HGGA, Volume *2*

Supplemental information

Variants in *LSM7* impair LSM complexes assembly, neurodevelopment in zebrafish and may be associated with an ultra-rare neurological disease

Alexa Derksen, Hung-Yu Shih, Diane Forget, Lama Darbelli, Luan T. Tran, Christian Poitras, Kether Guerrero, Sundaresan Tharun, Fowzan S. Alkuraya, Wesam I. Kurdi, Cam-Tu Emilie Nguyen, Anne-Marie Laberge, Yue Si, Marie-Soleil Gauthier, Joshua L. Bonkowsky, Benoit Coulombe, and Geneviève Bernard

Supplemental Figures:

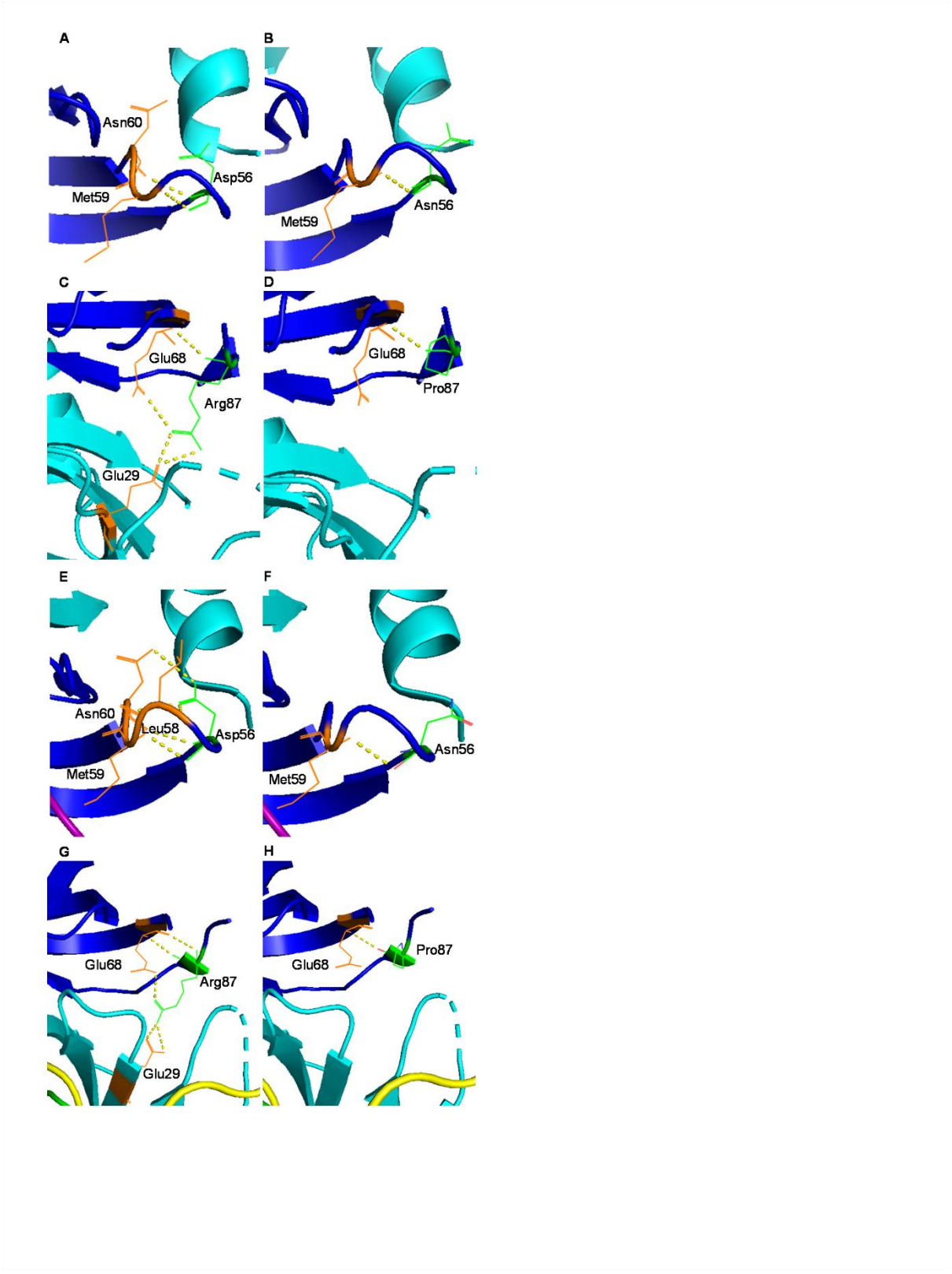

Figure S1. 3D representation of predicted mutational impact on LSM1-7 and LSM2-8 structures created with PyMOL.

(A-D) Pair-wise comparisons between the wild-type (left) and mutant (right) residues in LSM1-7 complex.

(A) Asp56 residue on yeast Lsm7 has polar contacts with Lsm7 residues Met59 and Asn60 in LSM1-7 complex.

(B) Asn56 mutant residue on yeast Lsm7 only has polar contacts with the Lsm7 residue Met59 in the LSM1-7 complex and has lost its interaction with the Asn60 residue.

(C) Arg87 residue on yeast Lsm7 has polar contacts with Lsm7 residue Glu68 and Lsm5 residue Glu29 in the LSM1-7 complex.

(D) Pro87 mutant residue on yeast Lsm7 only has polar contacts with Lsm7 residue Glu68 in the LSM1-7 complex and has lost its interaction with the Glu29 residue of Lsm5.

(E-H) Pair-wise comparisons between the wild-type (left) and mutant (right) residues in LSM2-8 complex.

(E) Asp56 residue on yeast Lsm7 has polar contacts with Lsm7 residues Leu58, Met59 and Asn60 in the LSM2-8 complex.

(F) Asn56 mutant residue on yeast Lsm7 only has polar contacts with Lsm7 residue Met59 in the LSM2-8 complex and has lost its interaction with Met59 and Asn60.

(G) Arg87 residue on yeast Lsm7 has polar contacts with Lsm7 residue Glu68 and Lsm5 residue Glu29 in the LSM2-8 complex.

(H) Pro87 mutant residue on yeast Lsm7 only has polar contacts with Lsm7 residue Glu68 in the LSM2-8 complex and has lost its interaction with the Lsm5 residue Glu29.

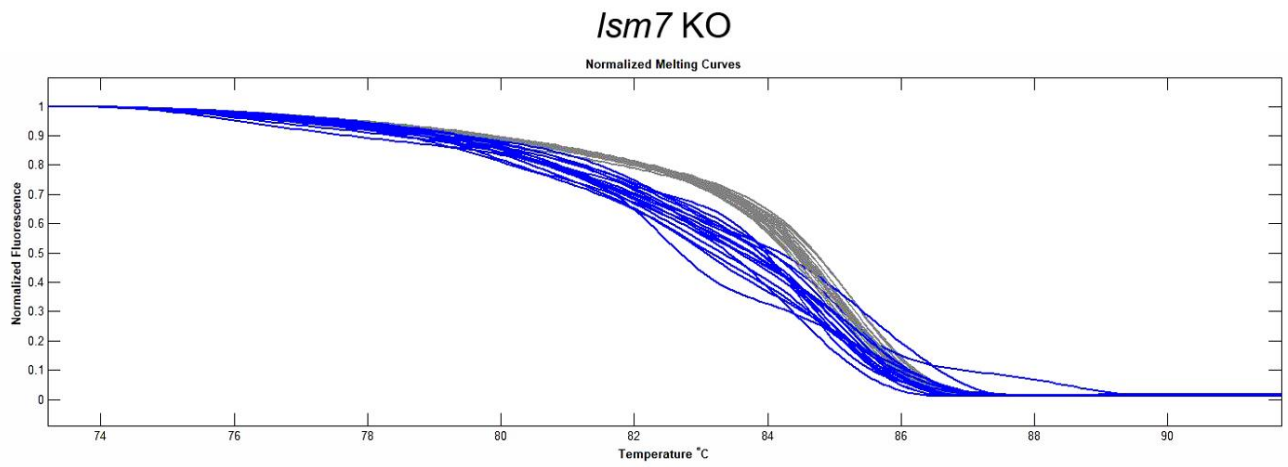

Figure S2. Lsm7 sgRNA could efficiently induce insertion and/or deletion formation of lsm7 loci.

HRMA analysis of wild-type embryos and Lsm7-Crispants. Gray curves, wild-type; blue curves, Lsm7-Crispants. X-axis, melt temperature (˚C); y-axis, normalized change in fluorescence with temperature.

Supplemental Table:

Table S1. *In silico* analysis using damage prediction algorithms and highest minor allele frequency in GnomAD

Variant	Mutation Taster	SIFT	CADD	Provean	PolyPhen2	MAF (GnomAD)
c.121G>A (p.Asp41Asn)	Disease Causing	Deleterious	Deleterious	Deleterious	Probably Damaging	3.87e-5
c.206G>C (p.Arg69Pro)	Disease Causing	Deleterious	Deleterious	Deleterious	Probably Damaging	Absent

Supplemental Material and Methods:

Affected individual ascertainment

Experiments involving human participants or data were conducted in compliance with all relevant ethical regulations. Informed consent was obtained from all participants. Approval for human subjects' research was obtained at the McGill University Health Centre Research Research Ethics Board (project number PED-11-105, 2019-4972). Clinical brain MRIs were obtained for individual 1 and longitudinal clinical assessment performed in order to determine disease course. Saliva samples were collected for DNA isolation from individual 1 and his mother and father. A dermal skin punch was also obtained from individual 1 after informed consent and was cultured in 10% fetal bovine serum and DMEM to obtain low passage primary fibroblasts. Consent for clinical next generation sequencing of both parents of individual 2 was obtained at the King Faisal Specialist Hospital and Research Centre in Riyadh, Saudi Arabia.

Whole exome sequencing and variant identification

Clinical whole exome sequencing was performed on affected individual 1, as a trio with biological parents by GeneDx. The exonic regions and flanking splice junctions of the genome were captured using a proprietary system developed by GeneDx and sequenced by massive parallel (NextGen) sequencing on an Illumina system with 100bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg 19 and analyzed for sequence variants using a custom developed analysis tool (Xome Analyzer). Additional sequencing technology and variant interpretation protocols have been previously described.[1] Capillary sequencing or

another appropriate method was used to confirm all potentially pathogenic variants identified in individual 1 and their relative samples. The general assertion criteria for variant classification are publicly available on GeneDx ClinVar submission page. A second analysis of the sequence variants was conducted in-house at the MyeliNeuroGene lab of the Research Institute of the McGill University Health Centre. Sequence variants were prioritized using the standard and guidelines for interpretation set out by the American College of Medical Genetics (ACMG).[2] Clinical whole exome sequencing was performed on the parents of individual 2 as previously described.[3]

Cell culture

Fibroblasts from individual 1 and two age- and sex-matched controls were grown in DMEM supplemented with 10% fetal bovine serum and maintained at 37°C and 5% CO_2. Each cell line was expanded and harvested at similar passages at least three consecutive times for use in molecular experiments. HEK293T cells were grown in DMEM media supplemented with 10% fetal bovine serum, 2mM glutamine and 1% penicillin-streptomycin and maintained at 37°C and 5% CO_2.

RT-qPCR

RNA was extracted from fibroblasts using TRIzol reagent according to manufacturer's instructions (Invitrogen). For each sample, 1µg of total RNA was reverse transcribed using M-MLV reverse transcriptase (Promega) and random hexamers (Invitrogen) according to manufacturer's protocol (Promega). Primers for RT-qPCR were designed, and efficiency tested according to the MIQE guidelines.[4, 5] Primers were as follows:

LSM7 forward 5'-GAAGCCAGTGGAATCCTGAAGG-3'; *LSM7* reverse: 5'-CCTCCGTGAGCTTGTACTGG-3'; *U6 snRNA* forward: 5'-CTCGCTTCGGCAGCACA-3'; *U6 snRNA* reverse: 5'-AACGCTTCACGAATTTGCGT-3'; *RPL30* forward: 5'-AAGGCAGGAAGATGGTGGCC-3'; *RPL30* reverse: 5'-GAGTCTGCTTGTACCCCAGGAC-3'; and *SDHA* forward: 5'-CAGCATGTGTTACCAAGCTG-3'; *SDHA* reverse: 5'-GGTGTCGTAGAAATGCCAAC-3'. Reverse transcription was performed in triplicate with a 1:4 dilution of cDNA using SsoAdvanced Universal SYBR Green Supermix (Biorad) on the Roche LightCycler 96. All data were normalized to *SDHA* or *RPL30* using the ΔΔCt method in accordance with the MIQE guidelines.[4, 5]

Western blot

Total protein was extracted from fibroblasts in ice cold RIPA buffer with protease inhibitors (Roche) and spun down for 20 minutes at maximum speed. Protein concentration was determined using the Bradford protein assay in cuvettes according to manufacturer's instructions (Biorad). 30ug of protein were separated on a 15% SDS-polyacrylamide gel and transferred to a nitrocellulose membrane using the Bio-Rad Trans-Blot Turbo Transfer System. The membranes were first blocked with 1X TBST with 5% w/v non-fat dry milk and then probed with primary antibodies against LSM7 (Abcam 241656, dilution 1:10,000) and against beta-tubulin (Abcam 6046, dilution 1:20,000) diluted in primary antibody buffer overnight at 4℃. The next day, a secondary polyclonal goat-anti rabbit IgG (H+L) (Novus biologicals, 1:5,000) antibody was used in 5% w/v non-fat milk in TBST and membranes were subsequently

visualized with Amersham ECL Western Blot Detecting Reagent according to manufacturer's protocol (GE Life Sciences).

Construction of Lsm1-7 and Lsm2-8 structural models

Structural models of the wild-type and mutant yeast Lsm1-7 (PDB ID 4M75)[6] and Lsm2-8 complexes (PDB ID 4M77)[6] were generated using PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

Protein affinity purification coupled to mass spectrometry

Plasmids for 3xFLAG-tagged LSM7 wild-type and the two mutants (p.Asp41Asn and p.Arg69Pro) were generated using the p3xFLAG-CMV-14 plasmid (Sigma). 0.4ug of the plasmids were transfected into HEK293 cells seeded at 80% confluence in 6-well plates using jetPRIME DNA transfection reagent kit according to manufacturer's protocol (Polyplus-Transfection). The HEK293 cells were maintained in culture in DMEM media supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin and 1% L-glutamine. Forty-eight hours after transfection, protein lysates were collected using a gentle lysis buffer to preserve protein-protein interactions and quantified in a 96-well plate using Bradford Reagent according to manufacturer's protocol (LifeTechnologies). Equal amounts of protein were purified using FLAG M2 magnetic beads according to standard procedures in 4 independent replicate experiments.[7, 8] The protein extracts were dried down in a speed-vac before being resolubilized and digested with trypsin. The tryptic peptides were then purified and identified using tandem mass spectrometry (LC-MS/MS) with an HPLC system coupled to Orbitrap fusion mass spectrometer

(Thermo Scientific) through a Nanospray Flex Ion Source. Protein database searching was performed using MaxQuant version 1.6.10.43[9-14] against the SwissProt human protein database downloaded on 4 April 2019. Default MaxQuant parameters were modified as follows: trypsin was used as the digestion enzyme, LFQ computation and "Match between runs" were activated. Known AP-MS protein contaminants including keratins were excluded. Protein intensities were analysed using Perseus version 1.6.10.43.[15-18] Proteins marked by MaxQuant as being "Only identified by site" and "Reverse" were excluded. Proteins not present in at least 3 out of 4 replicates of WT, p.Asp41Asn, or p.Arg69Pro were excluded. Missing LFQ values were replaced by normally distributed values with a mean downshifted by 1.8 and a standard deviation of 0.3 times the non-missing values. WT, p.Asp41Asn, and p.Arg69Pro proteins were compared against FLAG empty vector control samples and were labeled as high-confidence interactors when their adjusted p-values were under 0.05 and their intensity ratio over 2. P-values were corrected with a permutation-based approach using an s0 correction factor of 0.1 with 10,000 iterations.[19] A Student's *t*-test was performed between p.Asp41Asn and p.Arg69Pro against the WT samples. For the generation of the heatmap, log2 ratios were computed between the means of the WT, p.Asp41Asn, and p.Arg69Pro against WT for known interactors of LSM7.

Zebrafish ethics statement

Zebrafish experiments were performed in strict accordance of guidelines from the University of Utah Institutional Animal Care and Use Committee (IACUC), regulated under federal law (the Animal Welfare Act and Public Health Services Regulation Act)

by the U.S. Department of Agriculture (USDA) and the Office of Laboratory Animal Welfare at the NIH, and accredited by the Association for Assessment and Accreditation of Laboratory Care International (AAALAC).

Fish stocks and embryo raising

Adult fish were bred according to standard methods. Embryos were raised at 28.5°C in E3 embryo medium and staged by time and morphology. For *in situ* staining and immunohistochemistry, embryos were fixed in 4% paraformaldehyde (PFA) in 1x PBS overnight at 4°C, washed briefly in 1xPBS with 0.1% Tween-20, serially dehydrated, and stored in 100% MeOH at -20°C until use. Transgenic fish line and alleles used in this paper were the following: Tg(*olig2:dsRed)*^{vu19}.[20]

***Lsm7* sequence analysis and construct generation**

Human and zebrafish Lsm7 amino acid sequences were aligned using PRALINE. The construct for the *lsm7* riboprobe was amplified by PCR using Phusion® High-Fidelity DNA Polymerases (New England Biolabs, NEB) with primers: *lsm7* forward (5'-GGATCCGCCGCCACCATGGCGGACAAAGACAAGAAAAAGAAGGAG-3') and *lsm7* reverse (5'- CTCAAGGCGTCCAGTAACCAAAGCCC-3').

The constructs of human LSM7$^{full\text{-}length}$, LSM7^{D41N}, and LSM7^{R69P} for capped RNA synthesis were amplified from p3XFlag-CMV-LSM7$^{full\text{-}length}$, p3XFlag-CMV-LSM7^{D41N}, and p3XFlag-CMV-LSM7^{R69P}, respectively, by PCR using Phusion® High-Fidelity DNA Polymerases with the following primers: *LSM7* forward (5'-GAATTCGCCGCCACCATGGCGGATAAGGAGAAGAAGAAAAAG-3') and *LSM7*

reverse (5’- GAATTCCTACTTGTCATCGTCATCCTTGTAGTC -3’), and then subcloned into pCS2^{+} vector.

lsm7 CRISPR sgRNA construction and injection

We designed sgRNA target sites for the zebrafish *lsm7* gene (Ensembl Zv11: ENSDART00000081188.7) by looking for sequences corresponding to GGN$_{18}$nGG on the sense or antisense strand using the CRISPR design program CHOPCHOP.[21] Off-target effects were checked with NIH BLAST tool applied to the zebrafish genome (zv11). Off-target sequences that had significant matches of the final 23 nt of the target and NGG PAM sequence were discarded. DNA template for *lsm7* exon3 sgRNA was amplified by PCR using Phusion® High-Fidelity DNA Polymerases with *lsm7* exon3 specific forward primer (5’-GAAATTAATACGACTCACTATACCCGTTGTTGAATCTTGTGTGTTTTAGAGCTAGAAATAGC -3’) and universal reverse primer (5’-AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC -3’). The PCR program used for template synthesis was 98°C 30 seconds, 35 cycles of [98°C 10 seconds, 60°C 30 seconds, 72°C 15 seconds], 72°C 10 minutes, 10°C end. Subsequently, the DNA template was purified by QIAquick PCR Purification Kit (Qiagen). The *lsm7 exon3* sgRNA was synthesized by the HiScribe T7 RNA Synthesis kit (BioLabs) followed by RNA purification with RNA Clean & Concentrator-5 (Zymo research).

The *lsm7* exon3 sgRNA (460 pg) and Cas9 protein (460 pg, PNA BIO) were injected into one cell stage embryos. CRISPR efficiency was evaluated on individual 24 hpf

injected embryos after DNA extraction, PCR amplification of the target locus, and HRMA analysis. A *lsm7* exon3 sgRNA dose of 460 pg resulted in >90% mutagenesis in 24 hpf embryos, assayed by high-resolution melt analysis.[22] Embryos used for injection were derived from either wild-type AB parents or Tg(*olig2:dsRed)*^{vu19} as mentioned in the text.

Capped RNA synthesis

Capped RNA encoding the full coding sequence of LSM7$^{full\text{-}length}$, LSM7^{Asp41Asn}, or LSM7^{Arg69Pro} were prepared as per manufacturer's instructions and purified with Micro Bio-spin 6 columns (BioRad). Capped RNAs (460 pg) were injected for rescue experiments.

Immunohistochemistry and in situ hybridization

Immunohistochemistry was performed as previously described.[23] Antibodies used were rabbit anti-dsRed 1:250 (Clontech), Alexa 555 goat anti-rabbit 1:400 (ThermoFisher Scientific), and 4',6-diamidino-2-phenylindole (DAPI). The antisense digoxigenin-UTP labeled riboprobe for detecting *lsm7* transcript was synthesized according to manufacturer's instructions (Roche), and *in situ* hybridizations were performed as described previously.[24] The color reaction was carried out using NBT/BCIP substrate (Roche).

Morphological analysis

5 dpf fish larvae were fixed in 4% PFA in 1x PBS overnight at 4°C, washed briefly in 1xPBS with 0.1% Tween-20. One eye of an individual larva was dissected out and mounted laterally. Eyes were imaged, and the length was measured from anterior to posterior along the longest axis by ImageJ software.

Behaviour analysis

Larval behaviour analysis was performed on 7 dpf larvae in 96-well square bottom plates (Krackeler Scientific) using video analysis software (Noldus EthoVision). For spontaneous behaviour, animals were transferred at 6 dpf to a 96-well plate and kept at 28.5˚C overnight. At 7 dpf the plate was placed on the video imaging system and animals were allowed to adapt in the dark for 10 minutes, and then recording was performed for 5 minutes (1 minute dark and 4 minutes light).

Microscopy and image analysis

Immunostained embryos were transferred step-wise into 90% glycerol/10% PBS, mounted on a glass slide with a #0 coverslip, and imaged on a confocal microscope. Confocal stacks were projected in ImageJ, and images composed with Adobe Photoshop and Illustrator.

TUNEL quantification

Terminal deoxynucleotidyl transferase dUTP nick-end labeling (TUNEL) was performed on whole-mount larvae (ApopTag Fluorescein *In Situ* Apoptosis Detection Kit; Millipore)

as previously described.[25] Confocal imaging was performed, and images were rendered in ImageJ by compiling an average intensity projection of 120 µm (step size 5 µm) into a single z-stack image, for cell counting using Photoshop's (Adobe) count tool.

Statistical analysis

qPCR data were analyzed following the MIQE guidelines and expressed as average relative normalized expression with appropriate error bars (SEM). Normality of all data were assessed using the Shapiro-Wilks Test of Normality and the equality of variances were assessed using Levene's Test. Changes in *LSM7* mRNA levels in individual 1 compared to controls were analyzed using a parametric, unpaired *t*-test (two-tailed). Changes in LSM7 protein levels in individual 1 compared to controls were analyzed using a parametric, unpaired *t*-test (two tailed). Statistical significance was set at $p < 0.05$. Statistical analyses for zebrafish experiments were performed using Prism8 software (GraphPad). Student's *t*-test was used for two-way comparisons; comparisons between three or more groups was performed with ANOVA with post-hoc Tukey's HSD between individual means.

Supplemental References:

1. Retterer, K., Juusola, J., Cho, M.T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K.G., et al. (2016). Clinical application of whole-exome sequencing across clinical indications. Genet Med 18, 696-704.
2. 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.
3. Monies, D., Abouelhoda, M., Assoum, M., Moghrabi, N., Rafiullah, R., Almontashiri, N., Alowain, M., Alzaidan, H., Alsayed, M., Subhani, S., et al. (2019). Lessons Learned from Large-Scale, First-Tier Clinical Exome Sequencing in a Highly Consanguineous Population. Am J Hum Genet 105, 879.
4. Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M., and Nguyen, M. (2010). A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. Methods 50, S1-5.
5. Taylor, S.C., and Mrkusich, E.M. (2014). The state of RT-quantitative PCR: firsthand observations of implementation of minimum information for the publication of quantitative real-time PCR experiments (MIQE). J Mol Microbiol Biotechnol 24, 46-52.
6. Zhou, L., Hang, J., Zhou, Y., Wan, R., Lu, G., Yin, P., Yan, C., and Shi, Y. (2014). Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. Nature 506, 116-120.
7. Chen, G.I., and Gingras, A.C. (2007). Affinity-purification mass spectrometry (AP-MS) of serine/threonine phosphatases. Methods 42, 298-305.
8. Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.P., St-Denis, N.A., Li, T., Miteva, Y.V., Hauri, S., Sardiu, M.E., Low, T.Y., et al. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nature methods 10, 730-736.
9. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26, 1367-1372.
10. Cox, J., Michalski, A., and Mann, M. (2011). Software lock mass by two-dimensional minimization of peptide mass errors. J Am Soc Mass Spectrom 22, 1373-1380.
11. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13, 2513-2526.
12. Schaab, C., Geiger, T., Stoehr, G., Cox, J., and Mann, M. (2012). Analysis of high accuracy, quantitative proteomics data in the MaxQB database. Mol Cell Proteomics 11, M111 014068.
13. Tyanova, S., Temu, T., Carlson, A., Sinitcyn, P., Mann, M., and Cox, J. (2015). Visualization of LC-MS/MS proteomics data in MaxQuant. Proteomics 15, 1453-1456.

14. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc 11, 2301-2319.
15. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13, 731-740.
16. Rudolph, J.D., and Cox, J. (2019). A Network Module for the Perseus Software for Computational Proteomics Facilitates Proteome Interaction Graph Analysis. J Proteome Res 18, 2052-2064.
17. Tyanova, S., Albrechtsen, R., Kronqvist, P., Cox, J., Mann, M., and Geiger, T. (2016). Proteomic maps of breast cancer subtypes. Nat Commun 7, 10259.
18. Cox, J., and Mann, M. (2012). 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. BMC Bioinformatics 13 Suppl 16, S12.
19. Choquet, K., Pinard, M., Yang, S., Moir, R.D., Poitras, C., Dicaire, M.J., Sgarioto, N., Lariviere, R., Kleinman, C.L., Willis, I.M., et al. (2019). The leukodystrophy mutation Polr3b R103H causes homozygote mouse embryonic lethality and impairs RNA polymerase III biogenesis. Mol Brain 12, 59.
20. Kucenas, S., Takada, N., Park, H.C., Woodruff, E., Broadie, K., and Appel, B. (2008). CNS-derived glia ensheath peripheral nerves and mediate motor root development. Nat Neurosci 11, 143-151.
21. Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res 47, W171-W174.
22. Xing, L., Quist, T.S., Stevenson, T.J., Dahlem, T.J., and Bonkowsky, J.L. (2014). Rapid and efficient zebrafish genotyping using PCR with high-resolution melt analysis. J Vis Exp, e51138.
23. Bonkowsky, J.L., Wang, X., Fujimoto, E., Lee, J.E., Chien, C.B., and Dorsky, R.I. (2008). Domain-specific regulation of foxP2 CNS expression by lef1. BMC Dev Biol 8, 103.
24. Thisse, C., and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc 3, 59-69.
25. Lambert, A.M., Bonkowsky, J.L., and Masino, M.A. (2012). The conserved dopaminergic diencephalospinal tract mediates vertebrate locomotor development in zebrafish larvae. J Neurosci 32, 13488-13500.