SUPPLEMENTARY MATERIAL ONLINE

Robusto et al.

"The expanding spectrum of *PRPS1*-associated phenotypes: three novel mutations segregating with X-linked hearing loss and mild peripheral neuropathy"

Supplementary Methods

Supplementary Figure 1: Overview of exome data production.

Supplementary Table 1: Summary of SNPs identified in exome data.

Supplementary Table 2: Summary of indels identified in exome data.

Supplementary Table 3: Identification of candidate pathogenic variants.

Supplementary Table 4: Clinical and electrophysiological features from nerve conduction studies

Supplementary Figure 2: Allelic-specific expression in c.343A>G (p.M115V) and c.925G>T (p.V309F) carriers.

Supplementary Figure 3: Evolutionary conservation of amino-acid residues affected by novel *PRPS1* mutations

Supplementary Figure 4: Genomic context of novel mutations and associated phenotypes.

Supplementary Figure 5: *PRPS1*, *PRPS2*, *PPAT*, and *HGPRT* expression levels by qPCR.

SUPPLEMENTARY METHODS

Targeted DNA capture and exome sequencing

Briefly, 3 μg of genomic DNA were randomly sheared into fragments of about 150-200 bp by sonication (Covaris, Woburn, MA, USA), adaptor ligated, and purified. Extracted DNA fragments (~250 bp) were amplified by ligation-mediated PCR, purified, and hybridized to the Biotinylated RNA Library (BAITS), using the SureSelect 38M human exome kit (Agilent, Santa Clara, CA, USA). Exon-enriched DNA libraries were then sequenced on a HiSeq 2000 (Illumina, San Diego, CA, USA) as paired ends of 90 bp. Raw image files were processed by the Illumina base-calling Software v.1.7 with default parameters.

Read mapping, variant calling, and annotation

After removing duplicated reads, sequences were aligned to the human reference genome (hg18, NCBI build 36.3) using SOAPaligner/SOAP2 v.2.21 [1], with a tolerance of three mismatches in each read to reduce incorrect mapping. Genotype of each target base was determined by the SOAPsnp software v.1.03, with recommended parameters (http://soap.genomics.org.cn/soapsnp.html), whereas short insertions or deletions (indels) were identified with GATK [2]. The thresholds for calling SNPs and indels included the following: 1) the number of unique mapped reads supporting a SNP had to be ≥4 and ≤100; and 2) the consensus quality score had to be ≥20. The called variants were annotated according to their position and predicted functional impact on the protein. Then, since disease-causing mutations are expected to be rare in the general population, all changes were filtered to remove high-frequency variations against dbSNP130 (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/snp130.txt.gz), the 1000 Genome Project variant database (fttp://www.1000genome.org, Project pilot 1, release July 2010), eight HapMap exomes (http://snp.cshl.org), YH (http://yh.genomics.org.cn), and NHLBI GO Exome Sequencing Project (ESP, v.0.0.9 data release, November 2011). SIFT [3], Condel [4] and MutationTaster [5] were used to predict whether an amino-acid substitution affects protein function.

Allele-specific mRNA quantitation by Sanger sequencing

Total RNA was isolated using the Eurozol reagent (Euroclone, Wetherby, UK) under standard conditions, starting from PBMCs of two females, one carrying the c.343A>G (p.M115V) variant, and the other carrying the c.925G>T (p.V309F) variant, in the heterozygous state. RNA concentration was determined using the NanoDrop ND-1000 spectrophotometer.

Random nonamers (Promega, Madison, WI, USA) and Superscript III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) were used to perform first-strand complementary DNA (cDNA) synthesis starting from 500 ng of total RNA, according to the manufacturer's instructions. An aliquot (1 µL) of the total reverse-transcription reaction (20 µL) was used as template in a standard RT-PCR assay using exonic primer pairs to amplify the regions surrounding the variant under analysis (*PRPS1*_ex3cDNA_F: 5'-GTGGTTGTGGCGAAATCAAT-3', and *PRPS1*_ex3cDNA_R: 5'-ACCAGCATCAGGTGAGACAA-3' for the NM_002764.3:c.343A>G variant; *PRPS1*_ex7cDNA_F: 5'-TCCTGCTATTTCTCGCATCA-3', and *PRPS1*_ex7cDNA_R: 5'-CGGGTCTTCTGCTGAATTTG-3' for the NM_002764.3:c.925G>T variant).

PCR products were sequenced on both strands as described above. Peak areas corresponding to heterozygous positions were measured from sequence electropherograms using the ImageJ program v.1.47 (http://rsb.info.nih.gov/ij/). Peak-area ratios were used to evaluate the relative amount of the two mRNA species. To normalize for differences in ddNTPs incorporation efficiency and dye fluorescence emission, the same measurements were performed on electropherograms obtained by sequencing the same region from PCR-amplified genomic DNA of the same individual. As the amount of wild-type *vs.* mutant allele quantity in genomic DNA is expected to be 1:1, the peak area ratio measured from chromatograms obtained by sequencing PCR products from genomic DNA was used as a normalizing coefficient for RT-PCR sequence data, and a normalized ratio (mutant/wild-type base signal) was calculated [6].

X-inactivation analysis

Skewing of X-inactivation in female carriers of the c.343A>G (p.M115V) or c.925G>T (p.V309F) missense variants was investigated by analyzing a polymorphic repeat in the promoter region of either the *FMR1* (CGG repeat) or the *AR* gene (CAG repeat), for which the tested female was heterozygous, by modifying a previously described protocol [7]. Briefly, 100 ng of genomic DNA from each female carrier and her affected son was digested overnight at 37°C using a combination of 1 U *Rsa*I and 2.5 U of the methylation-sensitive restriction enzyme *Hpa*II (New England Biolabs, Beverly, MA, USA), as well as 1 U *Rsa*I alone as a negative control, in a total volume of 20 µL. To amplify the polymorphic trinucleotide repeats, an aliquot (5 µL) of the total digestion reaction was used as template for PCR using forward primers and fluorescein-labeled reverse oligonucleotides [7, 8], and the Go *Taq*® Hotstart Polymerase (Promega), in 10% DMSO. PCR reactions were separated on an ABI-3130XL sequencer and the peak areas measured by the Peak Scanner v.1.0 software (Applied Biosystems). Complete digestion of one of the two alleles was

confirmed by the absence of a PCR product in the *Rsa*I- and *Hpa*II-digested DNA of the hemizygous male control, whereas in all other samples at least one allele was amplified.

Evaluation of *PRPS1*, *PRPS2*, *PPAT*, and *HGPRT* expression levels by real-time qRT-PCR

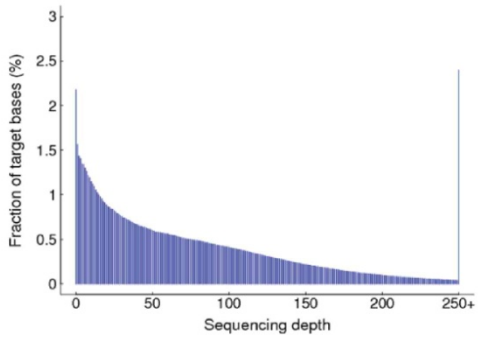
The expression levels of four genes coding for enzymes involved in the purine metabolism pathway (*PRPS1, PRPS2, PPAT, HGPRT*) were evaluated by real-time qRT-PCR. cDNA synthesis was performed as described above, starting from 500 ng of RNA extracted from PBMCs. All qRT-PCR reactions were performed in a final volume of 20 μL using the 2X SYBR Premix Ex *Taq* II (Takara Bio Inc, Otsu, Japan) on a Light Cycler 480 (Roche, Indianapolis, IN, USA). Oligonucleotide sequences and cycling conditions are available on request. Expression levels were normalized using two housekeeping genes (hydroxymethylbilane synthase, *HMBS*; β-actin, *ACTB*). Data were analyzed and rescaled using the GeNorm software [9]. Significance levels of t-tests, as well as of the one-way ANOVA analyses were assayed using the R software (http://www.r-project.org/).

References

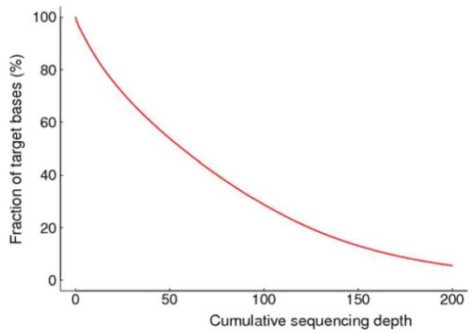
1. Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;**25**:1966-7.
2. DePristo M, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;**43**:491-98.
3. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;**4**:1073-81
4. González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* 2011;**88:**440-9.
5. Schwarz JM, Rödelsperger C, Schuelke M, et al. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010*;***7**:575-6.
6. Castaman G, Platè M, Giacomelli SH, et al. Alterations of mRNA processing and stability as a pathogenic mechanism in von Willebrand factor quantitative deficiencies. *J Thromb Haemost* 2010;**8**:2736-42.
7. Coene KL, Roepman R, Doherty D, et al. OFD1 is mutated in X-linked Joubert syndrome and interacts with LCA5-encoded lebercilin. *Am J Hum Genet* 2009;**85**:465-81.
8. Bodega B, Bione S, Dalprà L, et al. Influence of intermediate and uninterrupted FMR1 CGG expansions in premature ovarian failure manifestation. *Hum Reprod* 2006;**21**:952-7.
9. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;**3**:RESEARCH0034.

Supplementary Figure 1. Overview of Exome data production

A Distribution of per-base sequencing depth



B Cumulative depth distribution in target regions



C

Exome Capture Statistics	Family1, II-2
Target region (bp)	37,640,396
Raw reads	53,209,466
Raw data yield (Mb)	4,787.00
Reads mapped to genome	46,934,693
Reads mapped to target region	30,282,748
Data mapped to target region (Mb)	2,196.61
Mean depth of target region	58.36
Coverage of target region (%)	97.45%
Average read length (bp)	89.87
Fraction of target covered >=4X	92.09
Fraction of target covered >=10X	82.98
Fraction of target covered >=20X	71.22
Fraction of target covered >=30X	61.67
Fraction of target covered >=40X	53.45
Capture specificity (%)	67.05
Duplication rate	8.14
Mean depth of chrX	37.06
Mean depth of chrY	86.21
GC rate	42.10
Gender test result	M

Supplementary Figure 1. Overview of Exome data production.

A) The graph shows the distribution of per-base sequencing depth in target regions, which approximately follows a Poisson distribution, thus indicating that the captured exome region was evenly sampled. X-axis denotes sequencing depth, whereas Y-axis indicates the percentage of total target region under a given sequencing depth. B) The graph shows the cumulative depth distribution in target regions. X-axis denotes sequencing depth, whereas Y-axis indicates the fraction of bases that are covered at or above a given sequencing depth. C) Table summarizing the main exome data statistics.

Supplementary Table 1. Summary of SNVs identified in exome data.

Categories	II-2 (Family 1)
Number of genomic positions for calling SNVs[1]	106,141,607
Number of high-confidence genotypes[2]	94,302,377
Number of high-confidence genotypes in TR[2]	35,883,915
Total number of SNVs	64,772
Synonymous-coding	7,820
Missense	6,903
Nonsense	48
Readthrough	10
Splice site[3]	334
Intron	46,087
5'UTRs	670
3'UTRs	2,152
Intergenic	748
Homozygous	23,805
Heterozygous	40,967

SNV: Single Nucleotide Variant; TR: Target Region.
[1]Genomic positions for calling SNVs include captured target regions and its 200-bp flanking regions.
[2]Consensus genotype with quality score of at least 20.
[3]Intronic SNVs within 4 bp of exon/intron boundary.

Supplementary Table 2. Summary of indels identified in exome data.

	II-2 (Family 1)
Total number of indels	5,220
Ins-coding	74
Del-coding	112
Splice site	75
Intron	4,530
5'UTRs	171
3'UTRs	240
Intergenic	18
Total insertions	2,454
Total deletions	2,766
Heterozygous indels	3,201
Homozygous indels	2,019

Supplementary Table 3. Identification of candidate pathogenic variants.

Filter type	Nr of variants (genes)
NS/SS/I	6,855 (4,289)
Not in dbSNP130/1000Genomes	638 (551)
Not in ESP5400 exomes	281 (238)
Hom/compound Het	50 (26)
Within known NSHL genes	**1 (1)**

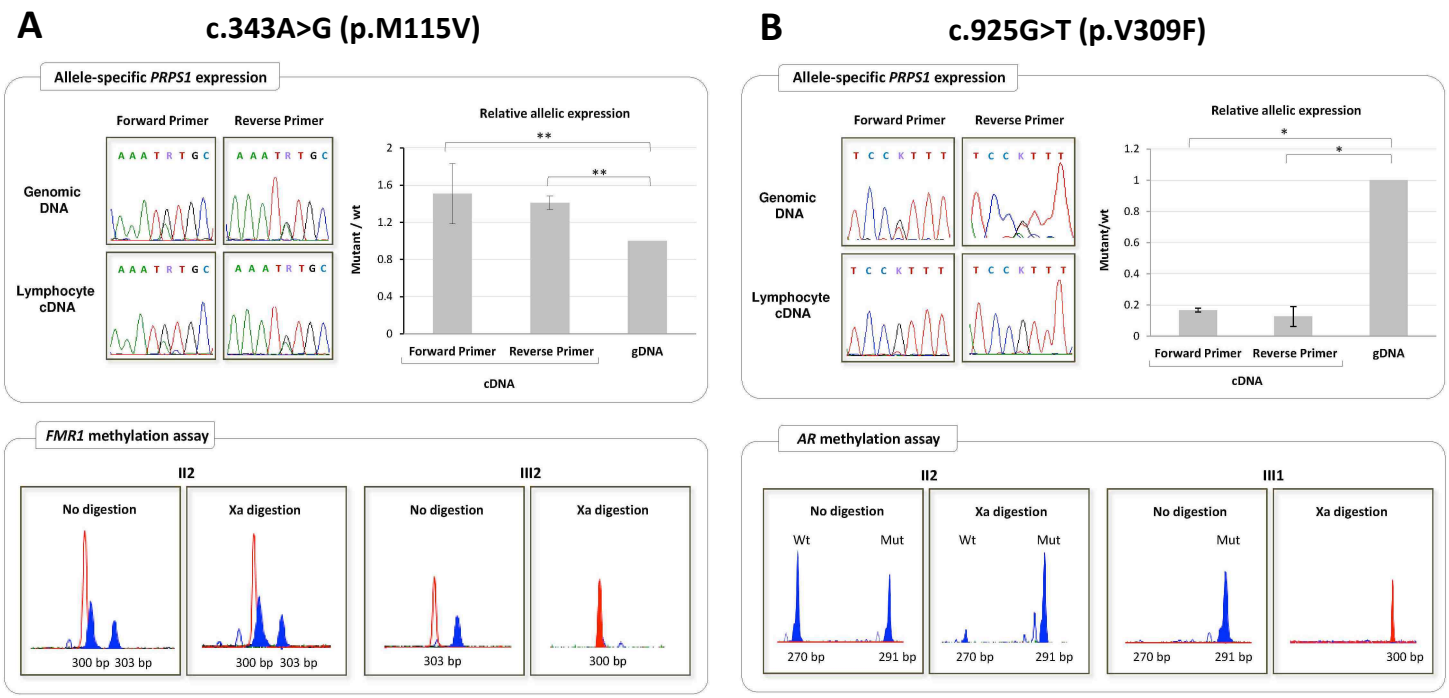
NS, non-synonymous variant; SS, splice-site variants; I, coding indel. Variants were filtered first by presence in dbSNP130 or 1000 Genomes (Project pilot 1, July 2010 data release) and then in 5,400 exomes obtained from the Exome Variant Server, NHLBI GO exome Sequencing Project (ESP, v.0.0.9 data release, November 2011, http://evs.gs.washington.edu/EVS/).

Supplementary Table 4. Clinical and electrophysiological features from nerve conduction studies.

Patient		Age /sex	Onset	Initial symptom	Motor involvement	Sensory involvement	Distal DTR	Pes cavus
Family 2	**II2**	37/F		None	None	None	Present	+
	III2	12/M	3	Gait disturbance	UL	None	Absent	+
	II6	60/F		None	None	Feet hypopallesthesia	Absent	+
	II4	49/F	45	Cramps	None	None	Present	+
	III4	9/M	8	Hand tingling paresthesia	None	Romberg positive	Absent	+
Family 3	**II4**	53/M		None	None	None	Absent	Flat foot
	II2	52/F		None	None	None	Reduced	
	III1	14/M	10	Fatigability	None	Romberg positive	Absent	Flat foot

Patient		UL CMAP/MCV	UL SNAP/SCV	LL CMAP/MCV	LL SNAP/SCV	EMG
Family 2	**II2**	14 / 61	11 / 66	9.9 / 47	18 / 58	Chronic denervation
	III2	7.8 / 55*	**2.9** / 52*	**2.7** / 47*	**3.3** / 70	Chronic denervation
	II6	11 / 50	11 / 50	10 / 48	10 / 48	Chronic denervation
	II4	16 / 54	18 / 59	11 / 50	11 / 56	Chronic denervation
	III4	6.7 / 57	13 / 55	6.0 /56	12 / 57	Chronic denervation
Family 3	**II4**	10 / 53	13 / 51	**1.5** / 42	14 / 51	Chronic denervation
	II2	17 / 66	15 / 65	**1.0** / 55	24 / 54	Normal
	III1	9 / 58	41 / 65	**4.0** / 54	20 / 50	Chronic denervation

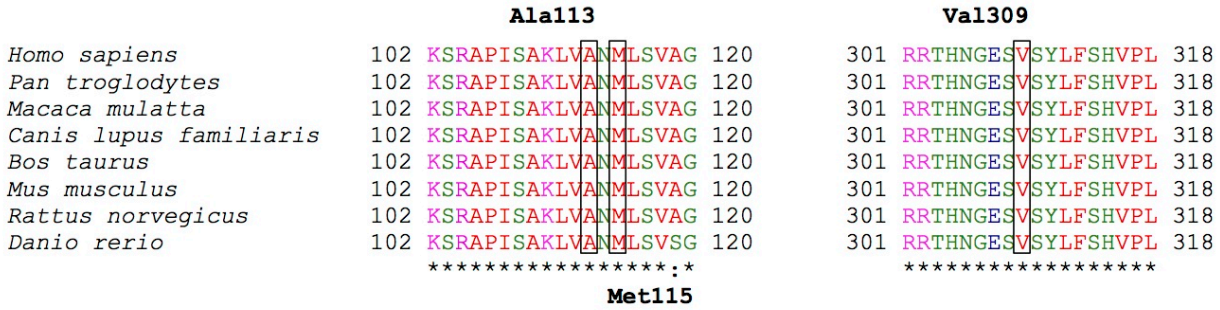
DTR, deep tendon reflexes; EMG, needle electromyography; UL, upper limbs; LL, lower limbs; CMAP, compound muscle action potential; MCV, motor conduction velocity; SNAP, sensory nerve action potential; SCV, sensory conduction velocity. For UL, values from the median nerve were recorded; *ulnar nerve; for LL, motor nerves from peroneal and nerves are reported; LL sensory nerve values are from the sural nerves. SNAP (uV), CMAP (mV), MCV and SCV (m/s). Values in bold typeface are abnormal.



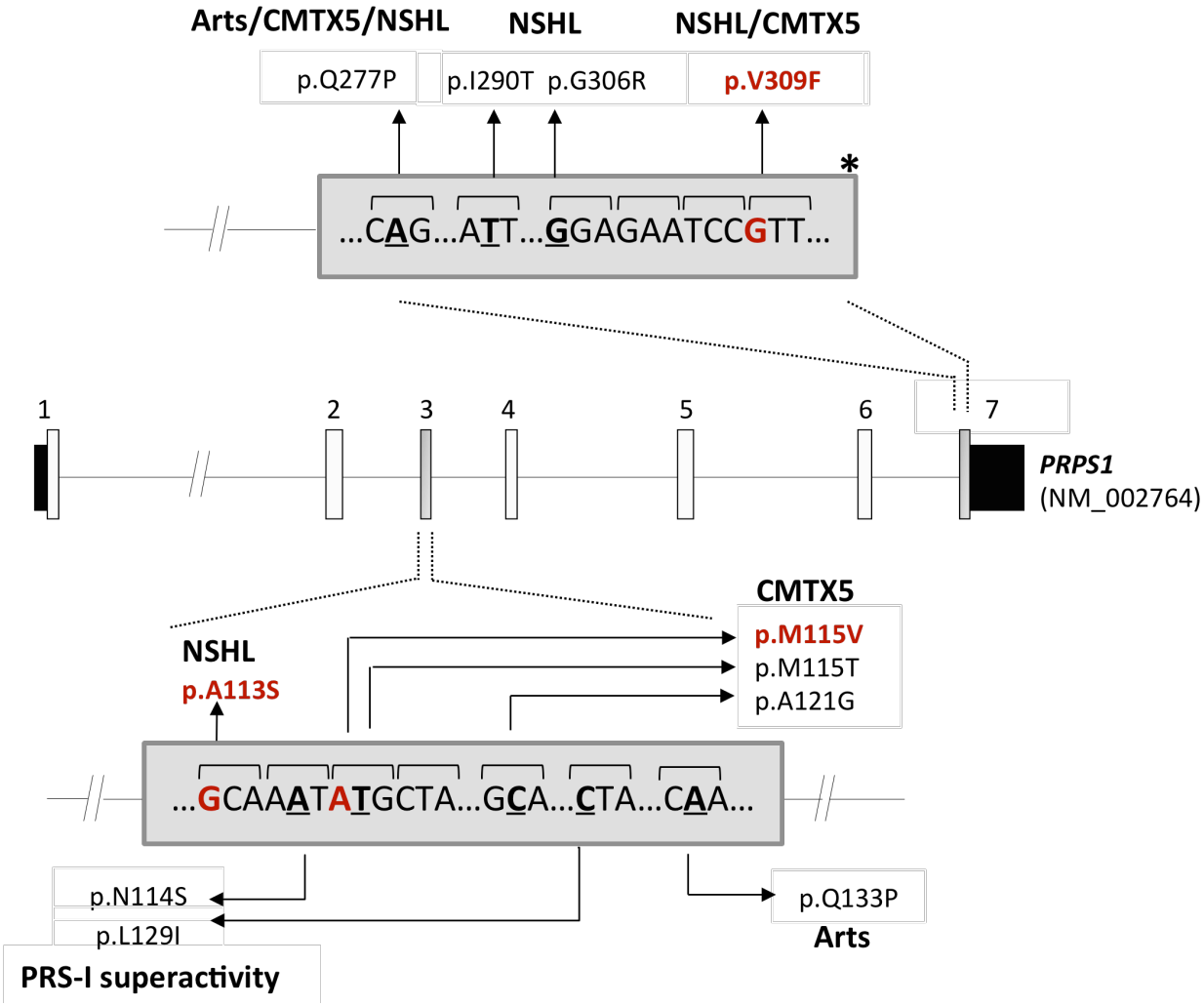
Supplementary Figure 2. *PRPS1* allelic-specific expression in c.343A>G (p.M115V) and c.925G>T (p.V309F) carriers.

Upper panels. On the left: DNA sequence electropherograms showing the region surrounding the variant position, PCR-amplified from genomic DNA or from lymphocyte cDNA of either the heterozygous individual carrying the c.343A>G mutation (A) or of the female carrier of the c.925G>T transversion (B). On the right: relative expression of the mutant *PRPS1* mRNA compared to the wild-type one. The ratio between the peak area of the mutant and the wild-type nucleotides at cDNA position 343 (A) and 925 (B) was measured. In order to account for differences in signal intensities, the peak-area ratio obtained from the cDNA sequence was normalized using the ratio obtained by sequencing the same region from genomic DNA. The genomic DNA level was set equal to 1. Sequencing was performed from both strands. Significance levels of t-tests are shown. **: P<0.01; *: P<0.05.

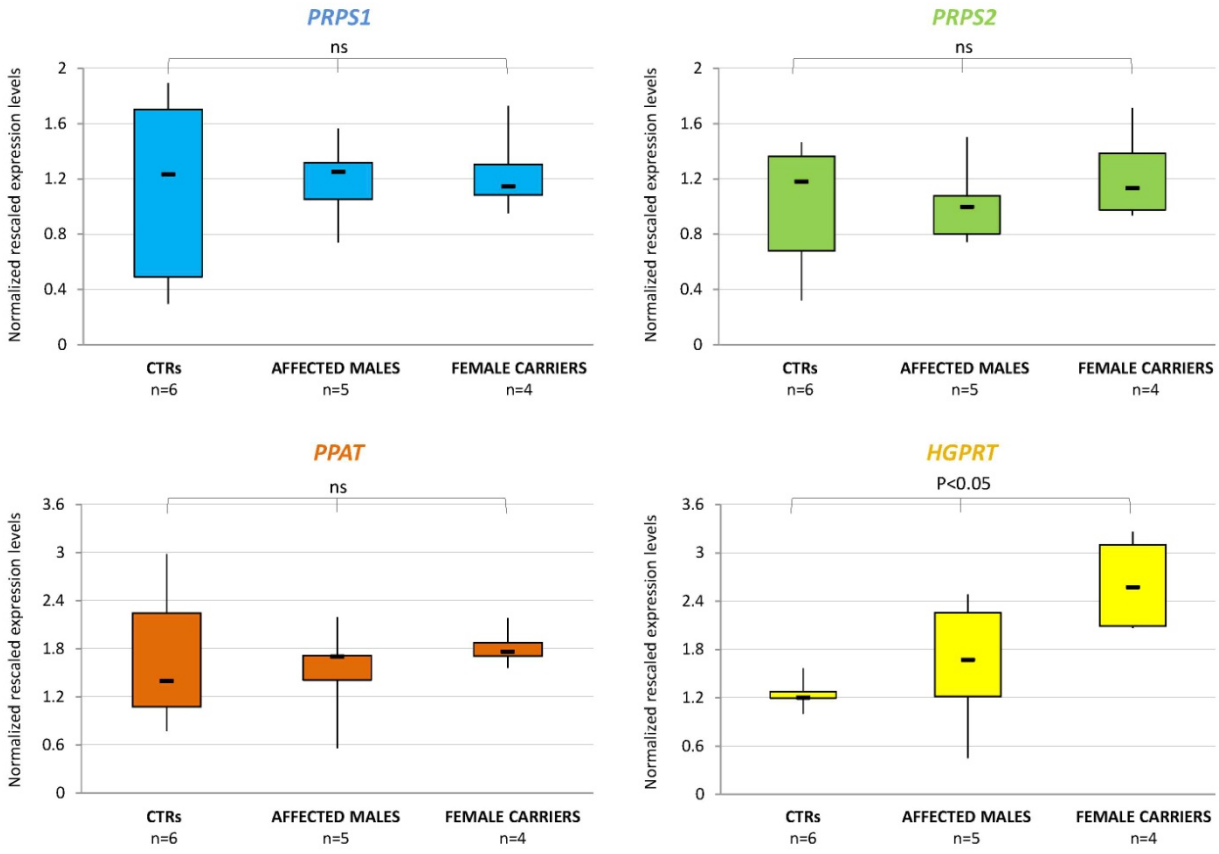
Bottom panels. GeneScan windows with peaks showing the X-inactivation patterns in the c.343A>G mutation carrier (A, individual II2) and in the c.925G>T carrier (B, individual II2). For comparison, the assays were performed also on their affected sons (A, III2; B, III1), who carry only the mutant *PRPS1* allele. Methylation assays were performed either on the *FMR1* CGG polymorphism (Family 2, A) or on the *AR* CAG polymorphic region (Familiy 3, B). No digestion: undigested genomic DNA; Xa digestion: DNA predigested with the methylation sensitive enzyme *Hpa*II, which only cuts restriction sites on the unmethylated, active X (Xa). GeneScan 500 ROX Size Standard is shown in red.



Supplementary Figure 3. **Evolutionary conservation of amino-acid residues affected by newly-identified *PRPS1* mutations.** The multiple protein sequence alignments of PRS-I orthologs, obtained by ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/), show the regions surrounding the three amino-acid residues that were found mutated in our families (boxed in black). Sequences used for the alignment are: *Homo sapiens*, NP_002755.1; *Pan troglodytes*, XP_001144504.1; *Macaca mulatta*, NP_001247511.1; *Canis lupus familiaris*, XP_850853.1; *Bos taurus*, NP_001039654.1; *Mus musculus*, NP_067438.1; *Rattus norvegicus*, NP_058939.1; and *Danio rerio*, NP_001070036.1. The colours of amino acids reflect their physicochemical properties. Asterisks indicate fully conserved (invariant) residues; semicolons indicate conservation between amino acids with strongly similar properties.



Supplementary Figure 4. Genomic context of novel mutations and associated phenotypes. Schematic representation of the *PRPS1* gene (NM_002764) highlighting the localization of the three novel pathogenic variants (bold) within exons 3 and 7 (grey blocks). The nucleotide sequence of the affected codon is reported. Neighboring known mutations, responsible for various *PRPS1*-associated phenotypes, are also indicated. Boxes indicate exons, lines indicate introns. Untranslated regions are colored in black. Not to scale. The stop codon is shown by an asterisk.



Supplementary Figure 5. *PRPS1*, *PRPS2*, *PPAT*, and *HGPRT* expression levels by qPCR.

To verify whether other modifier genes may contribute to the phenotypic variability among the analyzed families, we measured expression levels of *PRPS2*, PRPP amidotransferase (*PPAT*) and hypoxanthine-guanine phosphoribosyltransferase (*HGPRT*), all known to be involved in purine metabolism. Expression levels were evaluated by real-time semi-quantitative RT-PCR in PBMCs from five affected males -two carrying the p.(A113S), two carrying the p.(M115V), and one carrying the p.(V309F) variants- four female heterozygotes -three for the p.(M115V), one for the p.(V309F) variant- and six wild-type controls (CTRs: three males, three females). Results were normalized using the *HMBS* and *ACTB* housekeeping genes, and are presented as normalized rescaled values (calculated by the GeNorm software). Significance levels of the ANOVA tests are shown. ns: not significant; n: number of analyzed subjects.

No significant alteration of *PRPS2* and *PPAT* mRNA levels was measured, as well as of *PRPS1* itself. Interestingly, expression levels of *HGPRT*, which maps on the X chromosome, was increased (p=0.01) in the heterozygous female carriers in comparison either to the healthy controls or the affected males.