Mutation in *NSUN2*, which encodes an RNA methyltransferase, causes autosomal recessive intellectual disability

Muzammil Ahmad Khan[1,2], Muhammad Arshad Rafiq[1], Abdul Noor[1], Shobbir Hussain[3], Joana V. Flores[3], Verena Rupp[4], Akshita K. Vincent[1], Roland Malli[5], Ghazanfar Ali[2,6], Falak Sher Khan[2], Gisele E. Ishak[7], Dan Doherty[8], Rosanna Weksberg[9], Muhammad Ayub[10,11], Christian Windpassinger[1,4], Shahnaz Ibrahim[12], Michaela Frye[3], Muhammad Ansar[2], John B. Vincent[1,13]✉

[1]Molecular Neuropsychiatry & Development Lab, Neurogenetics Section, Centre for Addiction and Mental Health, Toronto, ON, M5T 1R8, Canada.

[2]Department of Biochemistry, Quaid-I-Azam University, Islamabad, 45320, Pakistan

[3]Wellcome Trust Centre for Stem Cell Research Tennis Court Road Cambridge, CB2 1QR United Kingdom

[4]Institute of Human Genetics, Medical University of Graz, Graz, A-8010, Austria;

[5]Institute of Molecular Biology and Biochemistry, Medical University of Graz, Graz, A-8010, Austria;

[6]Center of Excellence in Biotechnology Research, King Saud University, Riyadh, Kingdom of Saudi Arabia, 11451;

[7]Department of Radiology, Seattle Children's Hospital, University of Washington, Seattle, WA 98105, USA;

[8]Division of Genetics and Developmental Medicine, University of Washington, Seattle; Children's Hospital, Seattle, WA 98105, USA;

[9]Department of Genetics and Genome Biology, Research Institute, The Hospital for Sick Children, Toronto, ON, M5G 1X8, Canada;

[10]Tees, Esk and Wear Valleys NHS Foundation Trust, Darlington, DL2 2TS, United Kingdom;

[11]School for Health and Medicine, University of Durham, Stockton-On-Tees, TS17 6BH United Kingdom;

[12]Dept. of Paediatrics, Aga Khan University, Karachi, 75800, Pakistan

[13]Dept. of Psychiatry, University of Toronto, Toronto, Ontario, M5T 1R8, Canada.

✉ Correspondence should be addressed to: Dr. John B. Vincent, Neurogenetics Section, R-30, Centre For Addiction and Mental Health, 250 College Street, Toronto, ON, M5T 1R8, Canada; Tel: 416 535 8501 x6487; Fax: 416 979 4666; Email: john_vincent@camh.net

Abstract

Autosomal recessive causes of intellectual disability have, until very recently, been under-researched due to the high degree of genetic heterogeneity. However, now that genome-wide approaches can be applied to single multiplex consanguineous families, identification of genes harboring disease-causing mutations by autozygosity mapping is expanding rapidly. Here we have mapped a disease locus in a consanguineous family from Pakistan with intellectual disability (ID) and distal myopathy. We genotyped family members on genome-wide single nucleotide polymorphism microarrays, and used the data to determine a single 2.5 Mb homozygosity-by-descent (HBD) locus on 5p15.32-p15.31, and identified the missense change c.2035G>A; p.Gly679Arg at a conserved residue within *NSUN2*. This gene encodes a methyltransferase that catalyzes formation of 5-methylcytosine at C34 of tRNA-leu(CAA), and plays a role in spindle assembly during mitosis as well as chromosome segregation. In mouse brain we show that NSUN2 localizes to the nucleolus of Purkinje cells in the cerebellum. Confirmation of the effects of the mutation was performed using transfection of wild-type and mutant constructs into cells, followed by immunohistochemistry. We show that mutation to arginine at this residue causes the NSUN2 protein to fail to localize within the nucleolus. The intellectual disability combined with a unique profile of comorbid features presented here makes this an important genetic discovery, and the involvement of *NSUN2* highlights the role of RNA methyltransferase in human neurocognitive development.

Intellectual disability (ID), also called mental retardation (MR), is a neurodevelopmental disorder that can have a devastating impact on the affected individuals and their families. It is believed to occur at a frequency of ~1-3% within the population[1], and may often be caused by abnormalities at the genetic level[2,3]. ID may present as the sole clinical feature (non-syndromic [MIM #249500]), or may be present with additional clinical or dysmorphological features (syndromic).

We ascertained a consanguineous family from a farming community in the Khairpur district, within Sindh province in Pakistan, where the first-cousin parents had seven children, with 3 of the 5 female siblings affected (Figure 1). A male cousin, offspring of the mother's brother, was also reported to be affected, but was not available for the study. Appropriate informed consent was obtained for all participants in the study including unrelated healthy Pakistani controls and NSID individuals of European descent, and institutional research ethics approval was obtained through Quaid-I-Azam University and through Centre for Addiction and Mental Health. The affected family members were assessed by a consultant pediatric psychiatrist. The Vinelands II Adaptive Behavior Scale and Diagnostic Interview for Social and Communication Disorders (DISCO[4]) were used as a framework to obtain information about early development, schooling and academic achievement, current level of social functioning and adjustment and support and help required for different activities of daily living. Based on this information a consensus diagnosis was reached of moderate ID, with intelligence quotient (IQ) in the range of 40 – 50, for all three affected individuals. No features of autism were noted.

Neurological assessment was performed by a consultant neurologist. The oldest female (II:3) had significant delay to her development, walking and talking at five years. Her speech was limited to just a few words, and was dysarthric. Sensory examination and fundoscopy were

not possible as she was not cooperative. The individual had a long face, somewhat long and pointed nose and chin. Height and weight were below the 5th centile. Her fingers were tapering, but no hyper extensibility noted. Muscle tone was increased in all limbs (power grade 5) and reflexes were brisk, and plantars were equivocal. Other features observed are shown in Table 1. She is toilet trained, and can feed and dress/undress herself, and can help with household chores.

The second girl (II:4) also had significant delay, talking at 4 years, although age of walking was less delayed. Her speech is also limited and dysarthric, and is unable to recall her name when asked. Like her elder sister, she also had a long and pointed face and was below the 5th centile for height and weight. Muscle tone was increased in all limbs, reflexes were brisk, and plantars were equivocal. Other features observed are shown in Table 1. She is toilet trained, and can feed and dress/undress herself, but is unable to perform household chores. Coordination and sensory exam was grossly normal.

The third girl (II:5) walked at 5 years, but had no speech at age of examination (6 years). Head and body measures were below the 5th centile. Unlike II:3 and II:4, her gait was normal, tone was normal and reflexes were brisk with unsustained clonus. Fundus examination could not be performed, but other cranial nerve examination and systemic examination showed as normal. Other features observed are shown in Table 1.

Computed tomography of the brain was performed for two affected individuals, which showed normal ventricles and cerebral volume and was generally normal. Gray and white matter differentiation was preserved, and posterior fossa was unremarkable. Whole blood count was within normal range, however creatine phosphokinase (CPK) and lactose dehydrogenase (LDH) levels were measured in two individuals (II:3 and II:4) and indicated elevated LDH levels in both

(447 and 463 IU/L respectively; ref 122-234 IU/L), but elevated CPK only in the older sister (II:3).

Photographs of all affected individuals were assessed for dysmorphic features by an experienced clinical geneticist, and are shown in Supplementary Materials, Figure 1. Clinical examination of affected individuals is summarized in Table 1. Overall, the picture is of moderate intellectual disability, small head and body size, with signs of distal neuropathy as indicated by *pes cavus*, broad gait, and tight Achilles tendons; however nerve conductance velocities, wave latencies and amplitude were apparently normal. Other variable features include long nose/face, hypoplastic nails, partial $4^{th}/5^{th}$ toe syndactyly and hyper-separation between $1^{st}/2^{nd}$ toes, and absent or atrophied ovaries.

Genomic DNA was extracted from peripheral blood leukocytes by standard methods. DNA samples of three affected and one unaffected were analyzed using the Affymetrix GeneChip Mapping 500K array, using just the *Nsp*I chip, allowing us to genotype ~260,000 SNPs. Microarray analysis was performed at the London Regional Genomics Centre (LRGC, University of Western Ontario). Homozygosity mapping, performed using the dChip analyzer[5-7], identified a ~5 Mb homozygosity-by-descent (HBD) region on 5p15.32-p15.31 flanked by SNPs rs2259 and rs2914296 (7.657 Mb). This region overlaps with a locus previously identified in an Iranian family as harboring an NS-ARID-associated mutation, and designated MRT5, however no gene harboring disease-causing mutations has yet been reported for this locus[8]. The MRT5 locus was defined by SNP markers rs1824938 (5.092 Mb) and rs60701 (10.734 Mb). Thus, the common region shared between our Pakistani family and the Iranian family was from rs1824938 (5.092 Mb) to rs2914296 (7.657 Mb) - a 2.565 Mb critical region.

Additional genotype data from microsatellite markers on 5p also verified this, and linkage analysis gave maximum lod score of 2.77 for marker D5S406 (see Table 2). All known coding genes within this 2.565 Mb critical region were screened for mutations by sequencing. We identified a homozygous substitution, NM_017755.5:c.2035G>A (GRCh37/hg19 chr5:6,600,308C>T), within exon 19 of *NSUN2* [MIM 610916], which is one of the 8 known genes within this locus. This substitution would result in the missense change, p.Gly679Arg. This substitution is neither a known SNP in any SNP databases, nor in the 1000 Genomes Project, nor in the NHLBI Exome Sequencing Project, ESP5400 release (5379 subjects). We also confirmed this by sequencing over 200 Pakistani control individuals. In addition to c.2035G>A, a 250 bp insertion just following exon 9 was also identified in the members of the family. However, after genotyping Pakistani controls by polymerase chain reaction amplification across exon 19, it was apparent that this is a relatively common polymorphism. All primer sequences are available upon request. The 679^{Gly} residue is highly conserved across evolution of the animal kingdom (Figure 1E), and *in silico* analyses using POLYPHEN and SIFT identify the substitution as being "possibly damaging" and "not tolerated" respectively. A screening set of 45 NSID individuals of European descent (negative for mutations at *FMR1* and *MECP2*) did not show any mutations in *NSUN2*.

NSUN2 encodes a methyltransferase that catalyzes the intron-dependent formation of 5-methylcytosine at C34 of tRNA-leu(CAA)[9]. It also functions in spindle assembly during mitosis as well as chromosome segregation[10]. Previous work on constructs of the mouse homologue Nsun2 carrying the missense change Lys190Met have shown ablation of methyltransferase catalytic activity[10]. Lys190Met has not been identified in any human subjects.

Using a cDNA clone for *NSUN2* in the pcDNA3.1-Myc vector, site-directed mutagenesis was performed to generate the c.2035G>A/p.Gly679Arg mutation. Wild type (WT) and mutant constructs were transfected into the human breast cancer cell line HCC1954, and also into COS7 (monkey kidney) cells. 24 hours later, cells were stained with antibodies to the Myc epitope in order to detect transfected proteins. Whilst the WT NSUN2 protein was detected in the nucleus and nucleolus of transfected HCC1954 cells (Figure 2A), the p.Gly679Arg mutant fails to localize to the nucleoli in most transfected cells (Figure 2B-C). Antibodies to the nucleolar marker protein, nucleophosmin (NPM1) were used to confirm co-localization in the nucleoli (Figure 2A,B) and co-labelling for endogenous NSUN2 protein confirmed nucleolar localization of endogenous NSUN2 (Figure 2C).

Instead of being localized to the nucleoli, the p.Gly679Arg NSUN2 mutant protein accumulated in the nucleoplasm in 80% of transfected cells (Figure 3A,B,D). In less than 10% of transfected cells, it was observed that the p.Gly679Arg mutant NSUN2 did localize to the nuceoli and at the same time showed intense staining within the cytoplasm (Figure 3A,C,D). Also of note was the fact that the mutant protein seemed to be largely excluded from the nucleoplasm in these cells (Figure 3C). Similar results were also seen for the COS7 cells, with the mutant NSUN2 protein localizing mainly in the cytoplasm. We conclude that the substitution of glycine to arginine at position 679 impairs the proper cellular localization of NSUN2.

In parallel, cDNA constructs for WT and mutant were generated in the pcDNA3.1 vector with GFP tag, and transfected into the human endothelial cell line EA.hy 926 (from umbilical vein). In EA.hy 926 cells WT NSUN2-GFP colocalizes with the nucleophosmin 1 antibody (Santa Cruz) in the nucleoli, whereas the p.Gly679Arg mutant NSUN2-GFP remains in the nucleoplasm (data not shown).

In order to gain some insight into NSUN2 function in the brain we sought to identify specific mouse neural cell types that displayed NSUN2 protein localization. To this end we dissected whole brain from three month old mice into cortical and cerebellar regions. Tissue was fixed for 24 hours in paraformaldehyde, paraffin sections were taken, antigen retreival perfomed and sections stained with affinity purified NSUN2 antibody (Covalab). Although we were able to observe NSUN2 staining sporadically in some cortical and brain stem neurons (Supplementary Materials, Figure 2), by far the most striking localization was observed in Purkinje cells of the cerebellum (Figure 4A,B). Higher magnification imaging revealed that NSUN2 localises to the nucleoli of Purkinje cells and that these nucleoli were often located between or adjacent to dense heterochromatic regions (Figure 4C,D). Interestingly, in addition to intellectual disability, *NSUN2* individuals in the current study display features such as poor speech/dysarthria and broad gait which have previously been associated with cerebellar defects thus making it likely that disruption of Purkinje cell function of NSUN2 contibutes to the phenotype of these individuals. Given the developmental phenotype in the Pakistani family, we have also demonstrated NSUN2 expression in cerebellum of 3 day old mice, using LacZ staining (see Supplementary Material, Figure 2), suggesting that NSUN2 is expressed in these tissues during development as well as adulthood. It is also worth noting that, although intellectual disability is not usually associated with cerebellar defects, but much more commonly with cortical defects, there are several reports that describe intellectual disability in association with cerebellar dysfunction[11-14]. Purkinje cells are a class of GABAergic neuron and are at the heart of the cerebellar circuitry, receiving more synaptic inputs than any other cell type in the brain. It is interesting that the *NSUN2* missense variant identified in the ID family fails to localize to the nucleolus, as it hints at a mechanism by which proper Purkinje cell function might be inhibited

due to aberrant exclusion of NSUN2 from the nucleoli of these cells. In more recent years the nucleolus has become established as a multifunctional entity rather than simply a ribosomal RNA processing center, as previously envisioned[15]. A better understanding of the nucleolar functioning of NSUN2 in addition to identification of its methylation substrates will no doubt shed light on the cellular mechanisms disrupted in *NSUN2* mutation individuals.

In a recent study of *Nsun2* in mice, a knockout was generated that leads to ablation of *Nsun2* through deletion of exon 8[16]. Heterozygous mice appeared normal, with no visible phenotype. *Nsun2* -/- mice were also viable, and gross phenotype indicated weight loss (~30% reduction at 3 months old) and partial alopecia at ~10 months, suggesting a role for NSUN2 in skin homeostasis. *Nsun2* -/- males were sterile. The small size of the -/- mice may draw a parallel with the reduced growth in the affected individuals from the Khairpur family. Heights of the affected girls are below the mean for Pakistani girls (from the United Kingdom[17]) at ~10th centile for II:3 and 1st centile for II:4. Weight is strikingly low, below the 1st centile for II:3 and II:4 (according to standard growth charts[17]). Interestingly, in one of the affected individuals from the Khairpur family, ovaries were either atrophied or absent. However no information on sexual development or sterility in affected male family members was available. Studies of an *Nsun2* knockout mouse generated by the European Conditional Mouse Mutagenesis program (EUCOMM) has been reported by the Wellcome Trust Sanger Institute, and some preliminary data are available online from the Mouse Resources group. This mouse has a deletion of exon 6, resulting in a frameshift. Neurobehavioral and neurocognitive analysis of heterozygous and homozygous mice using hot plate tests, open field tests and a modified SHIRPA test[18] showed no overall difference to control mice, however male homozygotes are reported as hyperactive in calorimetric studies. Male homozygous mutant mice showed significantly reduced forepaw grip

strength, and abnormal humerus morphology is reported for both sexes. Body size was decreased for male and female homozygotes, and abnormal skull and teeth morphology was reported. An eye phenotype was also present in homozygotes, with abnormality of the cornea present as well as increased lens opacity. Fertility/fecundity are also reported as abnormal. These data were provided by and can be obtained from the Mouse Resources group at the Wellcome Trust Sanger Institute. More in-depth neurocognitive assessments would be particularly useful on this model. It may also be important to compare future clinical developments in the family for comparison with the mouse knockout, for instance, whether signs of alopecia develop later on.

NSUN2 is now the third RNA methyltransferase gene to be linked to ID. Previously, *FTSJ1* was identified in X-linked non-syndromic ID[19,20], and very recently *TRMT1* was identified as a cause of ARID[21]. Coexpression data from zebrafish, fruit fly and yeast have linked NSUN2 protein to both TRMT1 (STRING; coexpression score = 0.877), and to FTSJ1. Co-immunoprecipitation assays have also shown protein-protein interaction between NSUN2 and FTSJ1 (IntAct interaction database). Although NSUN2 was initially thought to have a very narrow range of tRNA targets, this range is broadening and is currently also thought to include mRNAs[22]. A role for NSUN2 in DNA methylation has not been excluded, however seems unlikely given its nucleolar localization.

Thus, our findings strongly support a role for mutations in *NSUN2* in ARID, for which additional clinical symptoms are apparent. A publication on gene mapping in Iranian families has also identified this region[8,23], and in this volume an article from this group also reports the identification of truncating mutations in the same gene, *NSUN2*, in several Iranian and one Kurdish consanguineous ARID families[24]. Thus, this is one of the few ARID genes for which there is validation in several independent families, and from more than one ethnic group or

geographic location. Given the similarities in the phenotypes of the families from this study and that of Abbasi-Moheb et al[24], we speculate that the missense mutation reported here has a similar effect to the truncation mutations in *NSUN2*, and that Gly679Arg results mainly in a loss of function due to failure of the protein to localize to the correct cellular organelles resulting in intellectual disability. We speculate further that the myopathy reported in the Pakistani family was not noted in either the Iranian or the Kurdish *NSUN2* truncating mutation families, and may result from a gain of function due to accumulation of NSUN2 protein in nucleoplasm and cytoplasm.

Supplemental Data

Supplemental Data include two figures and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgements

We wish to thank all the members of this family for their willing participation and cooperation with this study. We also wish to acknowledge Claire Simon, who assisted with the cell staining. MAK was supported by Indigenous Scholarship and International Research Support Program (IRSIP) from the Higher Education Commission (HEC) of Pakistan. JBV is a National Alliance for Research on Schizophrenia and Depression Independent Investigator. This research was supported by grants from the HEC (NRPU-1118), and the Canadian Institutes of Health Research (#MOP-102758). We declare that the authors have no competing interests for this article.

Web Resources

The URLs for data presented herein are as follows:

1000 Genome Project, http://www.1000genomes.org

CLUSTAL 2.1 multiple sequence alignment, http://www.ebi.ac.uk

NHLBI Exome Sequencing Project, http://evs.gs.washington.edu/EVS/Online Mendelian Inheritance in Man (OMIM), http://www.omim.org

POLYPHEN, http:// genetics.bwh.harvard.edu/pph

SIFT, http://sift.jcvi.org/

STRING, http://string-db.org/

UCSC Genome Bioinformatics, http://genome.ucsc.eduWellcome Trust Sanger Institute, Mouse Resources group, http://www.sanger.ac.uk/mouseportal

References

1. Roeleveld, N., Zielhuis, G.A., Gabreels, F. (1997). The Prevalence of Mental Retardation: A Critical Review of Recent Literature. Dev. Med. Child. Neurol. 39, 125-132.
2. Ropers, H.H. (2008). Genetics of intellectual disability. Curr. Opin. Genet. Dev. 18, 241-250.
3. Kaufman, L., Ayub, M., Vincent, J.B. (2010). The Genetic Basis of Non-syndromic Intellectual Disability: A Review. J. Neurodev. Dis. 2, 182-209.
4. Wing, L., Leekam, S.R., Libby, S.J., Gould, J., Larcombe, M. (2002). The Diagnostic Interview for Social and Communication Disorders: background, inter-rater reliability and clinical use. J. Child Psychol. Psychiatry 43, 307-325.

5. Li, C., Wong, W.H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc. Natl. Acad. Sci. U.S.A. 98, 31-36.
6. Zhao, X., Li, C., Paez, J.G., Chin, K., Jänne, P.A., Chen, T.H., Girard, L., Minna, J., Christiani, D., Leo, C., et al. (2004). An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. Cancer Res. 64, 3060-3071.
7. Zhao, X., Weir, B.A., LaFramboise, T, Lin, M., Beroukhim, R., Garraway, L., Beheshti, J., Lee, J.C., Naoki, K., Richards, W.G., et al. (2005). Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. Cancer Res. 65, 5561-5570.
8. Najmabadi, H., Motazacker, M.M., Garshasbi, M., Kahrizi, K., Tzschach, A., Chen, W., Behjati, F., Hadavi, V., Nieh, S.E., Abedini, S.S., et al. (2007). Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. Hum Genet 121, 43-48.
9. Brzezicha, B., Schmidt, M., Makalowska, I., Jarmolowski, A., Pienkowska, J., Szweykowska-Kulinska, Z. (2006). Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the pre-tRNA Leu (CAA). Nucleic Acids Res. 34, 6034-6043.
10. Hussain, S., Benavente, S.B., Nascimento, E., Dragoni, I., Kurowski, A., Gillich, A., Humphreys P, Frye M. (2009). The nucleolar RNA methyltransferase Misu (NSun2) is required for mitotic spindle stability. J. Cell Biol. 186, 27-40.
11. Schurig, V., Orman, A.V., Bowen, P. (1981). Nonprogressive cerebellar disorder with mental retardation and autosomal recessive inheritance in Hutterites. Am. J. Med. Genet. 9, 43-53.

12. Christianson AL, Stevenson RE, van der Meyden CH, Pelser J, Theron FW, van Rensburg PL, Chandler M, Schwartz CE. (1999). X-linked severe mental retardation, craniofacial dysmorphology, epilepsy, ophthalmoplegia, and cerebellar atrophy in a large South African kindred is localized to Xq24-q27. J. Med. Genet. 36, 759-766.
13. Ventura, P., Presicci, A., Perniola, T., Campa, M.G., Margari, L. (2006). Mental retardation and epilepsy in patients with isolated cerebellar hypoplasia. J. Child Neurol. 21, 776-781
14. Boycott, K.M., Bonnemann C, Herz J, Neuert S, Beaulieu C, Scott JN, Venkatasubramanian A, Parboosingh JS. (2009). Mutations in *VLDLR* as a cause for autosomal recessive cerebellar ataxia with mental retardation (dysequilibrium syndrome). J. Child Neurol. 24, 1310-1315.
15. Pederson, T., Tsai, R.Y.L. (2009). In search of nonribosomal nucleolar protein function and regulation. J. Cell Biol. 184, 771-776.
16. Blanco, S., Kurowski, A., Nichols, J., Watt, F.M., Aznar Benitah, S., Frye, M. (2011) The RNA-methyltransferase Misu (NSun2) Poises Epidermal Stem Cells to Differentiate. PLoS Genet. 7, e1002403.
17. Kelly, A.M., Shaw, N.J., Thomas, A.M.C., Pynsent, P.B., Baker, D.J. (1997). Growth of Pakistani children in relation to 1990 growth standards. Arch. Dis. Child 77, 401-405.
18. Rogers, D.C., Fisher, E.M., Brown, S.D., Peters, J., Hunter, A.J. and Martin, J.E. (1997). Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. Mamm. Genome 8, 711–713.
19. Freude, K., Hoffmann, K., Jensen, L.R., Delatycki, M.B., des Portes, V., Moser, B., Hamel, B., van Bokhoven, H., Moraine, C., Fryns, J.P. et al. (2004). Mutations in the

FTSJ1 gene coding for a novel S-adenosylmethionine-binding protein cause nonsyndromic X-linked mental retardation. Am. J. Hum. Genet. 75, 305-309.

20. Ramser, J., Winnepenninckx, B., Lenski, C., Errijgers, V., Platzer, M., Schwartz, C.E., Meindl, A., Kooy, R.F. (2004). A splice site mutation in the methyltransferase gene FTSJ1 in Xp11.23 is associated with non-syndromic mental retardation in a large Belgian family (MRX9). J. Med. Genet. 41, 679-683.
21. Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Zemojtel, T., Abedini, S.S., Chen, W., Hosseini, M., Behjati, F., Haas, S., et al. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 478, 57-63.
22. Squires, J.E., Patel, H.R., Nousch, M., Sibbritt, T., Humphreys, D.T., Parker, B.J., Suter, C.M., Preiss, T. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res. [epub ahead of print]
23. Kuss, A.W., Garshasbi, M., Kahrizi, K., Tzschach, A., Behjati, F., Darvish, H., Abbasi-Moheb, L., Puettmann, L., Zecha, A., Weissmann, R., et al. (2011). Autosomal recessive mental retardation: homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots. Hum. Genet. 129, 141-148.
24. Abbasi-Moheb et al. Mutations in *NSUN2*: a frequent cause for autosomal recessive intellectual disability in Middle Eastern populations (*Submitted back-to-back with current paper*)
25. Lindner, T.H., Hoffmann, K. (2005). easyLINKAGE: a PERL script for easy and automated two-/multi-point linkage analyses. Bioinformatics 21, 405-407.
26. Seelow, D., Schuelke, M., Hildebrandt, F., Nürnberg, P. (2009). HomozygosityMapper--an interactive approach to homozygosity mapping. Nucleic Acids Res. 37, 593-599.

Table 1. Anthropometric measures and clinical details for affected members of the family. Abbreviations used are: IQ- intelligence quotient; OFC- occipitofrontal circumference; LDH – lactose dehyrogenase; CPK – creatine phosphokinase; F - female. Comparison of clinical features between this family and the Iranian and Kurdish families described in Abbasi-Moheb et al (submitted) is shown in Table 1, Abbasi-Moheb et al. Photographs of affected individuals are shown in Supplementary Materials, Figure 1.

Family ID	Sex	Age	IQ	OFC (cm)	Height (cm)	Weight (Kg)	LDH (IU/l)	CPK (IU/l)	Epilepsy	Gait
II:3	F	14	40-50	50	152	32	447[b]	294[c]	-	Broad
II:4	F	13	40-50	49	136	26.3	NK[a]	NK[a]	-	Broad
II:5	F	6	40-50	46	NK[a]	NK[a]	463 [b]	60[c]	-	Normal

Family ID	Feet/toes	Eyes	Menses	Pelvic ultrasound
II:3	Bilateral *pes cavus*; left Achilles tight	Right eye strabismus	No (Tanner stage 3)	Small midline uterus, thin endometrium, ovaries not visible
II:4	left foot *pes cavus*; bilateral equinus position and tight Achilles; hyper-separation between $1^{st}/2^{nd}$ toes	Fundoscopy, eye movement & pupillary response normal	Yes (Tanner stage 4)	NK[a]
II:5	Partial, $4^{th}/5^{th}$ toe, bilateral syndactyly	Intermittent esotropia; fine horizontal nystagmus	No	NK[a]

[a]NK: not known or not measured
[b]: reference level: 122-234 IU/l
[c]: reference level: 34-145 IU/l

Table 2. A. Two-point linkage analysis for markers across chromosome 5p. The EasyLINKAGE program[25] was used. Physical position according to UCSC Feb 2009 (GRCh37/hg19) assembly. For reference, the common HBD locus for the Pakistani family and the Iranian family[8] extends from SNPs rs1824938 (5.092 Mb) to rs2914296 (7.657 Mb), and *NSUN2* is located between 6.599 and 6.633 Mb.

Markers	**Pos-cM**	**Phy-Pos Mb**	**LOD Score at Recombination Fraction, θ**			
			0.00	**0.05**	**0.2**	**0.4**
D5S1981	1.7200	1.155	0.4327	0.3972	0.2597	0.0585
D5S406	**11.8500**	**4.994**	**2.7729**	**2.4887**	**1.6231**	**0.5141**
D5S2505	14.3000	5.817	1.6695	1.4721	0.8628	0.1437
D5S580	17.8700	8.141	0.4327	0.3972	0.2597	0.0585
D5S630	19.6710	9.561	0.4327	0.3972	0.2597	0.0585

Figure 1. A. Pedigree of family MR14 from Khairpur district. Filled circles indicate affected girls. **B.** HomozygosityMapper analysis[25] for microarray SNP data: Genome-wide. Significant regions of HBD are seen only on 5p and 14q. The 14q locus was excluded because one of the unaffected siblings was also homozygous at this locus, whereas at the 5p locus, unaffected the sibling was genotyped as heterozygous. **C.** Ideogrammatic representation of the critical autozygous or HBD locus on 5p15.31, as determined from this study and in relation to the MRT5 locus identified by Najmabadi et al.[8]. **D.** c.2035G>A substitution encoding the Gly679Arg change in a heterozygous carrier and affected homozygote. **E.** ClustalW alignment of NSUN2 across multiple species showing conservation of the Gly679 residue in vertebrates and also in non-vertebrate animal species.

Figure 2. Wild type (WT) **(A)** and mutant **(B)** constructs for *NSUN2* in the vector pcDNA-Myc transfected into the human breast cancer cell line HCC1954. A. WT NSUN2 (A) but not mutant NSUN2 (B) protein co-localizes with nucleophosmin, a nucleolar marker. (C) Co-staining for endogenous NSUN2 confirms exclusion of mutant NSUN2 (Myc-labelled) from the nucleoli. Arrows indicate nucleoli (A-C). Co-staining with DAPI shows nuclear localization.

Figure 3. (A) Immunostaining showing nucleolar localization of wild-type (WT) NSUN2 in HCC1954 cells compared to cellular localization of NSUN2 carrying the 679^{Arg} variant (Gly679Arg) in the nucleoplasm (B) and cytoplasm (C). Co-staining with DAPI shows nuclear localization. (D) Quantification of cellular localizations shown in (A-C). (E) Nucleolar localization of GFP-tagged wild type NSUN2 in HeLa cells versus (F) nuclear and cytoplasmic localization in 679^{Arg} mutant NSUN2 (Gly679Arg). White arrow heads indicate nucleoli.

Figure 4. Sections of mouse cerebellum, (A) labeled with an antibody to NSUN2. Higher magnification of (A; insert) shows NSUN2 protein in Purkinje cells (B). NF: nerve fibers; GCL: granule cell layer; ML: molecular layer. (C) Sections of the cerebellum labeled for L7/Pcp-2, as a marker for Purkinje cells. (D) Co-localization of NSUN2 with nucleophosmin (Npm1) in nucleoli of Purkinje cells. Nuclei are counter stained with DAPI.