep issues, vision problems, and other organ anomalies were also associated with *MSL2* variants. With the invaluable participation of the individuals, a sensitive and specific DNA methylation episignature has been established as a molecular biomarker. Moreover, the generation of *MSL2* patient-derived iPSCs allowed us to molecularly characterize three *MSL2* variants proving invaluable in further disease modeling and research.

Deciphering the functional and phenotypic outcomes of *MSL2* variants is challenged by the multifunctional nature of the protein. MSL2 has two evolutionarily conserved and functional domains: a RING domain which is both responsible for ubiquitin ligase activity and is embedded within the portion of the protein required for MSL1 interaction as well as a CXC domain capable of DNA targeting.[10,47,49,50]

In our cohort, the four members whose variants were predicted to lead to a monoallelic loss and/or disruption of the MSL1-interacting region (individuals 1, 2, 8, 9) were empirically more likely to have behavioral abnormalities and autistic features, seizures, sleep issues, and dysmorphisms, but less multi-systemic involvement, excepting gastrointestinal anomalies (Table 2). Two of these variants were predicted to undergo nonsense-mediated decay and were both associated with seizures.

Same-variant phenotypes were highly similar, regardless of individual sex. While warning that females were underrepresented in our cohort and that sex-based tendencies would not necessarily be representative of the MSL2-related disorder, we still attempted a comparison. Our observations suggested that, overall, dysmorphisms, CTD

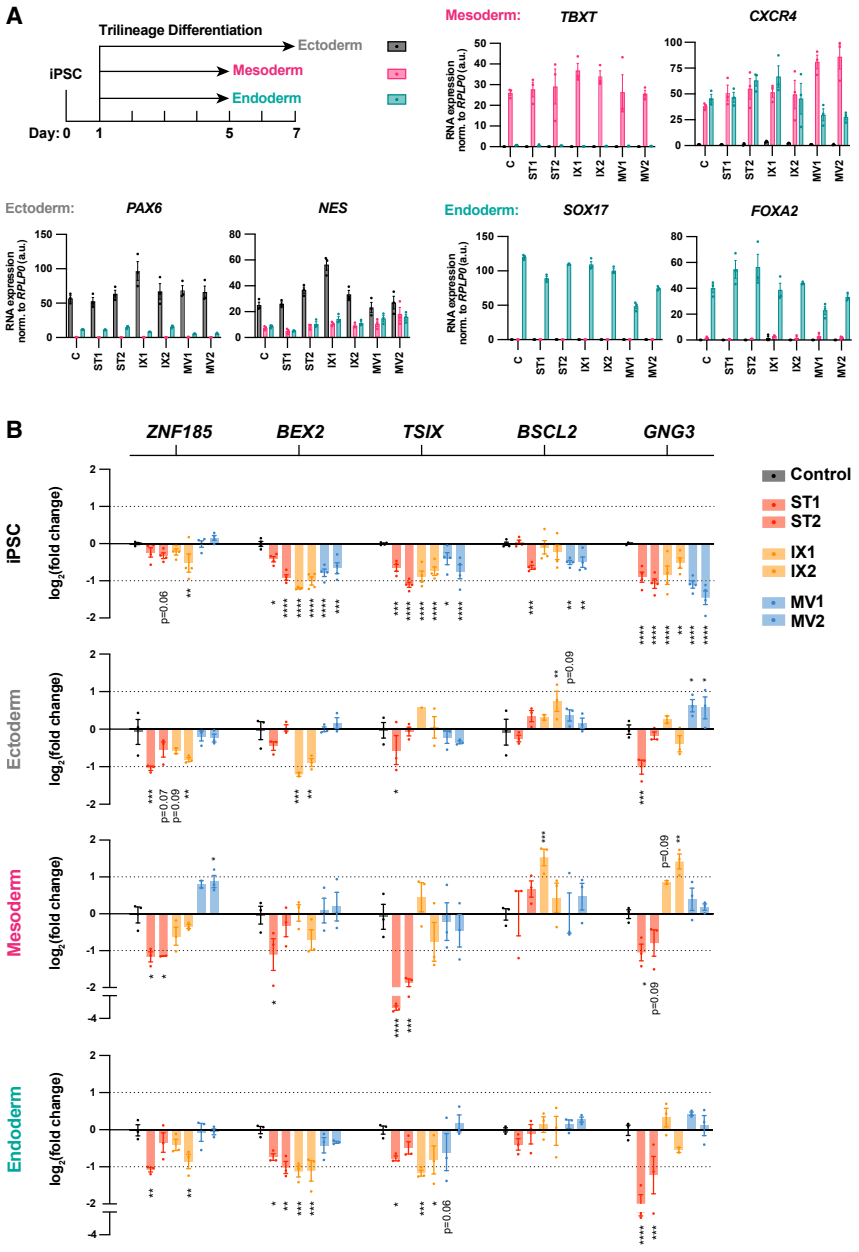


Figure 6. Analysis of the expression of MSL2 target genes in affected-individual-derived iPSCs and early germ layers

(A) Top, schematic of the differentiation timeline of iPSCs into all three germ layers. Bottom, RT-qPCR results showing the relative RNA expression (normalized to *RPLP0*, arbitrary units [a.u.]) of lineage-specific marker genes for ectoderm (*PAX6*, *NES*), mesoderm (*TBXT*, *CXCR4*), and endoderm (*SOX17*, *FOXA2*). Note that the expression of each marker gene is enriched for its respective lineage across all affected-individual-derived clones compared to the control indicating successful differentiation onto all germ layers. $n = 3$ per clone per differentiation.

(B) RT-qPCR results showing the log2(fold change) of relative RNA expression of putative mammalian targets of MSL2 (*ZNF185*, *BEX2*, *TSIX*, *BSCL2*, *GNG3*) in patient-derived iPSCs (bar graphs in respective colors assigned to individuals) compared to control (gray). Expression levels are compared in iPSCs (top, $n = 4$ per clone) as well as upon trilineage differentiation to ectoderm, mesoderm, and endoderm ($n = 3$ per clone per lineage). Two-way ANOVA with post-hoc Fisher's LSD test, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$. Data are represented as the mean ± SEM.

signs, gastrointestinal anomalies, and vision and sleep problems were more common in males. OFC anomalies, psychiatric concerns, and, most importantly, neuromotor impairment (specifically coordination issues and motor delay) were more common in females. Further research on bigger, more equally distributed cohorts remains imperative.

A study of more than 2,500 simplex families cited *MSL2 de novo* variants as possible contributors to ASD, although the gene did not reach statistical significance.[51] It was subsequently confirmed as a candidate with recurrent *de novo* variants in an ASD cohort.[16] In 2021, a case study of an individual whose diagnosis was facilitated by genotypic matchmaking through the electronic health record was published. The 26-year-old woman with a *de novo* p.Ser232Thrfs*10 variant presented with mild ID, ASD, OCD, high myopia, and joint hypermobility. These clinical traits were all seen among our four individuals sharing that same variant although these individuals did not have identical phenotypes. The 26-year-old woman's medical team searched for *MSL2* matches and found two individuals with similar phenotypes: first, a 15-year-old male (c.73C>T [p.Pro25Ser], *de novo*) with the same clinical features but a formal hypermobile Ehlers-Danlos syndrome (EDS) diagnosis for the joint hypermobility; second, a 13-year-old female (p.Thr217Aspfs*2, codon coordinates unavailable) who had DI, ASD, OCD, attention deficit disorder, visual and language processing disorder, high myopia, and hypermobile EDS.[18] These phenotypes corroborate the common traits of autistic features and behavioral abnormalities, visual anomalies, and CTD signs seen in our cohort. The missense p.Arg15Leu (individual 2) and p.Pro25Ser (aforementioned 15-year-old male[18]) variants are located in the MSL1 interaction domain. These individuals had similar phenotypes, specifically in terms of autistic features and behavioral abnormalities, which might suggest a stronger correlation between those traits and missense variants in the MSL1 interaction domain. Finally, a study from 2021 reported an association between NDDs and *MSL2* based on two *de novo* protein-truncating variants (p.Leu192Valfs*3 and p.Ser486Ilefs*11, codon coordinates unavailable) and

on frequent co-expression/interaction of MSL2 with known candidate genes.[17] Our study provides a substantial cohort with *de novo* protein-truncating variants exhibiting NDD-associated phenotypes (ASD, behavioral/psychiatric problems, speech disorders, etc.), suggesting the previously identified *MSL2* variants are linked to the MSL2-related disorder. We also provide evidence that start-loss or missense MSL2 variants may be pathogenic leading to the MSL2-related disorder phenotype. Finally, we propose that variants predicted to lead to partial MSL2 loss are associated with seizures.

By analyzing blood DNA from six individuals, we have identified a sensitive and specific episignature associated with *MSL2* variants. These findings further confirm a shared molecular pathophysiology in these individuals in association with this disorder. This finding also suggests that, in the future, this test may be useful in resolving ambiguous cases, including those with VUSs (Figures 3 and 4). No additional samples with *MSL2* variants were available for DNA methylation analysis at the time of the study. Current ongoing work is focused on collecting additional samples from affected individuals with a broader range of variant types for further episignature refinement. One point to consider even in this limited cohort is that a more robust methylation pattern was observed in case 18 relative to the rest of the cohort, suggesting a possibility of episignature heterogeneity, which we hope to address as part of the future work on an expanded cohort. It would be also interesting to establish the episignature of MRXSBA and compare it to the MSL2-related disorder to see if the two disorders have a convergent or divergent profile.

We have derived *MSL2* patient-specific iPSCs from three of our cohort members. These clones were genetically stable and were functionally pluripotent showing that our *MSL2* patient-derived iPSC lines can be faithfully used for further investigations. The generation of these iPSCs allowed us to characterize the molecular nature of the MSL2-related disorder. *De novo* truncating variants of *MSL2* lead to lower levels of intact MSL2 in the nucleus; thus, loss of functional MSL2 levels may be linked to haploinsufficiency. The manifestation of pathological nonsense variants in the MSL1 interaction domain also supports MSL2 haploinsufficiency. An additional factor contributing to the pathology could be that truncated MSL2 retaining the MSL1 interaction domain may lead to malfunctional MSL complex. Similarly, the *MSL2* missense variant p.Met317Val is likely still expressed and localized to the nucleus (Figure 5D). Considering that this individual also shows similar phenotypes, missense *MSL2* variant p.Met317Val is expected to be dysfunctional. A closer biochemical analysis of MSL2 is needed to determine whether the truncated or missense variants are stably expressed and/or partially functional to better understand the molecular etiology of the disease.

Our findings in *MSL2* patient-derived iPSCs mirror previous findings in mouse embryonic stem, neural progenitor, and human THP-1 cells, which similarly do not exhibit global decreases in H4K16ac upon loss of MSL2.[4,48,52] Our previous work in mouse cells suggests that instead of eliciting bulk effects on H4K16ac, MSL2 fine-tunes levels of this histone modification at a specific subset of developmental target genes.[48] Therefore, a potential underlying mechanism of pathology could be a dysregulation of local H4K16ac at the direct chromatin targets of MSL2, which are linked to fetal and neural development.[53,54] This is supported by our results showing that *MSL2* variants resulted in the dysregulation of putative MSL2 targets in iPSCs and differentiated cells. This provides the evidence that common genes are under the regulation of MSL2 in mice and humans. These target genes were affected more in the truncating variants lacking the CXC domain. The variability in lineage-specific differential gene expression among proband-derived samples may reflect the varying vulnerabilities of different tissues to *MSL2* variants, consistent with the findings that MSL2 chromatin targeting is dynamically regulated in different cell types.[5] The variant- and germ-layer-specific dysregulations implied by our *in vitro* studies support the heterogeneity and the multisystemic involvement seen in our cohort's phenotypes. In addition, our recent work showed that MSL2 ensures bi-allelic expression of dosage-sensitive developmental genes.[4,48,52] It is possible that *MSL2* variants lead to the disturbance in the control of allelic gene dosage which may partly explain the heterogeneity of symptoms observed in the MSL2-related disorder. Indeed, mice lacking one or both copies of *Msl2* show variable embryonic phenotypes linked to neurodevelopment.

MSL3 variants are associated with MRXSBA characterized by global DD/ID, feeding difficulties, and hypotonia (see last two columns of Table 2), which are all features seen in our *MSL2* cohort but are not very specific.[14,15] MRXSBA is also characterized by spasticity, while ataxic gait is commonly seen as well. There was only one proband in our cohort who presented with each of these features (individuals 13 and 10, respectively). However, one highly specific clinical trait of particular interest seen with both conditions is the dilatation of the aorta. Further exploration and *in vitro* studies would be required in order to elucidate why some MSL2-related disorder probands exhibit CTD signs and why aorta dilation has been observed, as in MRXSBA. It could be that the expression of specific genes (collagens, fibrillins, TGF-β signaling) is affected in both conditions. Respiratory and gastrointestinal issues as well as macrocephaly are much more common in MRXSBA than in our cohort. Finally, in MRXSBA, specific movement disorders such as brady-/hypokinesia and dystonia are frequently seen, as is hearing loss. These were not seen in our cohort. Despite being members of the same histone acetyltransferase (HAT)-containing complex, variants in *MSL2* do not have a bulk effect on the enzymatic activity of MOF, whereas *MSL3* variants strongly attenuate its ability to catalyze global H4K16ac. The molecular differences between *MSL2*- and *MSL3*-associated syndromes are also interesting from a therapeutic perspective. Treatment of female *MSL3*-affected-individual-derived

fibroblasts with histone deacetylase inhibitors (HDACi) to pharmacologically raise their H4K16ac levels led to a remarkable reversal of gene expression defects, suggesting this as a potential therapeutic avenue.[14] It is unclear whether a similar approach would be suitable for the MSL2-related disorder given that HDACi often has global rather than localized effects on histone acetylation levels.[55]

MOF variants have also recently been reported in an NDD characterized by DD/ID, dysmorphisms, as well as epilepsy, EEG/MRI defects, behavioral abnormalities including autistic traits, and cardiac anomalies.[13] Additional neurodevelopmental features commonly seen in our cohort, such as hypotonia and coordination issues, were not as recurrent in the MOF-related disorder. Where investigated, the molecular phenotype of NDD-associated variants in KATs and components of KAT-associated complexes is almost always reported to manifest at the level of enzyme activity. Variants in *KAT6A* (MIM: 601408), *KAT6B* (MIM: 605880), *BRPF1* (MIM: 602410), *CREBBP*, *KAT3A* (MIM: 600140), and *KAT5* (MIM: 601409) have all been demonstrated to exhibit reduced *in vivo* or *in vitro* histone acetyltransferase activity.[56–60] Similarly, case-mimicking mutations in MOF result in reduced *in vitro* H4K16 acetylation activity of the protein,[13] and primary human dermal fibroblasts taken from individuals with *de novo* heterozygous variants in *MSL3* (causative of MRXSBA) exhibit reduced bulk H4K16ac.[14] NDD-associated deleterious variants in KATs and associated proteins that do not affect acetyltransferase activity at the molecular level have, therefore, not been well explored in the literature. On the other hand, MSL2-associated variants have no influence on global H4K16ac levels (Figure 5D). This makes our *MSL2* cohort a unique category of individuals with *KAT*-associated variants with no noticeable global effect on KAT activity. We hypothesize that MSL2 targets the MSL complex to specific developmental genes for H4K16ac regulation. Further studies in patient-derived samples exploring the changes in the chromatin binding profile of MSL2 as well as changes in H4K16ac at MSL2-bound loci will improve our understanding of the molecular mechanism of the MSL2-related disorder pathology.

Despite the different effects of pathogenic *MSL3* and *MOF* variants compared to *MSL2* variants toward H4K16ac levels, there are shared neurodevelopmental phenotypes among MSL-complex syndromes, namely ID, DD, and epilepsy. Remarkably, these phenotypes emerge as recurrent characteristics in pathogenic variants of diverse epigenetic regulators.[60–67] This suggests a distinctive vulnerability of the nervous system to disruptions in finely tuned dynamics of developmental gene expression by epigenetic regulators. Although the molecular origins may differ, the resulting neuronal misdevelopment appears to converge on shared neurological symptoms. Conversely, expanding these clinical phenotypes into subcategories to capture the intricate distinctions in underlying mechanisms might prove challenging, largely due to the symptomatic heterogeneity observed within each syndrome. Furthermore, while lysine acetylation is a key molecular function of KATs and KAT-associated complexes, recent work has suggested that many histone-modifying enzymes, including KATs, have both enzymatic and nonenzymatic functions.[68–70] Although further work will be required to substantiate this hypothesis, we speculate that the MSL complex may have non-histone targets or non-catalytic functions that are indispensable during early human development and linked to the MSL2-related disorder pathology. It is also conceivable that MSL2 alone may have complex-independent functions.

Taken together, *MSL2* variants lead to a neurodevelopmental syndrome characterized by DD and ID with a high prevalence of motor disease, dysmorphisms, and psychiatric conditions including ASD, as well as with a characteristic blood episignature and a distinct, *MSL2*-specific, molecular etiology compared to other MSL complex-related diseases.

Data and code availability

Datasets used in this study that are available publicly are previously described.[71] Anonymized data for each subject is described in the study. The individual genomic and epigenomic or any other personally identifiable data for other samples in the EpiSign Knowledge Database (EKD) are prohibited from deposition in publicly accessible databases due to institutional and ethics restrictions. Specifically, these include data and samples submitted from external institutions to London Health Sciences EKD that are subject to Institutional Material and Data Transfer agreements, data submitted to London Health Sciences for episignature assessment under Research Services Agreements, and research study cohorts under Institutional Research Ethics Approval (Western University REB 106302; and REB 116108). Some of the software packages used in this study are publicly available as described in the material and methods. EpiSign™ is a commercial software and is not publicly available.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2024.05.001.

Acknowledgments

This work was funded by the government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-188) awarded to B.S. Data for individual 4 was obtained with support from the Neurological Foundation of New Zealand to L.S.B. Data for individual 24 were obtained by the Duke Research Sequencing Clinic supported by the Duke University Health System (Duke Pro00032301). This study was supported by the German Research Foundation (DFG) under Germany's Excellence Strategy (CIBSS – EXC-2189 – Project ID 390939984), CRC 992 (A02), CRC 1425

(P04), CRC 1381 (B3), and the Gottfried Wilhelm Leibniz Prize awarded to A.A., as well as NIH-NINDS funding (K23NS119666) awarded to S.S. Data for individual 25 were obtained through the French Genomic Initiative (Plan France Médecine Génomique 2025) and the AURAGEN laboratory. The graphical abstract was generated using BioRender.com.

Author contributions

R.K. designed, performed, and analyzed the functional and molecular experiments involving patient-derived iPSCs, wrote the corresponding parts of the paper, and contributed to writing and revising the rest of the paper. M.C.B. contacted the clinicians, helped collect and verify the clinical data, performed the phenotypical and the HPO ontology analyses, wrote the clinical parts of the paper, and contributed to writing and revising the rest of the paper. S.H. performed the episignature studies, wrote the corresponding parts of the paper, and revised it. A.N. assisted with the experiments involving patient-derived iPSCs. J. Reilly, M.A.L., R.R., J. Kerkhof, and H.M. contributed to the generation and analysis of the episignature. M.S. contributed to writing and revising the paper. A.K.P., K.M., C.Z., G.V., A.R., J.M.S., M.R.M., L.S.B., G.P., A.A.-E.-H., J.D., V.H., S.S., E.B., N.S.H., M.v.d.B., U.H., N.H., H.B.F., S.M., I.P.C.K., Y.L., A.T., R.Y., S.A., M.R., V.M., A.C., K.W.B., V.S., J.A.S., A.P., M.I., M.P.C., J.F., I.A.R., J. Kierstein, J.J.S., F.N.A., X.M., B.A., F.B.-K., K.P., A.-B.T., J.Q., A.B., N.C., L.J., C.R., D.A.C., K.G.M., and K.A.M. are clinicians or other members of the respective teams treating each proband included in this study. They provided the clinical data, proband and/or family consent, images and/or samples where applicable, and they reviewed and approved the manuscript before submission. J. Rousseau managed the proband samples and the fibroblast reprograming and revised the paper. B.S. coordinated the episignature generation and contributed to writing and revising the paper. A.A. coordinated the functional and molecular experiments involving patient-derived iPSCs and contributed to writing and revising the paper. P.M.C. coordinated the project, helped collect and verify the clinical data, managed the proband samples, and contributed to writing and revising the paper.

Declaration of interests

B.S. is a shareholder in EpiSign Inc, involved in commercial uses of EpiSign™ technology D.A.C. and K.G.M. are employees of GeneDx, LLC.

Received: September 23, 2023
Accepted: May 1, 2024
Published: May 29, 2024

References

1. Reichard, J., and Zimmer-Bensch, G. (2021). The Epigenome in Neurodevelopmental Disorders. Front. Neurosci. *15*, 776809.
2. Park, J., Lee, K., Kim, K., and Yi, S.-J. (2022). The role of histone modifications: from neurodevelopment to neurodiseases. Signal Transduct. Target. Ther. *7*, 217.
3. Valencia, A.M., Sankar, A., van der Sluijs, P.J., Satterstrom, F.K., Fu, J., Talkowski, M.E., Vergano, S.A.S., Santen, G.W.E., and Kadoch, C. (2023). Landscape of mSWI/SNF chromatin remodeling complex perturbations in neurodevelopmental disorders. Nat. Genet. *55*, 1400–1412.
4. Radzisheuskaya, A., Shliaha, P.V., Grinev, V.V., Shlyueva, D., Damhofer, H., Koche, R., Gorshkov, V., Kovalchuk, S., Zhan, Y., Rodriguez, K.L., et al. (2021). Complex-dependent histone acetyltransferase activity of KAT8 determines its role in transcription and cellular homeostasis. Mol. Cell *81*, 1749–1765.e8.
5. Chelmicki, T., Dündar, F., Turley, M.J., Khanam, T., Aktas, T., Ramírez, F., Gendrel, A.-V., Wright, P.R., Videm, P., Backofen, R., et al. (2014). MOF-associated complexes ensure stem cell identity and Xist repression. Elife *3*, e02024.
6. Ravens, S., Fournier, M., Ye, T., Stierle, M., Dembele, D., Chavant, V., and Tora, L. (2014). Mof-associated complexes have overlapping and unique roles in regulating pluripotency in embryonic stem cells and during differentiation. Elife *3*, e02104.
7. Cai, Y., Jin, J., Swanson, S.K., Cole, M.D., Choi, S.H., Florens, L., Washburn, M.P., Conaway, J.W., and Conaway, R.C. (2010). Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. J. Biol. Chem. *285*, 4268–4272.
8. Chatterjee, A., Seyfferth, J., Lucci, J., Gilsbach, R., Preissl, S., Böttinger, L., Mårtensson, C.U., Panhale, A., Stehle, T., Kretz, O., et al. (2016). MOF Acetyl Transferase Regulates Transcription and Respiration in Mitochondria. Cell *167*, 722–738.e23.
9. Karoutas, A., Szymanski, W., Rausch, T., Guhathakurta, S., Rog-Zielinska, E.A., Peyronnet, R., Seyfferth, J., Chen, H.-R., de Leeuw, R., Herquel, B., et al. (2019). The NSL complex maintains nuclear architecture stability via lamin A/C acetylation. Nat. Cell Biol. *21*, 1248–1260.
10. Hallacli, E., Lipp, M., Georgiev, P., Spielman, C., Cusack, S., Akhtar, A., and Kadlec, J. (2012). Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in Drosophila. Mol. Cell *48*, 587–600.
11. Kadlec, J., Hallacli, E., Lipp, M., Holz, H., Sanchez-Weatherby, J., Cusack, S., and Akhtar, A. (2011). Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. Nat. Struct. Mol. Biol. *18*, 142–149.
12. Keller, C.I., and Akhtar, A. (2015). The MSL complex: juggling RNA-protein interactions for dosage compensation and beyond. Curr. Opin. Genet. Dev. *31*, 1–11.
13. Li, L., Ghorbani, M., Weisz-Hubshman, M., Rousseau, J., Thiffault, I., Schnur, R.E., Breen, C., Oegema, R., Weiss, M.M., Waisfisz, Q., et al. (2020). Lysine acetyltransferase 8 is involved in cerebral development and syndromic intellectual disability. J. Clin. Invest. *130*, 1431–1445.
14. Basilicata, M.F., Bruel, A.-L., Semplicio, G., Valsecchi, C.I.K., Aktaş, T., Duffourd, Y., Rumpf, T., Morton, J., Bache, I., Szymanski, W.G., et al. (2018). De novo mutations in MSL3 cause an X-linked syndrome marked by impaired histone H4 lysine 16 acetylation. Nat. Genet. *50*, 1442–1451.
15. Brunet, T., McWalter, K., Mayerhanser, K., Anbouba, G.M., Armstrong-Javors, A., Bader, I., Baugh, E., Begtrup, A., Bupp, C.P., Callewaert, B.L., et al. (2021). Defining the genotypic and phenotypic spectrum of X-linked MSL3-related disorder. Genet. Med. *23*, 384–395.
16. Du, Y., Li, Z., Liu, Z., Zhang, N., Wang, R., Li, F., Zhang, T., Jiang, Y., Zhi, X., Wang, Z., and Wu, J. (2020). Nonrandom occurrence of multiple de novo coding variants in a proband

indicates the existence of an oligogenic model in autism. Genet. Med. *22*, 170–180.
17. Zhang, Y., Wang, T., Wang, Y., Xia, K., Li, J., and Sun, Z. (2021). Targeted sequencing and integrative analysis to prioritize candidate genes in neurodevelopmental disorders. Mol. Neurobiol. *58*, 3863–3873.
18. Brokamp, E., Koziura, M.E., Phillips, J.A., 3rd, Tang, L.A., Cogan, J.D., Rives, L.C., Robertson, A.K., Duncan, L., Bican, A., Peterson, J.F., et al. (2021). One is the loneliest number: genotypic matchmaking using the electronic health record. Genet. Med. *23*, 1830–1832.
19. Levy, M.A., McConkey, H., Kerkhof, J., Barat-Houari, M., Bargiacchi, S., Biamino, E., Bralo, M.P., Cappuccio, G., Ciolfi, A., Clarke, A., et al. (2022). Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. HGG Adv. *3*, 100075.
20. Aref-Eshghi, E., Kerkhof, J., Pedro, V.P., Groupe DI France, Barat-Houari, M., Ruiz-Pallares, N., Andrau, J.-C., Lacombe, D., Van-Gils, J., Fergelot, P., et al. (2020). Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. Am. J. Hum. Genet. *106*, 356–370.
21. Levy, M.A., Relator, R., McConkey, H., Pranckeviciene, E., Kerkhof, J., Barat-Houari, M., Bargiacchi, S., Biamino, E., Palomares Bralo, M., Cappuccio, G., et al. (2022). Functional correlation of genome-wide DNA methylation profiles in genetic neurodevelopmental disorders. Hum. Mutat. *43*, 1609–1628.
22. Sadikovic, B., Levy, M.A., Kerkhof, J., Aref-Eshghi, E., Schenkel, L., Stuart, A., McConkey, H., Henneman, P., Venema, A., Schwartz, C.E., et al. (2021). Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. Genet. Med. *23*, 1065–1074.
23. Kerkhof, J., Squeo, G.M., McConkey, H., Levy, M.A., Piemontese, M.R., Castori, M., Accadia, M., Biamino, E., Della Monica, M., Di Giacomo, M.C., et al. (2022). DNA methylation episignature testing improves molecular diagnosis of Mendelian chromatinopathies. Genet. Med. *24*, 51–60.
24. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum. Mutat. *36*, 928–930.
25. Zhang, C., Jolly, A., Shayota, B.J., Mazzeu, J.F., Du, H., Dawood, M., Soper, P.C., Ramalho de Lima, A., Ferreira, B.M., Coban-Akdemir, Z., et al. (2022). Novel pathogenic variants and quantitative phenotypic analyses of Robinow syndrome: WNT signaling perturbation and phenotypic variability. HGG Adv. *3*, 100074.
26. Calame, D.G., Guo, T., Wang, C., Garrett, L., Jolly, A., Dawood, M., Kurolap, A., Henig, N.Z., Fatih, J.M., Herman, I., et al. (2023). Monoallelic variation in *DHX9*, the gene encoding the DExH-box helicase DHX9, underlies neurodevelopment disorders and Charcot-Marie-Tooth disease. Am. J. Hum. Genet. *110*, 1394–1413.
27. Greene, D., Richardson, S., and Turro, E. (2017). ontologyX: a suite of R packages for working with ontological data. Bioinformatics *33*, 1104–1106.
28. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.
29. Aref-Eshghi, E., Bend, E.G., Colaiacovo, S., Caudle, M., Chakrabarti, R., Napier, M., Brick, L., Brady, L., Carere, D.A., Levy, M.A., et al. (2019). Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. Am. J. Hum. Genet. *104*, 685–700.
30. Aref-Eshghi, E., Rodenhiser, D.I., Schenkel, L.C., Lin, H., Skinner, C., Ainsworth, P., Paré, G., Hood, R.L., Bulman, D.E., Kernohan, K.D., et al. (2018). Genomic DNA Methylation Signatures Enable Concurrent Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. Am. J. Hum. Genet. *102*, 156–174.
31. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.
32. Ho, D.E., Imai, K., King, G., and Stuart, E.A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. J. Stat. Softw. *42*.
33. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). 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.
34. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.
35. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinf. *13*, 86.
36. Peters, T.J., Buckley, M.J., Statham, A.L., Pidsley, R., Samaras, K., V Lord, R., Clark, S.J., and Molloy, P.L. (2015). De novo identification of differentially methylated regions in the human genome. Epigenet. Chromatin *8*, 6.
37. Phipson, B., Maksimovic, J., and Oshlack, A. (2016). missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics *32*, 286–288.
38. Villegas, J., and McPhaul, M. (2005). Establishment and culture of human skin fibroblasts. Curr. Protoc. Mol. Biol. *Chapter 28*. Unit 28.3.
39. Benabdallah, B., Désaulniers-Langevin, C., Goyer, M.-L., Colas, C., Maltais, C., Li, Y., Guimond, J.V., Tremblay, J.P., Haddad, E., and Beauséjour, C. (2021). Myogenic progenitor cells derived from human induced pluripotent stem cell are immune-tolerated in humanized mice. Stem Cells Transl. Med. *10*, 267–277.
40. Juric-Sekhar, G., and Hevner, R.F. (2019). Malformations of Cerebral Cortex Development: Molecules and Mechanisms. Annu. Rev. Pathol. *14*, 293–318.
41. Lautarescu, A., Craig, M.C., and Glover, V. (2020). Prenatal stress: Effects on fetal and child brain development. Int. Rev. Neurobiol. *150*, 17–40.
42. Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nat. Rev. Mol. Cell Biol. *20*, 406–420.
43. Zhou, X., Edmonson, M.N., Wilkinson, M.R., Patel, A., Wu, G., Liu, Y., Li, Y., Zhang, Z., Rusch, M.C., Parker, M., et al. (2016). Exploring genomic alteration in pediatric cancer using ProteinPaint. Nat. Genet. *48*, 4–6.

44. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res. *49*, D412–D419.
45. Abad-Navarro, F., de la Morena-Barrio, M.E., Fernández-Breis, J.T., and Corral, J. (2018). Lost in translation: bioinformatic analysis of variations affecting the translation initiation codon in the human genome. Bioinformatics *34*, 3788–3794.
46. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47*, D886–D894.
47. Copps, K., Richman, R., Lyman, L.M., Chang, K.A., Rampersad-Ammons, J., and Kuroda, M.I. (1998). Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. EMBO J. *17*, 5409–5417.
48. Valsecchi, C.I.K., Basilicata, M.F., Semplicio, G., Georgiev, P., Gutierrez, N.M., and Akhtar, A. (2018). Facultative dosage compensation of developmental genes on autosomes in Drosophila and mouse embryonic stem cells. Nat. Commun. *9*, 3626.
49. Fauth, T., Müller-Planitz, F., König, C., Straub, T., and Becker, P.B. (2010). The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. Nucleic Acids Res. *38*, 3209–3221.
50. Wu, L., Zee, B.M., Wang, Y., Garcia, B.A., and Dou, Y. (2011). The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in crosstalk with H3 K4 and K79 methylation. Mol. Cell *43*, 132–144.
51. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.
52. Sun, Y., Wiese, M., Hmadi, R., Karayol, R., Seyfferth, J., Martinez Greene, J.A., Erdogdu, N.U., Deboutte, W., Arrigoni, L., Holz, H., et al. (2023). MSL2 ensures biallelic gene expression in mammals. Nature *624*, 173–181.
53. Han, C., Liu, H., Liu, J., Yin, K., Xie, Y., Shen, X., Wang, Y., Yuan, J., Qiang, B., Liu, Y.-J., and Peng, X. (2005). Human Bex2 interacts with LMO2 and regulates the transcriptional activity of a novel DNA-binding complex. Nucleic Acids Res. *33*, 6555–6565.
54. Fernández-Marmiesse, A., Sánchez-Iglesias, S., Darling, A., O'Callaghan, M.M., Tonda, R., Jou, C., and Araújo-Vilar, D. (2019). A de novo heterozygous missense BSCL2 variant in 2 siblings with intractable developmental and epileptic encephalopathy. Seizure *71*, 161–165.
55. Mackmull, M.-T., Iskar, M., Parca, L., Singer, S., Bork, P., Ori, A., and Beck, M. (2015). Histone Deacetylase Inhibitors (HDACi) Cause the Selective Depletion of Bromodomain Containing Proteins (BCPs). Mol. Cell. Proteomics *14*, 1350–1360.
56. Arboleda, V.A., Lee, H., Dorrani, N., Zadeh, N., Willis, M., Macmurdo, C.F., Manning, M.A., Kwan, A., Hudgins, L., Barthelemy, F., et al. (2015). De novo nonsense mutations in KAT6A, a lysine acetyl-transferase gene, cause a syndrome including microcephaly and global developmental delay. Am. J. Hum. Genet. *96*, 498–506.
57. Simpson, M.A., Deshpande, C., Dafou, D., Vissers, L.E.L.M., Woollard, W.J., Holder, S.E., Gillessen-Kaesbach, G., Derks, R., White, S.M., Cohen-Snuijf, R., et al. (2012). De novo mutations of the gene encoding the histone acetyltransferase KAT6B cause Genitopatellar syndrome. Am. J. Hum. Genet. *90*, 290–294.
58. Yan, K., Rousseau, J., Littlejohn, R.O., Kiss, C., Lehman, A., Rosenfeld, J.A., Stumpel, C.T.R., Stegmann, A.P.A., Robak, L., Scaglia, F., et al. (2017). Mutations in the Chromatin Regulator Gene BRPF1 Cause Syndromic Intellectual Disability and Deficient Histone Acetylation. Am. J. Hum. Genet. *100*, 91–104.
59. Lopez-Atalaya, J.P., Gervasini, C., Mottadelli, F., Spena, S., Piccione, M., Scarano, G., Selicorni, A., Barco, A., and Larizza, L. (2012). Histone acetylation deficits in lymphoblastoid cell lines from patients with Rubinstein-Taybi syndrome. J. Med. Genet. *49*, 66–74.
60. Humbert, J., Salian, S., Makrythanasis, P., Lemire, G., Rousseau, J., Ehresmann, S., Garcia, T., Alasiri, R., Bottani, A., Hanquinet, S., et al. (2020). De Novo KAT5 Variants Cause a Syndrome with Recognizable Facial Dysmorphisms, Cerebellar Atrophy, Sleep Disturbance, and Epilepsy. Am. J. Hum. Genet. *107*, 564–574.
61. Verberne, E.A., Goh, S., England, J., van Ginkel, M., Rafael-Croes, L., Maas, S., Polstra, A., Zarate, Y.A., Bosanko, K.A., Pechter, K.B., et al. (2021). JARID2 haploinsufficiency is associated with a clinically distinct neurodevelopmental syndrome. Genet. Med. *23*, 374–383.
62. Zhang, L.X., Lemire, G., Gonzaga-Jauregui, C., Molidperee, S., Galaz-Montoya, C., Liu, D.S., Verloes, A., Shillington, A.G., Izumi, K., Ritter, A.L., et al. (2020). Further delineation of the clinical spectrum of KAT6B disorders and allelic series of pathogenic variants. Genet. Med. *22*, 1338–1347.
63. Bell, S., Rousseau, J., Peng, H., Aouabed, Z., Priam, P., Theroux, J.-F., Jefri, M., Tanti, A., Wu, H., Kolobova, I., et al. (2019). Mutations in ACTL6B Cause Neurodevelopmental Deficits and Epilepsy and Lead to Loss of Dendrites in Human Neurons. Am. J. Hum. Genet. *104*, 815–834.
64. Nixon, K.C.J., Rousseau, J., Stone, M.H., Sarikahya, M., Ehresmann, S., Mizuno, S., Matsumoto, N., Miyake, N., DDD Study, and Baralle, D., et al. (2019). A Syndromic Neurodevelopmental Disorder Caused by Mutations in SMARCD1, a Core SWI/SNF Subunit Needed for Context-Dependent Neuronal Gene Regulation in Flies. Am. J. Hum. Genet. *104*, 596–610.
65. Cogné, B., Ehresmann, S., Beauregard-Lacroix, E., Rousseau, J., Besnard, T., Garcia, T., Petrovski, S., Avni, S., McWalter, K., Blackburn, P.R., et al. (2019). Missense Variants in the Histone Acetyltransferase Complex Component Gene TRRAP Cause Autism and Syndromic Intellectual Disability. Am. J. Hum. Genet. *104*, 530–541.
66. Machol, K., Rousseau, J., Ehresmann, S., Garcia, T., Nguyen, T.T.M., Spillmann, R.C., Sullivan, J.A., Shashi, V., Jiang, Y.-H., Stong, N., et al. (2019). Expanding the Spectrum of BAF-Related Disorders: De Novo Variants in SMARCC2 Cause a Syndrome with Intellectual Disability and Developmental Delay. Am. J. Hum. Genet. *104*, 164–178.
67. Snijders Blok, L., Rousseau, J., Twist, J., Ehresmann, S., Takaku, M., Venselaar, H., Rodan, L.H., Nowak, C.B., Douglas, J., Swoboda, K.J., et al. (2018). CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language. Nat. Commun. *9*, 4619.
68. Morgan, M.A.J., and Shilatifard, A. (2020). Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. Nat. Genet. *52*, 1271–1281.
69. Xie, G., Lee, J.-E., Senft, A.D., Park, Y.-K., Jang, Y., Chakraborty, S., Thompson, J.J., McKernan, K., Liu, C., Macfarlan, T.S., et al. (2023). MLL3/MLL4 methyltransferase activities control early

embryonic development and embryonic stem cell differentiation in a lineage-selective manner. Nat. Genet. *55*, 693–705.

70. Ciabrelli, F., Rabbani, L., Cardamone, F., Zenk, F., Löser, E., Schächtle, M.A., Mazina, M., Loubiere, V., and Iovino, N. (2023). CBP and Gcn5 drive zygotic genome activation independently of their catalytic activity. Sci. Adv. *9*, eadf2687.

71. Aref-Eshghi, E., Kerkhof, J., Pedro, V.P., Groupe DI France, Barat-Houari, M., Ruiz-Pallares, N., Andrau, J.C., Lacombe, D., Van-Gils, J., Fergelot, P., et al. (2020). Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. Am J Hum Genet *106*, 356–370. https://doi.org/10.1016/j.ajhg.2020.01.019.

The American Journal of Human Genetics, Volume *111*

Supplemental information

MSL2 variants lead to a neurodevelopmental syndrome with lack of coordination, epilepsy, specific dysmorphisms, and a distinct episignature

Remzi Karayol, Maria Carla Borroto, Sadegheh Haghshenas, Anoja Namasivayam, Jack Reilly, Michael A. Levy, Raissa Relator, Jennifer Kerkhof, Haley McConkey, Maria Shvedunova, Andrea K. Petersen, Kari Magnussen, Christiane Zweier, Georgia Vasileiou, André Reis, Juliann M. Savatt, Meghan R. Mulligan, Louise S. Bicknell, Gemma Poke, Aya Abu-El-Haija, Jessica Duis, Vickie Hannig, Siddharth Srivastava, Elizabeth Barkoudah, Natalie S. Hauser, Myrthe van den Born, Uri Hamiel, Noa Henig, Hagit Baris Feldman, Shane McKee, Ingrid P.C. Krapels, Yunping Lei, Albena Todorova, Ralitsa Yordanova, Slavena Atemin, Mihael Rogac, Vivienne McConnell, Anna Chassevent, Kristin W. Barañano, Vandana Shashi, Jennifer A. Sullivan, Angela Peron, Maria Iascone, Maria P. Canevini, Jennifer Friedman, Iris A. Reyes, Janell Kierstein, Joseph J. Shen, Faria N. Ahmed, Xiao Mao, Berta Almoguera, Fiona Blanco-Kelly, Konrad Platzer, Ariana-Berenike Treu, Juliette Quilichini, Alexia Bourgois, Nicolas Chatron, Louis Januel, Christelle Rougeot, Deanna Alexis Carere, Kristin G. Monaghan, Justine Rousseau, Kenneth A. Myers, Bekim Sadikovic, Asifa Akhtar, and Philippe M. Campeau

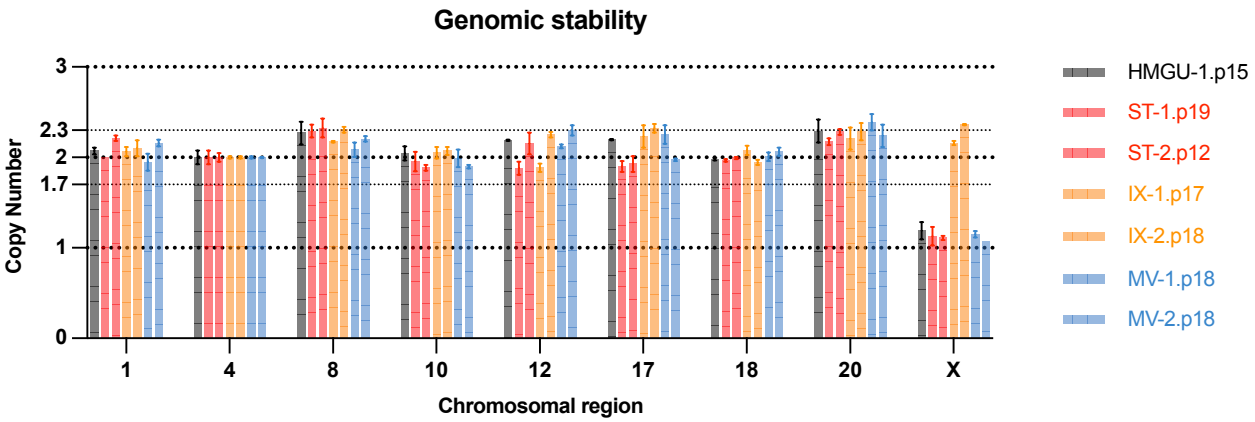


Figure S1. Genetic stability test for patient-derived iPSCs, related to Figure 5.

qPCR-based hPSC genetic analysis of the control and patient-derived iPSCs, analyzing the copy number of chromosomal regions where the majority of the karyotypic abnormalities in iPSCs are observed. Results are normalized to the results of the male control genomic DNA with stable copy number of the chromosomal regions reported. Note that analysis of the female IX line show two X chromosomes while the other male lines have only one as expected. n=2 per line. Data are represented as the mean ± SEM.

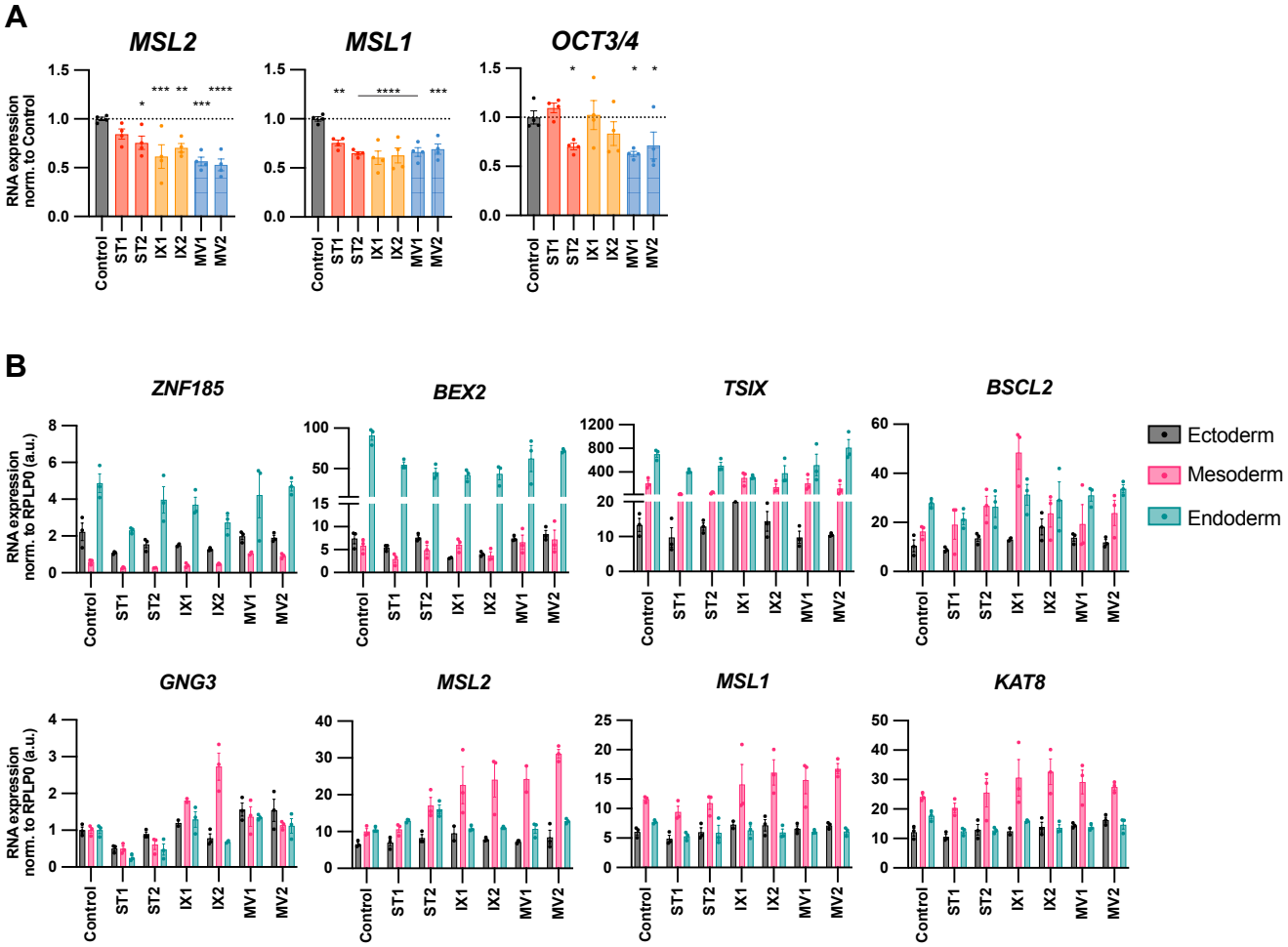


Figure S2. Supplemental RT-qPCR results, related to Figure 6.

(A) RT-qPCR results showing the relative RNA expression (normalized to *RPLP0*) of *MSL2*, *MSL1* and pluripotency marker *OCT3/4* (*POU5F1*) in patient-derived iPSCs compared to control. n=4 per line. One-way ANOVA with post-hoc Fisher's LSD test, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.
(B) RT-qPCR results showing the relative RNA expression (normalized to *RPLP0*, arbitrary units, a.u.) of putative mammalian MSL2 targets and MSL complex members (*MSL2*, *MSL1* and *MOF* [*KAT8*]) in three germ layers differentiated from patient-derived and control iPSCs. n=3 per line per lineage. Data are represented as the mean ± SEM.

Table S1. List of Primers

Gene_Forward/Reverse	Primer Sequence
Hs.RPLP0_F	CCTCGTGGAAGTGACATCGT
Hs.RPLP0_R	CTGTCTTCCCTGGGCATCAC
Hs.OCT3/4_F	CCTGAAGCAGAAGAGGATCACC
Hs.OCT3/4_R	AAAGCGGCAGATGGTCGTTTGG
Hs.NESTIN_F	TCAAGATGTCCCTCAGCCTGGA
Hs.NESTIN_R	AAGCTGAGGGAAGTCTTGGAGC
Hs.PAX6_F	CTGAGGAATCAGAGAAGACAGGC
Hs.PAX6_R	ATGGAGCCAGATGTGAAGGAGG
Hs.TBXT_F	CCTTCAGCAAAGTCAAGCTCACC
Hs.TBXT_R	TGAACTGGGTCTCAGGGAAGCA
Hs.CXCR4_F	CTCCTCTTTGTCATCACGCTTCC
Hs.CXCR4_R	GGATGAGGACACTGCTGTAGAG
Hs.SOX17_F	ACGCTTTCATGGTGTGGGCTAAG
Hs.SOX17_R	GTCAGCGCCTTCCACGACTTG
Hs.FOXA2_F	GGAACACCACTACGCCTTCAAC
Hs.FOXA2_R	AGTGCATCACCTGTTCGTAGGC
Hs.ZNF185_F	AGACACAGGCACCGTTTATCGC
Hs.ZNF185_R	CTGTTTGAGCCAGGAGTGGATC
Hs.BEX2_F	TTGGAGAGCCACAGGCAAGGAT
Hs.BEX2_R	AGTGCTGACTGCCCGCAAACTA
Hs.BSCL2_F	GTCTGTCTTCCTCTATGGCTCC
Hs.BSCL2_R	CCTTAGTCAGCGAGACATTGGC
Hs.GNG3_F	GATTGAAGCCAGCTTGTGTCGG
Hs.GNG3_R	GGGTTCTCCGAAGTGGGCACA
Hs.TSIX_F	GGTAGTGGTCACAAGGTGCTCA
Hs.TSIX_R	CATCCAGATGGCCTGAAGCAAC
Hs.MSL1_F	AGGAGGAGACTGTAGCAAGGTG
Hs.MSL1_R	CCTTAGAGGCTCTACTGAGTGG
Hs.MSL2_F	ACAGGCTCTTGCCTTACTTCCG
Hs.MSL2_R	ATGTTGGCAGGTGGAGTTGGTG
Hs.KAT8_F	CTGCTGAAGTGATCCAGTCTCG
Hs.KAT8_R	CGGTTCTTGTCTACCCACTCGT