Identification of a DLG3 stop mutation in the MRX20 family

Identification of a *DLG3* stop mutation in the MRX20 family

Jolien Huyghebaert[1], Ligia Mateiu[1], Ellen Elinck[1], Kirsten Esther Van Rossem[1], Bregje Christiaenssen[1], Claudio Peter D'Incal[1], Michael K. McCormack[2,3], Alice Lazzarini[4], Geert Vandeweyer[1], R. Frank Kooy[1]*

Affiliations:

1) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium.
2) Department of Psychiatry, Rutgers University-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, USA.
3) Department of Cell Biology and Neurosciences, Virtua Health College of Medicine and Life Sciences of Rowan University, Stratford, New Jersey 08084, USA.
4) Department of Neurology, Rutgers University-Robert Wood Johnson Medical School, New Brunswick, New Jersey 08903, USA.

*Corresponding author:

Prof. dr. R. Frank Kooy

Department of Medical Genetics

University of Antwerp

Prins Boudewijnlaan 43/6

2650 Edegem

Belgium

Tel.: +32 3 2759760

E-mail: Frank.Kooy@uantwerpen.be

Abstract

Here, we identified the causal mutation in the MRX20 family, one of the larger X-linked pedigrees that have been described in which no gene had been identified up till now. In 1995, the putative disease gene had been mapped to the pericentromeric region on the X chromosome, but no follow-up studies were performed. Here, whole exome sequencing (WES) on two affected and one unaffected family member revealed the c.195del/p.(Thr66ProfsTer55) mutation in the *DLG3* gene (NM_021120.4) that segregated with the affected individuals in the family. *DLG3* mutations have been consequently associated with intellectual disability and are a plausible explanation for the clinical abnormalities observed in this family. In addition, we identified two other variants co-segregating with the phenotype: a stop gain mutation in *SSX1* (c.358G>T/p.(Glu120Ter)) (NM_001278691.2) and a nonsynonymous SNV in *USP27X* (c.56A>G/p.(Gln19Arg)) (NM_001145073.3). RNA sequencing revealed 14 differentially expressed genes (p-value < 0.1) in 7 affected males compared to 4 unaffected males of the family, including four genes known to be associated with neurological disorders. Thus, in this paper we identified the c.195del/p.(Thr66ProfsTer55) mutation in the

54 *DLG3* gene (NM_021120.4) as likely responsible for the
55 phenotype observed in the MRX20 family.
56
57 Key words: MRX20, XLID, *DLG3*, Intellectual Disability

Introduction

According to the fifth edition of the 'Diagnostic and statistical manual of mental disorders (DSM-5), intellectual disability (ID) is a neurodevelopmental disorder characterized by deficits in cognition and adaptive function with an onset during the developmental period [1]. It is estimated that 1 – 3% of the population is affected, with a male to female ratio of 1.6:1 [2]. The sex difference in frequency is commonly attributed to the excess of ID genes on the X chromosome [3]. Although the X chromosome covers only 5% of the human genome, it contains 15% of the genes currently known to be associated with ID [3]. Before whole exome sequencing (WES) facilitated the analysis of all genes in the genome, families with X-linked inheritance were prioritized for disease-gene identification studies because an X-linked pattern of inheritance facilitated the identification of affected families. Subsequent linkage analysis narrowed down the region of interest to a specific region of the X chromosome, further reducing the number of potential candidate-genes. A classification system was put in place for families with a LOD-score above 2, indicating significant linkage to the X chromosome, and such families were numbered in sequential order of discovery as MRX (for non-syndromic) or MRXS (for syndromic) families [3, 4, 5]. This distinction between

syndromic and non-syndromic, in retrospect, has been often arbitrary since even despite a careful clinical evaluation, the syndromic features common to all members of any family can be difficult to recognize and may be age-dependent [4]. In total, 105 families with MRX received MRX numbers, after which this tradition was no longer continued. For 67 of these families, a causative gene has been reported, significantly aided by large-scale initiatives such as Euro-MRX and GenCodys [3]. X-linked families thus contributed significantly to the discovery of MRX genes, of which the total number is estimated to be 141 according to the latest update [3].

Here, we studied the MRX20 family, a large pedigree in which initial linkage studies mapped the putative disease gene to a 55.6 cM interval in the pericentromeric region of the X chromosome, between the short tandem repeat polymorphism markers DXS1068 (Xp11.4-p21, hg19: chrX:38908118-38908368) and DXS454 (Xq21.1-q23, hg19: chrX:97986121-97986265) [6]. This family presents with ID but no obvious other clinical manifestations were found. In this study, we were able to identify a causal mutation in the *Discs Large MAGUK Scaffold Protein 3 (DLG3)* gene. Further, we found two other variants in the linked region, which may or may not play an additional role in the disease manifestation in this family.

Methods

Collection of patient data

V.3 was referred to AL and MKM for genetic counseling regarding a family history of ID. Her brother, V.2, was examined by clinical geneticist Cheryl S. Reid, M.D.. At 21 years of age he presented as an affable young man with obesity, a mildly gynecoid habitus, mild micrognathia and dysarthric speech but had no physical findings suggestive of a specific syndrome. Available medical records showed a birth weight of 2.3 kg and an initial referral for neurologic evaluation only at an age of 5 years and 8 months for learning difficulties and hyperactivity.

Family history revealed three maternal uncles of the proband reported to be similarly affected, as were three cousins of these uncles, thus likely three obligate carrier females (II.3, III.3 and III.6). The limited medical records still extant for IV.2 and IV.4, were accessed and reported below and in the original publication [6]. Both wards of the state and 50-year residents of group homes, they appeared more severely debilitated than V.2 who lived at home nurtured by a supportive family where he had managed to hold supervised employment.

Mindful of the power presented by the extensive pedigree to identify a causative gene, relevant blood samples for DNA analysis were deposited at the Human Genetic Mutant Cell Repository at the Coriell Institute (Camden, New Jersey, USA) with informed consent.

Post-mortem cerebellar tissue of an unaffected 9-year old child was obtained from the institute Born-Bunge vzw IBB NeuroBioBank with approval of the Ethics Committee of the Antwerp University Hospital.

Cell culture

Epstein-Barr virus transformed lymphoblastoid cell lines of 19 family members were obtained from Coriell Institute: III.1 (Coriell 400718), III.4 (Coriell 400719), III.5 (Coriell 400717), IV.1 (Coriell 400706), IV.2 (Coriell 400705), IV.3 (Coriell 400703), IV.4 (Coriell 400708), IV.5 (Coriell 400707), IV.6 (Coriell 400709), IV.7 (Coriell 400715), IV.8 (Coriell 400716), IV.9 (Coriell 400710), IV.10 (Coriell 400711), IV.11 (Coriell 400704), IV.12 (Coriell 400712), IV.13 (Coriell 400701), V.1 (Coriell 400714), V.2 (Coriell 400700), V.3 (Coriell 400713). All cell lines were cultured in RPMI (Life Technologies, Carlsbad, California, USA), supplemented with 15% fetal bovine serum (Life Technologies), 1% penicillin/streptomycin (Life Technologies), 1% sodium

pyruvate (Life Technologies), and 1% GlutaMAX (Life Technologies).

DNA extraction and WES

Genomic DNA was extracted using the DNeasy® Blood & Tissue Kit (Qiagen, Hilden, Germany) following manufacturer's instructions. WES was executed on DNA of two affected members (IV.3 and V.2) and one unaffected member (III.3) by BGI (Copenhagen, Denmark). Sequencing was performed using TruSeq DNA sample preparation (Illumina, San Diego, California, USA) and SureSelect Human All Exon V5 kit (Agilent, Santa Clara, California, USA) according to the standard protocols. Sequencing was performed on an Illumina HiSeq 4000 using a 2 × 150 bp sequencing run. Data-analysis was done using an in-house pipeline as described before [7]. Data filtering and annotation of variants in the linkage interval was performed with VariantDB [7, 8].

Sanger Sequencing validation

Sanger sequencing was performed with primers listed in Supplementary Table 1.

X-inactivation experiments

An X-inactivation assay was performed on the genomic DNA of following female family members: III.1, III.5, IV.1, IV.6, IV.12, IV.13 and V.3 based on the protocol described by Jones et al., 2014 [9]. X-skewing was determined by fragment analysis of the *AR* gene and the *RP2* gene. Fragments were analyzed on an ABI3130XL (Applied Biosystems, Waltham, Massachusetts, USA) in the presence of an internal sizing standard (ROX). Amplicon sizes were determined using GeneMarker v2.6.4 (SoftGenetics, State College, Pennsylvania, USA). Calculation of the X-inactivation ratio was performed using the areas under the allele peaks with or without *HpaII* cleavage. The ratio of X-inactivation is interpreted as follows: <80:20 is random; 80:20 to 90:10 is moderately skewed; >90:10 is highly skewed.

DLG3 expression analysis

DLG3 mRNA levels were quantified in lymphoblastoid cell lines using Real-Time PCR (RT-PCR) as described below. Protein expression was evaluated according to the methods as previously published [10]. Briefly, 20 µg of protein was separated using SDS-PAGE and transblotted to a nitrocellulose membrane which was incubated with N-terminal SAP102 (Invitrogen, Waltham, Massachusetts, USA; PA5-51626, 1/1000 dilution) and C-terminal SAP102 (Abcam, Cambridge, UK; ab288436, 1/1000 dilution) primary antibodies.

198

199 **RNA extraction and sequencing**

200 RNA was extracted using the Quick-RNA™ Miniprep Kit (Zymo
201 Research, Irvine, California, USA) following manufacturer's
202 instructions. RNA sequencing was performed on following
203 affected males; IV.2, IV.3, IV.4, IV.9, IV.10, IV.11, V.2, and
204 unaffected males; IV.5, IV.7, IV.8, V.1, of the family. Fragment
205 analysis was performed using the RNA kit (DNF-471), standard
206 sense RNA analysis kit (15nt) of Agilent and RNA sequencing
207 was performed using the QuantSeq 3' mRNA-Seq Library Prep
208 Kit FWD for Illumina (Lexogen, Vienna, Austria) following
209 manufacturer's instructions. The RNA-seq data was analyzed by
210 the trimming and cleaning with bbduk [11], the alignment with
211 STAR [12] and the feature extraction with Subread
212 featureCounts [13]. Differential expression analysis for the
213 protein coding genes was conducted using DESeq2 [14] in R and
214 Benjamini-Hochberg adjusted p-values controlling for false
215 discovery rate at 10%.

216

217 **Pathway enrichment analysis**

218 Pathway analysis was performed using the Web-based Gene set
219 analysis toolkit WebGestalt with the GSEA method, and KEGG
220 database [15].

221

Identification of enriched transcription factors among differentially expressed genes

Enriched transcription factors among the fourteen differentially expressed genes (p-value < 0.1) observed through RNA sequencing were searched for with the plugin IRegulon v1.3 in Cytoscape [16]. Following settings were used: motif collection of 10K [9,713 position weight matrices (PWMs)], track collection of 1,120 ChiP-seq tracks, the putative regulatory region of 20 kb centered around transcription start site (TSS), motif ranking database 20 kb centered around TSS (seven species), and track ranking database of 20 kb centered around TSS (ChiP-seq-derived). In addition, we used a normalized enrichment score (NES) threshold of 5.0, a ROC threshold for AUC calculation of 0.03, and a rank threshold of 5,000. For transcription factor (TF) prediction, the maximum False Discovery Rate (FDR) on the motif similarity threshold was 0.001. To strengthen the link between the targets of the top ranked enriched transcription factor we used the TFlink database, that provides comprehensive information on transcription factors and their targets [17].

RT-PCR validation

Real-time PCR (RT-PCR) was used to examine the differential expression. 1 μg of total RNA was converted to cDNA using the

Superscript III First-Strand Synthesis System (Invitrogen). The primer design was performed using an in-house automated pipeline [18] with primers listed in Supplementary Table 1. Quantitative PCR was performed using the qPCR Mastermix Plus for SYBR Green I – no ROX (Eurogentec, Seraing, Belgium) following manufacturer's instructions on a CFX384 Real-time system (Bio-Rad, Hercules, California, USA). Statistical analysis was performed using the qBASE+ software (CellCarta, Montreal, Canada). The data was normalized to *ACTB*, *UBC* and *YWHAZ*, and stability of these reference genes were checked with the qBASE+ software. Statistical analysis was performed in GraphPad Prism 9.0 using a two-tailed Mann-Whitney U test.

Results

Patient Data

ID was observed in seven males across two generations in a family originating from the USA (Figure 1) [6]. More extensive clinical information was retrieved from individuals IV.2, IV.3 and V.2 of which relevant abnormalities are briefly summarized here.

V.2 demonstrated mild ID, learning difficulties and hyperactivity. There were no complications during pregnancy.

His psychomotor development was delayed as he did not walk until 19 months of age. At 9 years of age, impairment in both gross and fine motor function was observed as well as microcephaly (10^{th} percentile). At 18 months of age, alternating exotropia was observed, which was confirmed at a check at 9 years of age where he exhibited with 10° of exotropia. During consultation at 21 years of age, no exotropia could be observed. No facial dysmorphia was observed apart from mild micrognathia.

IV.2 presented with severe ID and a mental age of 4 years and 11 months (IQ ratio 27 on Stanford-Binet intelligence Scale) at age 55.

IV.3 presented with moderate ID. At 46 years of age he scored 51 for Verbal IQ, 52 for Performance IQ and 48 for full scale IQ at the Wechsler Adult Intelligent Scale (WAIS). At 54 years of age, he scored 36 on Stanford-Binet intelligence Scale and his mental age was 5 years and 9 months.

Individuals IV.4, IV.9, IV.10 and IV.11 were diagnosed with ID without further specifications.

Gene identification/Mutation identification

A single base pair mutation in the *DLG3* gene was identified using WES; c.195del/p.(Thr66ProfsTer55) (NM_021120.4). Segregation of this stop mutation was confirmed in all affected individuals of this family and obligate carrier females using Sanger sequencing (Figure 1 and 2). The mutation was not identified in the unaffected relatives. Loss-of-function (LoF) mutations in *DLG3* have consistently been associated with ID and several other families had been described **[19, 20, 21, 22, 23, 24, 25, 26]**. Beyond the published mutations, ClinVar reports 17 additional pathogenic or likely pathogenic LoF or splice site variants. We conclude that this *DLG3* mutation on its own may explain the clinical presentation of this family.

In addition, we found a LoF variant (stop gain), c.358G>T/p.(Glu120Ter) (NM_001278691.2), in the *SSX1* gene with a CADD-Phred 1.4 score of 33.0 and a variant of unknown significance (VUS), c.56A>G (nonsynonymous SNV)/p.Gln19Arg (NM_001145073.3), in the *USP27X* gene with a CADD-Phred 1.4 score of 20.2, both segregating with the disease in the same manner as the *DLG3* mutation (Figure 1 and 2).

X-inactivation assay

Skewed X chromosome inactivation occurs frequently in families with X-linked ID and has been previously observed in a

family with a deleterious *DLG3* mutation [26, 27]. Here, moderately skewed X-inactivation was observed in the only available obligate carrier based upon two independent markers. Apart from the obligate carrier, we determined the skewing pattern for six additional females in the family, of which four presented with a nonrandom X-inactivation pattern (Figure 1).

DLG3 expression

DLG3 encodes the SAP102 protein, a post-synaptic density protein. There are approximately 10 different transcripts reported originating from the *DLG3* gene (Ensemble and GTEx), of which four are protein coding [28, 29]. Two of these transcripts, ENST00000374360.8 and ENST00000194900.8, are translated to a protein of 90 - 93 kDa and are predominantly brain-specific. These isoforms contain 19 and 21 exons, respectively. The shorter transcripts, ENST00000374355.8 and ENST00000542398.1, contain 14 and 12 exons and are translated to proteins of 58 kDa an 42 kDa, respectively. These isoforms are more widely expressed throughout different human tissues. However, these transcripts lack exon 1 and hence the described mutation. We investigated *DLG3* expression at the mRNA and protein level to compare the expression of several isoforms in lymphoblastoid cell lines of

our family and an unaffected human brain, as a positive control. At the RNA level, we could detect expression of both the small and large transcripts in the control brain. However, in the lymphoblastoid cell lines only small quantities of the shorter transcript could be reliably detected (Supplementary Figure 1A). This is in line with the data available on GTEx [29]. At the protein level, we observed the canonical SAP102 isoform in the human brain, with both an N-terminal (detecting the large isoforms, Supplementary Figure 1B) and C-terminal antibody (detecting both the large and small isoforms, Supplementary Figure 1C). However, we were not able to detect this larger isoform in lymphoblastoid cell lines derived from both affected and unaffected family members of the MRX20 family. In addition, a smaller SAP102 isoform (approximately 42 kDa), detected by the C-terminal antibody, showed modest expression in the control human brain, but was too low to detect in lymphoblastoid cell lines (Supplementary Figure 1C).

Differential expression analysis

To unravel the impact of the WES identified variants, we conducted 3' mRNA sequencing and performed differential expression analysis on lymphoblaistoid cell lines derived from the males of this family. Here, we identified 14 genes with an adjusted p-value lower than 0.1, of which 10 showed an

adjusted p-value lower then 0.05 (Figure 3A). In the affected family members *WWTR1, HLA-DRA, HLA-DPA1, LDHA, CDCA4* and *PPP1R16B* were upregulated, in contrast to *TMEM51, BCL11A, TFE3, NMT2, FRY, DNAJC5, SELENOW* and *PEX26* which were downregulated. From these differentially expressed genes, four are known transcription factors (*WWTR1 [30],CDCA4 [31], BCL11A [32], TFE3 [33]*) and four are associated with neurological disorders (*BCL11A [34], TFE3 [35], FRY [36], DNAJC5 [37]*). Through pathway analysis using WebGestalt [15], we found a significant enrichment of the 'hematopoietic cell lineage' (Figure 3B, Supplementary Table 2). Subsequently, we confirmed a differential expression pattern using RT-PCR for the following five genes using the strict criteria of a two-tailed Mann-Whitney U test; *CDCA4, WWTR1, PEX26, LDHA, NMT2* (Figure 3C). Differential expression of genes *BCL11A, PPP1R16B* were found to be borderline significant with a p-value < 0.1 and > 0.05 (Supplementary Figure 2).

Identification of enriched transcription factors among differentially expressed genes

The iRegulon prediction tool was used to determine enriched transcription factors of the fourteen differentially expressed genes (p-value < 0.1) [16]. We reported all transcription factors with multiple target genes amongst our differentially expressed

genes and with a statistically significant normalized enrichment score (NES) above five (Supplementary Table 3). TFLink was used to provide extra evidence for the link between the most enriched transcription factor and its target genes [17]. Here, the four targets, namely *WWTR1*, *PEX26*, *BCL11A* and *TFE3*, which were enriched for the transcription factor SRF were found in additional databases compared to the original (Supplementary Table 4).

Discussion

The gene found to be causal for ID in this family is the *DLG3* gene on the X chromosome (For an overview of all known published causal variations/mutations in *DLG3;* see Supplementary Table 5 and Figure 4). The canonical form of this gene contains 19 exons and its encoded protein, synapse-associated protein 102 (SAP102), is the major member of the membrane-associated guanylate kinase (MAGUK) family expressed in neurons during the early brain development [28, 38]. MAGUKs are known to be central building blocks for the postsynaptic density, linking surface-expressed receptors to an intracellular signaling molecule [39]. The SAP102 protein was first described by Müller et al. (1996), who stated that this SAP102 protein contains three tandem PDZ domains, an scr

homology (SH3) domain and a guanylate kinase (GK) domain [28]. This widely expressed protein is found in dendrites as well as axons in the cytoplasm and postsynaptic density [38]. A knockout mouse model of *DLG3* revealed the importance of the SAP102 protein for NMDA receptor-driven plasticity, behavior and signal transduction [40]. The mutant mice revealed cognitive deficits with a specific spatial learning deficit, which could be overcome by additional training. Typically, MAGUKs are thought of as stabilizing synaptic proteins, but in contrast to others, the SAP102 is also known to play a role in clearing NMDARs from the synaptic site [41].

In the last decade mutations in the *DLG3* gene were identified in several patients diagnosed with mild to severe ID **[19, 20, 21, 22, 23, 24, 25, 26]**. Including the mutation described here, a total of ten stop mutations and two splice donor site mutations have been reported. A representation of all mutations published to date can be found in Figure 4 and Supplementary Table 5.

As stated before, the *DLG3* gene gives rise to approximately 10 different transcripts (Ensemble and GTEx), of which two (ENST00000374360.8 and ENST00000194900.8) are translated to the larger isoforms of the SAP102 protein and two

436 (ENST00000374355.8 and ENST00000542398.1) are translated
437 to shorter isoforms. The two splicing regions are called I1 and I2
438 and are situated respectively on the N-terminus and between
439 the SH3 and GK domains of SAP102 [28, 42]. The larger
440 isoforms, including the canonical form, are mainly expressed in
441 brain tissue, whereas the shorter isoforms are more widely
442 expressed throughout the body (GTEX) [28, 29]. The isoforms
443 share their C-terminal sequence but the N-terminal sequence is
444 unique to the larger isoforms. The variant found in this study in
445 *DLG3* is positioned at exon 1 and thus is solely harboured by the
446 larger brain-specific isoforms. We determined the expression of
447 *DLG3* both at RNA and at protein level in the Epstein-Barr virus
448 transformed lymphoblastoid cell lines originating from the
449 MRX20 family members, as well as a control brain sample
450 (Supplementary Figure 1). Of these experiments we can
451 conclude that Epstein-Barr virus transformed lymphoblastoid
452 cell lines express the shorter isoform and may express some
453 mRNA of the longer isoform, but lack detectable levels SAP102
454 protein. Our results are in contrast with the data presented by
455 Kumar et al. (2016) that showed expression of the SAP102
456 protein in lymphoblastoid cell lines on western blot. In an
457 attempt to reproduce the protein detection, we repeated the
458 western blot experiments using the extraction protocol as
459 described in their paper (data not shown), but nevertheless, no

protein band at the right size could be observed. We have no explanation for this apparent discrepancy in protein detection between the two studies, except that we used a different antibody as compared to the abovementioned study. Unfortunately, no brain tissue of this family is available to replicate this study on disease relevant tissue.

Nevertheless, using differential transcriptome analysis on lymphoblastoid cell lines, we found 14 genes which were differentially expressed in the affected family members (p-value < 0.1), thereby identifying the 'hematopoietic cell lineage'-pathway as significantly enriched. Enriched transcription factors for the differently expressed genes include the Serum Response Factor (SRF), an important known transcription factor in the brain [43]. The link between SRF and its involved targets was strengthened by the findings in an additional database (TFLink) which relies on chromatin immunoprecipitation assay data [17]. Furthermore, we should take into account that the transcriptomic analysis was performed on RNA extracted from Epstein-Barr virus transformed lymphoblastoid cell lines, which can affect cellular gene expression profiles and activities of cellular pathways [44]. Of note, while the clinical presentation of the affected individuals is most likely a result of the *DLG3* mutation, the

484 observed differential expression observed might be influenced
485 by two other variants co-segregating in this family.
486
487 The stop gain variant in the *SSX1* gene; c.358G>T/p.(Glu120Ter),
488 is unlikely causative on its own, as multiple hemizygous LoF
489 variants in non-neurological controls are present in gnomAD.
490 Exceptionally, gnomAD states that more LoF and missense
491 variants then expected are present in control populations
492 (pLI=0, $Z_{missense}$=-4.37). *SSX1* is a primate specific gene that is
493 mainly expressed in the brain and testis [45]. It competes with
494 SMARCB1 for nucleosome acidic patch binding. SMARCB1 is a
495 subunit of mSWI/SNF complex just as BCL11A which we found
496 to be downregulated in this family [46]. This complex is a
497 chromatin remodeling complex which plays an important role
498 in neurological disorders [47]. Remarkably, haploinsufficiency
499 of the *BCL11A* gene causes Logan-Dias syndrome [34]. Logan-
500 Dias syndrome was discovered by Dias et al. (2016) and is
501 known to be an intellectual developmental disorder with
502 persistence of fetal hemoglobin (HbF). Unfortunately, family
503 members were not available for the testing of HbF persistence.
504 A second variant co-segregating with the disease is a
505 nonsynonymous variant in the *ubiquitin-specific protease 27X*
506 *(USP27X)* gene; c.56A>G/p.(Gln19Arg). *USP27X* is reported as a
507 candidate gene for ID by Hu et al. [48]. However, independent

confirmation has not been reported. In addition, *USP27X* is important in the maintenance of neural stem/progenitor cells by regulating HES1 [49] and aberrant expression of *USP27X* resulted in reduced neuronal differentiation. However, the effect of the missense mutation with CADD 1.4 score of 20,2 in our family is difficult to predict.

In this study we have identified the causal mutation c.195del/p.(Thr66ProfsTer55) for the ID phenotype in the MRX20-family in the *DLG3* gene. Further, differential expression analysis revealed 14 significantly differentially expressed genes between affected and unaffected males in this family (p-value < 0.1). The differential expression pattern might be influenced by two other variants which are co-segregating with the phenotype in this family in the genes, *SSX1* and *USP27X*.

Acknowledgements

We thank Cheryl S. Reid for family consultation.

Author Contributions

Clinical examination and counseling of the family was performed by AL and MKM. JH was responsible for the

conceptualization and overview of the experiments under supervision of RFK and GV. Primers were developed by JH and EE. Experiments were executed by JH, EE, KEVR and BC. GV analyzed the WES data. Differential expression was analyzed by LM. Pathway enrichment analysis and identification of enriched transcription factors were performed by JH and LM. Analysis of RT-PCR data and preparation of the corresponding figures were performed by JH and CPD. Western Blotting was performed by CPD. JH drafted the manuscript which was reviewed and approved by all authors.

Data Availability

The datasets generated during the current study are not publicly available due consent restrictions, but are available from the corresponding author on reasonable request.

Competing Interests

The authors declare no conflicts of interest.

Ethical Approval

The cell lines used in this study were donated with informed consent to the Human Genetic Mutant Cell Repository at the Coriell Institute (Camden, New Jersey, USA) for research purposes.

Funding

The authors acknowledge the support of the Research Fund of the University of Antwerp OEC-Methusalem grant 'GENOMED'.

References

1. Association AP, American Psychiatric A, American Psychiatric A, American Psychiatric Association DSMTF. Diagnostic and statistical manual of mental disorders : DSM-5. 5th edition ed. Arlington, Va: American Psychiatric Association; 2013. xliv, 947 pages p.

2. Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. Ment Retard Dev Disabil Res Rev. 2002;8(3):117-34.

3. Neri G, Schwartz CE, Lubs HA, Stevenson RE. X-linked intellectual disability update 2017. Am J Med Genet A. 2018;176(6):1375-88.

4. van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. Annu Rev Genet. 2011;45:81-104.

5. Raymond FL. X linked mental retardation: a clinical guide. J Med Genet. 2006;43(3):193-200.

6. Lazzarini A, Stenroos ES, Lehner T, McKoy V, Gold B, McCormack MK, et al. Short tandem repeat polymorphism linkage studies in a new family with X-linked mental retardation (MRX20). Am J Med Genet. 1995;57(4):552-7.

7. Helsmoortel C, Vandeweyer G, Ordoukhanian P, Van Nieuwerburgh F, Van der Aa N, Kooy RF. Challenges and

opportunities in the investigation of unexplained intellectual disability using family-based whole-exome sequencing. Clin Genet. 2015;88(2):140-8.

8. Vandeweyer G, Van Laer L, Loeys B, Van den Bulcke T, Kooy RF. VariantDB: a flexible annotation and filtering portal for next generation sequencing data. Genome Med. 2014;6(10):74.

9. Jones JR. Nonrandom X chromosome inactivation detection. Curr Protoc Hum Genet. 2014;80:9 7 1-9 7

10. D'Incal CP, Cappuyns E, Choukri K, Szrama K, De Man K, Aa NVd, et al. In Search of the Hidden Protein: Optimization of Detection
Strategies for autism-associated Activity-Dependent Neuroprotective Protein (ADNP) mutants. Research Square. 2022.

11. Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner. 9th Annual Genomics of Energy & Environment Meeting; United States2014.

12. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

13. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.

14. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

15. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47(W1):W199-W205.

16. Janky R, Verfaillie A, Imrichova H, Van de Sande B, Standaert L, Christiaens V, et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. PLoS Comput Biol. 2014;10(7):e1003731.

17. Liska O, Bohar B, Hidas A, Korcsmaros T, Papp B, Fazekas D, et al. TFLink: an integrated gateway to access transcription factor-target gene interactions for multiple species. Database (Oxford). 2022;2022.

18. Iqbal Z, Vandeweyer G, van der Voet M, Waryah AM, Zahoor MY, Besseling JA, et al. Homozygous and heterozygous disruptions of ANK3: at the crossroads of neurodevelopmental and psychiatric disorders. Hum Mol Genet. 2013;22(10):1960-70.

19. Kumar R, Ha T, Pham D, Shaw M, Mangelsdorf M, Friend KL, et al. A non-coding variant in the 5' UTR of DLG3 attenuates protein translation to cause non-syndromic intellectual disability. Eur J Hum Genet. 2016;24(11):1612-6.

20. Matis T, Michaud V, Van-Gils J, Raclet V, Plaisant C, Fergelot P, et al. Triple diagnosis of Wiedemann-Steiner, Waardenburg and DLG3-related intellectual disability association found by WES: A case report. J Gene Med. 2020;22(8):e3197.

21. Philips AK, Siren A, Avela K, Somer M, Peippo M, Ahvenainen M, et al. X-exome sequencing in Finnish families with intellectual disability--four novel mutations and two novel syndromic phenotypes. Orphanet J Rare Dis. 2014;9:49.

22. Tzschach A, Grasshoff U, Beck-Woedl S, Dufke C, Bauer C, Kehrer M, et al. Next-generation sequencing in X-linked intellectual disability. Eur J Hum Genet. 2015;23(11):1513-8.

23. Tarpey P, Parnau J, Blow M, Woffendin H, Bignell G, Cox C, et al. Mutations in the DLG3 gene cause nonsyndromic X-linked mental retardation. Am J Hum Genet. 2004;75(2):318-24.

24. Zanni G, van Esch H, Bensalem A, Saillour Y, Poirier K, Castelnau L, et al. A novel mutation in the DLG3 gene encoding the synapse-associated protein 102 (SAP102) causes non-syndromic mental retardation. Neurogenetics. 2010;11(2):251-5.

25. Sandestig A, Green A, Aronsson J, Ellnebo K, Stefanova M. A Novel DLG3 Mutation Expanding the Phenotype of X-

652 Linked Intellectual Disability Caused by DLG3 Nonsense
653 Variants. Mol Syndromol. 2020;10(5):281-5.
654 26. Gieldon L, Mackenroth L, Betcheva-Krajcir E, Rump A,
655 Beck-Wodl S, Schallner J, et al. Skewed X-inactivation in a
656 family with DLG3-associated X-linked intellectual disability. Am
657 J Med Genet A. 2017;173(9):2545-50.
658 27. Plenge RM, Stevenson RA, Lubs HA, Schwartz CE,
659 Willard HF. Skewed X-chromosome inactivation is a common
660 feature of X-linked mental retardation disorders. Am J Hum
661 Genet. 2002;71(1):168-73.
662 28. Muller BM, Kistner U, Kindler S, Chung WJ, Kuhlendahl
663 S, Fenster SD, et al. SAP102, a novel postsynaptic protein that
664 interacts with NMDA receptor complexes in vivo. Neuron.
665 1996;17(2):255-65.
666 29. Consortium GT. Erratum: Genetic effects on gene
667 expression across human tissues. Nature. 2018;553(7689):530.
668 30. Lopez-Hernandez A, Sberna S, Campaner S. Emerging
669 Principles in the Transcriptional Control by YAP and TAZ.
670 Cancers (Basel). 2021;13(16).
671 31. Hayashi R, Goto Y, Ikeda R, Yokoyama KK, Yoshida K.
672 CDCA4 is an E2F transcription factor family-induced nuclear
673 factor that regulates E2F-dependent transcriptional activation
674 and cell proliferation. J Biol Chem. 2006;281(47):35633-48.

675 32. Liu H, Ippolito GC, Wall JK, Niu T, Probst L, Lee BS, et al.
676 Functional studies of BCL11A: characterization of the
677 conserved BCL11A-XL splice variant and its interaction with
678 BCL6 in nuclear paraspeckles of germinal center B cells. Mol
679 Cancer. 2006;5:18.
680 33. Beckmann H, Su LK, Kadesch T. TFE3: a helix-loop-helix
681 protein that activates transcription through the
682 immunoglobulin enhancer muE3 motif. Genes Dev.
683 1990;4(2):167-79.
684 34. Dias C, Estruch SB, Graham SA, McRae J, Sawiak SJ,
685 Hurst JA, et al. BCL11A Haploinsufficiency Causes an
686 Intellectual Disability Syndrome and Dysregulates
687 Transcription. Am J Hum Genet. 2016;99(2):253-74.
688 35. Lehalle D, Vabres P, Sorlin A, Bierhals T, Avila M,
689 Carmignac V, et al. De novo mutations in the X-linked TFE3
690 gene cause intellectual disability with pigmentary mosaicism
691 and storage disorder-like features. J Med Genet.
692 2020;57(12):808-19.
693 36. Paulraj P, Bosworth M, Longhurst M, Hornbuckle C,
694 Gotway G, Lamb AN, et al. A Novel Homozygous Deletion
695 within the FRY Gene Associated with Nonsyndromic
696 Developmental Delay. Cytogenet Genome Res.
697 2019;159(1):19-25.

698 37. Huang Q, Zhang YF, Li LJ, Dammer EB, Hu YB, Xie XY, et
699 al. Adult-Onset Neuronal Ceroid Lipofuscinosis With a Novel
700 DNAJC5 Mutation Exhibits Aberrant Protein Palmitoylation.
701 Front Aging Neurosci. 2022;14:829573.
702 38. Sans N, Petralia RS, Wang YX, Blahos J, 2nd, Hell JW,
703 Wenthold RJ. A developmental change in NMDA receptor-
704 associated proteins at hippocampal synapses. J Neurosci.
705 2000;20(3):1260-71.
706 39. Won S, Levy JM, Nicoll RA, Roche KW. MAGUKs:
707 multifaceted synaptic organizers. Curr Opin Neurobiol.
708 2017;43:94-101.
709 40. Cuthbert PC, Stanford LE, Coba MP, Ainge JA, Fink AE,
710 Opazo P, et al. Synapse-associated protein 102/dlgh3 couples
711 the NMDA receptor to specific plasticity pathways and
712 learning strategies. J Neurosci. 2007;27(10):2673-82.
713 41. Chen BS, Gray JA, Sanz-Clemente A, Wei Z, Thomas EV,
714 Nicoll RA, et al. SAP102 mediates synaptic clearance of NMDA
715 receptors. Cell Rep. 2012;2(5):1120-8.
716 42. Wei Z, Wu G, Chen BS. Regulation of SAP102 Synaptic
717 Targeting by Phosphorylation. Mol Neurobiol.
718 2018;55(8):6215-26.
719 43. Roszkowska M, Krysiak A, Majchrowicz L, Nader K,
720 Beroun A, Michaluk P, et al. SRF depletion in early life

721 contributes to social interaction deficits in the adulthood. Cell
722 Mol Life Sci. 2022;79(5):278.
723 44. Chaiwongkot A, Kitkumthorn N, Srisuttee R,
724 Buranapraditkun S. Cellular expression profiles of Epstein-Barr
725 virus-transformed B-lymphoblastoid cell lines. Biomed Rep.
726 2020;13(5):43.
727 45. Liu C, Si W, Tu C, Tian S, He X, Wang S, et al. Deficiency
728 of primate-specific SSX1 induced asthenoteratozoospermia in
729 infertile men and cynomolgus monkey and tree shrew models.
730 Am J Hum Genet. 2023;110(3):516-30.
731 46. McBride MJ, Mashtalir N, Winter EB, Dao HT, Filipovski
732 M, D'Avino AR, et al. The nucleosome acidic patch and H2A
733 ubiquitination underlie mSWI/SNF recruitment in synovial
734 sarcoma. Nat Struct Mol Biol. 2020;27(9):836-45.
735 47. Centore RC, Sandoval GJ, Soares LMM, Kadoch C, Chan
736 HM. Mammalian SWI/SNF Chromatin Remodeling Complexes:
737 Emerging Mechanisms and Therapeutic Strategies. Trends
738 Genet. 2020;36(12):936-50.
739 48. Hu H, Haas SA, Chelly J, Van Esch H, Raynaud M, de
740 Brouwer AP, et al. X-exome sequencing of 405 unresolved
741 families identifies seven novel intellectual disability genes. Mol
742 Psychiatry. 2016;21(1):133-48.
743 49. Kobayashi T, Iwamoto Y, Takashima K, Isomura A,
744 Kosodo Y, Kawakami K, et al. Deubiquitinating enzymes

regulate Hes1 stability and neuronal differentiation. FEBS J. 2015;282(13):2411-23.

Figure Legends

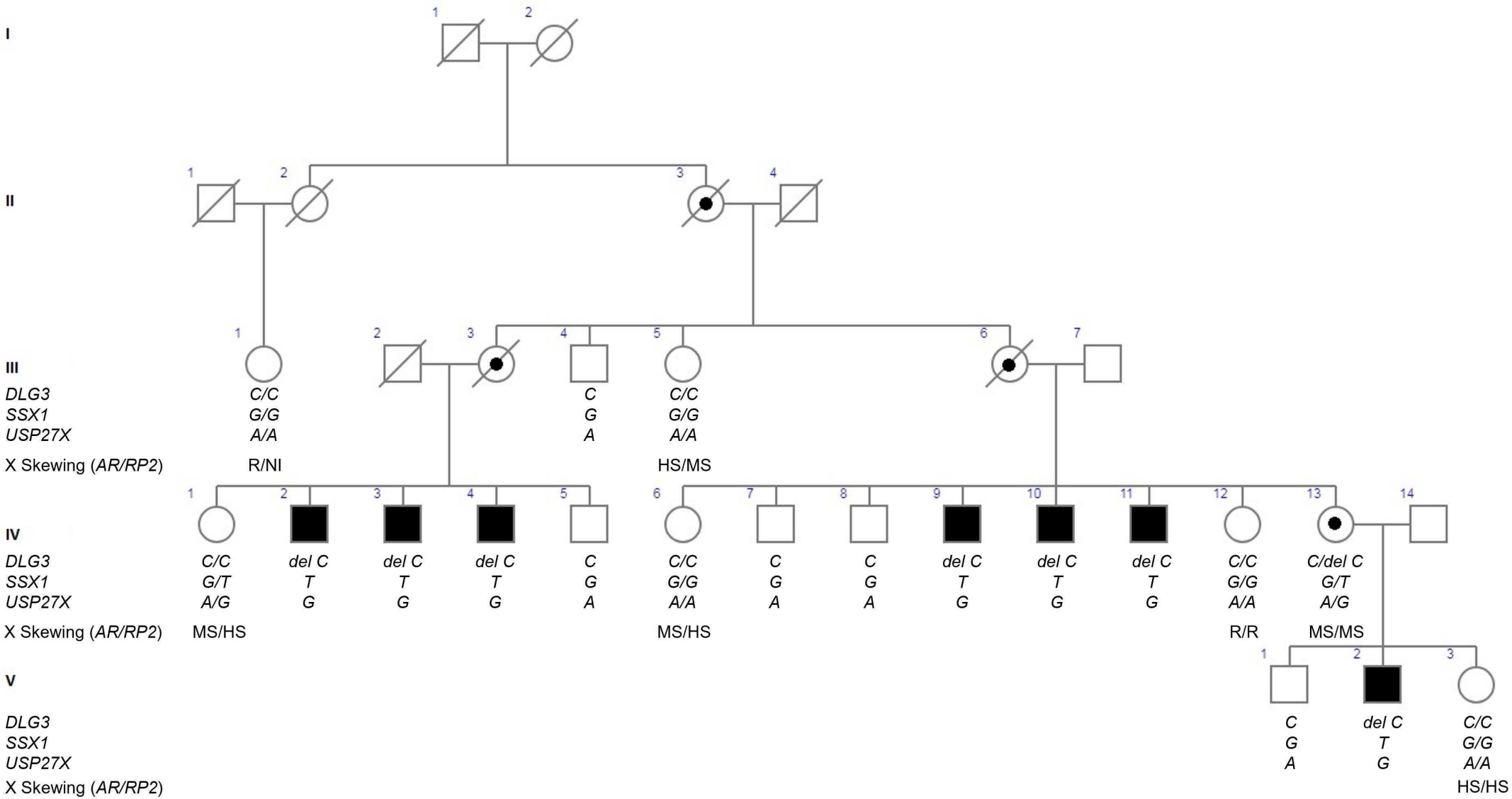
Figure 1: Pedigree of the MRX20 family. Pedigree is shown as in original publication [6]. Segregation of variants in *DLG3* *(*c.195del/p.(Thr66ProfsTer55)) (NM_021120.4), *SSX1* (c.358G>T/p.(Glu120Ter)) (NM_001278691.2) and *USP27X* (c.56A>G/p.(Gln19Arg)) (NM_001145073.3) are shown for all family members of which DNA was available as well as X-inactivation patterns of 7 women of the family. The ratio of X chromosome inactivation for the *AR* and *RP2* alleles were interpreted as follows: <80:20 = random (R); 80:20 to 90:10 = moderately skewed (MS); >90:10 = highly skewed (HS). NI: Non-informative.

Figure 2: DNA sequencing chromatogram of the variants found in *DLG3*, *SSX1* and *USP27X* in the MRX20 family. The sequence of the reverse strand is shown. An example of the wildtype allele, alternative allele (mutant allele for *DLG3*) and both alleles are represented. The chromosomal positions (GRCh38) of these variants are: NC_000023.11:g.70445396del, NC_000023.11:g.48263809G>T and NC_000023.11:g.49880363A>G, respectively.

Figure 3: Differential expression and enriched pathway analysis. A Differential expression analysis of the MRX20 family. Hierarchical clustering heatmap showing the differential expressed genes; all genes with an adjusted p-value < 0.1 are represented. * Represents the differentially expressed genes with an adjusted p-value < 0.05. **B** Enriched pathways based on the differential expression data. Bar chart shows enrichment ratio or NES of results with direction. Enriched pathways with a False Discovery Rate (FDR) ≤ 0.05 are shown in a darker shade. Analysis was performed using WebGestalt [14]. **C** Validations of differential expression of genes *CDCA4*, *WWTR1*, *PEX26*, *LDHA* and *NMT2* using RT-PCR normalized to *ACTB*, *UBC* and *YWHAZ* in the MRX20 family. Statistical significance was obtained by a two-tailed Mann-Whitney U test. * represents a p-value < 0.05 and ** represents a p-value < 0.01.

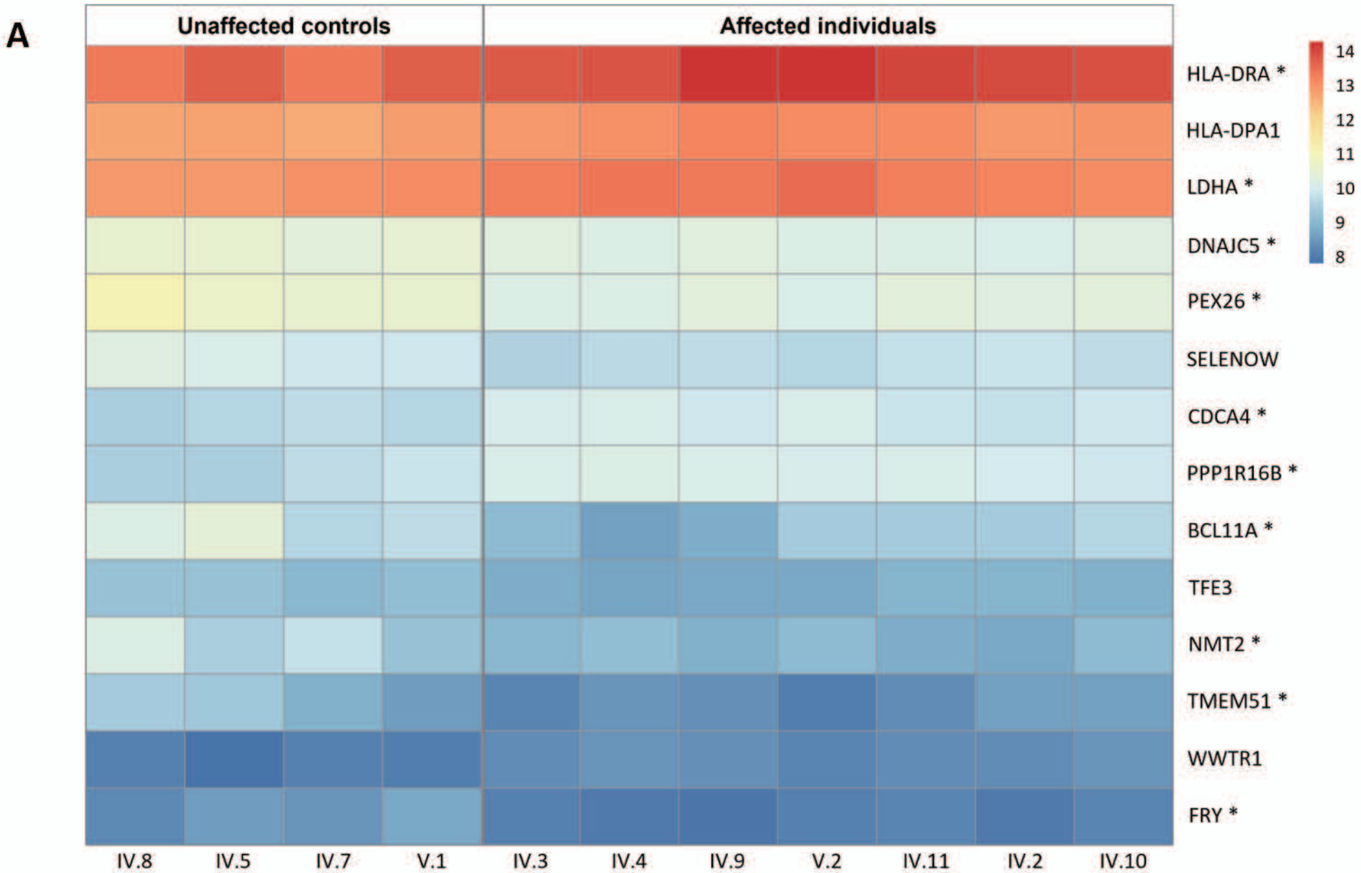
Figure 4: **Representation of the synapse-associate protein 102 (SAP102) encoded by the *Discs Large MAGUK Scaffold Protein 3* (*DLG3*) gene.** The protein contains three tandem PDZ domains, an scr homology (SH3) domain and a guanylate kinase (GK) domain. Known protein coding mutations are indicated by arrows (NM_021120.4/ NP_066943.2) (Supplementary Table 5

gives a representation of all published mutations in the *DLG3* gene).

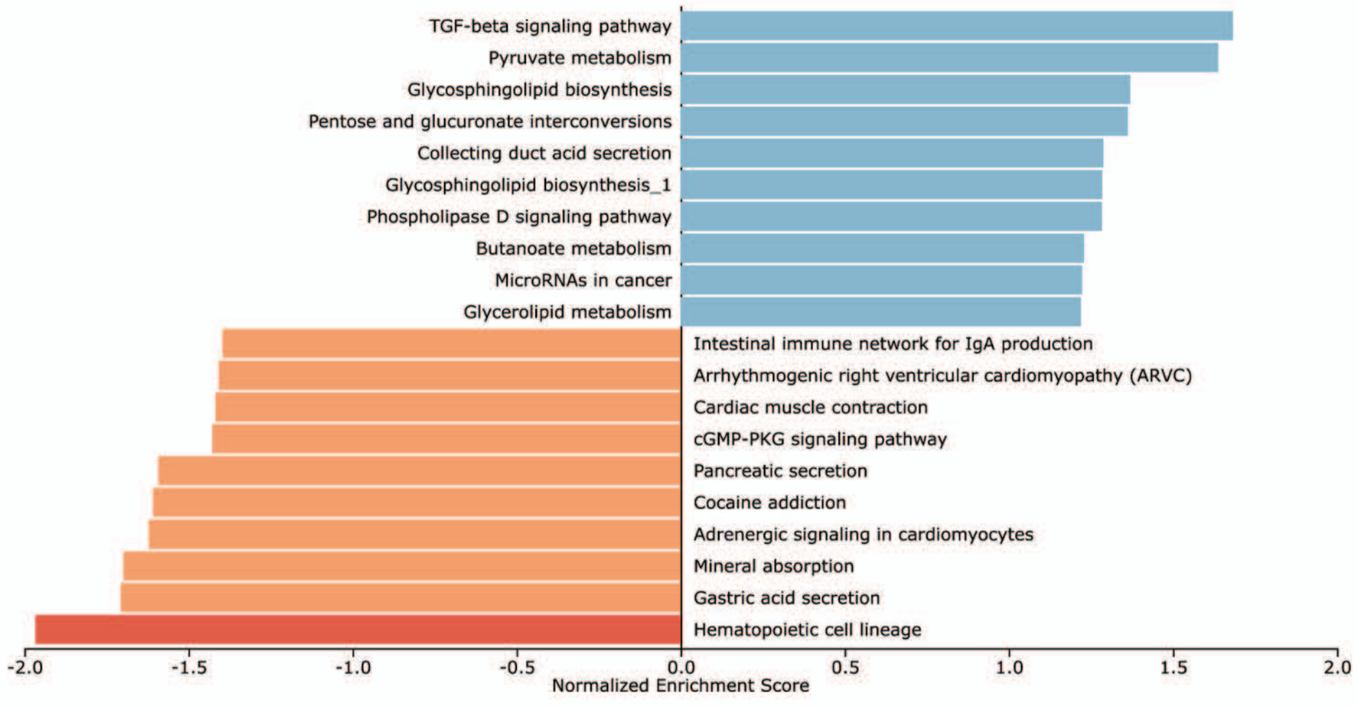

I
II
III
DLG3
SSX1
USP27X
X Skewing (AR/RP2)
C/C
G/G
A/A
R/NI
C
G
A
C/C
G/G
A/A
HS/MS
IV
DLG3
SSX1
USP27X
X Skewing (AR/RP2)
C/C
G/T
A/G
MS/HS
del C
T
G
del C
T
G
del C
T
G
C
G
A
C/C
G/G
A/A
MS/HS
C
G
A
C
G
A
del C
T
G
del C
T
G
del C
T
G
C/C
G/G
A/A
R/R
C/del C
G/T
A/G
MS/MS
V
DLG3
SSX1
USP27X
X Skewing (AR/RP2)
C
G
A
del C
T
G
C/C
G/G
A/A
HS/HS


DLG3
SSX1
USP27X
Wildtype allele
Alternative allele
Heterozygosity
R70 P69 T68 P67 T66 A65 G64 A63 Q62 S61
5′ CGAGGGGTGGGGGTGGCCCCCGCCTGCGAG 3′
5′ CGAGGGGTGGGGGTGCCCCCGCCTGCGAGG 3′
5′ CGAGGGGTGGGGGTGSCCCCSSCYKSRARG 3′
K124 S123 D122 N121 E120 D119 E118 A117 P116 K115
5′ CTTCRAAWCATTTTAGTCCTSTGCTRGCTT 3′
5′ CWTCRAAWCATTTTCGTCCTCTGCTRGCYT 3′
5′ CTTCRAAWCATTWWMGTCCTSTGCTGGCTT 3′
A22 E21 G20 Q19 E18 E17 K16 A15 I14
5′ AGCTTCTCCTTGCTCTTCTTTGGCAAT 3′
5′ AGCTTCTCCTCGCTCTTCTTTGGCAAT 3′
5′ AGCTTCTCCTYGCTCTTCTTTGGCAAT 3′

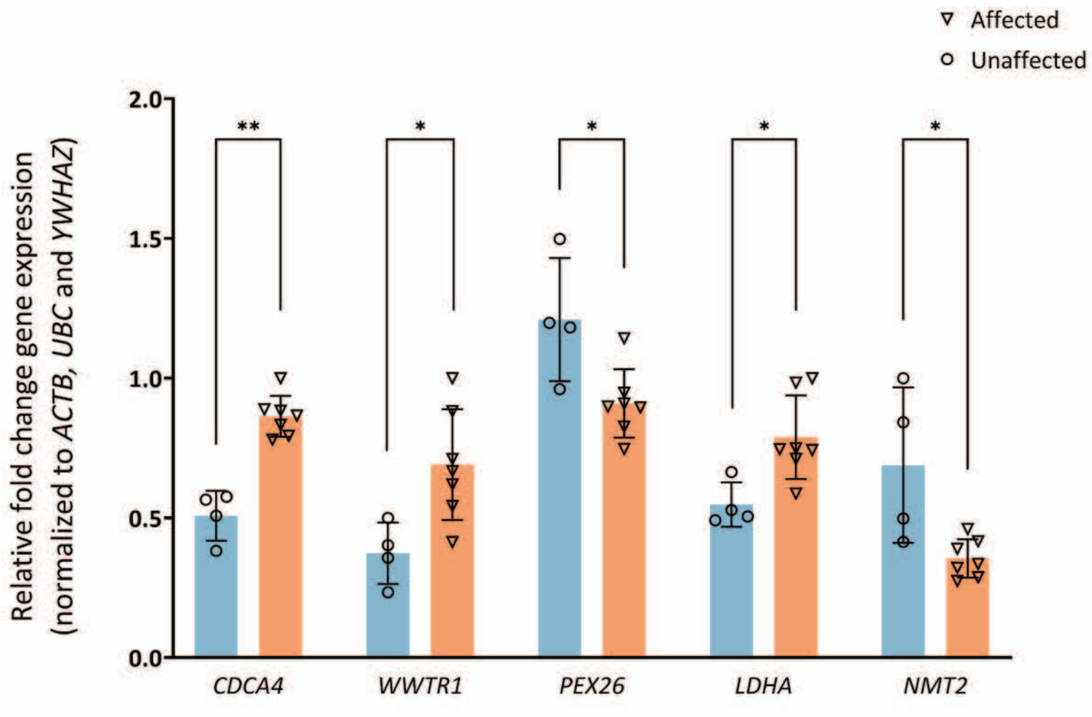
A

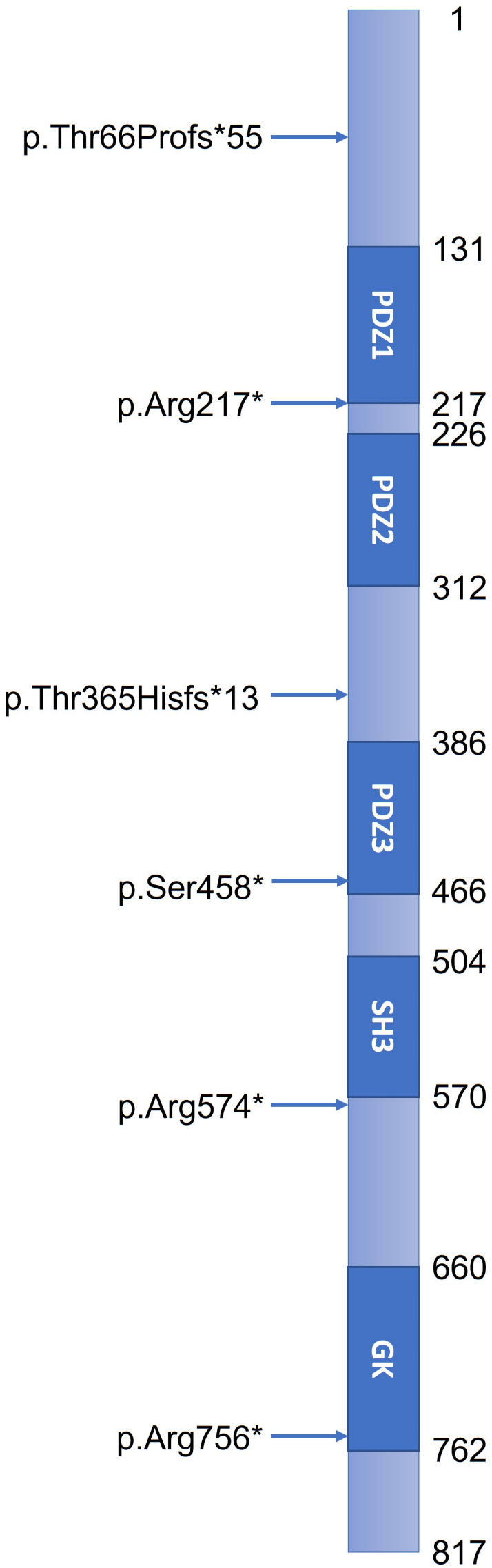

Unaffected controls
Affected individuals
HLA-DRA *
HLA-DPA1
LDHA *
DNAJC5 *
PEX26 *
SELENOW
CDCA4 *
PPP1R16B *
BCL11A *
TFE3
NMT2 *
TMEM51 *
WWTR1
FRY *
14
13
12
11
10
9
8
IV.8
IV.5
IV.7
V.1
IV.3
IV.4
IV.9
V.2
IV.11
IV.2
IV.10

B


TGF-beta signaling pathway
Pyruvate metabolism
Glycosphingolipid biosynthesis
Pentose and glucuronate interconversions
Collecting duct acid secretion
Glycosphingolipid biosynthesis_1
Phospholipase D signaling pathway
Butanoate metabolism
MicroRNAs in cancer
Glycerolipid metabolism
Intestinal immune network for IgA production
Arrhythmogenic right ventricular cardiomyopathy (ARVC)
Cardiac muscle contraction
cGMP-PKG signaling pathway
Pancreatic secretion
Cocaine addiction
Adrenergic signaling in cardiomyocytes
Mineral absorption
Gastric acid secretion
Hematopoietic cell lineage
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Normalized Enrichment Score

C


Affected
Unaffected
Relative fold change gene expression
(normalized to ACTB, UBC and YWHAZ)
2.0
1.5
1.0
0.5
0.0
**
*
*
*
*
CDCA4
WWTR1
PEX26
LDHA
NMT2


1
p.Thr66Profs*55
131
PDZ1
p.Arg217*
217
226
PDZ2
312
p.Thr365Hisfs*13
386
PDZ3
p.Ser458*
466
504
SH3
p.Arg574*
570
660
GK
p.Arg756*
762
817