UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Scienze Biologiche e Molecolari

XXVI Ciclo

Inherited hearing loss: from gene variants to mechanisms of disease

Michela Robusto

PhD Thesis

Scientific tutor: **Dott.ssa Giulia Soldà**

Academic year: 2012/2013

SSD: BIO/13

Thesis performed at the Department of Medical Biotechnology

and Translational Medicine

Index

ABBREVIATIONS AND NOTES

ESE: exonic splicing enhancer

ESS: exonic splicing silencer

HL: hearing loss

INDEL: insertion/deletion

miRNA: microRNA

NGS: next-generation sequencing

NMD: nonsense-mediated mRNA decay

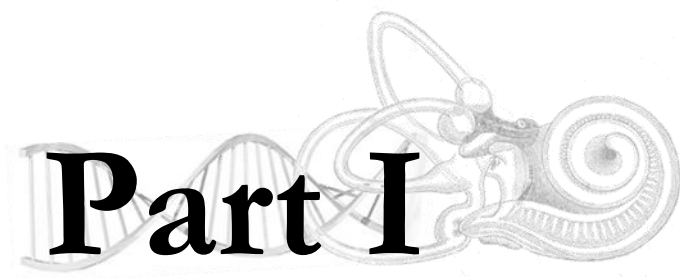**NSHL:** non-syndromic hearing loss

PBMC: peripheral blood mononuclear cells

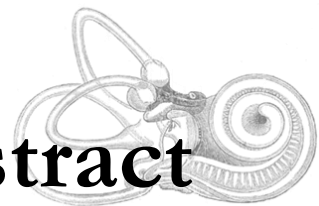**SNP**: single nucleotide polymorphism

UTR: untranslated region

WES: whole exome sequencing

Note. In this thesis, materials and methods are reported in the result section, below each figure.

Part I

Abstract

Nonsyndromic Sensorineural Hearing Loss (NSHL) is the most common sensory disorder worldwide, affecting at least 1 in 500 newborns and more than half individuals older than 80 years. It is estimated that about 60-70% of cases are due to genetic factors [Raviv *et al.*, 2010]. More than 70 genes have been associated with NSHL so far, but many others are still to be discovered. In this thesis, we investigated the genetic and molecular bases of NSHL by a double approach, consisting in: a) investigating the pathogenic mechanisms of mutations within the *MIR96* gene, and b) searching for new genes/mutations by Whole-Exome Sequencing (WES) in selected NSHL families.

As far as *MIR96* is concerned, we identified and characterized a novel mutation (the third described) within the *MIR96* gene, miR-96 (+57T>C), in an Italian family with autosomal dominant NSHL. MiR-96 is part of the conserved miR-183 microRNA (miRNA) family, which plays essential functions in the vertebrate inner ear. Point mutations within the seed region of miR-96 (miR-96-5p) cause autosomal dominant NSHL (AD-NSHL). The novel identified mutation replaces a highly conserved nucleotide and is predicted to reduce the stability of the pre-miRNA hairpin. *Ex vivo* assays in mammalian cells confirmed that both miR-96 and miR-96*(miR-96-3p) mature species were significantly reduced in the mutant, whereas the precursor level was unaffected. Moreover, miR-96 and miR-96* expression could be restored to normal levels by reconstituting the secondary structure of the pre-miR-96 hairpin, thus demonstrating that the mutation hinders the precursor processing. Finally, even though the mature miR-96 sequence is not altered, we demonstrated that the identified mutation significantly impacts on miR-96 regulation of selected targets. Taken together, these data provide further evidence of the involvement of miR-96 in human deafness and demonstrate that a quantitative defect of this miRNA may contribute to NSHL.

As far as WES is concerned, the application of NGS to the identification of novel genes/mutations in 6 genetically undiagnosed Italian families (NSHL1-4 and 6, 7), with recessive NSHL and at least two affected individuals, has enabled the molecular diagnosis in two families (NSHL4 and NSHL6), and highlighted a putative novel deafness-associated gene in another family (NSHL3). In particular, a novel missense mutation within the *PRPS1* gene was found in family NSHL4 and functionally characterized, together with other two mutations identified by candidate-gene screening in a larger X-linked NSHL cohort. In the NSHL3 family, WES pointed out a novel missense variant in *DIAPH2*, a gene not directly linked to NSHL but belonging to a family of proteins already involved in hearing loss. Studies aimed at the functional characterization of this mutation and at the clarification of the possible involvement of the gene in the pathogenesis of the disease are now being performed.

For the other three families (NSHL1, 2 and NSHL7), the search for pathogenic variations is still ongoing.

1. State of the Art

1.1 The auditory system

The auditory system is composed of three anatomical compartments: the outer, the middle, and the inner ear (Figure 1A). Sound waves impinging on the head are captured by the outer ear (auricle) and conveyed through the external auditory canal (acoustic duct) to the tympanic membrane. The vibrations of the tympanic membrane, caused by the airborne sound waves, are transmitted through the middle ear (tympanic cavity) to the inner ear by a chain of three movable bones. These auditory bones, or ossicles, consist of the malleus, which is connected to the tympanic membrane; the stapes, which is attached at its base to the oval window of the vestibule; and the incus, which is situated between the malleus and stapes and articulates with both. The sound vibrations of the tympanic membrane are propagated through a piston-like mechanical motion of these ossicles toward the base of the stapes, which moves in and out of the oval window of the vestibule, a central cavity in the inner ear. The mammalian inner ear is one of the most complex tiny organs in the body and is composed of two sensory organs: the cochlea, responsible for hearing (Figure 1B), and the vestibule, responsible for balance. While cochlear defects induce hearing loss, vestibular defects lead to disorientation, expressed as vertigo or dizziness in humans, and circling behavior (movement in circles) in both mice [Friedman et *al.*, 2007] and zebrafish [Whitfield, 2002].
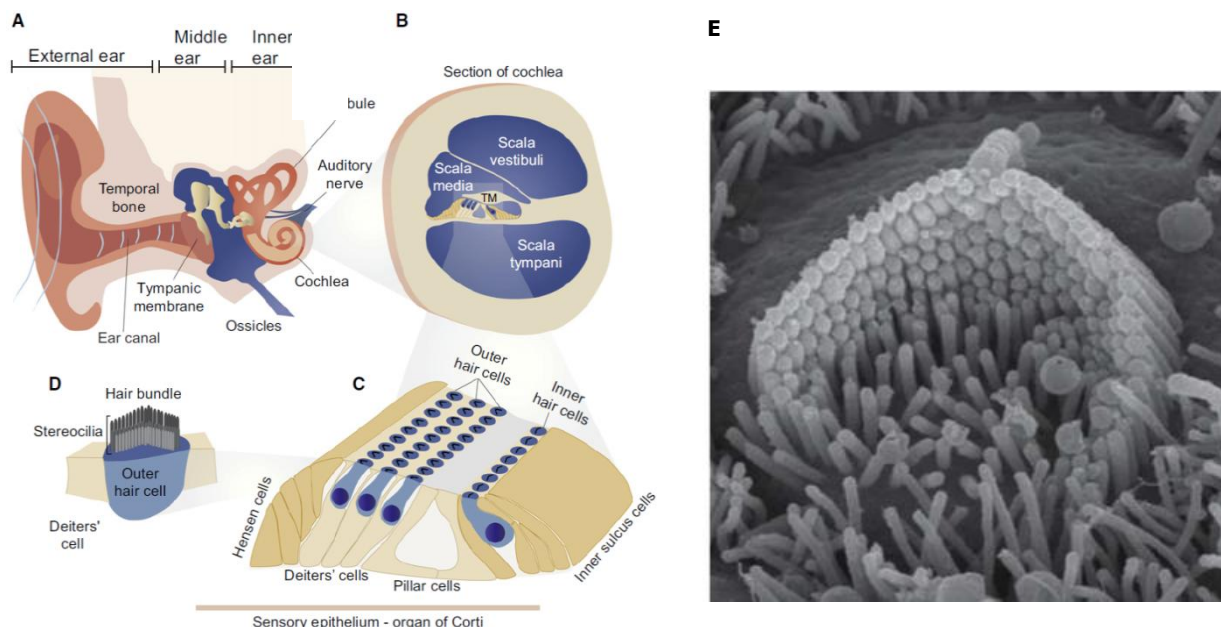
Figure 1. Schematic Illustration of the Mammalian Inner Ear. A. Representation of the three compartments of the human ear (external, middle, and inner) and of their connections. **B.** Cross-section of the cochlear duct. **C.** Structural organization of the organ of Corti. **D.** An enlargement of an hair bundle of an outer hair cell (OHC). **E.** Scanning electron microscopy (SEM) image showing the hair bundle of an OHC of a mouse, analogous to the scheme in D. Adapted from Dror and Avraham, 2010.

The sensory epithelia of the cochlea (organ of Corti, Figure 1C) contain supporting cells and mechanosensory receptor cells, which are called hair cells due to the hair-like stereocilia on their apical surface (Figure 1D and 1E). Proper development of the stereocilia is essential for hair-cell mechanotransduction, and, not surprisingly, defects in the participating proteins are causes of inherited deafness [Pauley *et al.*, 2008]. The hair and supporting cells are arranged in a precise cellular pattern required for proper mechanotransduction of auditory signals. For instance, in the organ of Corti the hair cells are arranged in four rows along the snail-shaped cochlea: three rows of outer hair cells (OHC) and a single row of inner hair cells (IHC), surrounded by supporting cells [Raphael and Altschuler, 2003] (Figure 1C).

Supporting cells express many factors required for the survival of hair cells or necessary for the hair cell environment, such as the connexin 26 and 30, encoded by the *GJB2* and *GJB6* genes [Nickel and Forge 2008], as well as *TECTA* and collagen genes, which encode components of the tectorial membrane [Richardson *et al.*, 2008]. Some of the sensory epithelium supporting cells are also responsible for controlling the ion homeostasis of the endolymph, the solution that fills the lumen to which the sensory epithelial cells face. The exact composition of the endolymph is crucial for the function and survival of the hair cells [Kikuchi *et al.*, 2000].

The tympanic membrane between the outer ear and the middle ear, as well as the tiny bones of the middle ear, convert auditory signals to cochlear fluid waves. The specialized stereocilia of cochlear hair cells move in response to the fluid movement in the inner-ear canals, causing the ion channels on the stereocilia plasma membrane to open. Thus, the cochlear hair cells convert auditory-derived mechanical stimuli into electrical signals [Raphael and Altschuler, 2003]. Indeed, afferent neurons innervate cochlear inner hair cells at glutamatergic synapses that transmit signals from the hair cells to the dendrites of the primary auditory neurons. The auditory neurons compose the auditory nerve, which in turn joins the vestibular nerve to form the vestibulocochlear nerve, or cranial nerve number VIII. The sound information propagates down the vestibulocochlear nerve, through intermediate stations, such as the cochlear nuclei and superior olivary complex of the brainstem and the inferior colliculus of the midbrain, being further processed at each waypoint. The information eventually reaches the thalamus, and from there it is relayed to the primary auditory cortex, located in the temporal lobe.

The vestibule, the second sensorial organ of the inner ear, is composed of three semicircular canals and five patches of sensory epithelia (utricular and saccular maculae, and three cristae), which also include several types of sensory hair and supporting cells. While the supporting cell population in the mammalian vestibule has a limited ability to transdifferentiate to new hair cells and thus may compensate for lost hair cells in this system, the mammalian cochlea has no ability to regenerate hair cells after birth [Raphael *et al.*, 2007]. Moreover, vestibular hair cells have a

different organization, a slightly different shape, and some differences in gene expression and regulation compared to the cochlear hair cells [Hertzano *et al.*, 2007]. Nonetheless, the mRNA and miRNA populations of cochlear and vestibular sensory epithelia seem to be very similar, with only minor differences [Friedman and Avraham, 2009]. The similarities between these two systems often lead to balance disorders in hearing impaired individuals [Gresty and Brookes, 1997].

1.2 Sensorineural hearing loss

Sensorineural hearing loss is the most common sensory disorder in humans, affecting at least 1 in 500 newborns and over half of individuals older than 80 years. It is estimated that about 278 million people worldwide have moderate to profound hearing loss in both ears (over 50 millions are living in Western countries, such as Europe and US), with severe consequences on public health and the quality of life of affected individuals [Nance, 2003]. For instance, hearing loss (HL) may have dramatic effects on language acquisition and educational progress when present in an infant, whereas it may severely influence the social and working lives of those affected when it becomes apparent in later childhood or in adult life. Similar to other sensory loss, hearing impairment has a wide spectrum of etiologies originating from both environmental and genetic factors (Table 1).

Among environmental factors, prolonged exposure to high intensity sound poses high risk for auditory function and can lower hearing thresholds. Acoustic trauma, as a result of a sudden loud noise, can lead to temporary and/or permanent hearing impairment [Dror and Avraham, 2010]. Moreover, different viral infections, as well as neonatal anoxia and hyperbilirubinemia, can cause permanent hearing defects. Eventually, long-term administration of ototoxic drugs such as aminoglycoside and gentamicin antibiotics has an adverse effect on the auditory system and accounts for hearing impairments [Yorgason *et al.*, 2006]. The relative contribution of genetic and environmental components may be determined by social factors such as population structure and consanguinity, infection control and immunization, and provision of neonatal and postnatal medical care [Morton, 1991]. Thus, in non-industrialized countries, environmental causes of hearing loss may outnumber those that are genetically determined, whereas in industrialized countries the importance of the genetic contribution to hearing loss has become more apparent. Epidemiological surveys of deaf patients have consistently shown that about 50% of childhood deafness can be attributed to genetic causes [Morton, 1991]. The etiology of the disease remains obscure in the remaining 25% of cases, even if in most of these, it is assumed to be of genetic

origin. Thus, genetic causes account for the largest proportion of all cases of early-onset hearing loss [Marres, 1998].

Table 1. Causes of hearing impairment

Genetic (syndromic and non-syndromic)
        Autosomal recessive
        Autosomal dominant
        X-linked
        Mitochondrial
        Chromosomal, *e.g.* Down syndrome and trisomies 13 and 18, Turner syndrome,
                22q11 deletions, mosaic trisomy 8

Environmental
        Ototoxic medication, *e.g.* aminoglycosides, platinum derivatives
        Prematurity
        Neonatal hypoxia
        Low birth weight
        Severe neonatal jaundice
        Head trauma
        Infection: prenatal, *e.g.* CMV, toxoplasmosis, rubella; postnatal, *e.g.* meningitis
        Noise exposure

CMV, citomegalovirus

Hearing loss can be further classified by several criteria, including the severity (average hearing thresholds in the range of 20-39 dB are defined as mild, 40-69 dB as moderate, 70-89 dB as severe, and >90 dB as profound deafness) [Petit *et al.*, 2001] and the age of onset. Indeed, hearing loss that occurs prior to speech acquisition is termed prelingual, either congenital or appearing after birth, whereas deafness that occurs later in life, after the acquisition of language, is defined postlingual. In most cases, prelingual hearing loss is severe but stable. About 1 child in 1000 is born with prelingual hearing loss, of whom approximately half have genetically determined monogenic forms of deafness; perinatal factors and infantile infections or trauma are responsible for the other half [Fraser, 1970, Morton, 1991]. Postlingual hearing loss is much more frequent than prelingual, affecting 10% of the population by the age of 60 years and 50% by the age of 80 years [Petit, 1996; Davis, 1989].

Another common classification of the disease is based on the defective anatomical site. Conductive hearing loss is characterized by external ear anomalies or abnormalities of the ossicles in the middle ear, sensorineural hearing loss is due to inner ear malfunction, and central hearing loss is caused by defects of the VIIIth nerve, the brain stem, or the cerebral cortex. Hearing loss can also be mixed.

The last element analyzed to characterize deafness is the association of the disease with additional clinical features. Indeed, when hearing loss is not linked to a clear pattern of other physical defects, it is referred to as nonsyndromic hearing loss (NSHL), which accounts for more than 70% of all hereditary cases of hearing impairments [Van Camp *et al.*, 1997]. On the contrary, syndromic deafness is characterized by additional manifestations, such as retinitis pigmentosa (e.g. Usher syndrome), euthyroid goiter and inner ear malformations (Pendred syndrome), craniofacial dysmorphia (Treacher-Collins syndrome), marfanoid body habitus (Stickler syndrome), renal anomalies (Alport syndrome), or the presence of long QT intervals (Jervell and Lange-Nielsen syndrome). Several hundred syndromes involving hearing loss have been described [Petit *et al.*, 2001]. Although useful when syndromic features can be recognized, this subdivision poses a dilemma when clinical manifestations are not fully developed. This is especially common in childhood, but may also be caused by variable gene expression. For example, the goiter in autosomal recessive Pendred syndrome may develop only in adulthood, if at all. With no other evident symptoms on physical examination, the diagnosis of Pendred syndrome is likely to be missed [Schrijver, 2004].

1.2.1 Genetic of hearing loss

Hearing loss is characterized by a high clinic and genetic heterogeneity, which reflects the anatomic and functional complexity of the inner ear, requiring the interaction of a diversity of proteins including ion channels, extracellular matrix, cytoskeletal proteins, and transcription factors [Schrijver, 2004]. It is estimated that approximately 1% of all human genes are involved in the hearing process [Friedman and Griffith, 2003]; to date, more than 60 genes and 110 chromosomal loci have been linked to the disease (Figure 2).

Inherited hearing loss can be transmitted as an autosomal recessive, autosomal dominant, X-linked, or mitochondrial trait (Table 1 and Figure 2). Allelic mutations in some genes can cause either recessive or dominant hearing loss, mutations in the same gene may cause syndromic or nonsyndromic hearing loss (i.e. *PRPS1*), and recessive hearing loss may be caused by a combination of two mutations in different genes from the same functional group (i.e. *GJB2* and *GJB6*) [Schrijver, 2004]. In fact, gene discovery in humans and protein characterization in animal models have revealed numerous molecular pathways in the inner ear. These include, but are not limited to, gene regulation, fluid homeostasis and mechanotransduction [Dror and Avraham, 2010]. For example, mutations in *GJB2* and *GJB6* (encoding connexin 26 and 30, respectively) are the most common cause of hearing loss in many populations. In addition to these, other genes

frequently found to bear deafness-causing mutations include *SLC26A4, COCH, MYO15A, POU3F4, OTOF, CDH23*, and *TMPRSS3* [Duman and Tekin, 2012].

Taking advantage of standardized nomenclature, a unified classification of the loci and genes for hearing impairment has been established (HUGO Gene Nomenclature Committee -HGNC-, http://www.genenames.org/) [Dror and Avraham, 2010]. Indeed, depending on the inheritance mode, the nonsyndromic genes or loci can be divided into DFNA (autosomal dominant deafness, ~15 to 20%), DFNB (autosomal recessive deafness, ~80%) and DFNX (X-linked deafness, ~1%) [ACMG, 2002; Hone and Smith, 2002]. Moreover, specific symbols are used for different forms of hearing loss including otosclerosis (OTSC), auditory neuropathy (AUNA), and mitochondrial (MRTNR, MTTS) genes. Each locus is associated with a number, designated by the chronological order of its discovery [Dror and Avraham, 2010].

Finally, starting from the evidence that approximately 98% of RNAs in mammalian cells do not code for proteins and that epigenetic factors are associated with hearing loss [Provenzano and Domann, 2007], it has been suggested that noncoding RNAs, such as microRNAs (miRNAs), may also be involved in inner ear development and hearing loss (see below).

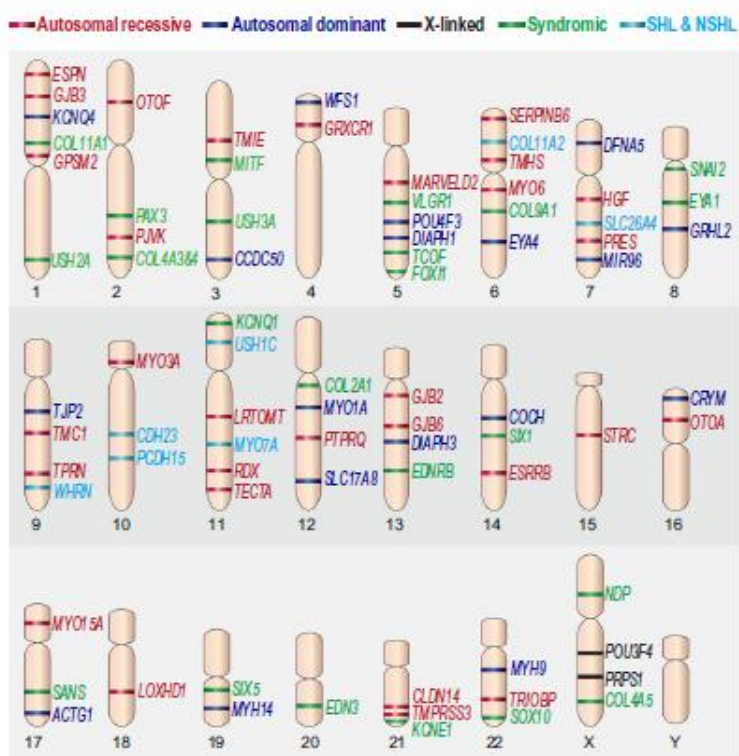**Figure 2. Known deafness genes.** Schematic representation of the chromosomal location of genes which mutations cause hearing impairment. The genes are classified as nonsyndromic autosomal recessive (red), nonsyndromic autosomal dominant (blue), X-linked (black), syndromic (green), and genes that are associated with both syndromic and nonsyndromic hearing loss (light blue). From Dror and Avraham, 2010.

Due to tremendous genetic heterogeneity, the identification of genes and gene defects that affect the process of hearing has been challenging. Over the past two decades linkage analysis with microsatellite markers has been used widely, allowing the chromosomal location of deafness genes to be mapped in families all over the world. Once the linkage region was elucidated, mutation analysis by Sanger sequencing often led to the identification of the causative mutation. However, despite the great contribution of linkage analysis methods, many deafness genes still remain to be discovered [Dror and Avraham, 2010].

1.2.2 Hearing loss treatment and prevention

The main goal of auditory genetic studies is to use the enormous knowledge gathered about deafness-causing genes and apply it to disease diagnosis and therapeutic treatment [Dror and Avraham, 2010].

Indeed, deafness is the only sensory defect that can be treated successfully even if the hearing impairment is complete. A cochlear implant study published in 2003 and performed in children of 8 to 9 years of age, who received their implants before the age of five, demonstrated that all children benefited from cochlear implantation in the areas of speech production, speech perception, and language [Schrijver, 2004]. There was a significant positive difference in cognitive and reading performance in children with identified *GJB2* mutations, which cause an isolated insult to the cochlea without damage to the VIIIth nerve or the central auditory system [Bauer *et al.*, 2003]. In this respect, specific knowledge about defective lesions along the auditory pathway based on genetic analyses has provided clinics with the advantage to predict the efficacy of different therapeutic approaches, such as determining compatibility for cochlear implants [Dror and Avraham, 2010].

Apart from surgical treatment, hearing loss is otherwise refractory to therapy. One notable exception seems to be the case of *PRPS1*-associated deafness. This gene codes for the phosphoribosylpyrophosphate synthetase 1 (PRS-I), an enzyme essential for the *de novo* purine synthesis, and has been previously associated with both syndromic (PRS-I superactivity, Arts syndrome, CMTX5) and nonsyndromic forms of X-linked deafness [Liu *et al.*, 2010]. Despite the importance of PRS-I, purine nucleotides can be produced by an alternative pathway utilizing S-adenosylmethionine (SAM) as a substrate. Indeed, dietary supplementation with SAM in patients with Arts syndrome has been reported to alleviate their symptoms. In addition, the progression of hearing impairment appears to have been stabilized [de Brouwer *et al.*, 2010]. Hence, SAM supplementation in the diet might represent a promising and safe therapy to delay the progression or the onset of symptoms in patients with *PRPS1* mutations.

The study of the genetic bases of hearing loss has contributed to the prevention of some forms of the disease. For example, mutations in the mitochondrial 12S rRNA have been shown to be associated with high risk for aminoglycoside-induced deafness [Fischel-Ghodsian, 1999]. Pharmacogenomic testing for such mutations is currently minimizing the risk for aminoglycoside-induced ototoxicity among groups of patients that are commonly treated with this drug [Bardien *et al.*, 2009; Veenstra *et al.*, 2007]. Understanding the molecular pathogenesis associated with drug susceptibility of these genetic mutations can help to promote the development of alternative therapies with similar efficacy but significant reduced toxicity [Dror and Avraham, 2010].

1.3   MiRNA involvement in NSHL: the miR-183 family

MicroRNAs (miRNAs) are 21-nucleotide-long, single stranded noncoding RNAs that mainly function as posttranscriptional regulators of gene expression in plant and animal cells. To date, 2578 human, 1908 mouse, and 255 zebrafish mature miRNAs have been identified (experimentally proven or homologous to experimentally verified miRNAs in other species), and their sequences are listed in the miRBase database (http://www.mirbase.org/; current release: 20, June 2013).

Once assembled into an RNA-induced silencing complex, each miRNA might inhibit the expression of hundreds of target messenger RNAs (mRNAs), by inducing translational repression and/or mRNA degradation [Huntzinger and Izaurralde., 2011]. Animal miRNAs recognize partially complementary binding sites, which are generally located in the 3' untranslated region (3'UTR) of target mRNAs. In particular, complementarity to the miRNA seed region, corresponding to nucleotides 2–8 at the 5' of the mature miRNA, is a major determinant in target recognition and is sufficient to trigger silencing [Bartel, 2009].

MiRNAs can be encoded in independent transcription units, in polycistronic clusters, or within the introns of protein-coding genes. They are transcribed, mostly by RNA polymerase II, as capped and polyadenylated primary miRNAs (pri-miRNAs), which contain extended hairpin structures. The pri-miRNAs are then processed by the Microprocessor complex into a hairpin precursor (pre-miRNA), which is exported to the cytoplasm and cleaved by the RNaseIII Dicer to generate a mature miRNA duplex [Krol *et al.*, 2010]. Usually, one strand of the duplex is preferentially selected for entry into the silencing complex to regulate gene expression, whereas the other strand, known as the passenger strand or miRNA∗, has typically been assumed to be degraded. However, recent evidence demonstrated that miRNA∗ species are often present at physiologically relevant levels, can associate with the silencing protein Argonaute and can inhibit target mRNAs in both cultured cells and transgenic animals [Ro *et al.*, 2007; Okamura *et al.*, 2008;

de Wit *et al.*, 2009; Marco *et al.*, 2010; Yang *et al.* 2011]. Hence, a precursor miRNA can give rise to two different mature miRNAs, referred to as -5p and -3p, depending on the strand of the hairpin from which it originates (Figure 3).

It is nowadays recognized that miRNAs play key roles in many, if not all, biological processes, including development [Lee *et al.,* 1993], differentiation [Dostie *et al.*, 2003; Chen *et al.*, 2004], cell proliferation [Bartel, 2004; Harfe, 2005], apoptosis [Brennecke *et al.*, 2003; Xu *et al.*, 2003], and organogenesis [Reinhart *et al.*, 2000]. Since the discovery of the strong impact of miRNAs on different cellular pathways, it has been hypothesized that mutations affecting miRNA function may have a pathogenic role in human diseases. In this respect, a large body of evidence has already shown that aberrant miRNA expression is implicated in most forms of human cancer [Bonci *et al.*, 2008, Visone and Croce, 2009], and in the pathogenesis of several complex human diseases, such as psoriasis [Sonkoly *et al.,* 2007], Alzheimer's disease, and metabolic disorders [Krutzfeldt and Stoffel, 2006]. Moreover, recent studies have established a clear link between miRNAs and hearing. In particular, point mutations within a miRNA belonging to the miR-183 family were identified as cause of nonsyndromic sensorineural hearing loss in both human and mouse [Mencía *et al.*, 2009; Lewis *et al.*, 2009].
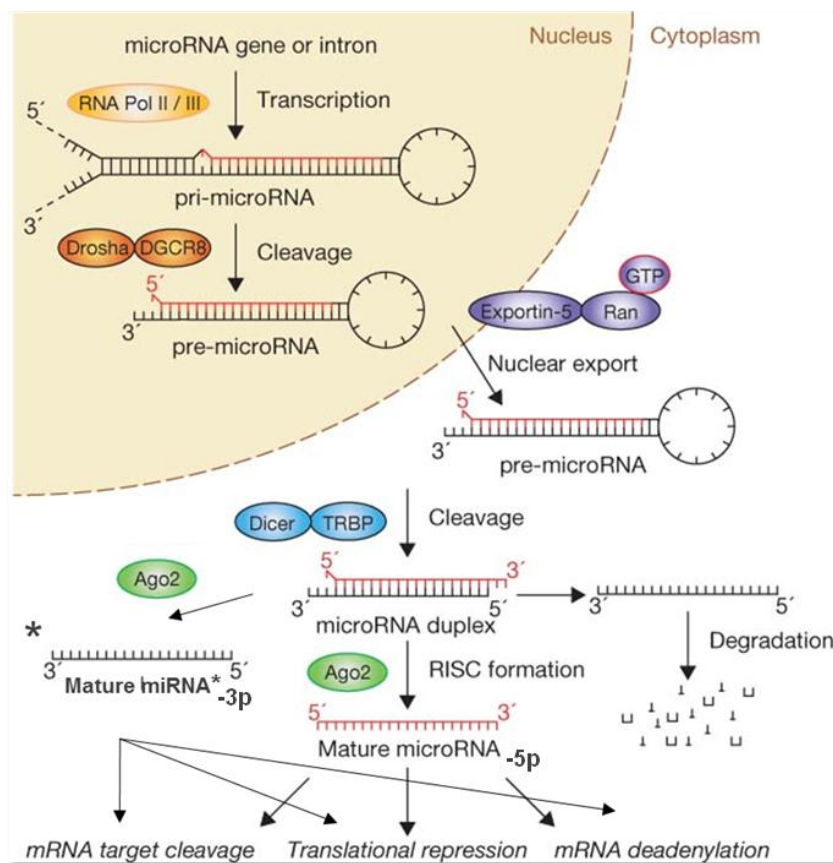

Figure 3. Biogenesis of miRNA. Adapted from Winter *et al.*, 2009.

The miR-183 family is composed of three miRNAs (miR-183, miR-96 and miR-182), which are coordinately expressed from a single genetic locus in vertebrates. Homologous miRNAs (miR-228 and mir-263b) are also present in invertebrates [Pierce *et al.*, 2008]. Importantly, this highly conserved family of miRNAs shows expression in ciliated neurosensory organs across phyla and has recently been demonstrated to contribute specifically to the differentiation and function of the mechanosensory hair cells in the vertebrate inner ear [Soukup *et al.*, 2009; Li *et al.,* 2010]. For instance, in zebrafish, the miR-183 family is predominantly expressed in the hair cells of the inner ear and of the lateral line, as well as in the olfactory and retinal sensory cells [Kapsimali *et al.*, 2007; Friedman and Avraham, 2009]. Overexpression of miR-96 or miR-182, but not of miR-183, was shown to induce duplicated otocysts, ectopic or expanded sensory patches, and extra hair cells. Conversely, knockdown of each of the three miRNAs led to a reduction in the number of hair cells in the inner ear and caused defects in semicircular canals, as well as the presence of abnormal neuromasts in the lateral line [Li *et al.*, 2010]. Collectively, these findings suggested both distinct and common roles for the three miRNAs in cell-fate determination in the inner ear.

In the human genome, the miR-183 family is clustered in a 4.5-kb region on chromosome 7q32, within a locus that has been linked to autosomal dominant NSHL (DFNA50, OMIM #613074). In 2009, two mutations in the seed region of miR-96 were detected in two Spanish families affected by autosomal dominant progressive NSHL: both mutations (+13G>A and +14C>A) affected nucleotides that are fully conserved among vertebrates (from fish to humans) and segregated with hearing loss in the corresponding families. The impact of these mutations on miR-96 processing and target recognition was analyzed in HeLa and NIH-3T3 cells, showing that both mutations result in reduced levels of mature miRNA and hinder its gene-silencing capacity [Mencía *et al.*, 2009]. Further evidence on the significance of miR-96 expression and function in the pathogenesis of hearing loss has been provided by genetic studies in the mouse, where a single nucleotide substitution in the seed region of the homolog of miR-96 was shown to cause hearing loss and hair cell defects in a murine model of sensorineural deafness, called diminuendo [Lewis *et al.*, 2009; Kuhn *et al.*, 2011]. This N-ethyl-N-nitrosurea (ENU)-induced mouse mutant show progressive hearing impairment in heterozygotes and profound deafness in homozygotes [Lewis *et al.*, 2009]. Moreover, the physiological development of mutant sensory hair cells is arrested at around the day of birth, before their bio-physical differentiation into IHC and OHC, and the remodeling of auditory nerve connections within the cochlea fail to occur. Taken together, these findings underline the key role of miR-96 in regulating the progression of the physiological and morphological differentiation of cochlear hair cells and, as such, in coordinating one of the most distinctive functional refinements of the mammalian auditory system [Kuhn *et al.*, 2011].

Finally, it is interesting to note that the work of Mencía and colleagues [Mencía *et al.*, 2009] represented the first evidence that point mutations in a miRNA can be responsible for a Mendelian trait.

1.4 Next-Generation Sequencing technologies and application to NSHL

Due to the high genetic heterogeneity of inherited hearing loss, molecular diagnosis is extremely challenging. Indeed, except for mutations in *GJB2*, which are found in up to 50% of families with autosomal recessive NSHL in some populations, most deafness mutations are private and are seen in only a single or few families [Duman and Tekin, 2012]. Moreover, many of NSHL genes consist of long and/or many exons, thus making extremely expensive and time-consuming - if not unfeasible - the mutational screening of candidate genes with conventional methods (e.g. Sanger sequencing alone or in combination with genome wide SNP genotyping) [Diaz-Horta *et al.*, 2012].

Next-generation sequencing (NGS) technologies have recently been introduced as an alternative approach to more traditional methods [Ng *et al.*, 2009]. In particular, the sequencing of the so-called exome, which includes all translated portion (about 1%) of the genome, was introduced as an efficient strategy to search for alleles underlying Mendelian disorders. Indeed, ~85% of disease-related mutations found so far are in the protein-coding regions or in the splice sites [Teer and Mullikin, 2010]. Therefore, the exome represents a highly enriched subset of the genome for the identification of variants with large effect size. Whole exome sequencing (WES) allows for a targeted capture and resequencing of nearly all exons of protein-coding genes in a single experiment, making possible the identification of genetic variations at a base-pair resolution. Moreover, this technology can be successfully applied to the analysis of small pedigrees and few patients' samples [Ng *et al.*, 2010]. Since its introduction, NGS has dramatically reduced the cost of sequencing on a per-base pair (bp) basis and increased the output of sequencing from a few hundred bp by each Sanger analysis to about 600 billion bp per NGS machine run [Clark *et al.*, 2011; Zhang *et al.*, 2011].

NGS is carried out by a high throughput sequencing of small DNA fragments. The first step in NGS consists in the fragmentation (mechanically or chemically) of the genomic DNA into small pieces, usually in the range of 300-500 bp [Borgstrom *et al.*, 2011]. Subsequently, platform-specific adapters are added to the ends of the DNA segments. One common feature shared by almost all current NGS platforms is that clonally amplified single DNA molecules, spatially separated in a defined microchamber, are sequenced in a massively parallel fashion [Lin *et al.*,

2012]. During NGS sequencing, results are generated by reading optical signals deriving from repeated cycles of polymerase-mediated fluorescent nucleotide extensions of four different colors (e.g., Illumina's HiSeq system), or from iterative cycles of fluorescently-labeled oligonucleotide ligation (e.g., ABI SOLiD system) [Lin *et al.*, 2012]. Alternatively, sequencing is based on the principle of pyrosequencing (e.g., Roche 454 system) [Margulies *et al.*, 2006]. The results of the sequenced segments are called "reads", which could be 25-100 bp from one or both ends [Lin *et al.*, 2012]. The massive capacity of NGS allows the sequencing of many randomly overlapping DNA fragments; therefore, each nucleotide in targeted regions may be included in many reads, leading to repeated analysis of the same element and to a consequent increase in the depth of coverage. Increased depth of coverage usually improves sequencing accuracy, because a consensus voting algorithm is used in determining the final nucleotide calls [Lin *et al.*, 2012].

Even if a clear regulatory oversight over NGS data quality control, data analysis standards, and standardization of data reporting has not yet been clearly defined, but only recently proposed (FGED-MINSEQE, version 1.0, June 2012), targeted sequencing has been shown to be a robust, effective technique; such approach is bringing a paradigm shift to biomedical research of Mendelian disorders and their clinical diagnoses, ultimately enabling personalized medicine based on one's genetic profile [Lin *et al.*, 2012]. Indeed, different targeted genomic capture methods and massive parallel sequencing approaches have already yielded many exciting findings [Brownstein *et al.*, 2011; Majewski *et al.*, 2011; Shearer *et al.*, 2010], and have been successfully applied to the discovery of novel NSHL genes/mutations at *DFNB82* and *DFNX4* loci [Walsh *et al.*, 2010; Schraders *et al.*, 2011; Huebner *et al.*, 2011].

In addition, the idea to combine the targeted capture strategy with NGS has led to the development of specific clinical platforms, such as OtoSCOPE, which enable the screening of all known deafness genes in a single analysis, thus certainly improving the success rate of the genetic diagnosis, even in isolated cases [Brownstein *et al.*, 2012].

2. Aim of the Project

Nonsyndromic sensorineural hearing loss (NSHL) is the most common sensory disorder in humans, affecting about 278 million people worldwide, with severe consequences on public health. To date more than 110 deafness-associated loci have been mapped, although the responsible genes have been identified in only a subset of these (www.hereditaryhearingloss.org), leaving the vast majority of patients with no definitive genetic diagnosis. Among known deafness genes, mutations in the *MIR96* gene have been recently reported in 2 NSHL families [Mencía *et al.*, 2009], providing the first evidence of a direct implication of a miRNA mutation in a Mendelian disease.

Due to the genetic heterogeneity of the disease, as well as the low frequency of most known pathogenic mutations, targeted genetic screening of all known NSHL-causing genes by traditional mutation analysis methods is unfeasible. Nowadays, whole-exome sequencing (WES) represents one of the most efficient strategy to search for rare variants underlying Mendelian diseases, and has already been applied to the discovery of novel genes/mutations responsible for autosomal (*DFNB82* locus) and X-linked (*DFNX4* locus) recessive NSHL [Walsh *et al.*, 2010; Schraders *et al.*, 2011; Huebner *et al.*, 2011].

With this as background, we sought to investigate the genetic and molecular bases of NSHL by a double approach, consisting in:

a) functionally dissecting the pathogenic mechanism underlying novel mutations within the microRNA *MIR96* gene and defining their effects on miRNA biogenesis and on regulation of selected targets;

b) applying WES to the identification of novel genes/mutations in selected NSHL families with a clear recessive pattern of inheritance, and -where possible- characterizing them by *in vitro* studies in mammalian cell lines.

3. Results & Conclusions: miR-96

MiR-96 is part of the conserved miR-183 family (including miR-96, miR-182, and miR-183), which is essential for differentiation and function of the vertebrate inner ear. Point mutations within the seed region of miR-96 (miR-96-5p) were reported to cause autosomal dominant (AD) NSHL by altering correct miRNA/target recognition [Mencía *et al.*, 2009]. We screened the entire miR-183 family in 882 genetically undiagnosed Italian NSHL patients and 836 normal-hearing Italian controls, and identified one novel mutation within *MIR96*, miR-96(+57T>C) (NR_029512.1:c.57T>C; NT_007933.15:g.67447397A>G), in an AD-NSHL family. The mutation is located in the stem region of the pre-miRNA and replaces a residue that is fully conserved throughout vertebrate evolution. This variation affects the mature miR-96* (miR-96-3p), which is processed from the complementary strand of the miR-96 precursor. In particular, the change occurs at position +6, within the miR-96* seed region. The mutation, which introduces a base-pairing mismatch, is predicted to create an enlarged RNA bulge in the pre-miR-96 stem, close to the Dicer cleavage site, and to reduce the stability of the hairpin.

3.1 Effects of the miR-96 (+57T>C) mutation on miRNA biogenesis

With this as background, we decided to examine the impact of the novel variant on both miR-96 and miR-96* expression and maturation. To this aim, an expression vector (psiUX) containing the genomic DNA fragments corresponding to either the wild-type or the miR-96 (+57T>C) sequences, were generated. Transfection experiments in HeLa cells, followed by real-time RT-PCR, showed that both miR-96 and miR-96* mature species were significantly reduced in the mutant (85% reduction, $P = 0.0006$, for miR-96 and 77%, $P = 0.019$, for miR-96*), whereas the precursor level was unaffected. Moreover, the introduction of a compensatory mutation, +23A>G, which reconstitutes the pairing of the +57 nucleotide, and hence the physiologic secondary structure of the precursor, restored miR-96 and miR-96* expression to normal levels, thus demonstrating that the mutation hinders the precursor processing, probably interfering with Dicer cleavage.

As comparison, similar experiments were performed on one of the already known miR-96 mutations, +13G>A, previously demonstrated to decrease the mature miR-96 levels [Mencía *et al.*, 2009]. In this case, the mutant did not affect miR-96* expression but, as expected, caused a significant reduction in miR-96 levels, which remained altered also in the presence of the +66C>T compensatory variant. This evidence suggests that the pathogenic mechanism underlying the +13G>A mutation is at least partially different from the one of the +57T>C variation, being independent from the correct folding of the hairpin precursor.

The same approach was used to verify whether presumably non-pathogenic variants within *MIR96* do not alter its normal processing/expression. Hence, the two known polymorphisms within the loop region of the miR-96 precursor (rs73159662, +42C>T; and rs41274239, +36T>C) were selected. No significant alteration of miR-96, miR-96* and pre-miR-96 levels was detected (Figure 4).
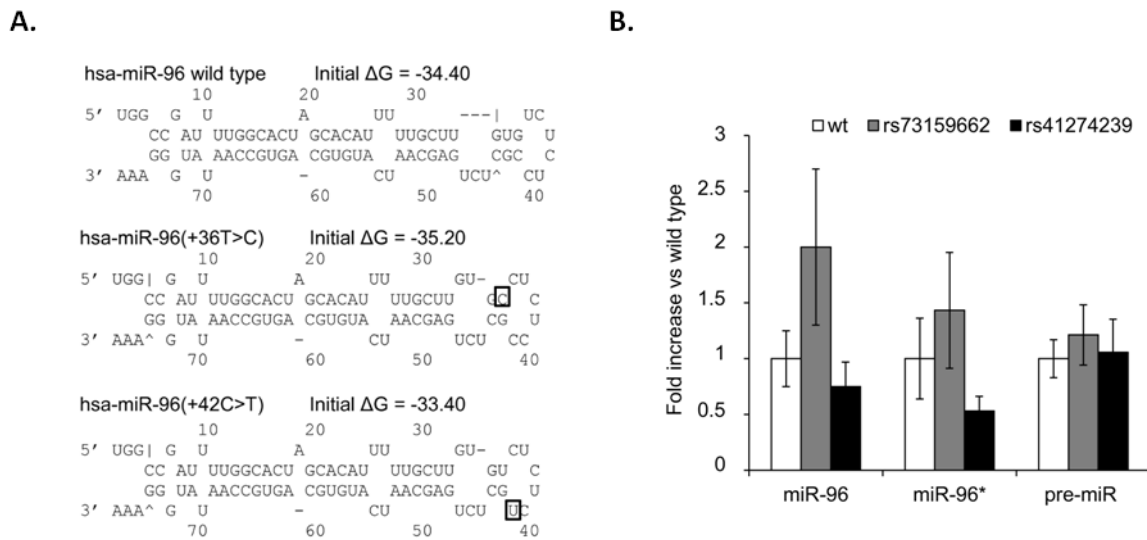
Figure 4. Effect of known *MIR96* polymorphisms on miRNA biogenesis. A. Predicted secondary structures of miR-96 precursors carrying the two known *MIR96* polymorphisms (rs73159662, +42C>T; and rs41274239, +36T>C) obtained by using the mfold algorithm (http://mfold.rna.albany.edu/?q=mfold). The nucleotide positions involved in the polymorphism are boxed. **B.** The effect of the analyzed polymorphisms on pre-miR-96 processing was evaluated by quantitative real-time RT-PCR using the ΔΔCt method. Variations in the expression levels of mature miRNAs and pre-miR-96 in the miR-96(+42C>T) and miR-96(+36T>C) mutants were compared to the wild-type samples (set as 1). Bars stand for mean ± SEM (represented as percentage of variation) of three independent experiments, each performed in triplicate in different days on different cell batches. The results were analyzed by unpaired t-test.

3.2 Does the miR-96 (+57T>C) mutation alter the miRNA function?

Since the +57T>C mutation directly impairs the seed-region of the mature miR-96* and concomitantly causes a quantitative defect in the miR-96 expression levels, we finally evaluated its effects on the regulation of both miR-96 and miR-96* mRNA targets by luciferase reporter assays. To this aim, eight genes, predicted as potential miR-96/96* targets by several bioinformatics programs, and reported in literature to be expressed in the auditory system, were selected for validation; three of them (*ACVR2B, CACNB4* and *MYRIP*) were predicted as common miR-96/miR-96* targets, whereas the other five (*ALCAM, BBS4, IQGAP2, PLS3* and

ROCK2) were targets of the miR-96* only. Luciferase reporter vectors (psi-CHECK2) containing the target 3' UTRs were constructed and used to co-transfect HeLa cells with a second vector (psiUX) expressing the wild-type or mutant pre-miR-96. This reporter system showed that three (*MYRIP, ACVR2B, CACNB4*) out of eight luciferase-3' UTR constructs were controlled by human miR-96/miR-96* and that the miR-96(+57T>C) mutation led, in all cases, to a significantly reduced silencing (~50%) of luciferase expression compared to the wild-type pre-miR-96 (Figure 5).
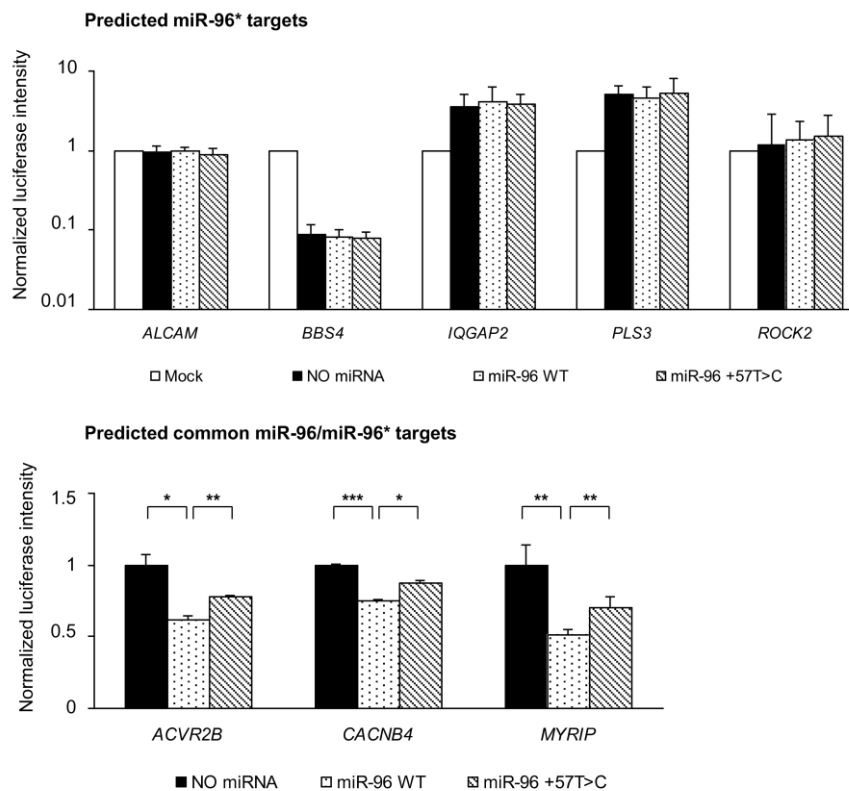
Figure 5. Validation of predicted miR-96/miR-96* targets by luciferase-based assays. Luciferase reporter assay data from HeLa cells transfected with the psiCHECK2 vector coupled to the 3'UTR regions of selected putative miR-96/miR-96* targets. Relative luciferase activity is expressed as mean ± SEM (six independent assays performed in triplicate for each target construct in different days and on different cell batches) and normalized to the luciferase activity of the respective psiCHECK2 plasmids (white bars). P values were calculated using unpaired t-test: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

To better discriminate between the relative effect of miR-96 and miR-96* on these targets, vectors carrying mutant 3' UTR deprived of the two miRNAs binding sites were produced by site-directed mutagenesis. This was particularly important in the case of *MYRIP*, whose 3' UTR contains an already validated miR-96 target site [Mencía *et al.*, 2009], and showed the most significant repression in the luciferase assay. The results of the experiment revealed that

MYRIP 3' UTR is regulated uniquely by miR-96, and that the predicted miR-96* binding site was not functional, being the construct lacking the miR-96* binding site still responsive to miR-96 repression whereas the construct deprived of miR-96 binding site was not. This evidence suggests that the impaired regulation of *MYRIP* in the miR-96(+57T>C) mutant is likely a consequence of the reduced levels of miR-96, caused by the defect in the hairpin processing. The result was confirmed by a luciferase assay performed in the presence of the +23A>G mutation, which was able to revert the phenotype, restoring the downregulation of *MYRIP* 3' UTR to a level similar to that of the wild-type miR-96 (data not shown).

These data led us to conclude that a quantitative defect in the production of this miRNA might be sufficient to deregulate its targets.

In conclusion, these data provide further evidence of the involvement of miR-96 in human deafness and demonstrate that a quantitative defect of this miRNA may contribute to NSHL (Figure 6).

The results of this study were published in Human Molecular Genetics (see Part II).
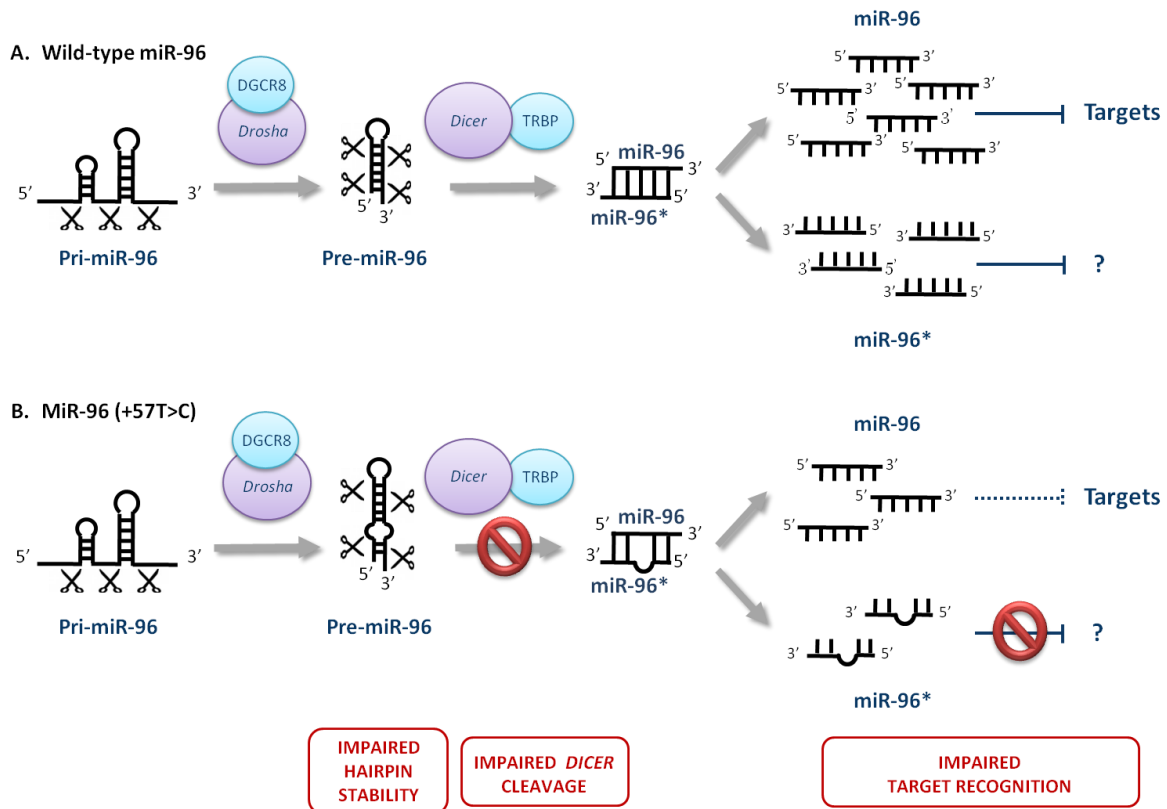
Figure 6. Schematic representation of the effects of the miR-96(+57T>C) mutation on miRNA biogenesis and function. Models of wild-type **(A)** and mutant **(B)** miR-96 maturation and target recognition. The stop symbols indicates a direct effect of the mutation on pre-miRNA processing as well as miR-96*/target recognition. Conversely, the dashed line points out an indirect effect of the mutation on miR-96 repression of its targets, which is a consequence of the reduction in its expression levels.

3.3 Conclusion and future perspectives

In this study, we identified and characterized a novel mutation (the third described) within the *MIR96* gene, miR-96 (+57T>C), in an Italian family with AD-NSHL. We demonstrated that this variant substantially impairs the production of mature miR-96 by altering the pre-miRNA secondary structure, and indirectly impacts on the normal regulation of miR-96 targets. Our results suggest that a quantitative defect of miR-96 can be sufficient to cause deafness independently from additional qualitative alterations and, together with the amount of data reported in literature, highlight the key role of miRNAs in the pathogenesis of diseases. Indeed, other studies have shown that common SNPs within miRNA genes can lead to impaired miRNA processing, and might be associated with increased susceptibility to human diseases.

In the case of the miR-96(+57T>C) mutation, however, a direct role of miR-96∗ in NSHL pathogenesis has still to be proved. Indeed, none of the putative miR-96∗ target sites was validated by luciferase-based assays (Figure 5). Profiling of miR-96∗ expression in a panel of human and mouse tissues evidenced that, although expressed at very low levels, miR-96∗ shows some degree of differential expression, suggesting that it might be regulated in a tissue-specific manner. Moreover, in the mouse organ of Corti, the miR-96∗ could be cloned, but not reliably quantitated by real-time RT–PCR.

More extensive and detailed experiments will be needed to define the role, if any, of this miRNA∗ species in the inner ear, as well as to evaluate the involvement of the mutant miR-96∗ in NSHL. To this purpose, we are performing in-situ hybridization assays, using miR-96∗-specific LNA probes, on mouse embryos at different developmental stages, 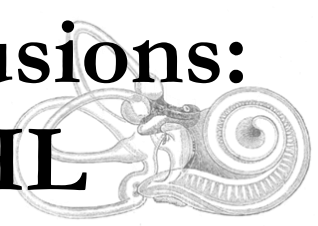to study miR-96∗ expression and to specifically define its possible involvement in the inner ear cell fate and differentiation. Moreover, expression experiments in UB/OC-2 cells are now being planned. This cell line is a conditional immortal cell line derived from hair cell progenitors from the H-2Kb-tsA58 transgenic mouse [Jat *et al.*, 1991; Holley and Lawlor, 1997]. This mouse carries a conditionally expressed, temperature-sensitive immortalizing gene that perpetuates cell division, preventing terminal differentiation at 33°C [Rivolta *et al.*, 1998]. UB/OC-2 cell line was immortalized at a late stage of differentiation and express a set of hair cell markers (i.e. alpha9AChR, Brn3c, POU4F3, and myosin VIIa) at relatively high levels, representing a valuable and straightforward model system for the study and characterization of deafness-associated microRNAs and genes. In particular, this strategy could help a more physiological evaluation of the effects of the +57T>C mutation on the expression and maturation of both miR-96 and miR-96*, in the identification of specific miR-96* targets and in the definition of the cellular and biological pathways in which the miR-96* is involved.

4. Results & Conclusions: identification of NSHL genes/mutations by WES

In recent years, the development of next generation sequencing (NGS) technologies, coupled with new DNA enrichment protocols, such as the selective capture of complete coding regions, has provided a powerful tool to identify in a cost- and time-effective way causative mutations underlying highly heterogeneous inherited diseases, such as NSHL. Here, we selected for exome sequencing six NSHL families (NSHL1-4 and 6, 7) with a recessive inheritance pattern of the disease and at least two affected individuals. All probands were negative for mutations in the *GJB2* and *GJB6* genes.

4.1 Data analysis flowchart

Exome sequencing was performed in collaboration with two different external service providers: the Beijing Genomic Institute (BGI), in the frame of the "1000 Mendelian disease project", and the Yale Center for Genome Analysis (YCGA, Yale University, USA). In both cases the sequencing was carried out using the Illumina HiSeq 2000 platform. The workflow followed for the NGS library preparation and sequencing is composed of five principal steps (Figure 7A): fragmentation, adaptor ligation, target capture by in-solution hybridization, bridge-PCR amplification, and sequencing. More in details: genomic DNA is randomly sheared into fragments of about 150-200 bp by sonication, adaptor ligated and purified. Extracted DNA fragments (~250 bp) are hybridized to a biotinylated RNA library for enrichment: hybridized fragments are bound to the strepavidin beads whereas non-hybridized fragments are washed out after 24h. Each captured library is amplified by ligation-mediated PCR and then sequenced.

Due to the high number of known deafness-causing loci and the low frequency of most mutations, we expected to find different causative genes/mutations in each family. For this reason, sequencing data originated from each family were analyzed separately.

After the sequencing, the obtained reads were passed through a quality control step and were mapped to the reference genome. Subsequently, single nucleotide variants (SNVs) and insertions/deletions (Indels) were identified and annotated according to their position and consequences on protein function, with the Annovar program (Figure 7B). Then, since disease-causing mutations are expected to be rare in the general population, all nonsynonymous changes were filtered against 1000 Genome variants databases (http://www.1000genomes.org), NHLBI GO Exome Sequencing Project (ESP, including data on 5400 exomes of Caucasian and African American origins, http://evs.gs.washington.edu/EVS/), and dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), in order to remove common variants. Moreover, for the analysis of consanguineous pedigrees, an additional step of analysis by "homozygosity mapping" (see Paragraph 4.4) was introduced in order to limit the search for candidate genes/variants to specific chromosomal region. Rare and potentially deleterious variants identified within known

NSHL loci in homozygous/hemizygous or compound heterozygous state, shared by affected siblings and not present in the healthy ones, were extracted and prioritized for validation.

In this study, WES on selected families was performed in three separate rounds of sequencing; for this reason the target enrichment was carried out using different capture kits.

In particular:

i. as a pilot study, the first round of sequencing was carried out on four index cases from different NSHL families (NSHL1-4), in order to maximize the chance to find novel disease genes. Target enrichment was performed using the SureSelect 38M capture kit (Agilent), which was the latest version available at the moment of sequencing. This platform, however, had the disadvantage of not completely covering all known NSHL genes. The obtained reads were mapped to the hg18 reference genome (March 2006 release) with SOAP2, and variants were called using the SOAPsnp program (http://soap.genomics.org.cn/soapsnp.html). This strategy let us to identify a novel pathogenic mutation within the *PRPS1* gene, which is responsible for the NSHL in family 4. In addition, we excluded several candidate variations, either in known or novel genes, in families 1-3, because of their lack of segregation with the phenotype in the corresponding pedigree.

ii. in the second round of sequencing, two additional siblings from the three NSHL families (NSHL1-3) remained without a definitive genetic diagnosis were subjected to WES. In this case, an updated enrichment kit was used (Sure Select 50M, Agilent), which was not available at the time of the first round of sequencing and allowed the capture of a broader set of targets, including previously missing deafness genes. The sequenced reads were mapped to the hg19 reference genome (February 2009 release) using bwa, and genotypes were called with both the SAMtools (http://samtools.sourceforge.net/) (for nucleotide variations) and GATK (http://www.broadinstitute.org/gatk/) (for Indels) softwares. By consequence, all data obtained from the pilot study were re-analyzed with the same pipeline, in order to make the results comparable. In this case, having sequenced few individuals from the same family, it was possible not only to prioritize candidate variants based on their location, frequency in the general population, and predicted effect on protein function, but also to combine/compare data to find variations shared by affected siblings and absent in healthy relatives, thus facilitating -at least in theory- the identification of pathogenic mutations. This approach led us to discover, in the NSHL3 family, a missense variant within *DIAPH2*, a gene not yet associated to hearing loss but extremely promising.

iii.   the third, and last, round of sequencing was carried out on two additional NSHL families (NSHL6 and 7). Also in this case, WES was performed on two siblings for each family. The Seq Cap v.2 capture kit (Roche NimbleGene) was used for target capture, whereas the data analysis pipeline was the same as the one used in the second round of sequencing. Two novel mutations were identified in the *TMPRSS3* gene, allowing us to obtain a definite molecular diagnosis for the NSHL6 family.

Finally, the actual presence of the identified potential disease-causing variants and their segregation with the pathology in the corresponding family was always confirmed by Sanger sequencing.
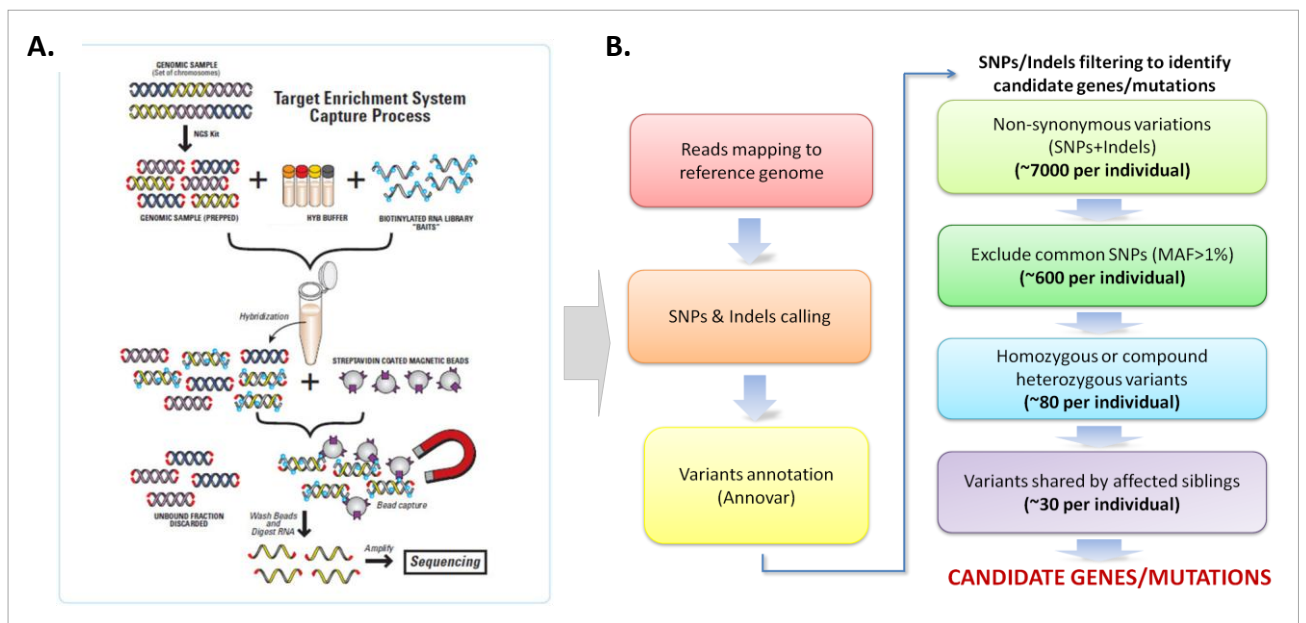

Figure 7. Whole exome sequencing (WES) pipeline. A. Schematic representation of targeted capture and exome sequencing using biotin-labeled in-solution hybridization probes. **B.** WES data analysis flowchart. Read mapping, variant calling, variant annotation, and filtering/prioritization steps are shown.

4.2 Mutations identified in NSHL-genes

4.2.1 NSHL6 family: the *TMPRSS3* gene

4.2.1.1 Exome sequencing identifies two novel mutations segregating with NSHL

We sequenced the exome of two individuals (II1 and II3) from an Italian family affected by postlingual bilateral NSHL, with a likely recessive pattern of inheritance (autosomal or X-linked). The probands are a 35 and a 23-years-old men with a normal-hearing sister (II2) (Figure 8A). Both probands had normal speech development, and are characterized by a down-sloping audiogram configuration, indicating a mild to severe impairment at low frequencies, and a profound hearing loss at high ones.

The analysis of the WES data, carried out following the previously described flowchart (Paragraph 4.1), revealed as most promising candidate variations two novel missense mutations within *TMPRSS3* (Table 2 and Figure 8B). The identified variants were a NM_024022:c.802T>C and a NM_024022:c.1307G>A nucleotide transitions, which result in the substitution of the Tryptophan in position 268 with an Arginine (NP_076927:p.Trp268Arg) and of the Arginine 436 with a Histidine (NP_076927:p.Arg436His), respectively. The mutations segregate with the NSHL phenotype in the patients' family, as confirmed by Sanger sequencing, being present in the compound heterozygous state in both analyzed probands (II1 and II3), but absent in the II2 healthy sister. Each parent carries one of the two variants in the heterozygous state (Figure 8C).

.

Table 2. Identification of candidate pathogenic variants in WES data of NSHL6 family.

Filter type	Nr of variants	
	II1	**II3**
NS/SS/I	9709	9722
Not in dbSNP132/1000Genomes	475	533
Not in ESP5400 exomes	310	350
Hom/compound Het	66	60
Shared by the affected siblings	16	
Within known NSHL genes	2	

NS, non-synonymous variant; SS, splice-site variant; I, coding indel. Variants present in dbSNP132, 1000 Genomes (February 2012 data release), and in 5400 exomes obtained from the Exome Variant Server (NHLBI GO Exome Sequencing Project), with an allele frequency threshold of 1%, were filtered out.

Both nucleotide substitutions were absent in a cohort of 127 Italian audiologically tested normal-hearing controls (mean age at withdrawal 32±9). Moreover, to better determine the incidence of the newly identified *TMPRSS3* mutations, 3500 exomes from an in-house database of Italian individuals with no familiarity for deafness were analyzed: while the p.Arg436His variant was identified in the heterozygous state in only one of these subjects, the p.Trp268Arg substitution was not found, suggesting that they may represent private variations rather than population-specific polymorphisms.

4.2.1.2 *TMPRSS3*

The *TMPRSS3* gene encodes a transmembrane serine protease which belongs to a subfamily of proteins also including TMPRSS1, TMPRSS2, TMPRSS4, TMPRSS5, MSPL and Enteropeptidase [Bugge *et al.*, 2009]. Like the other members of this family, TMPRSS3 is structurally defined by a transmembrane domain located near the amino terminus, a low density lipoprotein receptor A domain (LDLRA), which binds calcium [van Driel *et al.*, 1987] and low density lipoprotein [Sudhof *et al.*, 1985], a scavenger receptor cysteine rich domain (SRCR) that is involved in protein–protein interactions [Sarrias *et al.*, 2004], and a carboxy terminal serine protease domain (SP) from the S1 family of the SA clan of serine-type peptidases for which the prototype is chymotrypsin [Rawlings *et al.*, 2010].

The *TMPRSS3* gene, spanning approximately 24 kb on chromosome 21, is composed of 13 exons with the initiating codon in exon 2 [Charif *et al.*, 2012]. In human, there are four alternatively spliced transcripts (*TMPRSS3* a, b, c and d), encoding predicted polypeptides of 454, 327, 327 and 344 amino acids, respectively. Semi-quantitative RT–PCR revealed that all four transcripts show distinct patterns of expression, but *TMPRSS3*a, which is the isoform considered in this thesis, is the most abundantly and widely expressed one, also in fetal cochlea [Scott *et al.*, 2001]. In details, in the inner ear, *TMPRSS3* is expressed in the neuron bodies of the spiral ganglion, the stria vascularis and the epithelium of the organ of Corti [Duman and Tekin, 2012]. The function of the *TMPRSS3* gene in the auditory system is not fully understood, but it has been reported to play a crucial role in the morphological and functional maturation of the inner ear and maintenance of the correct composition of both perilymph and endolymph [Fasquelle *et al.*, 2011; Guipponi *et al.*, 2002]. Indeed, it has been demonstrated that *TMPRSS3* targets the epithelial amiloride sensitive sodium channel (ENaC), which mediates sodium reabsorption in the endolymph [Guipponi *et al.*, 2002]. Moreover, a recent study suggested that the gene dysfunction, due to a protein-truncating nonsense mutation, decreases the expression of Kcnma1 potassium

channels, which play a critical role in high-fidelity sound encoding and resting potential of inner hair cells [Molina *et al.*, 2013].

Mutations in this gene (MIM 605511) have been identified to segregate with hereditary nonsyndromic recessive deafness (*DFNB8/DFNB10* loci) [Bonne-Tamir *et al.*, 1996; Veske *et al.*, 1996], characterized by bilateral, postlingual, severe to profound hearing loss, with no described middle ear or vestibular deficits. The available clinical data of the analyzed families (to date about twenty-two) showed a characteristic audiological profile: in details, at a young age, *TMPRSS3*-associated hearing impairment was more pronounced at the high frequencies, and sooner or later, depending on the mutation, thresholds for the low frequencies deteriorated, eventually resulting in a flatter audiogram configuration (i.e., residual hearing). Therefore, a ski-slope audiogram configuration might be predictive of *TMPRSS3*-associated deafness [Weegerink *et al.*, 2011]. Environmental and/or genetic modifiers may influence the expressivity of hearing loss caused by *TMPRSS3* mutations [Fasquelle *et al.*, 2011].

4.2.1.3 Effects of the novel mutations on TMPRSS3 protein structure

Both identified mutations affect amino acids that are conserved in vertebrates and in the majority of related proteins belonging to the TMPRSS subfamily (Figure 9B). The effect of the two mutations on the protein structure were predicted using the Swiss-PDB Viewer program (http://spdbv.vital-it.ch) and molecular models of TMPRSS3 carrying the Trp268Arg and the Arg436His substitutions were built by homology modeling using epsin as template. Both amino-acid residues are located in the serine-protease domain (Figure 9A), within secondary structures (β sheets) relevant for the architecture of the catalytic domain and for the maintenance of the enzyme stability. In details, the identified variation within *TMPRSS3* exon 9 (p.Trp268Arg) is a not conservative substitution, which is predicted to lead to the generation of new hydrogen bonds with the surrounding amino acids, with a subsequent alteration of the local protein structure. In addition, the Tryptophan 268 is located close to an hydrophobic pocket, at the enzyme surface: the introduction of an Arginine residue at this position is responsible for a structural modification, thus resulting in the alteration of local electrostatic forces (Figure 9C). On the contrary, the mutation within *TMPRSS3* exon 12 (p.Arg436His), although conservative, causes a modification in the amino acidic interactions and a reduction in the local structure stability (Figure 9C).
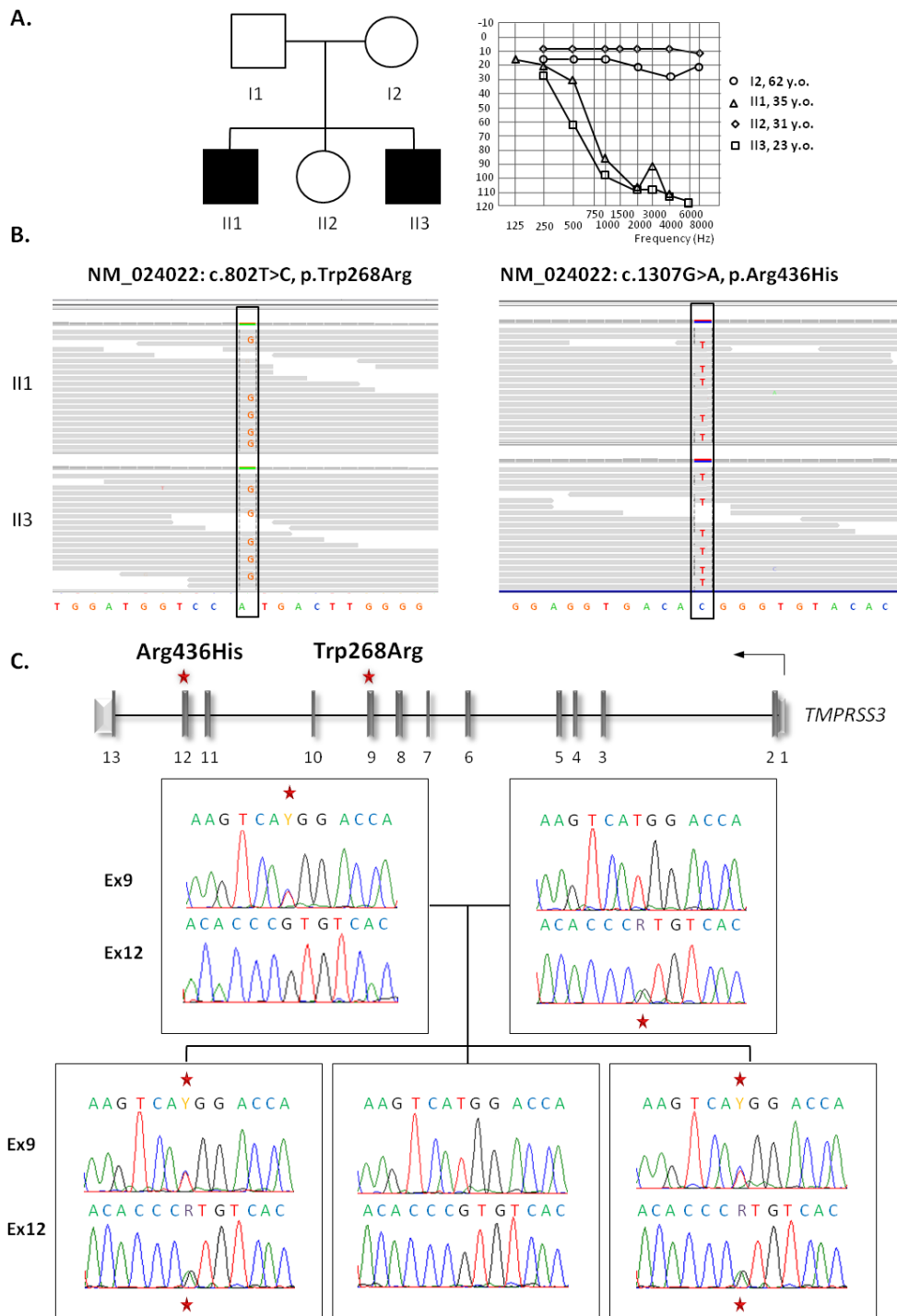
Figure 8. Exome sequencing identifies two novel missense mutations within *TMPRSS3*. A. Family tree and air conduction threshold measured by pure-tone audiometry (PTA) of NSHL6 family members (average hearing loss for the right and left ears). y.o.: years old. **B.** IGV (Integrative Genome Viewer) screenshots showing sequencing reads that support the two novel heterozygous variants (boxed in black) identified in both probands (II1 and II3). **C.** Schematic representation of the *TMPRSS3* gene and chromatograms indicating the genotype of all available family members. Non-coding exons (exon 1 and most of exon 13) are represented by light grey boxes. The two novel mutations, which localize within exon 9 and exon 12, are indicated by red stars.
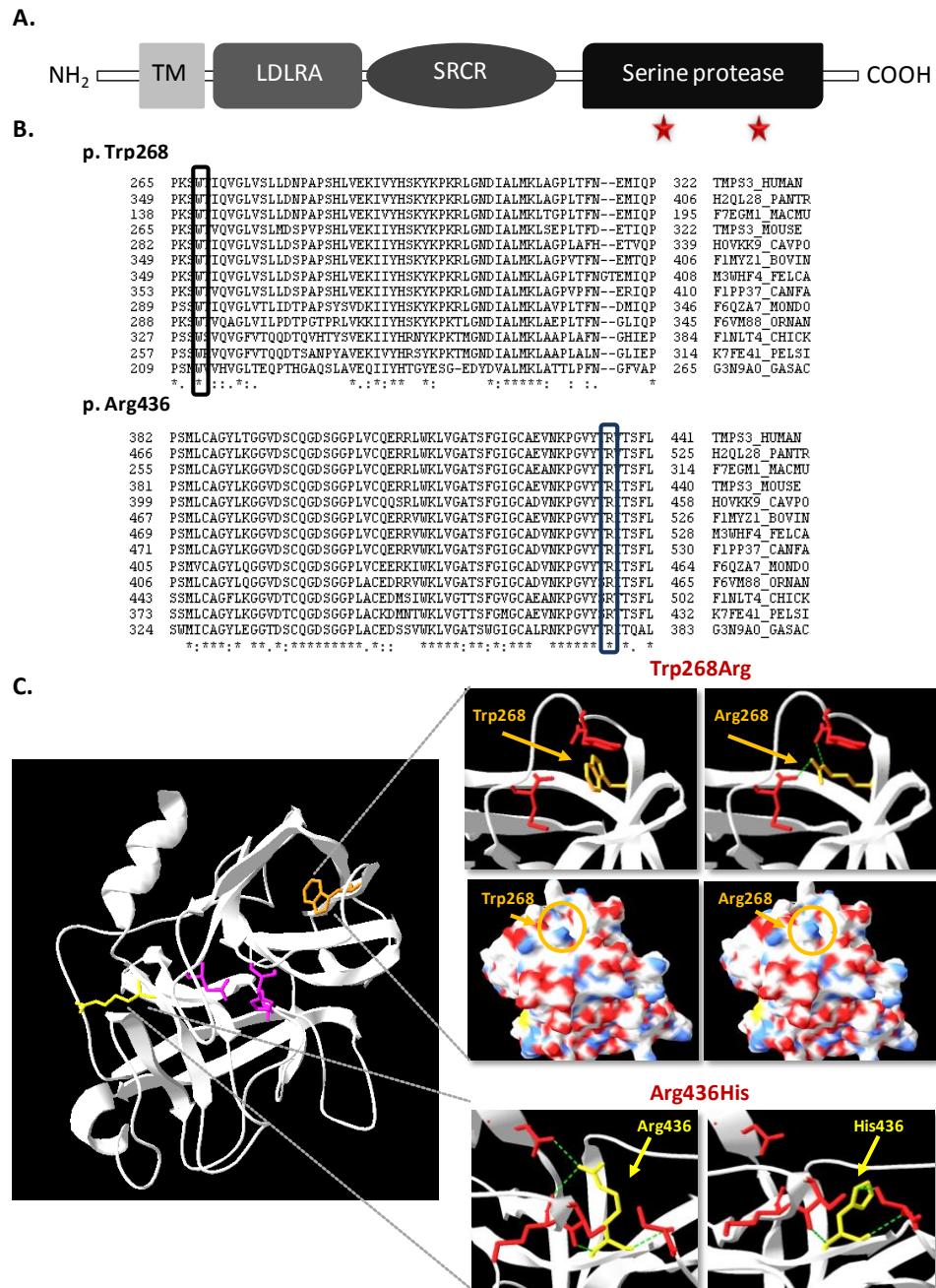
Figure 9. TMPRSS3 protein. A. Schematic representation of the TMPRSS3 protein, showing its four main structural domains, including the transmembrane (TM), low density lipoprotein receptor class A (LDLRA), scavenger receptor cysteine rich (SRCR), and serine protease domains. The localization of the newly identified mutations within the serine protease domain is indicated by red stars. **B.** Multiple alignment of TMPRSS3 amino acid sequences from different vertebrate species. Trp268 (boxed in black) and Arg436 (boxed in blue) residues are fully conserved. HUMAN (*Homo sapiens*, P57727); PANTR (*Pan troglodytes*, H2QL28); MACMU (*Macaca mulatta*, F7EGM1); MOUSE (*Mus musculus*, Q8K1T0); CAVPO (*Cavia porcellus*, H0VKK9); BOVINE (*Bos taurus*, F1MYZ1); FELCAN (*Felis catus*, M3WHF4); CANFA (*Canis familiaris*, F1PP37); MONDO (*Monodelphis domestica*, F6QZA7); ORNAN (*Ornithorhynchus anatinus*, F6VM88); CHICK (*Gallus gallus*, F1NLT4); PELSI (*Pelodiscus sinensis*, K7FE41); GASAC (*Gasterosteus aculeatus*, G3N9A0). **C.** Molecular modeling of the effects of the novel mutations on TMPRSS3 structure. The change of the Tryptophan in position 268 (indicated in orange) with an Arginine residue is expected to result in the alteration of the local protein structure and of the surface electrostatic forces. The Arginine 436 (in yellow) is substituted with an Histidine, with a consequent modification of the local interactions among surrounding amino acids. The Ser195, His57, Asp102 residues, which form the catalytic triad, are indicated in purple.

4.2.2 NSHL4 family: the *PRPS1* gene

4.2.2.1 Exome sequencing identifies a novel *PRPS1* mutation segregating with NSHL

We sequenced the exome of one individual (II2) from an Italian family (NSHL4) affected by postlingual bilateral profound NSHL, with a likely recessive mode of inheritance. The patient is a profoundly deaf 55-years-old man with an affected brother, one normal-hearing brother and one normal-hearing child (Figure 10A); the proband's mother is moderately deaf.

The analysis of the WES data (Table 3) revealed as the most promising candidate a novel missense mutation within exon 3 of *PRPS1,* a gene already known to be associated with X-linked syndromic and nonsyndromic hearing loss [de Brouwer *et al.*, 2010]. The variant was a NM_002764.3:c.337G>T nucleotide transversion, which is responsible for the substitution of the Alanine at position 113 with a Serine (NP_002755.1:p.A113S). The mutation segregates with deafness in the patient's family and was absent in a cohort of 123 Italian normal-hearing controls, and in 3500 exomes from an in-house database of Italian individuals with no familiarity for hearing loss.

Table 3. Identification of candidate pathogenic variants in WES data of NSHL4 family.

Filter type	Nr of variants
NS/SS/I	6855
Not in dbSNP130/1000Genomes	638
Not in ESP5400 exomes	281
Hom/compound Het	50
Within known NSHL genes	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, July 2010 data release) and then in 5400 exomes obtained from the Exome Variant Server, NHLBI GO Exome Sequencing Project.

4.2.2.2 *PRPS1*

PRPS1 codes for the phosphoribosylpyrophosphate synthetase 1 (PRS-I), a member of an evolutionarily conserved family of enzymes, which catalyzes the synthesis of phosphoribosyl

pyrophosphate (PRPP) from ATP and ribose-5-phosphate. These enzymes are essential for the *de novo* synthesis of nucleotides in both eukaryotes and prokaryotes [Becker, 2001].

The *PRPS1* gene has been associated with X-linked forms of phenotypically heterogeneous deafness (*DFNX1* locus). In general, hearing impairment in male patients with *PRPS1* mutations is bilateral, moderate to profound, and can be prelingual or postlingual, progressive or non-progressive. Female carriers may also be affected by unilateral or bilateral hearing impairment. So far, five *DFNX1* families were described, four with postlingual progressive NSHL, and one with congenital profound deafness [Liu *et al.*, 2010]. Four missense mutations in *PRPS1* (NM_002764.3:p.D65N, p.A87T, p.I290T, and p.G306R) were identified in these families as disease-causing. All of them result in a reduction of about half of PRS-I activity, as shown by *in vitro* enzymatic assays on patients' erythrocytes and cultured fibroblasts. None of these mutations are predicted to cause a major structural change in the PRS-I protein [Liu *et al.*, 2010].

Mutations in *PRPS1* might also result in a spectrum of syndromic conditions, including PRS-I superactivity (OMIM #300661; seven mutations so far identified) [Sperling *et al.*, 1972], Charcot-Marie-Tooth neuropathy type X-5 (CMTX5 or Rosenberg-Chutorian syndrome, OMIM #311070; two known mutations) [Kim *et al.*, 2007], and Arts syndrome (OMIM #301835; two mutations reported to date) [de Brouwer *et al.*, 2007]. In general, *PRPS1* gain-of-function mutations determine a loss of feedback regulation and hence enzyme superactivity, which in turn causes excessive uric acid production and gout. Conversely, loss-of-function mutations are responsible for NSHL, CMTX5, or Arts syndrome. In this case, the severity of the phenotype seems to correlate with the residual functionality of the enzyme, with milder mutations resulting only in hearing loss, and more severe genetic defects causing also peripheral neuropathy (CMTX5), or central nervous system involvement and increased liability to infections (Arts syndrome). Taken together, these findings suggest that the four *PRPS1*-related phenotypes are part of the same disease spectrum [de Brouwer *et al.*, 2010], thus stressing the importance of the functional characterization of pathogenic variants as well as the interpretation of genotype/phenotype correlations. In particular, CMTX5, which is characterized by a triad of symptoms: early-onset bilateral profound HL, peripheral neuropathy, and optic neuropathy, might be under-recognized by physicians and, thus, under-diagnosed [Liu *et al.*, 2013].

4.2.2.3 *PRPS1* mutations are a frequent cause of X-linked deafness

To analyze the recurrence of *PRPS1* mutations in X-linked sensorineural deafness, we screened by Sanger sequencing all seven *PRPS1* exons, including intron-exon boundaries, in 16 additional

unrelated male patients with familiarity for hearing loss, without reported neurological symptoms and a likely X-linked inheritance pattern.

Among these, we identified two additional novel missense variants. One was a NM_002764.3:c.343A>G transition in *PRPS1* exon3, leading to the NP_002755.1:p.M115V aminoacid change, which was found in a 12-years-old Italian patient affected by postlingual progressive bilateral hearing impairment. The proband has a normal-hearing brother, two affected cousins and two affected uncles. The proband's mother, obligate female carrier for the mutation, suffers from a mild form of deafness. The mutation segregates with hearing loss in the family (NSHL5; Figure 10B).

Finally, the third identified mutation was a NM_002764.3:c.925G>T transversion in *PRPS1* exon7, which results in the substitution of the Valine at position 309 with a Phenylalanine (NP_002755.1:p.V309F). This variant was found in a 14-years-old proband from a Peruvian family (NSHL13; Figure 10C) affected by postlingual bilateral progressive deafness, characterized by an audiometric profile very similar to that of the NSHL5 proband. Also in this case, the mutation segregates with X-linked hearing loss in the family, being present in an affected uncle from the proband's mother side, who has profound hearing loss.

4.2.2.4 The p.M115V and p.V309F mutations are associated with unrecognized peripheral neuropathy

As *PRPS1* mutations may cause peripheral neuropathy, all *PRPS1*-mutated patients underwent neurological evaluation at Clinical Centre for hereditary neuropathies (Ospedale San Raffaele, Milano). Among the three families, patients in NSHL5 (p.M115V) and NSHL13 (p.V309F) families showed clinical or subclinical signs of peripheral neuropathy. As expected, males and females showed significant differences at neurological examination: in females we only observed subclinical signs of peripheral neuropathy (pes cavus, reduced or absent deep tendon reflexes, chronic denervation in distal muscles of lower limbs), whereas males displayed more evident findings of neuropathy. Even in males, the neuropathy was mildly symptomatic, did not cause weakness or evident motor deficits (except for III2, NSHL5 family), whereas prevailed sensory signs and symptoms, such as Romberg positivity, absent deep tendon reflexes, paresthesias and cramps. Almost all patients, except the carrier female from NSHL13 family, showed chronic denervation at EMG evaluation, whereas only males showed mild/moderate axonal neuropathy. In NSHL5 family, we observed reduction of both motor and sensory nerve amplitudes, whereas in NSHL13 family only motor nerves were affected.

A.

B.

C.

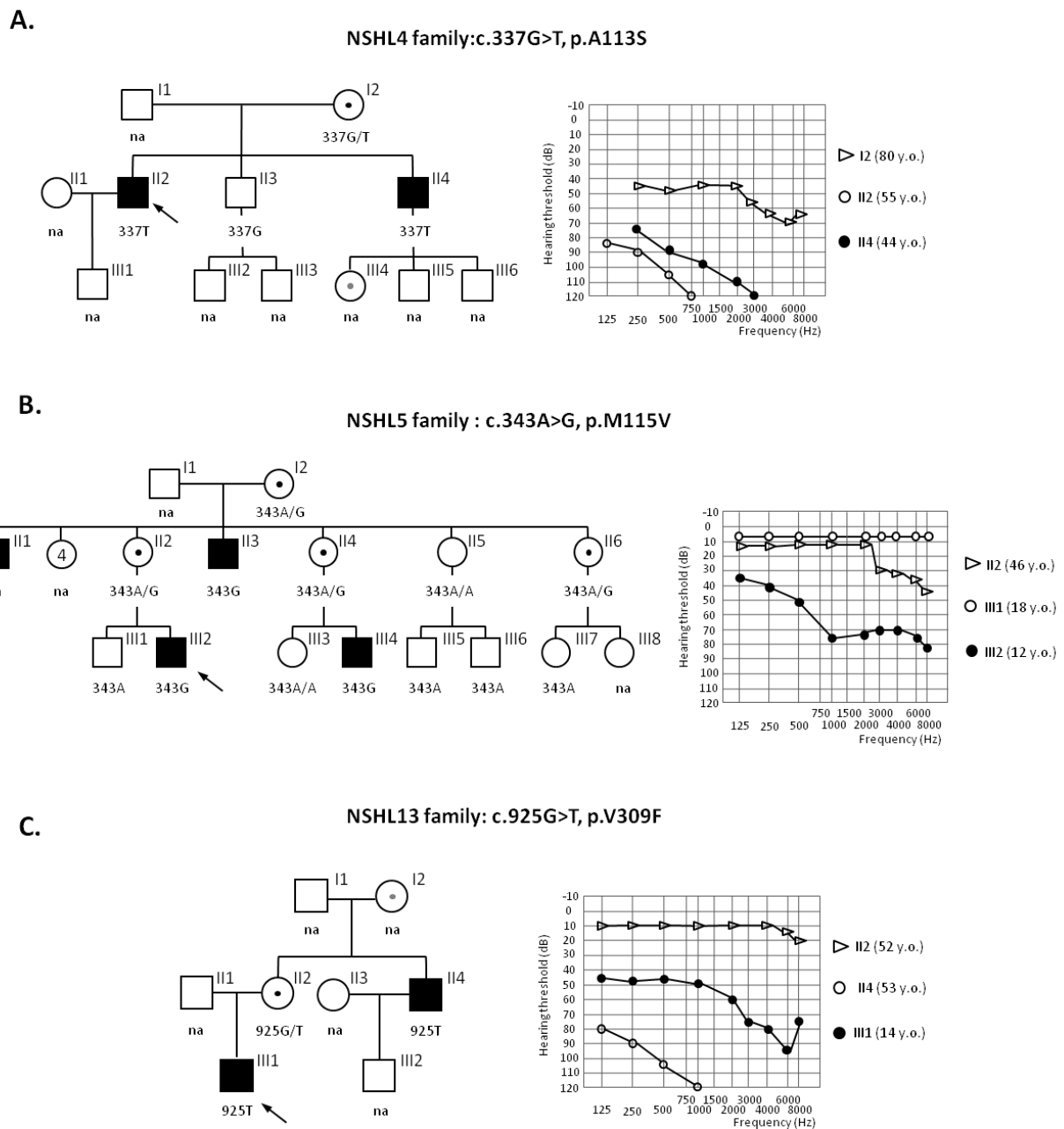**Figure 10. Identification of novel *PRPS1* mutations segregating with X-linked deafness, with or without neurological symptoms.** Pedigree of NSHL4 **(A)**, NSHL5 **(B)** and NSHL13 **(C)** families and corresponding audiograms, showing the average hearing loss for the right and left ears. The genotype of each individual is indicated below the corresponding symbols. The probands are pointed by an arrow. na: not analyzed.

4.2.2.5 The newly identified *PRPS1* mutations significantly reduce the enzyme activity *in vivo*

The open reading frame of human *PRPS1* consists of 957 nucleotides coding for 318 amino acids (GenBank accession number NM_002764). All three novel mutations affect amino acid residues that are highly evolutionarily conserved, from zebrafish to humans (Figure 11).

PRS-I is thought to be physiologically functional as a hexamer, which consists of three homodimers arranged in a propeller-like shape, each with an active site and two regulatory allosteric sites (I and II) [Eriksen *et al.*, 2000; Li *et al.*, 2007]. The active site comprises binding sites for both ATP and ribose-5-phosphate, is largely located at the interface of two subunits within one homodimer (dimer interface), and extends to the interface between homodimers (trimer interface). Allosteric site I is located at the trimer interface, and allosteric site II is at the dimer interface [de Brouwer *et al.*, 2010]. To predict the effects of the identified missense variants on the structure/function of PRS-I, molecular models of the protein carrying the p.A113S, the p.M115V, or the p.V309F change were hence built on the crystal structure of human PRS-I (PDB entry: 2H06; http://www.pdb.org) using the FoldX plugin for the Yet Another Scientific Artificial Reality Application (YASARA) program (http://www.yasara.org/).

The p.A113S mutation is located in an alpha-helix participating to the trimer interface (Figure 12). In particular, Ala113 is not contributing to the trimer interface, but is pointing towards the tightly packed hydrophobic core of the N-terminal domain of the corresponding subunit. Changing Ala113 to a more polar Serine side chain will destabilize the surrounding environment, affecting ATP-binding (Figure 12, panels A and B).

Met115 is located in the same alpha-helix, and is pointing towards the trimer interface. Two Methionine 115 side chains from two different PRS-I dimers partecipate in a hydrophobic interaction at the interface. The p.M115V substitution is thus predicted to disturb the packing, likely destabilizing the ATP-binding site and the allosteric site I, since both consist of amino acids of different PRS-I molecules coming together at the trimer interface (Figure 12, panels C and D).

Val309 is part of the allosteric site I and is located at the trimer interface near the center of the hexamer. The p.V309F is predicted to disturb the hexameric structure and thus the allosteric site I function (Figure 12, panels E and F).

To assess the consequences of these mutations on protein function, we measured PRS-I enzymatic activity in erythrocytes from five affected males (II2 and II4 from NSHL4 family; III2 and III4 from NSHL5 family; III1 from NSHL13 family), four female carriers (II2, II4 and II6 from NSHL5 family; II2 from NSHL13 family), as well as one unaffected male (III1 from NSHL5 family) and four unrelated healthy controls (CTR1-4, two males, two females).
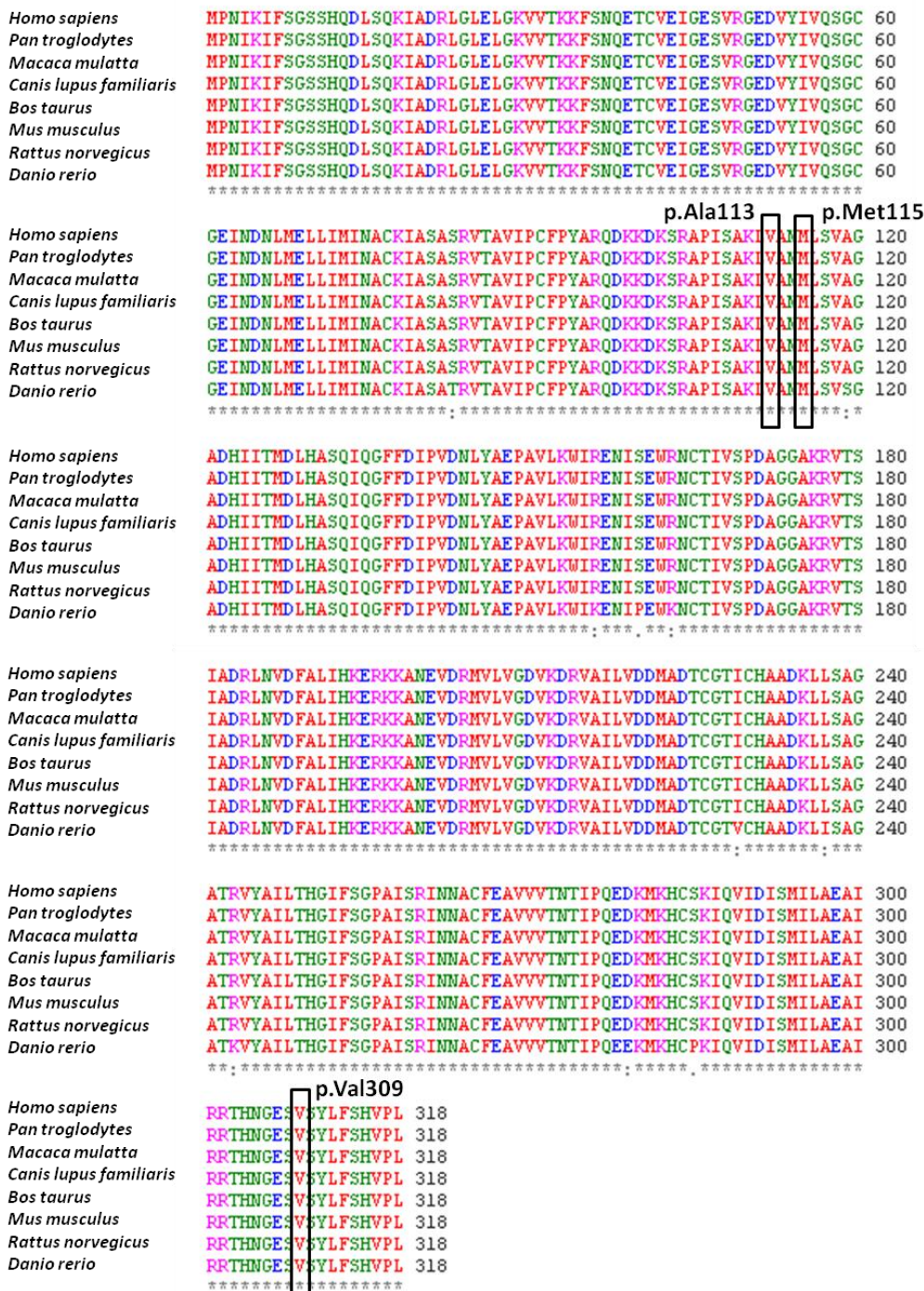
Figure 11. Conservation of the PRPS1 protein. Multiple alignment of PRPS1 protein sequences from different vertebrate species (ClustalW2 program, http://www.ebi.ac.uk/Tools/msa/clustalw2/). The newly identified mutations (boxed in black) affect amino acids fully conserved across evolution. *Homo sapiens*, 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; *Danio rerio*, NP_001070036.1. The colours of amino acids are according to their physicochemical properties. * indicates positions which have a single, fully conserved residue; : indicates conservation between groups of strongly similar properties; . indicates conservation between groups of weakly similar properties.
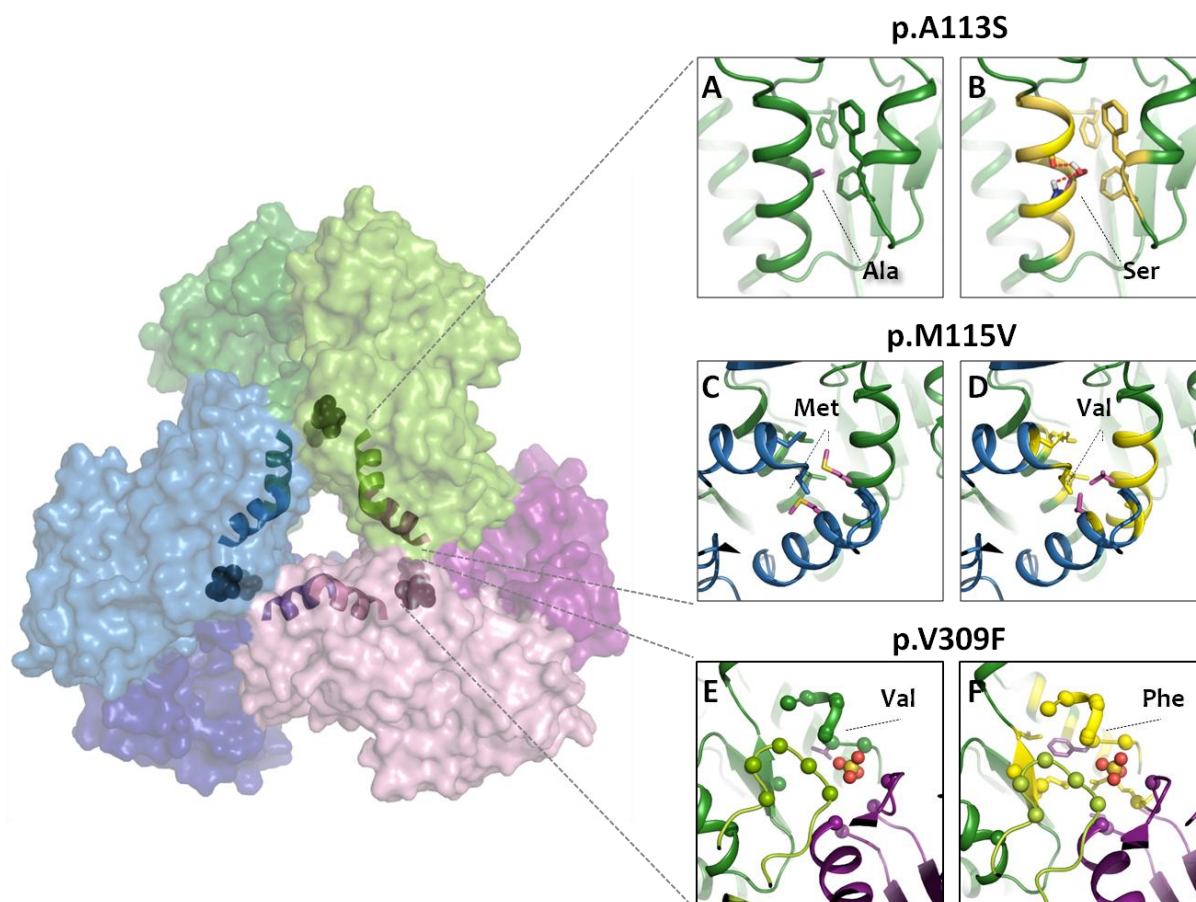
Figure 12. Structural analysis of the newly identified PRS-I mutations. On the left: Molecular surface of the three-dimensional structure of the PRS-I enzyme in its physiologic assembly, with the six different subunits indicated by separated colors. The 2 monomers forming a dimer are highlighted with different shades of the same color. The alpha-helices hosting p.A113S and p.M115V mutations are shown as ribbons. Residue Val309 is shown by spheres. **A-B.** Structural analysis of the p.A113S mutation. Left: Ala113 is completely buried in the hydrophobic core of the corresponding subunit, interacting with Phe92, Phe137, and Phe138. Right: Ser113 differs from Alanine in that one of the methylenic hydrogens is replaced by a hydroxyl group, through which this amino acid makes a new hydrogen bond. The larger and polar (hydrophilic) side chain of Ser113 makes this region to rearrange (highlight in yellow). **C-D.** Structural analysis of the p.M115V mutation. Left: Met115 is completely buried at the trimer interface, making hydrophobic interactions with the surrounding amino acids. Right: Val115 is a beta-branched residue, introducing a completely different steric hindrance, making this region to rearrange (highlight in yellow). **E-F.** Structural analysis of the p.V309F mutation. Left: Val309 is part of the allosteric site I and is located at the trimer interface near the center of the hexamer. Right: Phe309 is bulkier than Val and surrounding residues are predicted to move to avoid Van der Waals crashes. Residues of allosteric site I are indicated with spheres. A SO_4^{2-} ion present in the crystal structure is also shown. In all panels, the mutated residue is shown by stick: C atoms are colored magenta, O red, S yellow and H white. Residues predicted by FoldX to be perturbed are colored in yellow. Figures were produced using PyMOL (http://www.pymol.org).

The assay is based on HPLC measurement of AMP, which is produced in equimolar amounts with PRPP. The hemolysates were incubated for 60 minutes at 37°C with saturating amounts of substrates and an inhibitor of adenylate kinase (Ap5A) to prevent conversion of AMP to ADP. Aliquots of each reaction were terminated at five different times (0, 15, 30, 45, and 60 min), and proteins removed by filtration. The amount of ATP and AMP in the eluates was evaluated by high-performance liquid chromatography (HPLC) separation. Statistically significant differences in PRS-I activity in erythrocytes of affected males, female carriers, and wild-type controls were observed (ANOVA $p<2*10^{-5}$, at time point 60 min) (Figure 13). In particular, a significant reduction was evident in affected males, although the impairment in the enzymatic activity seemed less severe in patients carrying the p.A113S mutation than in the probands with the p.M115V or p.V309F substitutions. Indeed, PRS-I activity in patients II2 and II4 from NSHL4 family (p.A113S/-) was 5.29 and 7.33 nmoles/mg/hr, respectively, whereas in the p.M115V/- patients the enzyme activity was reduced to 2.11 and 1.06 nmoles/mg/hr and in the NSHL13 family proband (III1, p.V309F/-) to 0.84.

4.2.2.6 *PRPS1* allelic expression is unbalanced in the female carrier of the p.V309F mutation

The p.M115V heterozygous female (II2, NSHL5 family) suffers from a mild form of deafness whereas the p.V309F II2 carrier from NSHL13 family shows normal hearing (Figure 10, panels B and C). However, both these women showed a decreased PRS-I activity similar to that observed in patients hemizygous for the p.A113S substitution. Starting from this evidence, we investigated the possibility of an unbalanced expression of the wild-type (wt) and the mutant alleles. First, the relative expression of the two alleles was assessed by comparing the sequence of *PRPS1* mRNA at the mutant position, amplified by RT-PCR from PBMC RNA, with that obtained from genomic DNA. The results showed a preferential expression of the mutant *PRPS1* allele compared to the wild-type one in the c.343A>G carrier of NSHL5 family (60% mutant *vs* 40% wt allele) (Figure 14A); conversely, an almost-exclusive expression of the wild-type *PRPS1* allele in comparison to the mutant one in the c.925G>T carrier of NSHL13 family was observed (85% wt *vs* 15% mutant allele) (Figure 14B). However, while the unbalanced allelic expression observed in the individual from NSHL5 family is not due to a skewed inactivation of the X-chromosome, as determined by methylation assays, in the II2 carrier from NSHL13 family a mild skewed inactivation (≥75%) of the mutant X-chromosome is evident (Figure 14).

A.

B.

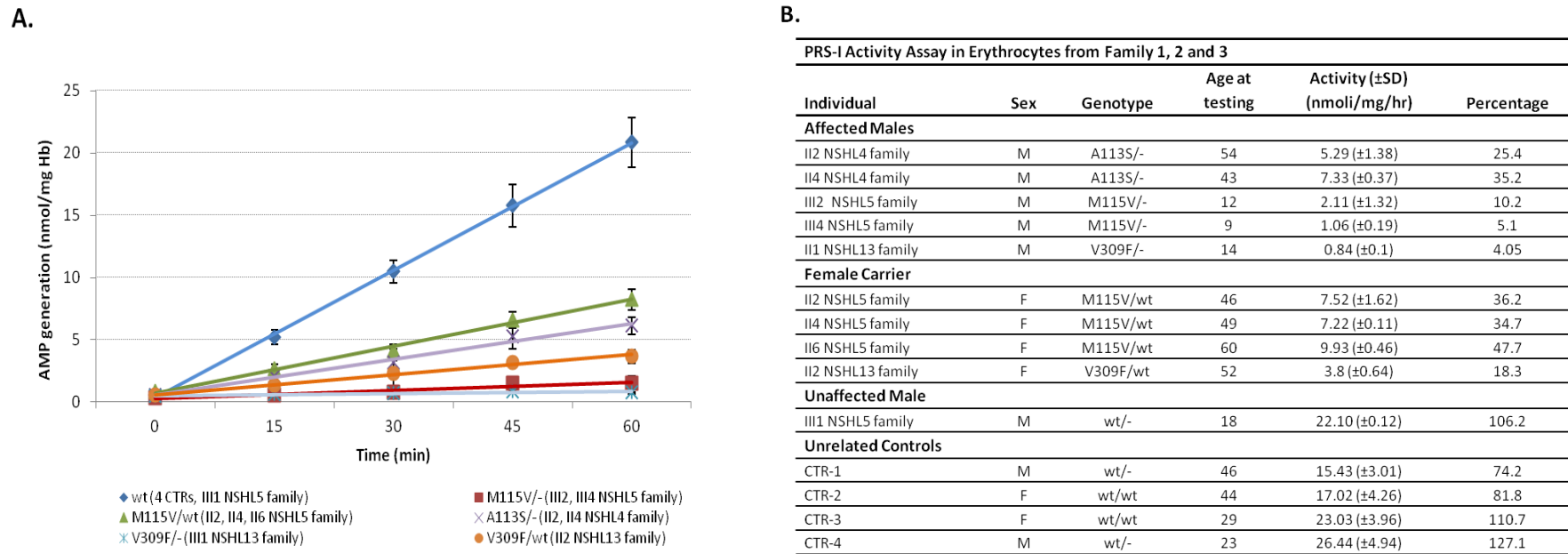| PRS-I Activity Assay in Erythrocytes from Family 1, 2 and 3					
Individual	Sex	Genotype	Age at testing	Activity (±SD) (nmoli/mg/hr)	Percentage
Affected Males					
II2 NSHL4 family	M	A113S/-	54	5.29 (±1.38)	25.4
II4 NSHL4 family	M	A113S/-	43	7.33 (±0.37)	35.2
III2 NSHL5 family	M	M115V/-	12	2.11 (±1.32)	10.2
III4 NSHL5 family	M	M115V/-	9	1.06 (±0.19)	5.1
II1 NSHL13 family	M	V309F/-	14	0.84 (±0.1)	4.05
Female Carrier					
II2 NSHL5 family	F	M115V/wt	46	7.52 (±1.62)	36.2
II4 NSHL5 family	F	M115V/wt	49	7.22 (±0.11)	34.7
II6 NSHL5 family	F	M115V/wt	60	9.93 (±0.46)	47.7
II2 NSHL13 family	F	V309F/wt	52	3.8 (±0.64)	18.3
Unaffected Male					
III1 NSHL5 family	M	wt/-	18	22.10 (±0.12)	106.2
Unrelated Controls					
CTR-1	M	wt/-	46	15.43 (±3.01)	74.2
CTR-2	F	wt/wt	44	17.02 (±4.26)	81.8
CTR-3	F	wt/wt	29	23.03 (±3.96)	110.7
CTR-4	M	wt/-	23	26.44 (±4.94)	127.1

Figure 13. Functional characterization of the novel *PRPS1* mutations. A. PRS-I activity in erythrocytes was evaluated by measuring the accumulation of AMP by HPLC in five affected males, four female carriers, as well as five healthy control individuals (wt). More in details, venous blood samples were collected in EDTA tubes and immediately after phlebotomy, peripheral blood mononuclear cells (PBMCs) and erythrocytes were separated by centrifugation on a Lympholyte Cell separation media gradient (Cederlane Laboratories Limited). Red cells were then washed three times in phosphate buffered saline, packed by centrifugation for 10 min at 2000 rpm, and stored at -20°C. On the day of the assay, each sample was thawed and treated with activated charcoal (3 mg/mL) for 15 min at 4°C. After centrifugation, the supernatant was assayed for PRS-1 activity, essentially as described by Torres and colleagues [Torres *et al.*, 1994]. Briefly: 100 µL of the charcoal-treated hemolysate were incubated for 60 min at 37°C with 1.9 mL of the pH 7.4 reaction buffer (50 mM Tris-HCl, 5 mM MgCl$_2$, 1 mM EDTA, 0.4 mM DTT, 0.5 mM ATP, 0.35 mM ribose-5-phosphate, 32 mM NaPi, and 0.25 mM diadenosine pentaphosphate, Ap5A). At five time points (0, 15, 30, 45, and 60 min), a 0.3 mL aliquot of each reaction was stopped by adding 30 µL of 100 mM EDTA, and centrifuged in Amicon Ultra Ultracel 10K Membrane filter (Millipore) at 4000 x g for 30 min at 4°C to remove proteins. Control reactions without ribose-5-phosphate were set up for each time-point using the hemolysate of a healthy control. ATP, ADP, and AMP in the filtrate were separated by HPLC. The graphic here reported show the correlation between ribose-5-phosphate-dependent AMP generation (y-axis) and the reaction incubation time (x-axis). Error bars represent the standard error of the mean (SEM) for genotypes represented by >1 individual (wt, n=5; M115V/-, n=2; M115V/wt, n=2; A113S/-, n=2), whereas they indicate the standard deviation (SD) for genotypes represented by a single individual. **B.** Summary of the functional data obtained from PRPP synthetase activity assay. The enzyme activity is expressed as nmoles of AMP per mg of Hb per hr. SD was calculated from three independent experiments performed in triplicate. The percentage was calculated as the ratio between the individual PRS-I activity and the mean of the PRS-I activity of the unaffected male and wt controls.
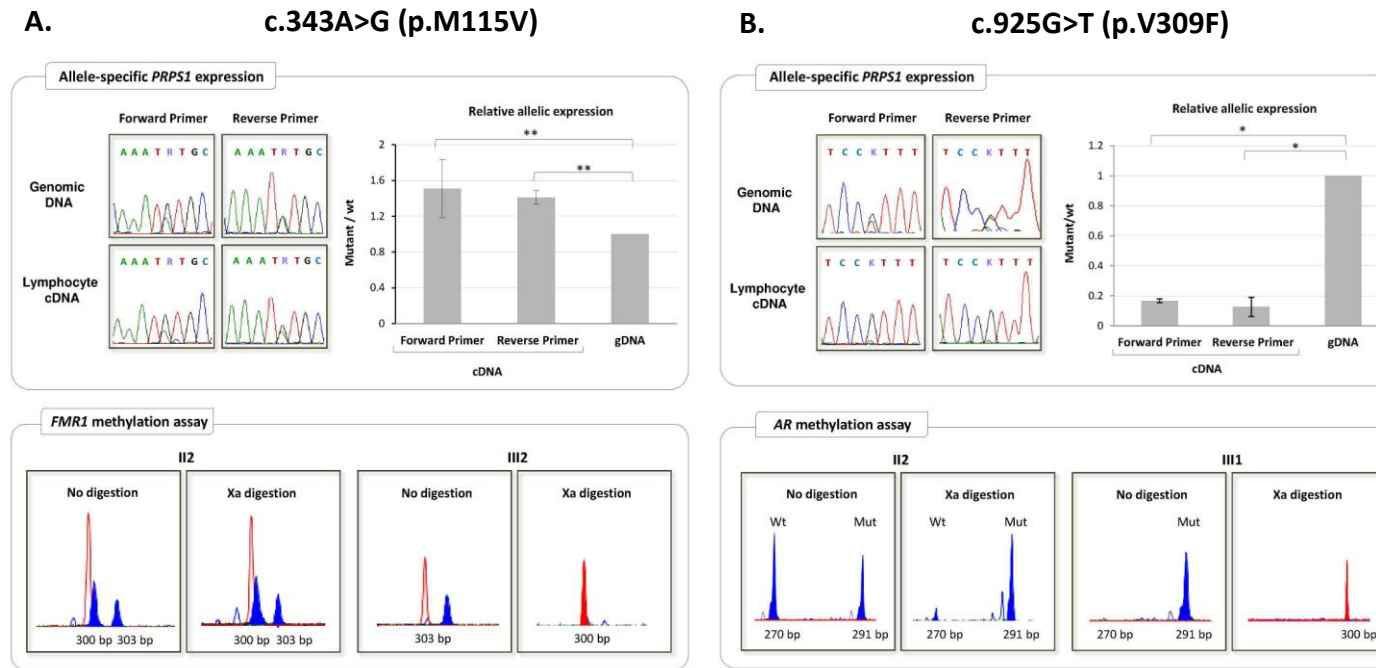
Figure 14. *PRPS1* allelic-specific expression in c.343A>G (p.M115V) and c.925G>T (p.V309F) carriers. *Upper panels*. On the left: DNA sequence electropherograms showing the region surrounding the variant position, PCR amplified from genomic DNA or from lymphocyte cDNA of either the heterozygous individual 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*.** Peak Scan 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 (NSHL5 family, **A**) or on the *AR* CAG polymorphic region (NSHL13 family, **B**), by modifying a previously described protocol [Coene *et al.*, 2009]. Briefly, 100 ng of genomic DNA 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), 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. PCR reactions were separated on an ABI-3130XL sequencer (Applied Biosystems) and the peak areas measured by the Peak Scanner v.1.0 software (Applied Biosystems). 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 HD Size Standard (Applied Biosystems) is shown in red.

4.2.2.7 *PRPS1, PRPS2, PPAT,* and *HGPRT* expression levels in *PRPS1* mutation carriers

To verify whether other modifier genes may contribute to the phenotypic variability among the three analyzed families, we measured expression levels of *PRPS2*, PRPP amidotransferase (*PPAT*), and hypoxanthine-guanine phosphoribosyltransferase (*HGPRT*), all known to be involved in the purine metabolism pathway [Liu *et al.*, 2013]. To this aim, we performed semi-quantitative real-time RT-PCR assays on RNA extracted from PBMCs of the five deaf patients (two with the p.A113S substitution, two with the p.M115V mutation and one with the p.V309F variation), four female carriers (three with the p.M115V amino acid change and one with the p.V309F substitution), and six controls (three males, three females). 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 carrier of the mutations in comparison either to the healthy controls or the affected males (Figure 15).
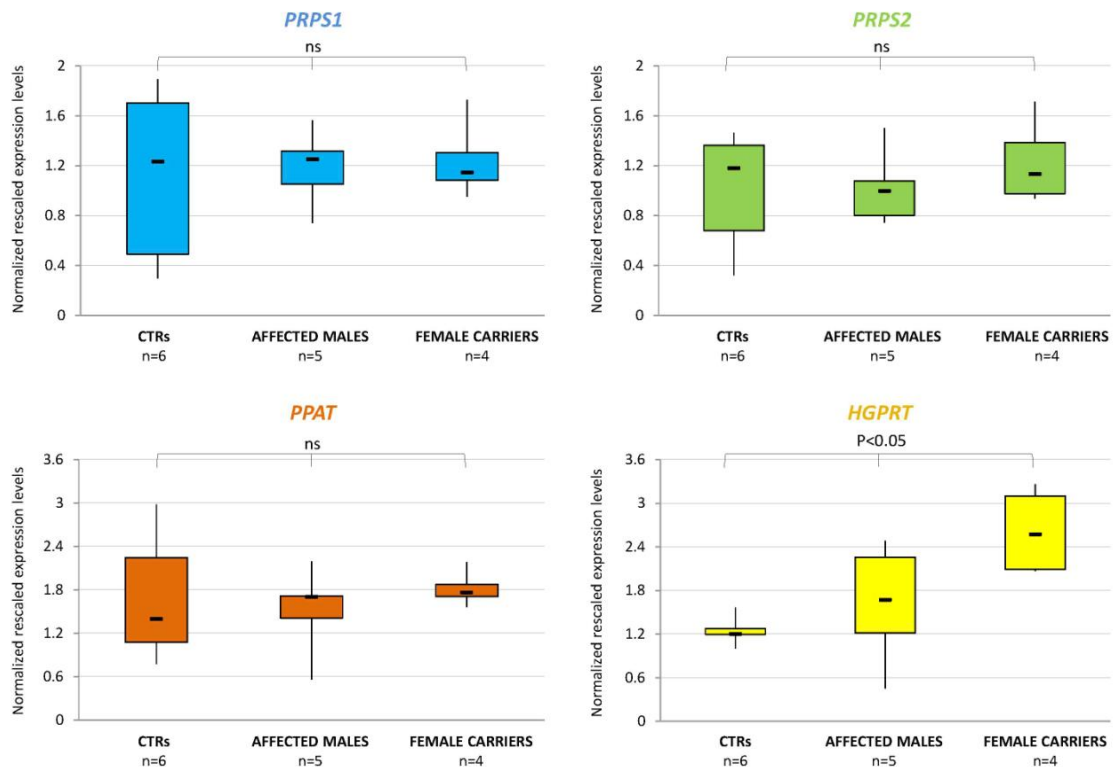
Figure 15. *PRPS1, PRPS2, PPAT,* and *HGPRT* expression levels by qPCR. *PRPS1*, *PRPS2*, *PPAT*, and *HGPRT* expression levels were measured by real-time semi-quantitative RT-PCR in PBMCs from five affected males (two carrying the p.A113S, two carrying the p.M115V, and one carrying the p.V309F mutation), four female heterozygotes (three for the p.M115V, one for the p.V309F variant), and six wild-type controls (CTRs). 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.

The results of this study are under submission to Journal of Medical Genetics (see Part III).

4.3 Mutations identified in novel genes: the NSHL3 family

4.3.1 Identification of a novel mutation within *DIAPH2*

The III3, III4 and III5 siblings (two affected and one normal-hearing brothers) from NSHL3 family were selected for exome sequencing (Figure 16). The WES data analysis (Table 4) highlighted the presence of a novel missense/splicing variant (NM_006729.4:c.868A>G, NP_006720.1:p.Ile290Val) in the *DIAPH*2 gene, located on the X chromosome.

Table 4. Identification of candidate pathogenic variants in WES data of NSHL3 family.

Filter type	Nr of variants
NS/SS/I	~10700
Not in dbSNP132/1000Genomes	~607
Not in ESP5400 exomes	~346
Hom/compound Het	~111
Shared by the III3 affected sibling (III3) and not present in the healthy brother (III4)*	54
Shared with the III5 affected sibling**	2
Within known NSHL genes	/

NS, non-synonymous variant; SS, splice-site variant; I, coding indel. Variants present in dbSNP132, 1000 Genomes (February 2012 data release), and in 5400 exomes obtained from the Exome Variant Server (NHLBI GO Exome Sequencing Project), with an allele frequency threshold of 1%, were filtered out.
* Subjects sequenced with the Sure Select 50M capture kit; ** Subject sequenced with the Sure Select 38M capture kit (pilot study).

The family tree is compatible with an X-linked transmission of hearing loss, and the segregation of the identified variation with the disease in the family was confirmed by Sanger sequencing. Indeed, the A>G transition was present in the hemyzygous state in the two affected siblings (III3 and III5) and in the heterozygous state in both the mother (II3) and grandmother (I4), whereas the normal-hearing brother (III4) and third-degree cousin (III7) carried the wild-type allele only. The presence/absence of the mutation was checked in all the other available family members (II2, II5, II6). Moreover, the variant was not identified in a cohort of 126 Italian audiologically-

tested, normal-hearing controls, as well as in an in-house database collecting all variants identified by WES in a large (>3000) cohort of Italian subjects.
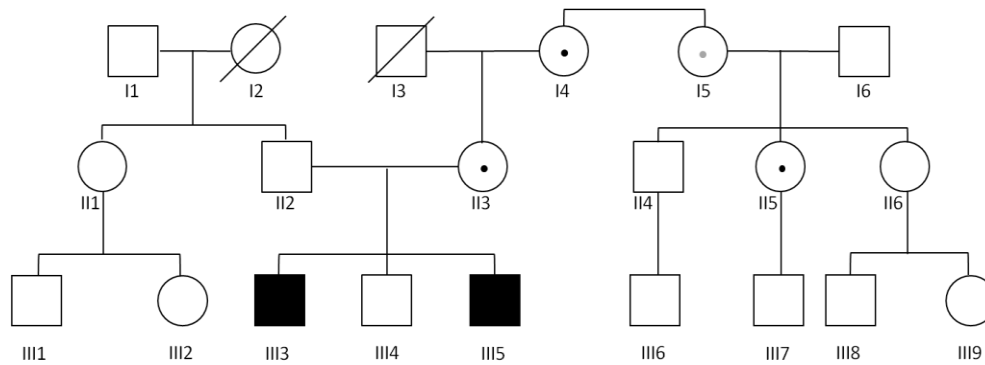
Figure 16. Pedigree of the NSHL3 family. ◦ = Obligate female carrier of the mutation. • = Heterozygous female carrier of the mutation, as confirmed by Sanger sequencing.

Given this interesting result, we checked the expression of *DIAPH2* mouse homolog (*Diap2*) in the inner ear, by performing an RT-PCR assay on a sample of organ of Corti derived from a post-natal day 4 mouse (P4) (Figure 17). The amplification of a product of the expected size (394 bp) supports the hypothesis of a possible involvement of *DIAPH2* in hearing loss.
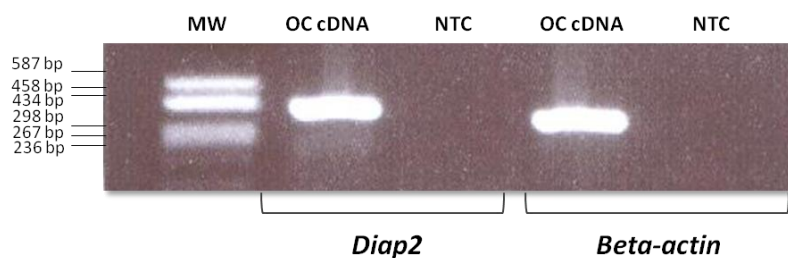
Figure 17. Expression of *DIAPH2* mouse homolog gene (*Diap2*) in the organ of Corti. Agarose gel showing the amplification of *Diap2* by RT-PCR, using primers located in exon 6 and 10, respectively. As positive control, the *beta-actin (Actb)* housekeeping gene was amplified. In both cases, RNA was extracted from the organ of Corti of a P4 mouse. In brief, the tissue was firstly mechanically disrupted in the presence of 700 μl of QIAzol Lysis Reagent (QIAGEN) and then homogenized using a QIAshredder homogenizer (QIAGEN). After a short (2 min) centrifugation at 13000 rpm, a classical phenol-chloroform extraction was performed. The obtained RNA was quantified, reverse transcribed to cDNA with random nonamers and the SuperScriptIII Reverse Transcriptase enzyme (Invitrogen), and finally PCR amplified. OC: organ of Corti; MW: molecular weight marker; NTC: no template control.

4.3.2 *DIAPH2* & the *Diaphanous-related formins* genes

Diaphanous-related formins (DRFs) are cytoskeleton proteins belonging to the family of formins, which act as Rho-GTPase effectors and are involved in the regulation and rearrangement of actin filaments, as well as in the cell-cell adhesion process, and in vescicular trafficking. Nevertheless, the precise role of DRFs in the cochlea is not fully understood.

In humans there are three different DRF proteins (here reported as DIAPH1, DIAPH2 and DIAPH3), encoded by *DIAPH1, DIAPH2* and *DIAPH3* genes*,* respectively. These three genes are characterized by a very similar genomic organization: they contain an almost identical number of exons (28 for both *DIAPH1* and *DIAPH3*, 27 for *DIAPH2*) and mainly differ for intron size. Although the protein sequence identity is limited (DIAPH3 exhibits 51,3% total amino acid identity with DIAPH1, and 57,3% with DIAPH2) [Katoh and Katoh, 2004], the protein structure and their domain organization is conserved. Indeed, these proteins contain 5 principal domains, known as GBD (GTPase binding domain), at the amino terminus, DID (Diaphanous inhibitory domain), FH1 (proline-rich formin homology 1 domain), FH2 (formin homology 2 domain), and DAD (Diaphanous autoregulatory domain), at the carboxyl terminal portion. In details, the FH1 domain is essential for protein-protein interaction (i.e. binding with SH3, WW domain-containing proteins and profilin) whereas the FH2 domain binds actin monomers [Wallar and Alberts, 2003]. Interaction between the N-terminal DID domain and the C-terminal DAD domain locks these formin proteins in an inactive conformation. Transition to an active state occurs when binding of a RhoGTPase to the GBD domain disrupts the autoinhibitory interaction [Young and Copeland, 2010]. Other domains found in DRFs include a dimerization domain (DD) and a coiled coil (CC) region, and some groups believe that the DD and DID together constitute a loosely defined formin homology 3 (FH3) domain [Goh and Ahmed, 2012]. In addition to that, DIAPH1 and DIAPH3 also have an N-terminal basic domain that binds phospholipids and allows them to localize to the plasma membrane [Ramalingam *et al.*, 2010].

The three human *DIAPH* genes are the orthologs of the *Drosophila melanogaster diaphanous* (*dia*) gene, which plays important roles both in the reproductive and in the auditory system. Indeed, *Drosophila* transgenic models overexpressing *dia* show a significant reduction in the auditory function [Schoen *et al.*, 2010]. Despite the evolutionary distance between *Drosophila* and human, the remarkable genetic and molecular similarities in the auditory organ suggest a conservation of the cellular mechanisms between species [Cosetti *et al.*, 2008]. Starting from this evidence, it is possible to speculate that in human these three Diaphanous-related proteins might have preserved a certain degree of functional redundancy, although evolving specific functions.

In particular, the *DIAPH1* gene, located on chromosome 5q31 (locus *DFNA1*), is associated to autosomal dominant high-frequency NSHL (Figure 18). To date, two different deafness-causing mutations have been described in the literature: 1) a splicing variation, which disrupts the donor splice site of intron 17, ultimately leading to a frameshift and to the generation of a premature termination codon [Lynch *et al.*, 1997]; 2) a missense mutation in the FH1 domain, which causes the substitution of the Proline in position 678 with a Serine residue [Baek *et al.*, 2012].

Similarly, a regulatory mutation in the 5'UTR of *DIAPH3*, on chromosome 13q21.2, is responsible for autosomal dominant nonsyndromic auditory neuropathy 1 (AUNA1), a particular type of hearing loss defined by the preservation of cochlear outer hair cell function and abnormal or absent auditory brainstem responses (ABR) (Figure 18).
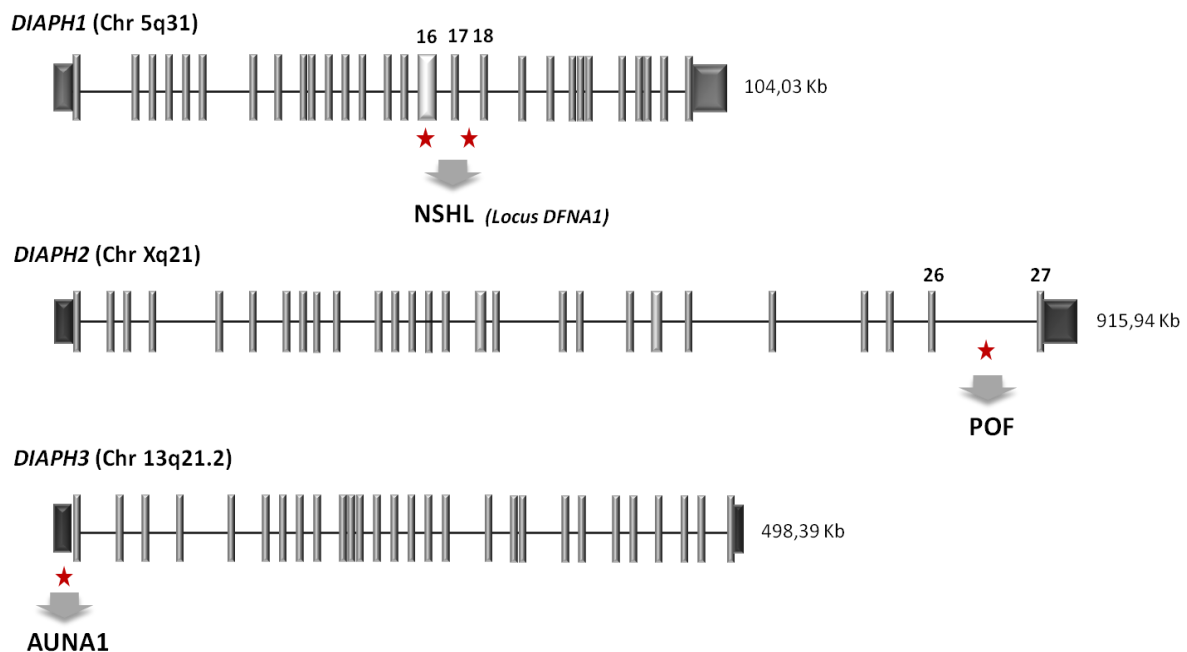
Figure 18. Localization and genomic structure of *DIAPH1* (NM_005219.4), *DIAPH2* (NM_006729) and *DIAPH3* (NM_001042517) genes. A schematic representation of the exon-intron structure of known *DIAPH* genes is shown; introns are represented as black lines while exons as grey bars. The associated diseases and causative mutations are indicated by red stars.

Finally, the *DIAPH2* gene is located on chromosome Xq21 and, differently from the other two *DIAPH* genes, is known to be involved in oogenesis and generation of ovarian follicles. Indeed, to date, one mutation within this gene has been identified as cause of premature ovarian failure (POF2A, OMIM 300108), a secondary hypergonadotropic amenorrhea occurring before the age of 40 and affecting 2–3% of females [Coulam 1982] (Figure 18). The reported mutation is a (X;12) translocation which involves the last intron of *DIAPH2* and is responsible for the

49

transduction of a modified protein or, alternatively, for the degradation of the truncated transcripts. In either case, the subsequent loss of the 3' UTR sequence and all its AT-rich motifs, implicated in the translational control of mRNAs, alters the pattern of events that lead to oocyte and ovary maturation [Bione *et al.*, 1998]. In human, there are two alternatively *DIAPH2* spliced transcripts (*DIAPH2-156* and *DIAPH2-12C*), which are characterized for the presence of an alternative last exon and encode predicted polypeptides of 1101 and 1096 amino acids, respectively. The *DIAPH2* gene is ubiquitously expressed, in rather low amounts, in human adult and fetal tissues [Bione *et al.*, 1998].

Although no experimental evidence clearly supports a *DIAPH2* involvement in inner ear function, its membership to the DRF family of protein and its role in the regulation of actin cytoskeleton makes it a plausible candidate for NSHL.

4.3.3 Effects of the p.Ile290Val mutation on DIAPH2 protein structure

To better unravel the functional implication of the newly identified mutation and to clarify its possible involvement in NSHL pathogenesis, we decided to study the impact of the amino acid substitution on the protein by bioinformatic predictions, using the YASARA program and the structure coordinates obtained from the SWISS-MODEL Repository (accession number: O60879). Interestingly, the amino acid affected by the mutation (Isoleucine 290) is highly conserved among vertebrates and is located within a region involved both in the mechanism of auto-inhibition of the protein (DID domain within the FH3 domain) and in the interaction with RhoGTPases (GBD domain). The substitution with a Valine residue causes a conservative and not severe amino acid change, which does not seem to affect the overall protein structure (Figure 19). However, we could not completely rule out the possibility that the variation might have some functional effects, for example by modifying protein-protein interaction between DIAPH2 and its partners.
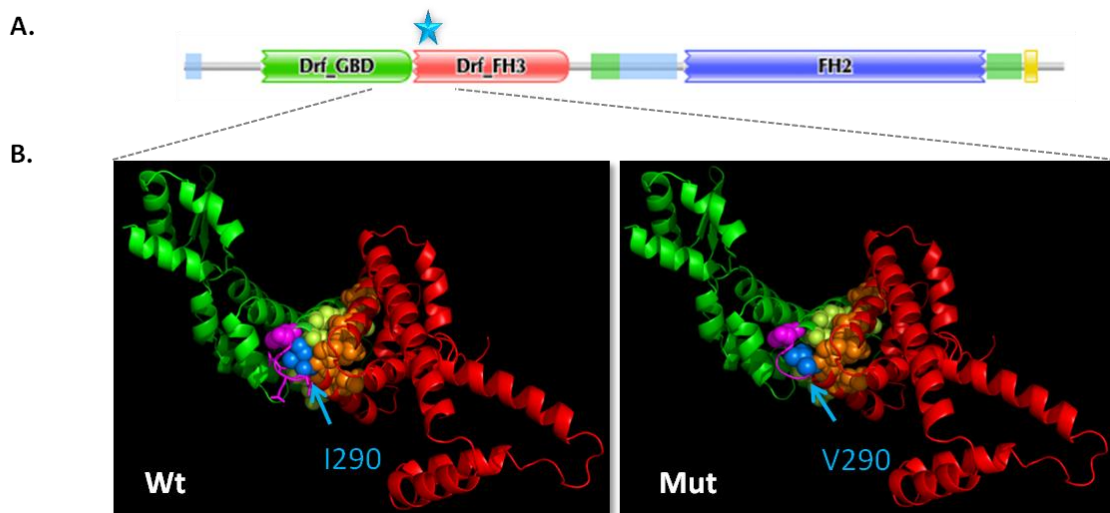
Figure 19. Molecular modeling of the p.Ile290Val substitution. A. Schematic representation of diaphanous-related formins (DRFs) domains. GBD (aa 98-284): GTPase-binding domain; FH3 (aa 289-482): it consists of the dimerization domain (DD) and the inhibitory domain (DID); FH2 (aa 628-1003): binds actin monomers. The position of the newly identified mutation is indicated by a blue star **B.** Tridimensional model of DIAPH2 GBD (green) and FH3 (red) domains. The amino acid involved in the mutation is indicated in blue. No significant alteration was identified in the presence of the mutation, also in terms of steric hindrance with the surrounding amino acids (van der Walls representation). Wt: wild-type protein; Mut: mutant protein

4.3.4   Does the mutation affect the splicing mechanism?

Given these results and considering that the mutation is located at only two base pairs upstream of the 3' end of exon 8, we decided to study its possible impact on *DIAPH2* splicing.

4.3.4.1   *In vivo* analysis

We first evaluated *DIAPH2* transcripts expression on whole blood RNA from all available NSHL3 family members (II2, II3, III3, III4, III5), performing an RT-PCR analysis with exonic primers specifically designed to amplify both exon 8 and flanking exons. Only one amplification product with size compatible with that of the wild-type *DIAPH2* transcript was detected in all analyzed samples, independently from the genotype, suggesting that the mutation does not alter exon 8 splicing (Figure 20).

However, these results could not completely exclude that a splicing alteration might occur in different cell types or tissues, such as the inner ear, or that aberrant (i.e. out-of-frame) splicing events occur but are not detectable because they are efficiently targeted by the nonsense-mediated decay (NMD) mechanism.
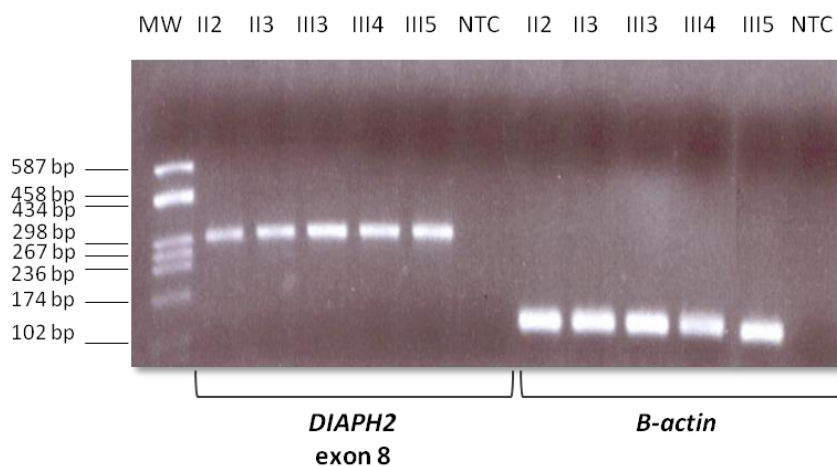
Figure 20. Agarose gel of an RT-PCR demonstrating the existence of one *DIAPH2* transcript. RT-PCRs were performed on cDNA obtained from PBMCs of all available family members, using primers located in exon 6 and 9, respectively (301 bp). As positive control, the *β-actin (ACTB)* housekeeping gene was amplified. MW: molecular weight marker; NTC: no template control.

4.3.4.2 Is *DIAPH2* allelic expression unbalanced in the female carrier of the mutation?

To verify whether transcripts deriving from the mutant *DIAPH2* allele were subjected to NMD, we investigated the possibility of an unbalanced expression between the wild-type and the mutant allele on the II3 heterozygous carrier of the variation. The relative expression of the two alleles was assessed by comparing the sequence of *DIAPH2* mRNA at the mutant position (NM_006729.4:c.868A>G), amplified by RT-PCR from PBMC RNA, with that obtained from genomic DNA. The results showed a co-expression of the two alleles (data not shown).

4.3.4.3 *In vitro* analysis

In order to explore the possibility of a cell- or tissue-specific splicing alteration, we decided to characterize the novel mutation by *in vitro* studies. To this aim, we clone the *DIAPH2* genomic region containing exon 8, with or without the newly identified mutation, and surrounding intronic sequences (~200 bp) into the hybrid alpha-globin-fibronectin EDB minigene plasmid (Figure 21A), kindly provided by Dr. Emanuele Buratti (International Centre for Genetic Engineering and Biotechnology, Trieste, Italy). The obtained plasmids were transiently transfected into 3 different cell lines (HeLa, Hek-293, and HepG2). The subsequent RT-PCR assays (Figure 21B) and the analysis of the generated transcripts (Figure 21C) showed, in HeLa and Hek-293 cells, the presence, besides the wild-type product, of an additional band, whose size

was compatible with the skipping of the entire exon 8, only in samples transfected with the mutant construct. The result was confirmed by Sanger sequencing. A low degree of exon 8 skipping was observed in HepG2. The different ratio of exon 8 skipping in the three cell lines analyzed, suggested a tissue specificity of this splicing event. All the obtained data were confirmed by competitive fluorescent RT-PCR and capillary electrophoresis analyses. Although promising, we were not able to reproduce these results anymore. The lack of reproducibility of these data was due to a technical problem: indeed, the cloning of *DIAPH2* exon 8 into the hybrid construct induced the skipping of the upstream plasmid exon (alpha-globin exon 3/ fibronectin exon 24 hybrid exon), independently from the presence of the mutation. In addition, the length of the hybrid alpha-globin/fibronectin exon was very similar to that of *DIAPH2* exon 8 (8-bp difference, see Figure 21D), thus making difficult the accurate quantification of *DIAPH2* exon8-skipping isoform. We conclude that the selected experimental system was not suitable for our study.

To overcome this problem and to better mimic the physiological condition, we decided to clone, into the pTARGET expression vector, a larger genomic *DIAPH2* sequence containing exons 6 to 9 (~6700 bp). The obtained plasmids were transiently transfected into HeLa, Hek-293 and MDCK cell lines. The subsequent RT-PCR analysis did not show any alternative splicing isoforms in the presence of the mutation (Figure 22). Therefore, we conclude that this nucleotide variation does not affect splicing, at least in the analyzed tissues.
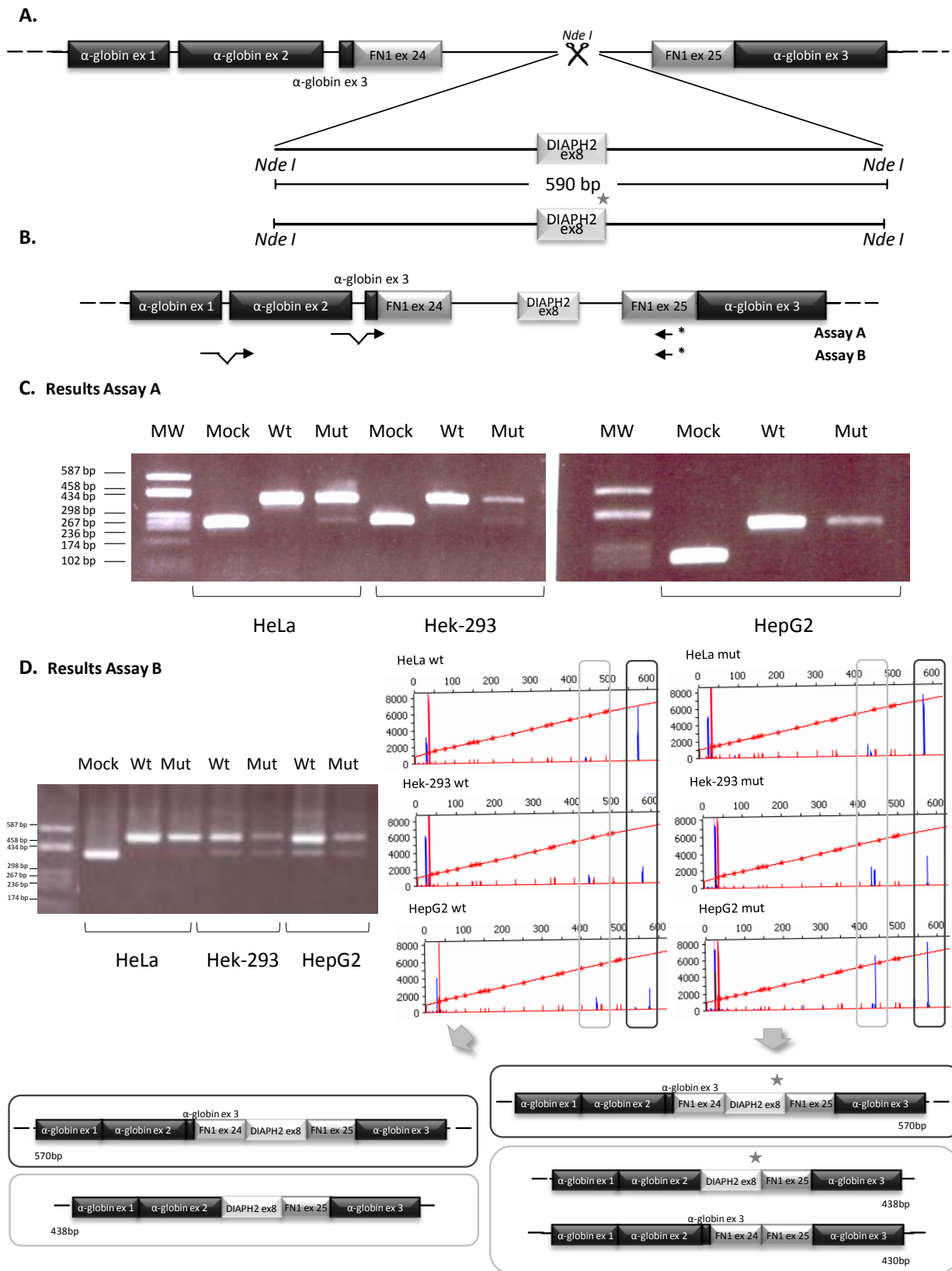
A.

B.

C. Results Assay A

HeLa   Hek-293   HepG2

D. Results Assay B

HeLa   Hek-293   HepG2

HeLa wt   HeLa mut
Hek-293 wt   Hek-293 mut
HepG2 wt   HepG2 mut

Figure 21. *In vitro* characterization of the c.868A>G *DIAPH2* mutation as a splicing variation using an hybrid minigene.A. To generate hybrid minigene constructs, the *DIAPH2* region containing exon 8, with or without the newly identified mutation, was cloned into a modified version of the alpha-globin-fibronectin EDB minigene plasmid (in which the alternatively-spliced EDB exon has been removed to generate a site for the insertion of exons under study) [Baralle *et al.*, 2003]. More in details, a short fragment (590 bp long) was PCR amplified from the genomic DNA of a healthy individual and then subjected to site-directed mutagenesis in order to generate the mutant version. After cloning, we obtained two different vectors: one containing the wild-type nucleotide (A) and the other carrying the mutant one (G). **B.** The position of the primers used for the RT-PCR assays is shown. *: indicates that the primer is FAM-labeled. **C.** In all experiments, 1 or 2 μg of each recombinant plasmid or of the empty vector (used as negative control, mock) were transiently transfected in 3 different human cell lines [HeLa (epithelial cervical carcinoma), HeK-293 (embryonic kidney) and HepG2 (liver carcinoma)]; all cell lines were cultured according to standard procedures. RNA extraction was performed 24 hours after transfections, and RT-PCR analysis was carried out using primers complementary to sequences in the flanking fibronectin exonic region (assay A, point B of this figure). The agarose gel of obtained RT-PCRs is shown. MW: molecular weight marker; Mock: cells transfected with the empty vector; Wt: cells transfected with the vector containing the wild-type nucleotide (A); Mut: cells transfected with the vector containing the mutation (G). The 392-bp-long band corresponds to the inclusion of exon 8 in the mature transcript; the 255-bp-long band is obtained when exon 8 is skipped from the mature transcript. **D.** RT-PCR analysis was carried out using a forward primer complementary to a sequence in flanking α globin and a reverse FAM-labeled primer complementary to the fibronectin exon 25 sequence (assay B, point B of this figure). The agarose gel of obtained RT-PCRs is shown. The 570-bp-long band corresponds to the inclusion of exon 8 in the mature transcript; the ~430-bp-long band corresponds to two different products, which differ of only 8 bp: the first one is the transcript without alpha globin exon3-fibronectin exon 24 hybrid exon (438 bp) while the other is the one without *DIAPH2* exon 8 (430 bp). An aliquot of RT-PCR was run on a 3130XL automatic DNA sequencer (Applied Biosystems), together with an internal size standard (ROX-500 HD, Applied Biosystems). The results were analyzed with Peak Scanner v.1.0 software (Applied Biosystems). A Peak Scanner window displaying fluorescence peaks corresponding to the different molecular species is shown. The red peaks represent the size standard, whereas the blue peaks represent the RT-PCR labeled products. The X-axis corresponds to Peak Scanner data points and the Y-axis represents fluorescence units (FUs). Schematic representations of the hybrid minigene construct and of its exonic composition, indicating the correspondence between the observed fluorescence peaks and the obtained splicing variants.

A.

B.
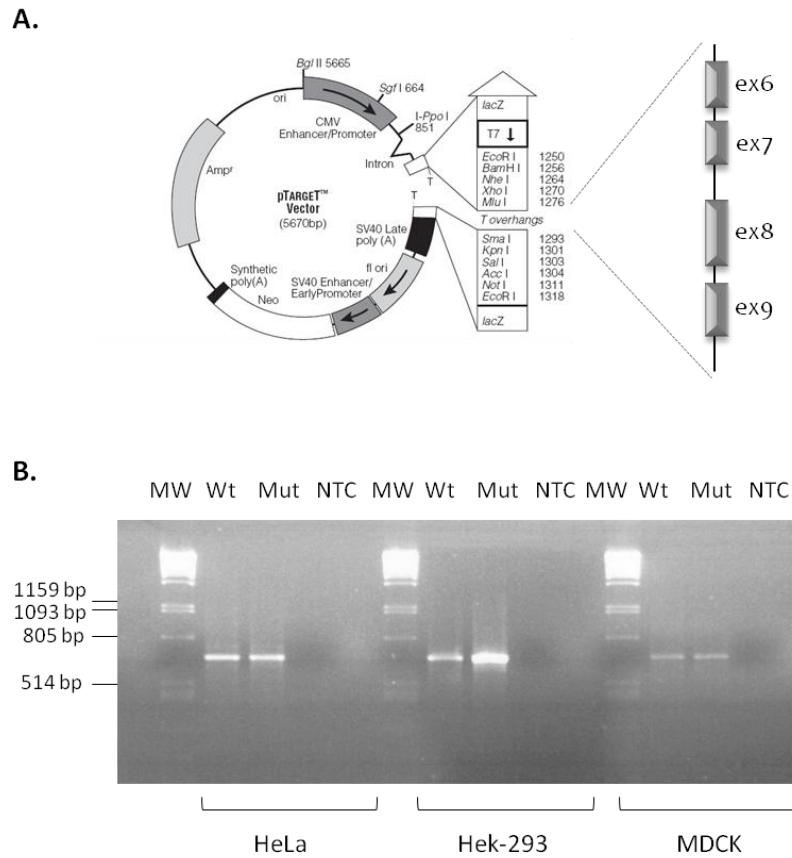
Figure 22. *In vitro* characterization of the effect of the c.868A>G variation on splicing using a *DIAPH2* minigene. A. Schematic representation of the pTARGET mammalian expression vector (Promega) and the portion of the *DIAPH2* gene cloned (exons 6, 7, 8 and 9). More in details, a fragment of about 6700 bp was PCR amplified from the genomic DNA of a healthy individual and then subjected to site-directed mutagenesis in order to obtained the mutant version. **B.** In all experiments, 1 or 2 μg of each recombinant plasmid were transiently transfected in 3 different human cell lines [HeLa, HeK-293, MDCK (Madin-Darby canine kidney cells)]. RNA extraction was performed 24 hours after transfections, and RT-PCR analysis was carried out using a forward primer complementary to the pTARGET T7 promoter sequence and a reverse oligonucleotide designed on *DIAPH2* exon 9. The agarose gel of RT-PCRs is shown. MW: molecular weight marker; Wt: cells transfected with the vector containing the wild-type nucleotide (A); Mut: cells transfected with the vector containing the mutation (G); NTC: no template control. The obtained ~660-bp-long band corresponds to the inclusion of exon 8 in the mature transcript, as confirmed by Sanger sequencing analysis.

4.4 Genetically undiagnosed cases: the dark side of the exomes

For what concerned the NSHL1, 2 and NSHL7 families (Figure 23) the search for pathogenic mutations is still ongoing. So far, 16 candidate variants (13 missense, 1 Indel, 1 nonsense, and 1 missense/splicing) identified by WES data analysis, both in known (2) and novel (11) genes, in the homozygous or compound heterozygous state, have been tested by Sanger sequencing (8 for the NSHL1 family, 4 for the NSHL2 family, and 4 for the NSHL7 family). Unfortunately, none of them was validated: 15 were excluded by the lack of segregation with the disease in the corresponding pedigree, being present in the family members independently from the phenotype; the remaining one was not confirmed by Sanger sequencing, representing a NGS false positive.

The difficulties in the identification of disease-causing variants may be due, at least in part, to the WES strategy and data analysis, described previously (Paragraph 4.1). For instance, for families NSHL1 and 2, the use of two different exome enrichment platforms (Agilent 38M and 50M SureSelect kit) made more challenging, and sometimes unfeasible, the direct comparison between the WES results. Starting from ths evidence, an optimized pipeline for variant filtering/prioritization is being set up to account for differences in the target region analyzed.

An additional problem is the identification of the insertion/deletion variants: in fact, the Indel "calling" algorithms are less reliable than the ones for the detection of single nucleotide substitutions, with many putative variants located within homopolymeric tracts and likely representing false positives [Albers *et al.*, 2011]. Another cause of false positive calls is represented by repeated sequences, which are responsible for an increased frequency of read misalignment. A clear example of this is *TRIOBP (Homo sapiens TRIO and F-actin binding protein)*, a gene already associated to autosomal recessive nonsyndromic hearing loss and identified as mutated in the NSHL2 family. Indeed, the analysis of the WES data for this family pointed out the presence of three single nucleotide variations within exon 5 of the gene; the involved region was characterized by the presence of a short tandem repeat element of 147 bp, thus resulting very difficult to cover without problems of read misalignment. The candidate variations were then rejected and classified as false positive.

A third element which complicates the identification of disease-causing mutations is likely correlated to the specific features of the filtering parameters used in the WES data analysis. Indeed, the higher is the stringency, the smaller is the number of the putative causative genes available for the experimental validation, with many variants representing no plausible candidates; on the contrary, when the stringency is too low, the highlighted mutations to study are too many. In this respect, a correction of the prioritization pipeline with optimal filters thresholds could probably help the research.
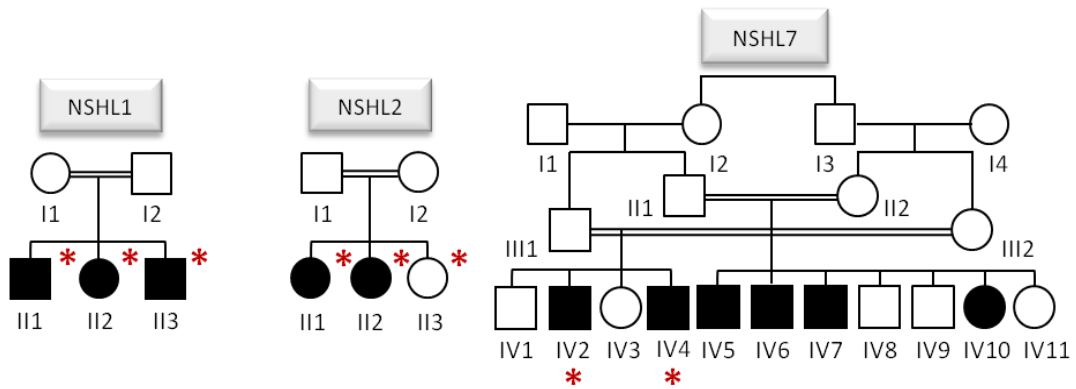
Figure 23. Pedigree of NSHL1, 2 and 7 families. For each family, the probands selected for exome sequencing are indicated by red asteriscs.

Although our standard analysis pipeline failed to detect the causative mutation in three families, a number of alternative strategies are being applied. For instance, as all the three "unresolved" families are consanguineous, we can exploit WES-derived data with the aim to reduce the number of candidate variants to validate by an "homozigosity mapping" approach. This strategy, also called EX-HOM (EXome HOMozygosity) combines in a single step the capacity of exome sequencing to identify all the coding variants present in a genome with the property of homozygosity mapping to limit the search for candidate genes to specific chromosomal regions [Pippucci *et al.*, 2011]. The rarer the autosomal recessive disorder, the higher the probability that affected children of consanguineous parents carry a mutation identical by descent in an autozygous genomic region. With this as background, sequencing two individuals, instead of only one, can improve the variant calling in scarcely captured regions, adding up the reads from the two independent experiments. Also, it is unlikely that genetic mechanism other than autozygosity might be causative for the disorder. For instance, it is virtually impossible that *de novo* mutations have occurred to cause the disease in both siblings [Pippucci *et al.*, 2011].

The application of this analysis to the WES data from NSHL7 family, for which the exome sequencing was performed on two affected siblings born to first-cousins parents (IV2 and IV4, Figure 23), let us to distinguish two region of homozygosity; the first one, of 12 Mb, on chromosome 5 and the second, of only 2 Mb, on chromosome 14. Unfortunately, none of the 5 candidate variants selected for the experimental validation (4 missense and 1 splicing mutations identified in 5 different novel genes*: PCDHB3, P4HA2, IK, PCDHB16, FSTL4*) segregated with the phenotype in the corresponding family, being present in the homozygous state in at least one of the healthy sons.

These results reveal how this strategy is sensitive to the limitations also affecting the WES technology, which, although considered a powerful screening test applicable to a broader range of genetic diseases (i.e. blindess, deafness, mithocondrial and movement disorders), has a limited diagnostic yield; in details, for what concerned hearing loss, the percentage of WES success in diagnosis has been assessed to less than 50 (44%) [Neveling *et al.*, 2013]. This problem could be correlated to different factors:

i. specific causative variants might have been missed due to the low coverage, even adding up the reads from two individuals [Pippucci *et al.*, 2011];

ii. standard exome sequencing data analysis pipeline is usually focused on the identification of point mutations (SNVs and Indels), thus missing certain class of variations, such as large insertions or deletions and large structural modifications;

iii. the analysis of regulatory mutations (i.e. variants within 3'/5' UTR or within splicing regulatory elements -ESE, branch site, ESS etc.-) need more detailed filtering/prioritization pipelines, which include additional algorithms/softwares, specifically designed to predict the variant's effect on gene splicing and/or expression (i.e. NetGene2 Server, Human Splicing Finder);

iv. the success in the identification of causative mutations is widely affected by their localization: variations within targets not completely included in the enrichment platform or within regions captured without uniformity (high GC content) will be missed [Sirmaci *et al.*, 2012];

v. the efficiency of exome sequencing, and of EX-HOM by consequence, largely depends on the capacity to filter out all the variants which are not related to the disorder [Pippucci *et al.*, 2011].

However, although these difficulties, the application of the EX-HOM analysis to the WES data from NSHL1 and NSHL2 families could probably help in the identification of the causative mutations or, at least, in the restriction of the candidate variations to specific chromosomal regions.

In addition to that, an optimization of our in-house developed bioinformatics pipeline of analysis, especially for what concerns the ins/del mutations and large structural variations, might facilitate the attribution of a definitive genetic diagnosis to all these, still unresolved, cases.

4.5 Conclusions and future perspectives

In this study, we report the application of whole exome sequencing technology on selected NSHL families for the identification of novel genetic determinants of deafness and, where possible, molecular characterization of their causal role.

In particular, such a strategy let us to discover the genetic defect underlying autosomal recessive NSHL in two Italian families (NSHL6 and NSHL4), thus increasing the mutational spectrum of the involved genes (*TMPRSS3* for the NSHL6 family, and *PRPS1* for the NSHL4 family).

For what *TMPRSS3* is concerned, two novel missense mutations (NM_024022:c.802T>C, NP_076927:p.Trp268Arg and NM_024022:c.1307G>A, NP_076927:p.Arg436His) were identified within exon 9 and 12, respectively; both variations are located within the serine-protease domain of *TMPRSS3*, affect evolutionarily conserved amino acids, and are predicted to alter the protein structure. Mutations in *TMPRSS3* are a relatively common cause of deafness in several populations. Indeed, to date, numerous genetic studies of *TMPRSS3*-related NSHL have been conducted in Middle Asian and Mediterranean countries, and various mutations predicted to disrupt the proteolytic activity of the enzyme during inner ear development have been identified [Lee *et al.*, 2013]. In details, up to now, more than 25 variations (23 missense/nonsense, 2 splicing, 2 small deletions, 2 small insertions and 1 complex rearrangement) have been reported, according to the Human Gene Mutation Database (HGMD, http://www.hgmd.cf.ac.uk/ac/index.php). However, the actual role of TMPRSS3 protein in the auditory system is currently elusive, stressing the importance of performing more detailed functional studies. To this aim, we are currently planning to use the zebrafish (*Danio rerio*) animal model. This is a tropical freshwater fish which represents an excellent system for the studies of the molecular bases of inner ear development and function and for the *in vivo* characterization of candidate genes pointed out by NGS. It combines rapid and accessible embryogenesis with the availability of genetic and genomic tools for systematic gene discovery and analysis. Zebrafish has a fairly typical and extensively described vertebrate inner ear, whose normal development and function is well conserved [Whitfield, 2002; Whitfield *et al.*, 2002]. A large collection of mutations affecting the ear and the lateral line, a related sensory system, have been isolated, providing different models for human deafness. In this context, the existence of a likely *TMPRSS3* hortologue in zebrafish genome, with about 56% sequence identity with the human protein, makes it a suitable model to better understand the role of this gene in hair-cell development, as well as the impact of the identified mutations on the protein function.

As far as *PRPS1* is concerned, the discover of three novel nucleotide substitutions, all responsible for a reduction in the PRS-I enzymatic activity -which correlates with clinical data-, highlights the

high recurrence of *PRPS1* genetic defects in X-linked deafness, suggesting that it may represent a major locus to be prioritized in genetic screening and to be included in gene panels for routine tests. Moreover, the results of our study stress the importance of the molecular diagnosis, which could help in a fine definition of the genotype-phenotype correlation and guide additional clinical evaluations (*i.e.* neurological examination), sometimes necessary to reveal subtle or subclinical symptoms that might otherwise escape recognition. In this respect, two of the three here analyzed families (NSHL5 and NSHL13 families) were initially diagnosed as having NSHL, but - following the identification of *PRPS1* as the gene responsible for deafness- were further evaluated for additional neurological symptoms, and found to suffer from various degree of peripheral neuropathy.

The description and characterization of novel *PRPS1* mutations has important consequences not only for diagnosis but also for therapeutic and prevention purposes. Indeed, it was reported that a dietary supplementation with S-adenosylmethionine (SAM), which replenishes both ATP and GTP independently from PRPP [de Brouwer *et al.*, 2010], might represent a promising therapy to postpone or slow down the onset and progression of neurological and audiological symptoms in patients with *PRPS1* mutations as well as mildly affected carrier females (see also Paragraph 1.1.2).

In addition to these results, the application of the WES strategy pointed out the presence of a novel missense variant in *DIAPH2*, a gene not directly associated to hearing loss but which could be a likely candidate pathogenic mutation in the corresponding family (NSHL3). However, the actual demonstration that a gene/mutation is a newly identified cause of NSHL is not trivial. First of all, the novel candidate gene should be expressed in tissues relevant for disease pathogenesis, and the identified mutation should segregate with the phenotype in the analyzed family. These points have been addressed for the newly reported p.Ile290Val variant in *DIAPH2*. More difficult is then to demonstrate that the candidate variation is deleterious and affects normal gene function. So far, we were not able to detect any splicing defect possibly associated with the mutation, nor to bioinformatically predict a clear alteration of protein structure. Therefore, additional functional analyses need to be performed. In particular, we are currently exploring the hypothesis that the p.Ile290Val variant affects DIAPH2 localization within the cell or its interaction with cytoskeleton proteins. To this aim, plasmids expressing an HA-tagged human DIAPH2 wild-type [Gasman *et al.*, 2003] or mutant (obtained by site-directed mutagenesis) isoform will be transfected into MDCK cells. Forty-eight (polarized cells) and twenty-four (not polarized cells) hours after transfection, an immunofluorescence assay against the HA-tag will be carried out. Moreover, co-localization studies with cytoskeleton-specific markers (i.e. phalloidin) are now being planned, togheter with immunoprecipitation assays aimed

at a better examination of the variation's consequences on protein-protein interactions and the autoinhibitory process.

In addition to that, an interesting perspective might be the functional characterization of the c.868A>G mutation as a splicing variation specifically in the inner ear. To this purpose, our idea is to use mouse cellular models, extremely useful for auditory hair cells, which are few in number, experimentally inaccessible, and not able to proliferate post-natally or *in vitro*. In this respect, several immortal cell lines from mammalian auditory sensory epithelia have been already established [Rivolta and Holley, 2002], providing important *in vitro* systems for studies of the development, function, and regeneration of hearing. For instance, in addition to the UB/OC-2 cells (Paragraph 3.3.), another important cell line is the OC-k3 one. This is an epithelial clonal line which derives from the organ of Corti of a P14 Immortomouse [Kalinec *et al.*, 1999] and expresses mRNA for markers of either hair cell (myo7a and alpha9AChR) or supporting cells (connexin26 and OCP-2). The value of taking cells at this late stage, some three weeks after terminal mitosis, is that they may express genes that are normally expressed later in development. Under current culture conditions the cells derived at earlier stages do not differentiate sufficiently far to represent this stage. Thus, OC-k3 are valuable models for gene expression and biochemical studies of inner ear specific proteins. Both UB/OC-2 and OC-k3 cell lines have been successfully applied for the functional characterization of deafness-causing mutations [Weiss *et al.*, 2003; Towers *et al.*, 2011].

From another side, as *DIAPH2* is conserved in zebrafish, we are currently trying, in collaboration with Prof. Del Giacco (Department of Biosciences, University of Milan), to knock-out the gene in this animal model by using the CRISPR technology [Hwang *et al.*, 2013]. This cutting-edge approach, very recently developed in zebrafish, allows the selective and stable inactivation of the gene of interest. In details, the CRISPR-Cas system is a bacterial adaptive mechanism that provides resistance to invading viruses and plasmids relying on the activity of the Cas9 endonuclease; Cas9 cuts the DNA recognizing a complex of short RNAs annealed in the vicinity of the target DNA. CRISPR-Cas system has been customized in order to function in zebrafish embryos producing stable knockout (KO) fish. Briefly, it is possible to generate a synthetic single-guide RNA (sgRNA), which contains both the short RNA recognized by the Cas9 endonuclease, and a 20-base stretch that is complementary to a sequence in the target gene, so that the Cas9 endonuclease activity can be precisely directed to the target site. The induced double-stranded break will be followed by error-prone non-homologous end-joining-mediated repair, which results in the inactivation of the target gene. To perform such gene KO, the target gene-specific *in-vitro* synthetized sgRNA will be coinjected, in the zygote, with the synthetic capped mRNA encoding for the Cas9 endonuclease. The microinjected embryos will be grown to

sexual maturity and then crossed to normal fish to obtain the F1 generation. The fish of the F1 will be genotyped; confirmed KO fish will be employed to generate heterozygous and homozygous fish for the knocked-out gene.

Besides functional analyses, we are also planning to extend our cohort of cases with a likely X-linked inheritance pattern of disease, currently composed of only 16 individuals, in order to test the incidence of *DIAPH2* mutations in Italian NSHL patients. In fact, the screening of the entire gene in these subjects, either by targeted resequencing or Sanger sequencing, could increase the chance to identify at least another family carrying a pathogenic mutation within *DIAPH2*, thus supporting the hypothesis of its involvement in hearing loss. However, we cannot exclude the possibility that mutations in this gene might be rare and, therefore, difficult to identify. The genetic analyses could be completed with a *DIAPH2* screening on DNA samples from patients affected by both hearing loss and POF.

At last, a new round of WES on additional NSHL families is now being performed, with the aim to identify novel causative genes/mutations and to finely characterized them.

In conclusion, the results of this project, on one side, significantly increase the knowledge on the genetic defects and molecular mechanisms underlying the pathogenesis of hereditary hearing loss, thus improving the genetic diagnosis and fostering the identification of novel disease-therapeutic targets; on the other, confirm, together with the amount of data already published, the usefulness of NGS for the rapid and cost-effective genetic screening of NSHL.

5. References

- ACMG. Genetics Evaluation Guidelines for the Etiologic Diagnosis of Congenital Hearing Loss. Genetic Evaluation of Congenital Hearing Loss Expert Panel. ACMG statement. Genet Med. 2002 May-Jun;4(3):162-71.

- Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R. Dindel: accurate indel calls from short-read data. Genome Res. 2011 Jun;21(6):961-73.

- Baek JI, Oh SK, Kim DB, Choi SY, Kim UK, Lee KY, Lee SH. Targeted massive parallel sequencing: the effective detection of novel causative mutations associated with hearing loss in small families. Orphanet J Rare Dis. 2012 Sep 3;7:60.

- Baralle M, Baralle D, De Conti L, Mattocks C, Whittaker J, Knezevich A, Ffrench-Constant C, Baralle FE. Identification of a mutation that perturbs NF1 agene splicing using genomic DNA samples and a minigene assay. J Med Genet. 2003 Mar;40(3):220-2.

- Bardien S, Human H, Harris T, Hefke G, Veikondis R, Schaaf HS, van der Merwe L, Greinwald JH, Fagan J, de Jong G. A rapid method for detection of five known mutations associated with aminoglycoside-induced deafness. BMC Med Genet. 2009 Jan 13;10:2.

- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004 Jan 23;116(2):281-97.

- Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009 Jan 23;136(2):215-33.

- Bauer PW, Geers AE, Brenner C, Moog JS, Smith RJ. The effect of GJB2 allele variants on performance after cochlear implantation. Laryngoscope. 2003 Dec;113(12):2135-40.

- Becker MA. Phosphoribosylpyrophosphate synthetase and the regulation of phosphoribosylpyrophosphate production in human cells. Prog Nucleic Acid Res Mol Biol. 2001;69:115-48.

- Bione S, Sala C, Manzini C, Arrigo G, Zuffardi O, Banfi S, Borsani G, Jonveaux P, Philippe C, Zuccotti M, Ballabio A, Toniolo D. A human homologue of the Drosophila melanogaster diaphanous gene is disrupted in a patient with premature ovarian failure: evidence for conserved function in oogenesis and implications for human sterility. Am J Hum Genet. 1998 Mar;62(3):533-41.

- Bonci D, Coppola V, Musumeci M, Addario A, Giuffrida R, Memeo L, D'Urso L, Pagliuca A, Biffoni M, Labbaye C, Bartucci M, Muto G, Peschle C, De Maria R. The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. Nat Med. 2008 Nov;14(11):1271-7.

- Bonné-Tamir B, DeStefano AL, Briggs CE, Adair R, Franklyn B, Weiss S, Korostishevsky M, Frydman M, Baldwin CT, Farrer LA. Linkage of congenital recessive deafness (gene DFNB10) to chromosome 21q22.3. Am J Hum Genet. 1996 Jun;58(6):1254-9.

- Borgström E, Lundin S, Lundeberg J. Large scale library generation for high throughput sequencing. PLoS One. 2011 Apr 27;6(4):e19119.

- Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell. 2003 Apr 4;113(1):25-36.

- Brownstein Z, Bhonker Y, Avraham KB. High-throughput sequencing to decipher the genetic heterogeneity of deafness. Genome Biol. 2012 May 29;13(5):245.

- Brownstein Z, Friedman LM, Shahin H, Oron-Karni V, Kol N, Abu Rayyan A, Parzefall T, Lev D, Shalev S, Frydman M, Davidov B, Shohat M, Rahile M, Lieberman S, Levy-Lahad E, Lee MK, Shomron N, King MC, Walsh T, Kanaan M, Avraham KB. Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in Middle Eastern families. Genome Biol. 2011 Sep 14;12(9):R89.

- Bugge TH, Antalis TM, Wu Q. Type II transmembrane serine proteases. J Biol Chem. 2009 Aug 28;284(35):23177-81.

- Charif M, Abidi O, Boulouiz R, Nahili H, Rouba H, Kandil M, Delprat B, Lenaers G, Barakat A. Molecular analysis of the TMPRSS3 gene in Moroccan families with non-syndromic hearing loss. Biochem Biophys Res Commun. 2012 Mar 23;419(4):643-7.

- Chen CZ, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. Science. 2004 Jan 2;303(5654):83-6.

- Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, Butte AJ, Snyder M. Performance comparison of exome DNA sequencing technologies. Nat Biotechnol. 2011 Sep 25;29(10):908-14.

- Coene KL, Roepman R, Doherty D, Afroze B, Kroes HY, Letteboer SJ, Ngu LH, Budny B, van Wijk E, Gorden NT, Azhimi M, Thauvin-Robinet C, Veltman JA, Boink M, Kleefstra T, Cremers FP, van Bokhoven H, de Brouwer AP. OFD1 is mutated in X-linked Joubert syndrome and interacts with LCA5-encoded lebercilin. *Am J Hum Genet* 2009;85:465-81.

- Cosetti M, Culang D, Kotla S, O'Brien P, Eberl DF, Hannan F. Unique transgenic animal model for hereditary hearing loss. Ann Otol Rhinol Laryngol. 2008 Nov;117(11):827-33.

- Coulam CB. Premature gonadal failure. Fertil Steril. 1982 Dec;38(6):645-55.

- Davis AC. The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. Int J Epidemiol. 1989 Dec;18(4):911-7.

- de Brouwer AP, van Bokhoven H, Nabuurs SB, Arts WF, Christodoulou J, Duley J. PRPS1 mutations: four distinct syndromes and potential treatment. Am J Hum Genet. 2010 Apr 9;86(4):506-18.

- de Brouwer AP, Williams KL, Duley JA, van Kuilenburg AB, Nabuurs SB, Egmont-Petersen M, Lugtenberg D, Zoetekouw L, Banning MJ, Roeffen M, Hamel BC, Weaving L, Ouvrier RA, Donald JA, Wevers RA, Christodoulou J, van Bokhoven H. Arts syndrome is caused by loss-of-function mutations in PRPS1. Am J Hum Genet. 2007 Sep;81(3):507-18.

- de Wit E, Linsen SE, Cuppen E, Berezikov E. Repertoire and evolution of miRNA genes in four divergent nematode species. Genome Res. 2009 Nov;19(11):2064-74.

- Diaz-Horta O, Duman D, Foster J 2nd, Sırmacı A, Gonzalez M, Mahdieh N, Fotouhi N, Bonyadi M, Cengiz FB, Menendez I, Ulloa RH, Edwards YJ, Züchner S, Blanton S, Tekin M. Whole-exome sequencing efficiently detects rare mutations in autosomal recessive nonsyndromic hearing loss. PLoS One. 2012;7(11):e50628.

- Dostie J, Mourelatos Z, Yang M, Sharma A, Dreyfuss G. Numerous microRNPs in neuronal cells containing novel microRNAs. RNA. 2003 May;9(5):631-2.

- Dror AA, Avraham KB. Hearing impairment: a panoply of genes and functions. Neuron. 2010 Oct 21;68(2):293-308.

▪ Duman D, Tekin M. Autosomal recessive nonsyndromic deafness genes: a review. Front Biosci (Landmark Ed). 2012 Jun 1;17:2213-36.

▪ Eriksen TA, Kadziola A, Bentsen AK, Harlow KW, Larsen S. Structural basis for the function of Bacillus subtilis phosphoribosyl-pyrophosphate synthetase. Nat Struct Biol. 2000 Apr;7(4):303-8.

▪ Fasquelle L, Scott HS, Lenoir M, Wang J, Rebillard G, Gaboyard S, Venteo S, François F, Mausset-Bonnefont AL, Antonarakis SE, Neidhart E, Chabbert C, Puel JL, Guipponi M, Delprat B. Tmprss3, a transmembrane serine protease deficient in human DFNB8/10 deafness, is critical for cochlear hair cell survival at the onset of hearing. J Biol Chem. 2011 May 13;286(19):17383-97.

▪ Fischel-Ghodsian N. Genetic factors in aminoglycoside toxicity. Ann N Y Acad Sci. 1999 Nov 28;884:99-109.

▪ Fraser GR. The causes of profound deafness in childhood. In: Sensorinerual hearing loss. Ciba Found Symp. 1970:5-40.

▪ Friedman LM, Avraham KB. MicroRNAs and epigenetic regulation in the mammalian inner ear: implications for deafness. Mamm Genome. 2009 Sep-Oct;20(9-10):581-603.

▪ Friedman LM, Dror AA, Avraham KB. Mouse models to study inner ear development and hereditary hearing loss. Int J Dev Biol. 2007;51(6-7):609-31.

▪ Friedman TB, Griffith AJ. Human nonsyndromic sensorineural deafness. Annu Rev Genomics Hum Genet. 2003;4:341-402.

▪ Gasman S, Kalaidzidis Y, Zerial M. RhoD regulates endosome dynamics through Diaphanous-related Formin and Src tyrosine kinase. Nat Cell Biol. 2003 Mar;5(3):195-204.

▪ Goh WI, Ahmed S. mDia1-3 in mammalian filopodia. Commun Integr Biol. 2012 Jul 1;5(4):340-4.

▪ Gresty M, Brookes G. Deafness and vertigo. Curr Opin Neurol. 1997 Feb;10(1):36-42.

▪ Guipponi M, Vuagniaux G, Wattenhofer M, Shibuya K, Vazquez M, Dougherty L, Scamuffa N, Guida E, Okui M, Rossier C, Hancock M, Buchet K, Reymond A, Hummler E, Marzella PL, Kudoh J, Shimizu N, Scott HS, Antonarakis SE, Rossier BC. The transmembrane serine protease (TMPRSS3) mutated in deafness DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro. Hum Mol Genet. 2002 Nov 1;11(23):2829-36.

▪ Harfe BD. MicroRNAs in vertebrate development. Curr Opin Genet Dev. 2005 Aug;15(4):410-5.

▪ Hertzano R, Dror AA, Montcouquiol M, Ahmed ZM, Ellsworth B, Camper S, Friedman TB, Kelley MW, Avraham KB. Lhx3, a LIM domain transcription factor, is regulated by Pou4f3 in the auditory but not in the vestibular system. Eur J Neurosci. 2007 Feb;25(4):999-1005.

▪ Holley MC, Lawlor PW. Production of conditionally immortalised cell lines from a transgenic mouse. Audiol Neurootol. 1997 Jan-Apr;2(1-2):25-35.

▪ Hone SW, Smith RJ. Medical evaluation of pediatric hearing loss. Laboratory, radiographic, and genetic testing. Otolaryngol Clin North Am. 2002 Aug;35(4):751-64.

▪ Huebner AK, Gandia M, Frommolt P, Maak A, Wicklein EM, Thiele H, Altmüller J, Wagner F, Viñuela A, Aguirre LA, Moreno F, Maier H, Rau I, Giesselmann S, Nürnberg G, Gal A, Nürnberg P, Hübner

CA, del Castillo I, Kurth I. Nonsense mutations in SMPX, encoding a protein responsive to physical force, result in X-chromosomal hearing loss. Am J Hum Genet. 2011 May 13;88(5):621-7.

- Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet. 2011 Feb;12(2):99-110.

- Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol. 2013 Mar;31(3):227-9.

- Jat PS, Noble MD, Ataliotis P, Tanaka Y, Yannoutsos N, Larsen L, Kioussis D. Direct derivation of conditionally immortal cell lines from an H-2Kb-tsA58 transgenic mouse. Proc Natl Acad Sci U S A. 1991 Jun 15;88(12):5096-100.

- Kalinec F, Kalinec G, Boukhvalova M, Kachar B. Establishment and characterization of conditionally immortalized organ of corti cell lines. Cell Biol Int. 1999;23(3):175-84.

- Kapsimali M, Kloosterman WP, de Bruijn E, Rosa F, Plasterk RH, Wilson SW. MicroRNAs show a wide diversity of expression profiles in the developing and mature central nervous system. Genome Biol. 2007;8(8):R173.

- Katoh M, Katoh M. Identification and characterization of human DIAPH3 gene in silico. Int J Mol Med. 2004 Mar;13(3):473-8.

- Kikuchi T, Adams JC, Miyabe Y, So E, Kobayashi T. Potassium ion recycling pathway via gap junction systems in the mammalian cochlea and its interruption in hereditary nonsyndromic deafness. Med Electron Microsc. 2000;33(2):51-6.

- Kim HJ, Sohn KM, Shy ME, Krajewski KM, Hwang M, Park JH, Jang SY, Won HH, Choi BO, Hong SH, Kim BJ, Suh YL, Ki CS, Lee SY, Kim SH, Kim JW. Mutations in PRPS1, which encodes the phosphoribosyl pyrophosphate synthetase enzyme critical for nucleotide biosynthesis, cause hereditary peripheral neuropathy with hearing loss and optic neuropathy (cmtx5). Am J Hum Genet. 2007 Sep;81(3):552-8.

- Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet. 2010 Sep;11(9):597-610.

- Krützfeldt J, Stoffel M. MicroRNAs: a new class of regulatory genes affecting metabolism. Cell Metab. 2006 Jul;4(1):9-12.

- Kuhn S, Johnson SL, Furness DN, Chen J, Ingham N, Hilton JM, Steffes G, Lewis MA, Zampini V, Hackney CM, Masetto S, Holley MC, Steel KP, Marcotti W. miR-96 regulates the progression of differentiation in mammalian cochlear inner and outer hair cells. Proc Natl Acad Sci U S A. 2011 Feb 8;108(6):2355-60.

- Lee J, Baek JI, Choi JY, Kim UK, Lee SH, Lee KY. Genetic analysis of TMPRSS3 gene in the Korean population with autosomal recessive nonsyndromic hearing loss. Gene. 2013 Dec 15;532(2):276-80.

- Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993 Dec 3;75(5):843-54.

- Lewis MA, Quint E, Glazier AM, Fuchs H, De Angelis MH, Langford C, van Dongen S, Abreu-Goodger C, Piipari M, Redshaw N, Dalmay T, Moreno-Pelayo MA, Enright AJ, Steel KP. An ENU-

induced mutation of miR-96 associated with progressive hearing loss in mice. Nat Genet. 2009 May;41(5):614-8.

- Li H, Kloosterman W, Fekete DM. MicroRNA-183 family members regulate sensorineural fates in the inner ear. J Neurosci. 2010 Mar 3;30(9):3254-63.

- Li S, Lu Y, Peng B, Ding J. Crystal structure of human phosphoribosylpyrophosphate synthetase 1 reveals a novel allosteric site. Biochem J. 2007 Jan 1;401(1):39-47.

- Lin X, Tang W, Ahmad S, Lu J, Colby CC, Zhu J, Yu Q. Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. Hear Res. 2012 Jun;288(1-2):67-76.

- Liu X, Han D, Li J, Han B, Ouyang X, Cheng J, Li X, Jin Z, Wang Y,Bitner-Glindzicz M, Kong X, Xu H, Kantardzhieva A, Eavey RD, Seidman CE, Seidman JG, Du LL, Chen ZY, Dai P, Teng M, Yan D, Yuan H. Loss-of-function mutations inthe PRPS1 gene cause a type of nonsyndromic X-linked sensorineural deafness,DFN2. Am J Hum Genet. 2010 Jan;86(1):65-71.

- Liu XZ, Xie D, Yuan HJ, de Brouwer AP, Christodoulou J, Yan D. Hearing loss and PRPS1 mutations: Wide spectrum of phenotypes and potential therapy. Int J Audiol. 2013 Jan;52(1):23-8.

- Lynch ED, Lee MK, Morrow JE, Welcsh PL, León PE, King MC. Nonsyndromic deafness DFNA1 associated with mutation of a human homolog of the Drosophila gene diaphanous. Science. 1997 Nov 14;278(5341):1315-8.

- Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? J Med Genet. 2011 Sep;48(9):580-9.

- Marco A, Hui JH, Ronshaugen M, Griffiths-Jones S. Functional shifts in insect microRNA evolution. Genome Biol Evol. 2010;2:686-96.

- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2006 May 4;441(7089):120.

- Marres HA: Congenital abnormalities of the inner ear. Diseases of the Ear. Edited by Ludman H, Wright T. Bath, Arnold & Oxford University Press, 1998, pp 288–296

- Mencía A, Modamio-Høybjør S, Redshaw N, Morín M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T, Moreno F, Moreno-Pelayo MA. Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet. 2009 May;41(5):609-13.

- Molina L, Fasquelle L, Nouvian R, Salvetat N, Scott HS, Guipponi M, Molina F, Puel JL, Delprat B. Tmprss3 loss of function impairs cochlear inner hair cell Kcnma1 channel membrane expression. Hum Mol Genet. 2013 Apr 1;22(7):1289-99.

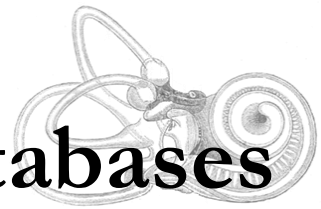- Morton NE. Genetic epidemiology of hearing impairment. Ann N Y Acad Sci. 1991;630:16-31.

- Nance WE. The genetics of deafness. Ment Retard Dev Disabil Res Rev. 2003;9(2):109-19.

- Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg EJ, Mensenkamp AR, Rodenburg RJ, Yntema HG, Spruijt L, Vermeer S, Rinne T, van Gassen KL, Bodmer D, Lugtenberg D, de Reuver R, Buijsman W, Derks RC, Wieskamp N, van den Heuvel B, Ligtenberg MJ, Kremer H, Koolen DA, van de Warrenburg BP, Cremers FP, Marcelis CL, Smeitink JA, Wortmann SB, van Zelst-Stams WA, Veltman JA, Brunner HG, Scheffer H, Nelen MR. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. Hum Mutat. 2013 Dec;34(12):1721-6.

- Ng SB, Nickerson DA, Bamshad MJ, Shendure J. Massively parallel sequencing and rare disease. Hum Mol Genet. 2010 Oct 15;19(R2):R119-24.

- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009 Sep 10;461(7261):272-6.

- Nickel R, Forge A. Gap junctions and connexins in the inner ear: their roles in homeostasis and deafness. Curr Opin Otolaryngol Head Neck Surg. 2008 Oct;16(5):452-7.

- Okamura K, Phillips MD, Tyler DM, Duan H, Chou YT, Lai EC. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. Nat Struct Mol Biol. 2008 Apr;15(4):354-63.

- Pauley S, Kopecky B, Beisel K, Soukup G, Fritzsch B. Stem cells and molecular strategies to restore hearing. Panminerva Med. 2008 Mar;50(1):41-53.

- Petit C, Levilliers J, Marlin S, Hardelin J-P: Hereditary hearing loss. The Metabolic and Molecular Bases of Inherited Disease, ed 8. Edited by Scriver CR, Beaudet AL, Sly WS, Valle D. New York, McGraw-Hill, 2001, 4:6281–6328.

- Petit C. Genes responsible for human hereditary deafness: symphony of a thousand. Nat Genet. 1996 Dec;14(4):385-91.

- Pierce ML, Weston MD, Fritzsch B, Gabel HW, Ruvkun G, Soukup GA. MicroRNA-183 family conservation and ciliated neurosensory organ expression. Evol Dev. 2008 Jan-Feb;10(1):106-13.

- Pippucci T, Benelli M, Magi A, Martelli PL, Magini P, Torricelli F, Casadio R, Seri M, Romeo G. EX-HOM (EXome HOMozygosity): a proof of principle. Hum Hered. 2011;72(1):45-53.

- Provenzano MJ, Domann FE. A role for epigenetics in hearing: Establishment and maintenance of auditory specific gene expression patterns. Hear Res. 2007 Nov;233(1-2):1-13.

- Ramalingam N, Zhao H, Breitsprecher D, Lappalainen P, Faix J, Schleicher M. Phospholipids regulate localization and activity of mDia1 formin. Eur J Cell Biol. 2010 Oct;89(10):723-32.

- Raphael Y, Altschuler RA. Structure and innervation of the cochlea. Brain Res Bull. 2003 Jun 15;60(5-6):397-422.

- Raphael Y, Kim YH, Osumi Y, Izumikawa M. Non-sensory cells in the deafened organ of Corti: approaches for repair. Int J Dev Biol. 2007;51(6-7):649-54.

- Raviv D, Dror AA, Avraham KB. Hearing loss: a common disorder caused by many rare alleles. Ann N Y Acad Sci. 2010 Dec;1214:168-79.

- Rawlings ND, Barrett AJ, Bateman A. MEROPS: the peptidase database. Nucleic Acids Res. 2010 Jan;38(Database issue):D227-33.

- Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000 Feb 24;403(6772):901-6.

- Richardson GP, Lukashkin AN, Russell IJ. The tectorial membrane: one slice of a complex cochlear sandwich. Curr Opin Otolaryngol Head Neck Surg. 2008 Oct;16(5):458-64.

- Rivolta MN, Grix N, Lawlor P, Ashmore JF, Jagger DJ, Holley MC. Auditory hair cell precursors immortalized from the mammalian inner ear. Proc Biol Sci. 1998 Sep 7;265(1406):1595-603.

- Rivolta MN, Holley MC. Cell lines in inner ear research. J Neurobiol. 2002 Nov 5;53(2):306-18.

- Ro S, Park C, Young D, Sanders KM, Yan W. Tissue-dependent paired expression of miRNAs. Nucleic Acids Res. 2007;35(17):5944-53.

- Sarrias MR, Grønlund J, Padilla O, Madsen J, Holmskov U, Lozano F. The Scavenger Receptor Cysteine-Rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. Crit Rev Immunol. 2004;24(1):1-37.

- Schoen CJ, Emery SB, Thorne MC, Ammana HR, Sliwerska E, Arnett J, Hortsch M, Hannan F, Burmeister M, Lesperance MM. Increased activity of Diaphanous homolog 3 (DIAPH3)/diaphanous causes hearing defects in humans with auditory neuropathy and in Drosophila. Proc Natl Acad Sci U S A. 2010 Jul 27;107(30):13396-401.

- Schraders M, Haas SA, Weegerink NJ, Oostrik J, Hu H, Hoefsloot LH, Kannan S, Huygen PL, Pennings RJ, Admiraal RJ, Kalscheuer VM, Kunst HP, Kremer H. Next-generation sequencing identifies mutations of SMPX, which encodes the small muscle protein, X-linked, as a cause of progressive hearing impairment. Am J Hum Genet. 2011 May 13;88(5):628-34.

- Schrijver I. Hereditary non-syndromic sensorineural hearing loss: transforming silence to sound. J Mol Diagn. 2004 Nov;6(4):275-84.

- Scott HS, Kudoh J, Wattenhofer M, Shibuya K, Berry A, Chrast R, Guipponi M, Wang J, Kawasaki K, Asakawa S, Minoshima S, Younus F, Mehdi SQ, Radhakrishna U, Papasavvas MP, Gehrig C, Rossier C, Korostishevsky M, Gal A, Shimizu N, Bonne-Tamir B, Antonarakis SE. Insertion of beta-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness. Nat Genet. 2001 Jan;27(1):59-63.

- Shearer AE, DeLuca AP, Hildebrand MS, Taylor KR, Gurrola J 2nd, Scherer S, Scheetz TE, Smith RJ. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. Proc Natl Acad Sci U S A. 2010 Dec 7;107(49):21104-9.

- Sirmaci A, Edwards YJ, Akay H, Tekin M. Challenges in whole exome sequencing: an example from hereditary deafness. PLoS One. 2012;7(2):e32000.

- Sonkoly E, Wei T, Janson PC, Sääf A, Lundeberg L, Tengvall-Linder M, Norstedt G, Alenius H, Homey B, Scheynius A, Ståhle M, Pivarcsi A. MicroRNAs: novel regulators involved in the pathogenesis of psoriasis? PLoS One. 2007 Jul 11;2(7):e610.
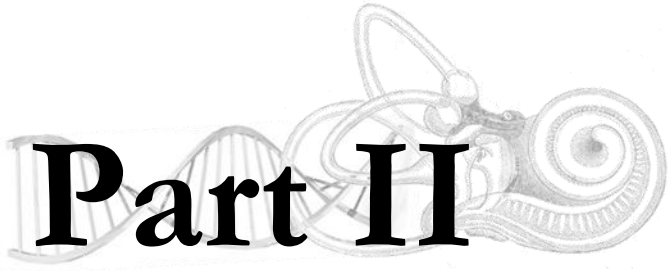
- Soukup GA, Fritzsch B, Pierce ML, Weston MD, Jahan I, McManus MT, Harfe BD. Residual microRNA expression dictates the extent of inner ear development in conditional Dicer knockout mice. Dev Biol. 2009 Apr 15;328(2):328-41.

- Sperling O, Eilam G, Sara-Persky-Brosh, De Vries A. Accelerated erythrocyte 5-phosphoribosyl-1-pyrophosphate synthesis. A familial abnormality associated with excessive uric acid production and gout. Biochem Med. 1972 Aug;6(4):310-6.

- Südhof TC, Goldstein JL, Brown MS, Russell DW. The LDL receptor gene: a mosaic of exons shared with different proteins. Science. 1985 May 17;228(4701):815-22.

- Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. Hum Mol Genet. 2010 Oct 15;19(R2):R145-51.

- Torres RJ, Mateos FA, Puig JG, Becker MA. A simplified method for the determination of phosphoribosylpyrophosphate synthetase activity in hemolysates. *Clin Chim Acta* 1994;224:55-63.

- Towers ER, Kelly JJ, Sud R, Gale JE, Dawson SJ. Caprin-1 is a target of the deafness gene Pou4f3 and is recruited to stress granules in cochlear hair cells in response to ototoxic damage. J Cell Sci. 2011 Apr 1;124(Pt 7):1145-55.

- Van Camp G, Willems PJ, Smith RJ. Nonsyndromic hearing impairment: unparalleled heterogeneity. Am J Hum Genet. 1997 Apr;60(4):758-64.

- van Driel IR, Goldstein JL, Südhof TC, Brown MS. First cysteine-rich repeat in ligand-binding domain of low density lipoprotein receptor binds Ca2+ and monoclonal antibodies, but not lipoproteins. J Biol Chem. 1987 Dec 25;262(36):17443-9.

- Veenstra DL, Harris J, Gibson RL, Rosenfeld M, Burke W, Watts C. Pharmacogenomic testing to prevent aminoglycoside-induced hearing loss in cystic fibrosis patients: potential impact on clinical, patient, and economic outcomes. Genet Med. 2007 Oct;9(10):695-704.

- Veske A, Oehlmann R, Younus F, Mohyuddin A, Müller-Myhsok B, Mehdi SQ, Gal A. Autosomal recessive non-syndromic deafness locus (DFNB8) maps on chromosome 21q22 in a large consanguineous kindred from Pakistan. Hum Mol Genet. 1996 Jan;5(1):165-8.

- Visone R, Croce CM. MiRNAs and cancer. Am J Pathol. 2009 Apr;174(4):1131-8.

- Wallar BJ, Alberts AS. The formins: active scaffolds that remodel the cytoskeleton. Trends Cell Biol. 2003 Aug;13(8):435-46.

- Walsh T, Shahin H, Elkan-Miller T, Lee MK, Thornton AM, Roeb W, Abu Rayyan A, Loulus S, Avraham KB, King MC, Kanaan M. Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. Am J Hum Genet. 2010 Jul 9;87(1):90-4.

- Weegerink NJ, Schraders M, Oostrik J, Huygen PL, Strom TM, Granneman S, Pennings RJ, Venselaar H, Hoefsloot LH, Elting M, Cremers CW, Admiraal RJ, Kremer H, Kunst HP. Genotype-phenotype correlation in DFNB8/10 families with TMPRSS3 mutations. J Assoc Res Otolaryngol. 2011 Dec;12(6):753-66.

- Weiss S, Gottfried I, Mayrose I, Khare SL, Xiang M, Dawson SJ, Avraham KB. The DFNA15 deafness mutation affects POU4F3 protein stability, localization, and transcriptional activity. Mol Cell Biol. 2003 Nov;23(22):7957-64.

- Whitfield TT, Riley BB, Chiang MY, Phillips B. Development of the zebrafish inner ear. Dev Dyn. 2002 Apr;223(4):427-58.

- Whitfield TT. Zebrafish as a model for hearing and deafness. J Neurobiol. 2002 Nov 5;53(2):157-71.

- Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol. 2009 Mar;11(3):228-34.

- Xu P, Vernooy SY, Guo M, Hay BA. The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. Curr Biol. 2003 Apr 29;13(9):790-5.

- Yang JS, Phillips MD, Betel D, Mu P, Ventura A, Siepel AC, Chen KC, Lai EC. Widespread regulatory activity of vertebrate microRNA* species. RNA. 2011 Feb;17(2):312-26.

- Yorgason JG, Fayad JN, Kalinec F. Understanding drug ototoxicity: molecular insights for prevention and clinical management. Expert Opin Drug Saf. 2006 May;5(3):383-99.

- Young KG, Copeland JW. Formins in cell signaling. Biochim Biophys Acta. 2010 Feb;1803(2):183-90.

- Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. J Genet Genomics. 2011 Mar 20;38(3):95-109.

6. Websites &
links to databases

- 1000 Genome variants databases: http://www.1000genomes.org

- ClustalW2 program: http://www.ebi.ac.uk/Tools/msa/clustalw2

- dbSNP database: http://www.ncbi.nlm.nih.gov/SNP/)

- GATK software: http://www.broadinstitute.org/gatk/

- Hereditary Hearing loss Homepage: www.hereditaryhearingloss.org

- HUGO Gene Nomenclature Committee: http://www.genenames.org/

- Human Gene Mutation Database (HGMD): http://www.hgmd.cf.ac.uk/ac/index.php

- mfold program: http://mfold.rna.albany.edu/?q=mfold

- miRBase database: http://www.mirbase.org/

- NHLBI GO Exome Sequencing Project: http://evs.gs.washington.edu/EVS/

- Protein Data Bank (PDB): http://www.pdb.org

- PyMOL program: http://www.pymol.org

- SAMtools software: http://samtools.sourceforge.net/

- SOAPsnp program: http://soap.genomics.org.cn/soapsnp.html

- Swiss-PDB Viewer program: http://spdbv.vital-it.ch/

- Yet Another Scientific Artificial Reality Application (YASARA) program: http://www.yasara.org/

Part II

Content

Soldà G, **Robusto M**, Primignani P, Castorina P, Benzoni E, Cesarani A, Ambrosetti U, Asselta R, Duga S. *A novel mutation within the MIR96 gene causes non-syndromic inherited hearing loss in an Italian family by altering pre-miRNA processing.* Hum Mol Genet. 2012 21:577-85. **[Published article]**

A novel mutation within the *MIR96* gene causes non-syndromic inherited hearing loss in an Italian family by altering pre-miRNA processing

Giulia Soldà[1],∗, Michela Robusto[1], Paola Primignani[3], Pierangela Castorina[4], Elena Benzoni[3], Antonio Cesarani[2,4], Umberto Ambrosetti[2,4], Rosanna Asselta[1] and Stefano Duga[1]

[1]Dipartimento di Biologia e Genetica per le Scienze Mediche and [2]Dipartimento di Scienze Chirurgiche Specialistiche, Università degli Studi di Milano, Milan, Italy, [3]Medical Genetics Laboratory and [4]UO Audiology, Fondazione IRCCS Cà Granda, Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milan, Italy

The miR-96, miR-182 and miR-183 microRNA (miRNA) family is essential for differentiation and function of the vertebrate inner ear. Recently, point mutations within the seed region of miR-96 were reported in two Spanish families with autosomal dominant non-syndromic sensorineural hearing loss (NSHL) and in a mouse model of NSHL. We screened 882 NSHL patients and 836 normal-hearing Italian controls and identified one putative novel mutation within the miR-96 gene in a family with autosomal dominant NSHL. Although located outside the mature miR-96 sequence, the detected variant replaces a highly conserved nucleotide within the companion miR-96∗, and is predicted to reduce the stability of the pre-miRNA hairpin. To evaluate the effect of the detected mutation on miR-96/mir-96∗ biogenesis, we investigated the maturation of miR-96 by transient expression in mammalian cells, followed by real-time reverse-transcription polymerase chain reaction (PCR). We found that both miR-96 and miR-96∗ levels were significantly reduced in the mutant, whereas the precursor levels were unaffected. Moreover, miR-96 and miR-96∗ expression levels could be restored by a compensatory mutation that reconstitutes the secondary structure of the pre-miR-96 hairpin, demonstrating that the mutation hinders precursor processing, probably interfering with Dicer cleavage. Finally, even though the mature miR-96 sequence is not altered, we demonstrated that the identified mutation significantly impacts on miR-96 regulation of selected targets. In conclusion, we provide further evidence of the involvement of miR-96 mutations in human deafness and demonstrate that a quantitative defect of this miRNA may contribute to NSHL.

INTRODUCTION

MicroRNAs (miRNAs) are ∼21-nucleotide (nt)-long, single-stranded noncoding RNAs that mainly function as post-transcriptional regulators of gene expression. Once assembled into an RNA-induced silencing complex, each miRNA might inhibit the expression of hundreds of target messenger RNAs (mRNAs), by inducing translational repression and/or mRNA degradation (1). Animal miRNAs recognize partially complementary binding sites, which are generally located in the 3′ untranslated region (3′UTR) of target mRNAs. In particular, complementarity to the miRNA seed region, corresponding to nts 2–8 at the 5′ of the mature miRNA, is a major determinant in target recognition and is sufficient to trigger silencing (2).

MiRNAs are initially transcribed from endogenous genes as long primary transcripts (pri-miRNAs), which contain extended hairpin structures. The pri-miRNAs are then processed by the Microprocessor complex into a hairpin precursor (pre-miRNA), which is exported to the cytoplasm and cleaved by the RNaseIII Dicer to generate a mature miRNA duplex (3). Usually, one strand of the duplex is preferentially selected for entry into the silencing complex to regulate gene expression, whereas the other strand, known as the passenger

∗To whom correspondence should be address at: Department of Biology and Genetics for Medical Sciences, Via Viotti 3/5, 20133, Milan, Italy. Tel: +39 0250315852; Fax: +39 0250315864; Email: giulia.solda@unimi.it

strand or miRNA*, has typically been assumed to be degraded. However, recent evidence demonstrated that miRNA* species are often present at physiologically relevant levels, can associate with the silencing protein Argonaute and can inhibit target mRNAs in both cultured cells and transgenic animals (4–8).

The miR-183 family is composed of three miRNAs (miR-183, miR-96 and miR-182), which are coordinately expressed from a single genetic *locus* in vertebrates. Homologous miRNAs (miR-228 and mir-263b) are also present in invertebrates (9). Importantly, this highly conserved family of miRNAs shows expression in ciliated neurosensory organs across *phyla* and has recently been demonstrated to contribute specifically to the differentiation and function of the mechanosensory hair cells in the vertebrate inner ear (10,11). For instance, in zebrafish, the miR-183 family is predominantly expressed in the hair cells of the inner ear and of the lateral line, as well as in the olfactory and retinal sensory cells (12,13). Overexpression of miR-96 or miR-182, but not of miR-183, was shown to induce duplicated otocysts, ectopic or expanded sensory patches, and extra hair cells. Conversely, knockdown of each of the three miRNAs led to a reduction in the number of hair cells in the inner ear and caused defects in semicircular canals, as well as the presence of abnormal neuromasts in the lateral line (11).

In the human genome, the miR-183 family is clustered in a 4.5 kb region on chromosome 7q32, within a *locus* that has been linked to autosomal dominant non-syndromic hearing loss (NSHL) (DFNA50, OMIM #613074). In 2009, two mutations in the seed region of miR-96 were detected in two Spanish families affected by autosomal dominant progressive NSHL. Both mutations (+13G>A and +14C>A) affect nts that are fully conserved among vertebrates (from fish to humans) and segregated with hearing loss in the affected families. The impact of these mutations on miR-96 processing and target recognition was analyzed in HeLa and NIH-3T3 cells, showing that both mutations result in reduced levels of mature miRNA and hinder its gene-silencing capacity (14). Further evidence on the significance of miR-96 expression and function in the pathogenesis of hearing loss has been provided by genetic studies in the mouse, where a single nt substitution in the seed region of the homolog of miR-96 was shown to cause progressive hearing loss and hair cell defects in a murine model of sensorineural deafness, called diminuendo (15,16). The work of Mencía *et al.* (14) represented the first evidence that point mutations in a miRNA can be responsible for a Mendelian trait. However, up to now, the mutational screening of the miR-183 family has been performed only in the original study—including 567 Spanish families with inherited hearing loss (14)—and in a single replication study on 150 American families with autosomal dominant NSHL, where no mutations were found (17). Therefore, a positive replication of the findings by Mencía *et al.* is still lacking.

In this study, we have identified a novel NSHL-causing mutation within the *MIR96* gene by screening a large case–control Italian population. At variance with previously reported mutations, this variation does not affect the miR-96 seed region, but alters the companion miR-96* sequence. Through a series of functional studies, we have demonstrated that this variant substantially impairs the production of mature miR-96 by altering the pre-miRNA secondary structure, and

indirectly impacts on the normal regulation of miR-96 targets. Our results suggest that a quantitative defect of miR-96 can be sufficient to cause deafness. However, as we have shown that the novel mutation also reduces the expression levels of miR-96* and potentially alters the recognition of its targets, a contribution of this miRNA species to NSHL pathogenesis cannot be ruled out.

RESULTS

A novel NSHL-causing mutation within the *MIR96* gene

A cohort composed of 882 genetically undiagnosed Italian NSHL patients, including sporadic and familial cases, and 836 normal-hearing Italian controls was screened for mutations in miR-96, miR-182 and miR-183 genes. A strategy based on a pre-screening by high-resolution melting (HRM) followed by direct sequencing was chosen. Neither of the two previously reported miR-96 mutations was found, suggesting that they might represent private mutations. Instead, five heterozygous nt variations were identified in NSHL cases (Supplementary Material, Table S1): two of them, miR-96(+42C>T) and miR-182 (+106G>A), were previously reported as single nt polymorphisms (SNPs rs73159662 and rs76481776, respectively), and were found also in controls. Two novel *MIR183* variants, miR-183(+3G>T) and miR-183(-27C>T), were detected, each in one proband: both were located outside the mature miRNA sequence at non-conserved nt residues, were absent in normal-hearing controls, but did not segregate with hearing loss in the corresponding families (the variants were absent in at least one affected relative in the corresponding family). The last variant identified, miR-96(+57T>C) (NR_029512.1:c.57T>C; NT_007933.15:g.67447397A>G) is located in the stem region of the pre-miRNA and replaces a residue that is fully conserved throughout vertebrate evolution, from fish to primates (Fig. 1A). This variation affects the mature miR-96*, which is processed from the complementary strand of the miR-96 precursor. In particular, the change occurs at position +6, within the miR-96* seed region (Fig. 1A). The miR-96* has been experimentally detected in humans, mouse, platypus and zebrafish (mirBase, http://www.mirbase.org/) (18), suggesting that also the miRNA* species has been maintained throughout vertebrate evolution, although its sequence is only partially conserved. However, very little is known about miR-96* expression and function.

The miR-96(+57T>C) mutation was found in the heterozygous state in a patient with a family history of autosomal dominant progressive NSHL (with the age of onset ranging from ∼25 to 40) and was absent in all 839 normal-hearing controls. The proband is a profoundly deaf 56-year-old woman with an affected brother, two normal hearing sisters and three normal-hearing children (Fig. 1B and C). The proband's mother and grandmother were also profoundly deaf, but with a various degree of hearing loss (the grandmother being slightly more severe; Supplementary Material, Fig. S1). The proband shows non-syndromic, bilateral, sensorineural deafness that onset as a mild hearing impairment at ∼25 years and slowly progressed first to a severe form at the age of 45 and then to a profound form in the sixth decade (Fig. 1D). She presents a down-sloping audiometric
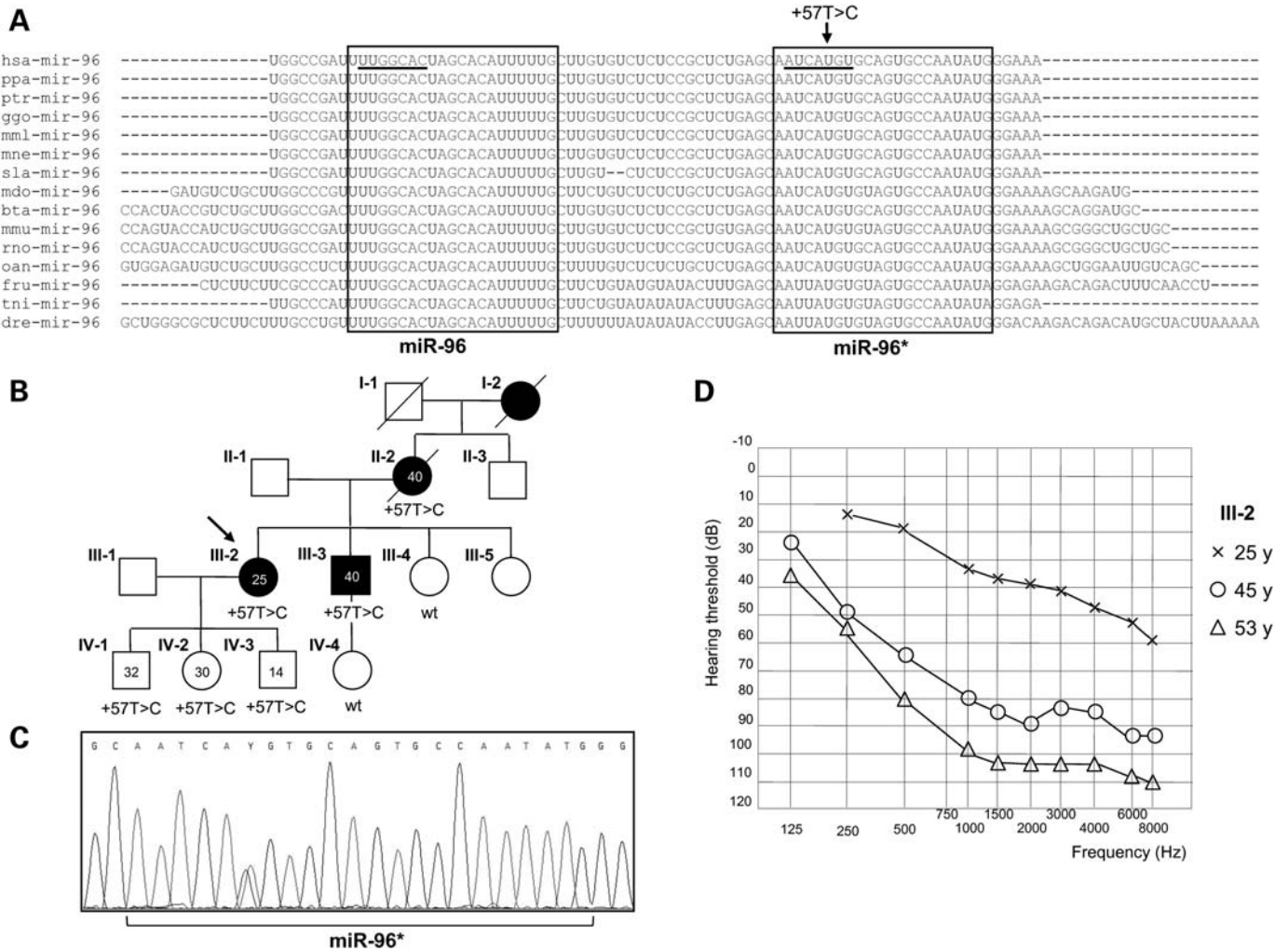
Figure 1. The novel pre-miR-96(+57T>C) mutation identified in an Italian family with autosomal dominant non-syndromic hearing loss. (**A**) Multiple alignment of pre-miR-96 sequences (annotated in miRBase) from different vertebrate species obtained using the CLUSTALW software. The nt position +57 within the human pre-miR-96 is indicated by an arrow. Mature miR-96 and miR-96* sequences are boxed, and their seed regions are underlined. hsa, *Homo sapiens*; ppa, *Pan paniscus*; ptr, *Pan troglodytes*; ggo, *Gorilla gorilla*; mml, *Macaca mulatta*; mne, *Macaca nemestrina*; sla, *Saguinus labiatus*; mdo, *Monodelphis domestica*; bta, *Bos taurus*; mmu, *Mus musculus*; rno, *Rattus norvegicus*; oan, *Ornithorhynchus anatinus;* fru, *Fugu rubipres*; tni, *Tetraodon nigroviridis*; dre, *Danio rerio*. (**B**) Pedigree of the Italian NSHL family carrying the miR-96(+57T>C) mutation. Black symbols indicate affected subjects. The numbers within the symbols represent: age of onset of NSHL (black symbols) and present age (empty symbols). Presence or absence of the mutation is indicated below the genetically analyzed subjects. (**C**) Electropherogram depicting the pre-miR-96 sequence surrounding the mutated nt. (**D**) Audiograms showing the progression of the hearing impairment in the proband (III-2). The age at which each audiometric record was obtained is indicated. Each graph point represents the average hearing loss for the right and left ears.

profile in which all frequencies are affected (Fig. 1D). Neither the proband nor any affected individuals of her family referred visual or olfactory problems; however, episodes of vertigo were reported in the proband (III-2), her mother (II-2) and her brother (III-3). The miR-96(+57T>C) variant segregates with the phenotype within the family, being present in all affected individuals and absent in the normal hearing individuals III-4 and IV-4. However, the variation is also present in the three normal-hearing proband's children (Fig. 1C and Supplementary Material, Fig. S1): this may be due either to the age-relatedness of hearing loss in the family (all non-penetrants are below the average age of onset of the disease among affected relatives) or to incomplete penetrance.

The +57T>C mutation impairs pre-miRNA processing into mature miR-96 and miR-96*

As the miR-96(+57T>C) variation is located in the stem region of the miRNA precursor, we used the mfold program (19) to examine how the newly identified nt variation could alter the predicted RNA secondary structure of pre-miR-96. The miR-96(+57T>C) substitution introduces a base-pairing mismatch, decreasing the free energy value and creating an enlarged RNA bulge in the pre-miR-96 stem, close to the Dicer cleavage site (Fig. 2A). In this frame, we decided to examine the impact of the novel variant on both miR-96 and miR-96* expression and maturation. To this aim, expression vectors (psiUX) containing genomic DNA fragments spanning both miR-183 and miR-96
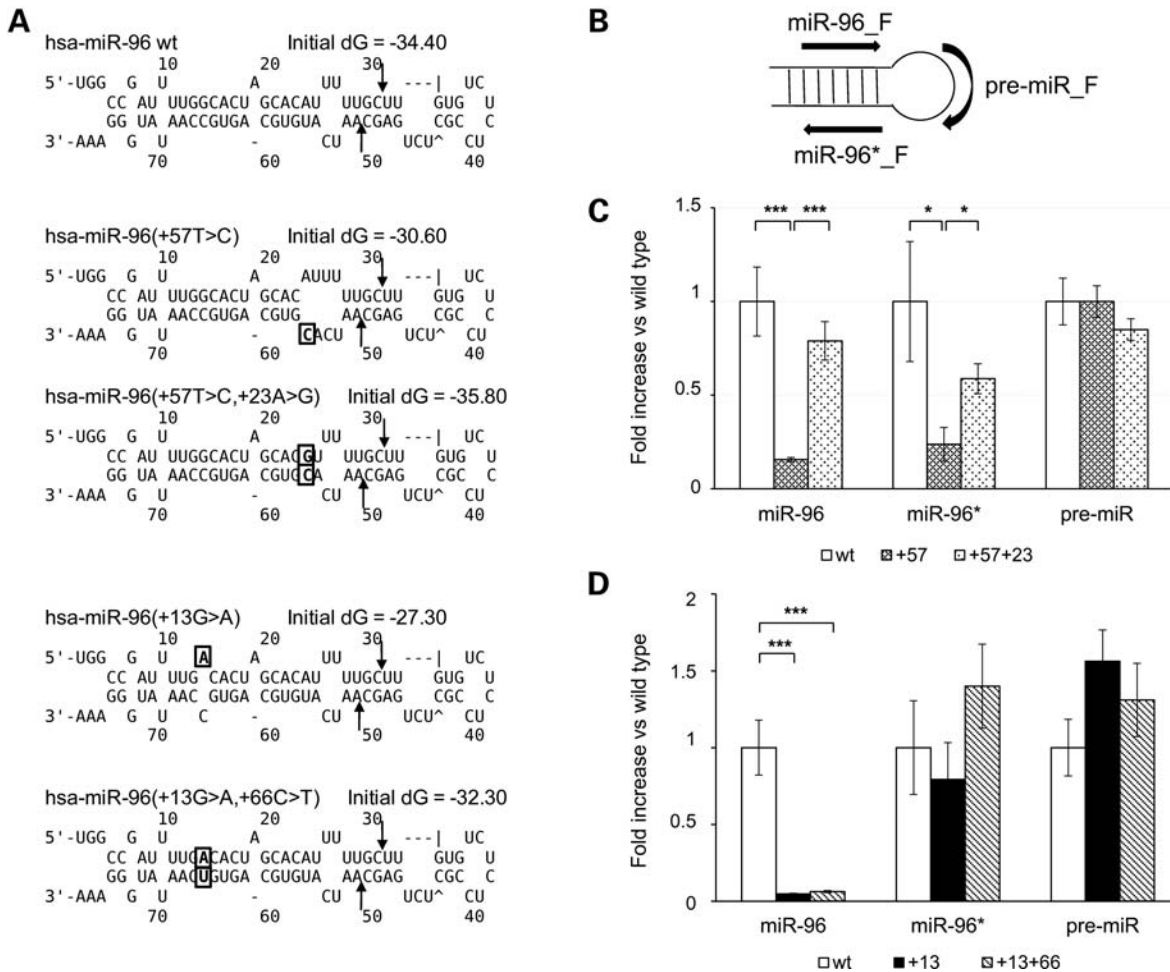
Figure 2. Different functional consequences of the miR-96(+57T>C) and the known miR-96(+13G>A) mutations on miR-96/miR-96* biogenesis. (**A**) Predicted secondary structures of wild-type, mutant (+57T>C and +13G>A) and double-mutant (+23A>G+57T>C and +13G>A+66C>T) miR-96 precursors obtained by using the mfold algorithm. The nt positions involved in the mutations are boxed. The double mutants were created to restore the pre-miR-96 secondary structure, in order to verify whether the defects in miRNA expression were dependent on correct folding of the hairpin precursor. Dicer cleavage sites are indicated by arrows. (**B**) The positions of the 5′ primers used for real-time RT–PCR to quantify the levels of mature miRNA and pre-miRNA species. Primer sequences are listed in Supplementary Material, Table S2. (**C**) The miR96(+57T>C) impairs the processing of pre-miR-96 to its mature forms, and is rescued by a compensatory mutation (+23A>G) restoring the correct hairpin folding. (**D**) The known miR96(+13G>A) impairs mature miR-96, but not miR-96* levels, independently from the correct folding of the miR-96 precursor. The effect of the analyzed mutants on pre-miR-96 processing was evaluated by quantitative real-time RT–PCR using the ΔΔCT method. Variations in the expression levels of mature miRNAs and pre-miR-96 in the miR-96(+57T>C) and miR-96(+23A>G+57T>C) (C), or the known miR-96(+13G>A) and miR-96(+13G>A+66C>T) (D) mutants were compared with the wild-type samples (set as 1). Bars stand for mean ± SEM (represented as percentage of variation) of six independent experiments, each performed in triplicate in different days on different cell batches and with different plasmid preparations. The results were analyzed by unpaired *t*-test (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

precursors, and corresponding to either the miR-96 wild-type or the miR-96(+57T>C) sequences, were generated and used to transiently transfect HeLa cells. Total RNA was extracted 24 h after transfection, reverse transcribed and the expression levels of mature miR-96, miR-96* as well as of pre-miR-96 were evaluated by real-time reverse-transcription polymerase chain reaction (RT–PCR), with specific custom-designed assays (Fig. 2B). Standard curves were generated to verify that the RT–PCR assays had comparable amplification efficiencies (see Materials and Methods and Supplementary Material, Fig. S2). In all experiments, miR-183 was used as internal control to normalize for transfection efficiency. Interestingly, expression levels of both miR-96 and miR-96* were significantly reduced in the mutant (+57T>C) compared with the wild-

type (85% reduction, $P = 0.0006$ for miR-96 and 77% reduction, $P = 0.019$, for miR-96*, respectively), whereas the precursor levels were unaffected by the mutation (Fig. 2C). Similar experiments were performed using a third expression vector carrying one of the known miR-96 mutations (+13G>A), previously shown, by Northern blot experiments, to decrease the level of miR-96 (14). In this case, the mutant showed a significant reduction in mature miR-96 ($P < 0.0001$) but, unexpectedly, miR-96* levels were unaffected (Fig. 2D), suggesting an at least partially different pathogenic mechanism.

The same approach was used to verify whether presumably non-pathogenic variants within *MIR96* impact the normal levels of the mature miRNA. Hence, we selected the two known polymorphisms within *MIR96* that are currently

annotated in dbSNP (rs73159662, +42C>T; and rs41274239, +36T>C). Both are located within the loop region of the miR-96 precursor, and were detected in the heterozygous state in healthy controls during our genetic screening (Supplementary Material, Fig. S3 and Table S1). No significant alteration of miR-96, miR-96* and pre-miR-96 levels was measured by real-time RT–PCR, although the rs73159662 mutant tends to produce slightly elevated levels of both mature miRNA species (Supplementary Material, Fig. S3).

The miR-96(+57T>C) mutation is rescued by reconstituting the precursor secondary structure

The miR-96(+57T>C) mutation causes a reduction in both mature miRNAs but does not alter the expression level of the miRNA precursor, thus suggesting a defect in the Dicer-mediated cleavage.

This may be due to the modification of the normal pre-miR-96 secondary structure close to the Dicer cleavage site, as predicted by mfold (see above and Fig. 2A). If this is true, a second mutation that would reconstitute the pairing of the +57 nt and hence the physiologic secondary structure of the miR-96 precursor (Fig. 2A) should restore miR-96 and miR-96* levels. To verify this hypothesis, we transfected HeLa cells with the double-mutant vector psiUX-miR-96(+23 A>G+57T>C), and evaluated the expression levels of miR-96, miR-96* and pre-miR-96. Both miR-96 and miR-96* levels were significantly increased in the double mutant compared with the +57T>C mutant ($P = 0.0003$ and $P = 0.02$, respectively) restoring mature miRNA expression at levels not significantly different from the wild-type (Fig. 2C). These results confirm that reduced miR-96/miR-96* levels, caused by the miR-96(+57T>C) mutation, are a consequence of the altered precursor secondary structure.

As comparison, we next verified if also the effect of the known miR-96(+13G>A) mutation could be rescued by restoring the correct folding of the miR-96 precursor. In principle, as this mutation does not impair the precursor expression and affects only one of the mature miRNA species (Fig. 2D), it should not directly interfere with either the Microprocessor-mediated or the Dicer-mediated cleavage steps. Indeed, transfection of a double-mutant vector carrying a compensatory mutation, psiUX-miR-96(+13G>A+66C>T), confirmed this hypothesis: miR-96 levels were still markedly reduced in the double mutant compared with the wild-type ($P < 0.0001$), and not significantly different from the +13G>A mutant, whereas the miR-96* and pre-miRNA levels remained unaffected (Fig. 2D). These results indicate that the known +13G>A mutation reduces mature miR-96 levels in a way that is independent from the correct folding of the hairpin precursor, pointing again to different pathogenic mechanisms for the +13G>A and the +57T>C mutations.

Impact of the +57T>C mutation on miRNA target regulation

We next evaluated the effect of the miR-96(+57T>C) mutation on the regulation of both miR-96 and miR-96* mRNA targets by luciferase reporter assays. First, we selected eight candidate target genes that: (i) contained either a 7-nt perfect match or multiple 6-nt binding sites for the miR-96* and/or for the miR-96 seed region; (ii) were predicted as potential miR-96/miR-96* targets by several bioinformatics programs (i.e. PITA, RNAhybrid, miRANDA/mirSVR) (20–22); (iii) were reported in the literature to be expressed in the vertebrate inner ear/auditory system (see below—Materials and Methods). Three of them (*ACVR2B*, *CACNB4* and *MYRIP*) were predicted as common miR-96/miR-96* targets, whereas the other five (*ALCAM*, *BBS4*, *IQGAP2, PLS3* and *ROCK2*) were targets of the miR-96* only (Supplementary Material, Fig. S4). To validate the candidate targets, we co-transfected a vector expressing either the wild-type or the mutant pre-miR-96 hairpin into HeLa cells, together with a construct containing the luciferase reporter cDNA coupled to the 3′UTR of each target gene (psiCHECK2-3′UTR). This reporter system showed that three (*MYRIP*, *ACVR2B*, *CACNB4*) out of eight luciferase-3′UTR constructs were controlled by human miR-96/miR-96* and that the miR-96(+57T>C) mutation led, in all cases, to a significantly reduced silencing (~50%) of luciferase expression compared with the wild-type pre-miR-96 (Fig. 3A and Supplementary Material, Fig. S5). In particular, the *MYRIP* 3′UTR, which contains an already validated miR-96 target site (14) as well as a predicted miR-96* site, showed the most significant regulation. To discriminate between the relative effect of miR-96 and miR-96* on *MYRIP* 3′UTR regulation, we created two mutant 3′UTR constructs in which the miR-96 or the miR-96* binding sites were selectively disrupted. While the construct lacking the miR-96*-binding site was still significantly responsive to miR-96 repression ($P < 0.01$), the mutant deprived of miR-96-binding site was not (Fig. 3A), indicating that *MYRIP* 3′UTR is regulated uniquely by miR-96, and that the predicted miR-96* binding site is not functional. Therefore, the impaired regulation of *MYRIP* 3′UTR in the miR-96(+57T>C) mutant is likely a consequence of the reduced levels of miR-96, caused by the defect in the hairpin processing. To confirm this hypothesis, we repeated the luciferase assays using the double miR-96(+23 A>G+57T>C) mutant, and observed that in this case the downregulation of *MYRIP* 3′UTR was restored to a level similar to that of the wild-type miR-96 (Fig. 3B). Hence, we can conclude that a quantitative defect in miR-96 production might be sufficient to impair the regulation of at least some of its downstream targets.

DISCUSSION

Here we report a novel mutation within the *MIR96* gene that causes autosomal dominant post-lingual progressive NSHL in an Italian family. This is the third mutation described in this gene in humans, and the first in a different population from that of the original report (14). The +57(T>C) variant is also the first mutation in the *MIR96* gene that does not affect the miR-96 seed region, neither its mature sequence. Instead, our results indicate that this mutation alters the correct maturation of the precursor miR-96, leading to a decrease in the level of both its mature forms. In particular, the +57(T>C) mutation is predicted to create an enlarge bulge in the secondary structure of miR-96 hairpin, 5 nts 3′
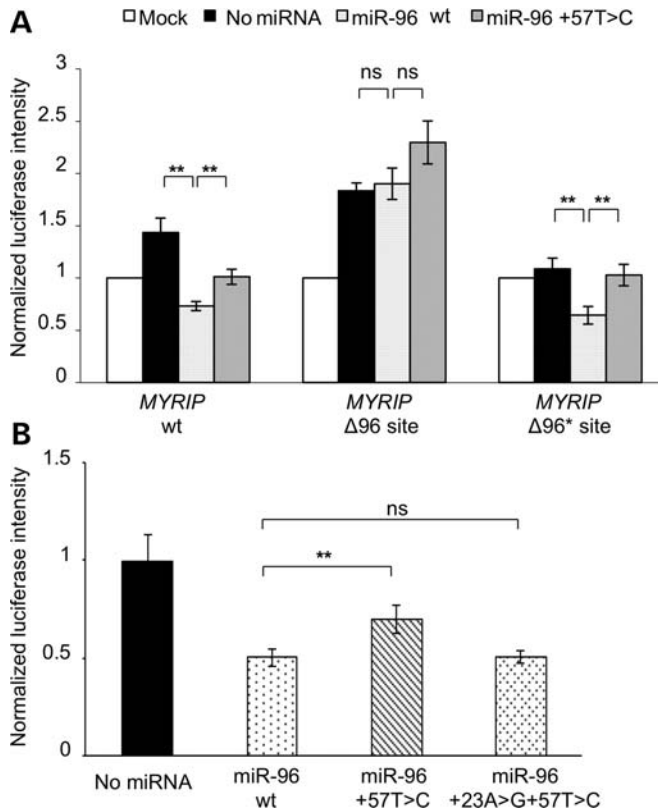
Figure 3. The miR96(+57T>C) mutation impairs the regulation of *MYRIP* by reducing mature miR-96 levels. (**A**) Downregulation of *MYRIP* 3′UTR is only dependent on miR-96, and is impaired in the miR96(+57T>C) mutant. The *MYRIP* 3′UTR reporter vectors (wild-type, or Δ96 and Δ96* mutants) were cotransfected with plasmids expressing either the wild-type or the +57T>C mutant pre-miR-96. Mock samples were transfected with the empty psiCHECK2 and psiUX vectors. (**B**) Restoration of correct pre-miR-96 folding rescues the regulation of *MYRIP* 3′UTR. The wild-type *MYRIP* 3′UTR reporter vector was cotransfected with the wild-type, the +57T>C or the double-mutant (+23A>G+57T>C) pre-miR-96 expression constructs. Relative luciferase activity is expressed as mean ± SEM of six independent experiments, each performed in triplicate in different days on different cell batches and with different plasmid preparations. The luciferase activity of the empty psiCHECK2 plasmid (A) or of the corresponding psiCHECK2-3′UTR plasmid (B) is set as 1. The results were analyzed by unpaired *t*-test (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

to the Dicer cleavage site (Fig. 2A), suggesting that it might interfere with Dicer processing. Other studies have reported that common SNPs within miRNA genes can lead to impaired miRNA processing, and might be associated with increased susceptibility to human diseases (23–25). Interestingly, one SNP within the miR-146a, which confers an increased risk to develop thyroid cancer, not only hindered the pre-miRNA processing, but also affected the seed region of the companion miR-146a*, similarly to the miR-96(+57T>C) mutation. The miR-146a* was expressed in both normal and tumor thyroid tissues and the altered regulation of the miRNA* targets in heterozygotes compared with homozygotes was suggested as the possible mechanism underlying tumor predisposition (26).

In the case of the miR-96(+57T>C) mutation, however, a direct role of miR-96* in NSHL pathogenesis has still to be proved. In fact, none of the putative miR-96* target sites was validated by luciferase-based assays (Supplementary

Material, Fig. S5). Profiling of miR-96* expression in a panel of human and mouse tissues evidenced that, although expressed at very low levels, miR-96* shows some degree of differential expression, suggesting that it might be regulated in a tissue-specific manner (Supplementary Material, Fig. S6). In the mouse Organ of Corti, the miR-96* could be cloned, but not reliably quantitated by real-time RT–PCR. More extensive and detailed experiments will be needed to define the role, if any, of this miRNA* species in the inner ear, as well as to evaluate the involvement of the mutant miR-96* in NSHL.

Even though mutations within miR-96 are not a common cause of deafness, as evidenced by this and previous works (14,17), the description of novel causative variants might help in elucidating the pathogenic mechanisms underlying the DFNA50-associated phenotype. In particular, the functional analysis of the +57(T>C) mutation points to a contribution of quantitative defects in miR-96 to hearing loss pathogenesis, independently from additional qualitative defects (i.e. changes in the actual mature miR-96 sequence). Indeed, tight regulation of miR-96 levels within the vertebrate inner ear seems to be crucial for the correct function and development of hair cells, as demonstrated by knockdown experiments in zebrafish (11) and by the phenotypic differences between the heterozygous and the homozygous diminuendo mice (15,16). Interestingly, also the three affected families (the two Spanish and the here reported Italian one) show some degree of phenotypic variation, suggesting that the pathogenic mechanisms underlying each family's phenotype might be at least partially different. In particular, the family carrying the +57(T>C) mutation is characterized by a late onset (between 25 and 40 years) and a slow progression of the hearing impairment. In addition, the presence of the mutation in three young individuals of the last generation (Fig. 1B), who currently have normal hearing at audiological analysis (Supplementary Material, Fig. S1), suggests incomplete penetrance, although they might still develop the disease in the future. The phenotypic differences with the previously reported families might be partly explained by molecular findings: a quantitative defect of miR-96 expression may lead to a delayed or reduced penetrance of the hearing impairment, whereas mutations directly impairing miR-96 target recognition would probably result in an earlier defect of hearing. In this respect, it should be recalled that the *diminuendo* mouse, which carries a mutant miR-96 seed region, does not have a major effect on miR-96 expression levels (15).

In conclusion, deciphering the different pathogenic mechanisms linking *MIR96* mutations to progressive hearing loss will be crucial for helping to develop new personalized therapeutic approaches for people carrying these mutations.

MATERIALS AND METHODS

Patients selection

This study was approved by the local Ethical Committee of the University of Milan and was performed according to the Declaration of Helsinki and to the Italian legislation on sensible data recording. Signed informed consent was obtained from all participants and from parents of subjects younger than

18 years. Clinical history ruled out environmental factors as cause of the hearing loss and physical examination did not reveal any evidence of syndromic features.

A total of 882 genetically undiagnosed NSHL patients (familial 30%, and sporadic 70%) were recruited. All patients underwent ear, nose and throat, and audiological examinations. For adults and collaborative children, hearing levels were determined by pure-tone audiometry, in accordance with International Standard Organization (ISI 8253-1-3) protocols. The hearing impairment diagnosis of small children (<1 year) was obtained through the auditory brainstem responses and the observation of their behaviors, whereas in older children a behavioral audiometry was performed. In addition, evoked otoacoustic emissions and tympanometry with acoustic reflex thresholds were evaluated. Average thresholds in the range of 21–40 dB were defined as mild, 41–70 dB as moderate, 71–95 dB as severe and >95 dB as profound hearing loss. Finally, specific questions were asked to evaluate olfactory, visual and vestibular functions.

All recruited patients did not carry mutations within gap-junction proteins connexins 26 and 30 (*GJB2*, *GJB6*), and the mitochondrial 12S rRNA (*MTRNR1*) genes.

A total of 120 individuals with no familiarity for hearing loss and instrumentally verified normal auditory function (mean age at withdrawal 32 ± 9) were also collected as controls. Additional 716 controls with a mean age >50 years and declared normal auditory function were also recruited.

Mutational screening

Genomic DNA extraction from peripheral blood was performed using a semi-automatic extractor (Fujifilm Europe GmBH, Düsseldorf, Germany), whereas DNA from buccal swabs from small children was purified according to the QIAamp DNA Mini Kit extraction protocol (Qiagen, Hilden, Germany). DNA samples were quantified on a Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), standardized for concentration, arrayed into 96-well plates and stored at −20°C.

Although both previously reported miR-96 mutations were found in families with autosomal dominant NSHL, we chose to perform the mutational screening on all available genetically undiagnosed NSHL patients, independently from the mode of inheritance. Three sets of primer pairs were designed to PCR amplify the genomic DNA fragments containing the precursor sequences and flanking regions of *MIR183*, *MIR96* and *MIR182* (Supplementary Material, Table S1). PCR amplimers were screened for mutations by HRM on a LigthCycler 480 using the HRM Master kit (Roche, Basel, Switzerland) and a touch-down protocol (thermal profiles are available on request). Amplicons were analyzed with the Gene Scanning Software (Roche). Samples showing a melting profile different from the wild-type control were further analyzed by direct sequencing, using the BigDye Terminator Cycle Sequencing Ready Reaction Kit v1.1 (Applied Biosystems, Foster City, CA, USA) and an ABI-3130XL sequencer, as previously described (27). The Variant Reporter software was used for mutation detection (Applied Biosystems).

Computational methods

The mfold program (19) was used to evaluate the possible impact of the +57T>C mutation on the secondary structure of pre-miR-96.

We used the TargetRank program (28) to predict potential miR-96, miR-96*T (wild-type), and miR-96*C (mutant) targets and selected those genes that were either potential targets of the wild-type miR-96* only, or targeted by both the miR-96 and the wild-type miR-96*, but not the mutant miR96*. Among these, we then selected for validation by luciferase assays eight genes that: (i) contained either a 7-nt perfect match or multiple 6-nt binding sites for the miR-96* and/or miR-96 seed region; (ii) were predicted as potential miR-96/miR-96*T targets by additional bioinformatics programs (i.e. PITA, RNAhybrid, miRANDA/mirSVR) (20–22); (iii) were expressed in the inner ear/auditory system (15,29–34).

Expression vectors

To generate vectors for miRNA expression, genomic DNA fragments including both the miR-183 and the miR-96 precursors, and corresponding to either the miR-96 wild-type or the miR-96(+57T>C) sequences, were amplified from genomic DNA using modified primers (containing a *Kpn*I or a *Xho*I site at their 5′ends; Supplementary Material, Table S2), inserted into the psiUX vector (kindly provided by prof. I. Bozzoni, Università di Roma La Sapienza) (35) and then verified by sequencing. The psiUX-miR-96(+13G>A) vector, carrying one of the known miR-96 mutations (14), the double mutants, psiUX-miR-96(+23A>G+57T>C) and psiUX-miR-96(+13G>A+66C>T) and the vectors carrying the two known *MIR96* polymorphisms (rs73159662, +42C>T; and rs41274239, +36T>C) were obtained by site-directed mutagenesis using the QuikChange kit (Stratagene, La Jolla, CA, USA), following the manufacturer's instructions.

The entire 3′UTRs of putative target genes were amplified from genomic DNA and directionally cloned into a psi-CHECK2 vector (Promega, Madison, WI, USA), downstream of the renilla luciferase gene. Constructs carrying mutant versions of miRNA binding sites (Δ96 and Δ96* binding sites) were obtained by site-directed mutagenesis.

Ex-vivo analysis of miRNA expression and biogenesis

HeLa cells were cultured in Dulbecco's modified Eagle medium containing 2 mM L-glutamine, 10% fetal bovine serum and antibiotics (100 U/ml penicillin and 100 μg/ml streptomycin; Euroclone, Wetherby, UK) and grown at 37°C in a humidified atmosphere of 5% CO_2 and 95% air, according to the standard procedures.

In each transfection experiment, an equal number of cells (250 000) were transiently transfected in six-well plates with the Fugene HD reagent (Promega) and 4 μg of plasmid DNA, following the manufacturer's instructions. Twenty-four hours after transfection, cells were washed twice with phosphate-buffered saline and total RNA was extracted using the EUROzol reagent (Euroclone), according to the manufacturer's instructions. RNA concentration was quantified by Nanodrop ND-1000 spectrophotometer (NanoDrop

Technologies). The miRNA First Strand Synthesis kit (Agilent Technologies, Palo Alto, CA, USA) was used to reverse transcribe miRNAs, starting from 300 ng of total RNA, according to the manufacturer's instructions. An aliquot (1 μl) of a 1:5 dilution of the RT reaction was used as template in a standard real-time RT–PCR amplification, using the universal reverse primer (Agilent Technologies) and miRNA-specific forward primers (Supplementary Material, Table S2). Although in principle the primers for the amplification of mature miR-96 and miR-96* can also amplify the pre-miRNA, real-time RT–PCR conditions were set up to favor the amplification of the mature forms and the analysis of melting curves confirmed the presence of a single amplification product (data not shown). The specific 5′ primer used to amplify mature miR-96* was designed downstream of the +57 nt position, so that the same primer could be used to amplify both the wild-type and the +57T>C mutant. Conversely, standard curves were generated to verify that the RT–PCR assays to detect: (i) wild-type miR-96, miR-96(+13G>A) or miR-96(+23A>G); (ii) wild-type miR-96* and miR-96*(+66 T>C); (iii) wild-type pre-miR-96, pre-miR-96(+42C>T) and pre-miR-96(+36T>C) had comparable amplification efficiencies (Supplementary Material, Fig. S2). In all cases, real-time RT–PCR assays were performed in triplicate on a LightCycler 480 (Roche), using a touch-down thermal profile (PCR conditions are available on request). MiRNA expression levels were quantified by the $\triangle\triangle CT$ method (36) using the miR-183, co-expressed by the same construct, as internal normalization reference. An unpaired, one-tailed *t*-test was performed to test for significant differences between the mutants and the wild-type samples.

MiRNA target validation by luciferase assays

Transient transfections were performed by using the Fugene HD reagent (Roche) and 2 μg of psiCHECK2-3′UTR constructs together with 0.2 μg of the wild-type, the mutant (+57T>C) or the double-mutant (+23A>G+57T>C) psiUX-miR-96 vector. In this case, the psiUX constructs used for miRNA expression contained only the miR-96 precursor, thus avoiding the confounding effect of cloning the related miR-183. Cells co-transfected with the empty psiCHECK2 and psiUX vectors served as mocks. Forty-eight hours post-transfection, renilla and firefly luciferase activities were measured in cell lysates using the Dual-Luciferase Assay System (Promega) on a Wallac Victor 1420 Reader (Perkin Elmer Life Sciences, Waltham, MA, USA). The results were analyzed by unpaired *t*-test.

MiRNA expression profile

Expression levels of miR-96 and miR-96* in a panel of RNAs from human (First Choice total RNA; Ambion, Austin, TX, USA) and adult mouse tissues were evaluated by semi-quantitative real-time RT–PCR, using specific assays (Supplementary Material, Table S2). The organs of Corti from wild-type P4 mice (background 129S5) were kindly provided by prof. Karen Steel (Wellcome Trust Sanger Institute, Hinxton, UK). Total RNA extraction was performed as reported (15), and first-strand cDNA synthesis was carried

out as described above. All assays were performed in triplicate on a LightCycler 480 (Roche), using a touch-down thermal profile (PCR conditions are available on request). MiRNAs expression levels were quantified by the $\triangle\triangle CT$ method using either the U6 RNA (human) or the snoRNA 142 (mouse) as internal normalization reference.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

REFERENCES

1. Huntzinger, E. and Izaurralde, E. (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.*, **12**, 99–110.
2. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
3. Krol, J., Loedige, I. and Filipowicz, W. (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, **11**, 597–610.
4. Ro, S., Park, C., Young, D., Sanders, K.M. and Yan, W. (2007) Tissue-dependent paired expression of miRNAs. *Nucleic Acids Res.*, **35**, 5944–5953.
5. Okamura, K., Phillips, M., Tyler, D., Duan, H., Chou, Y. and Lai, E. (2008) The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution. *Nat. Struct. Mol. Biol.*, **15**, 354–363.
6. de Wit, E., Linsen, S.E., Cuppen, E. and Berezikov, E. (2009) Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res.*, **19**, 2064–2074.
7. Marco, A., Hui, J.H., Ronshaugen, M. and Griffiths-Jones, S. (2010) Functional shifts in insect microRNA evolution. *Genome Biol. Evol.*, **2**, 686–696.
8. Yang, J.S., Phillips, M.D., Betel, D., Mu, P., Ventura, A., Siepel, A.C., Chen, K.C. and Lai, E.C. (2011) Widespread regulatory activity of vertebrate microRNA* species. *RNA*, **17**, 312–326.
9. Pierce, M., Weston, M., Fritzsch, B., Gabel, H., Ruvkun, G. and Soukup, G. (2008) MicroRNA-183 family conservation and ciliated neurosensory organ expression. *Evol. Dev.*, **10**, 106–113.
10. Soukup, G., Fritzsch, B., Pierce, M., Weston, M., Jahan, I., McManus, M. and Harfe, B. (2009) Residual microRNA expression dictates the extent of inner ear development in conditional Dicer knockout mice. *Dev. Biol.*, **328**, 328–341.
11. Li, H., Kloosterman, W. and Fekete, D. (2010) MicroRNA-183 family members regulate sensorineural fates in the inner ear. *J. Neurosci.*, **30**, 3254–3263.

12. Kapsimali, M., Kloosterman, W., de Bruijn, E., Rosa, F., Plasterk, R. and Wilson, S. (2007) MicroRNAs show a wide diversity of expression profiles in the developing and mature central nervous system. *Genome Biol.*, **8**, R173.

13. Friedman, L. and Avraham, K. (2009) MicroRNAs and epigenetic regulation in the mammalian inner ear: implications for deafness. *Mamm. Genome*, **9–10**, 581–603.

14. Mencía, A., Modamio-Høybjør, S., Redshaw, N., Morín, M., Mayo-Merino, F., Olavarrieta, L., Aguirre, L., del Castillo, I., Steel, K., Dalmay, T. *et al.* (2009) Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat. Genet.*, **41**, 609–613.

15. Lewis, M., Quint, E., Glazier, A., Fuchs, H., De Angelis, M., Langford, C., van Dongen, S., Abreu-Goodger, C., Piipari, M., Redshaw, N. *et al.* (2009) An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. *Nat. Genet.*, **41**, 614–618.

16. Kuhn, S., Johnson, S.L., Furness, D.N., Chen, J., Ingham, N., Hilton, J.M., Steffes, G., Lewis, M.A., Zampini, V., Hackney, C.M. *et al.* (2011) miR-96 regulates the progression of differentiation in mammalian cochlear inner and outer hair cells. *Proc. Natl Acad. Sci. USA*, **108**, 2355–2360.

17. Hildebrand, M., Witmer, P., Xu, S., Newton, S., Kahrizi, K., Najmabadi, H., Valle, D. and Smith, R. (2010) miRNA mutations are not a common cause of deafness. *Am. J. Med. Genet. A*, **152A**, 646–652.

18. Soares, A., Pereira, P., Santos, B., Egas, C., Gomes, A., Arrais, J., Oliveira, A., Moura, G. and Santos, M. (2009) Parallel DNA pyrosequencing unveils new zebrafish microRNAs. *BMC Genomics*, **10**, 195.

19. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

20. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. and Segal, E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.

21. Krüger, J. and Rehmsmeier, M. (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.*, **34**, W451–W454.

22. Betel, D., Koppal, A., Agius, P., Sander, C. and Leslie, C. (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.*, **11**, R90.

23. Duan, R., Pak, C. and Jin, P. (2007) Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. *Hum. Mol. Genet.*, **16**, 1124–1131.

24. Sun, G., Yan, J., Noltner, K., Feng, J., Li, H., Sarkis, D.A., Sommer, S.S. and Rossi, J.J. (2009) SNPs in human miRNA genes affect biogenesis and function. *RNA*, **15**, 1640–1651.

25. Jazdzewski, K., Murray, E.L., Franssila, K., Jarzab, B., Schoenberg, D.R. and de la Chapelle, A. (2008) Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc. Natl Acad. Sci. USA*, **105**, 7269–7274.

26. Jazdzewski, K., Liyanarachchi, S., Swierniak, M., Pachucki, J., Ringel, M.D., Jarzab, B. and de la Chapelle, A. (2009) Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. *Proc. Natl Acad. Sci. USA*, **106**, 1502–1505.

27. Guella, I., Pistocchi, A., Asselta, R., Rimoldi, V., Ghilardi, A., Sironi, F., Trotta, L., Primignani, P., Zini, M., Zecchinelli, A. *et al.* (2010) Mutational screening and zebrafish functional analysis of GIGYF2 as a Parkinson-disease gene. *Neurobiol. Aging*, **11**, 1994–2005.

28. Nielsen, C.B., Shomron, N., Sandberg, R., Hornstein, E., Kitzman, J. and Burge, C.B. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, **13**, 1894–1910.

29. May-Simera, H.L., Ross, A., Rix, S., Forge, A., Beales, P.L. and Jagger, D.J. (2009) Patterns of expression of Bardet-Biedl syndrome proteins in the mammalian cochlea suggest noncentrosomal functions. *J. Comp. Neurol.*, **514**, 174–188.

30. Mothe, A.J. and Brown, I.R. (2001) Expression of mRNA encoding extracellular matrix glycoproteins SPARC and SC1 is temporally and spatially regulated in the developing cochlea of the rat inner ear. *Hear. Res.*, **155**, 161–174.

31. McCullar, J.S., Ty, S., Campbell, S. and Oesterle, E.C. (2010) Activin potentiates proliferation in mature avian auditory sensory epithelium. *J. Neurosci.*, **30**, 478–490.

32. Kuhn, S., Knirsch, M., Rüttiger, L., Kasperek, S., Winter, H., Freichel, M., Flockerzi, V., Knipper, M. and Engel, J. (2009) Ba2+ currents in inner and outer hair cells of mice lacking the voltage-dependent Ca2+ channel subunits beta3 or beta4. *Channels*, **3**, 366–376.

33. Robertson, N.G., Khetarpal, U., Gutiérrez-Espeleta, G.A., Bieber, F.R. and Morton, C.C. (1994) Isolation of novel and known genes from a human fetal cochlear cDNA library using subtractive hybridization and differential screening. *Genomics*, **23**, 42–50.

34. Daudet, N. and Lebart, M.C. (2002) Transient expression of the t-isoform of plastins/fimbrin in the stereocilia of developing auditory hair cells. *Cell Motil. Cytoskeleton*, **53**, 326–336.

35. Denti, M.A., Rosa, A., Sthandier, O., De Angelis, F.G. and Bozzoni, I. (2004) A new vector, based on the PolII promoter of the U1 snRNA gene, for the expression of siRNAs in mammalian cells. *Mol. Ther.*, **10**, 191–199.

36. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.

Part III

Content

Robusto M*, Fang M*, Castorina P, Previtali S C, Benzoni E, Caccia S, De Cristofaro R, Yu C, Liu X, Li W, Primignani P, Ambrosetti U, Asselta R, Xu X$, Duga S$, Soldà G. *"Mutations in PRPS1 are a potentially overlooked genetic cause for X-linked hearing loss and may result in a continuum spectrum of phenotypes"*. **[Under submission to Journal of Medical Genetics]**

*shared first authorship

$co-corresponding

Mutations in *PRPS1* are a potentially overlooked genetic cause for X-linked hearing loss and may result in a continuum spectrum of phenotypes

Michela Robusto[1]*, Mingyan Fang[2]*, Pierangela Castorina[3], Stefano C. Previtali[6], Elena Benzoni[3], Sonia Caccia[1], Raimondo De Cristofaro[5], Cong Yu[2], Xuanzhu Liu[2], Wangsheng Li[2], Paola Primignani[4], Umberto Ambrosetti[3], Rosanna Asselta[1], Xun Xu[2]$, Stefano Duga[1]$, Giulia Soldà[1]

*shared first authorship

$co-corresponding

1. Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, Milan, Italy;
2. BGI-Shenzhen, Shenzhen, China;
3. UO Audiologia, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy;
4. Laboratory of Medical Genetics, Molecular Genetic Sector, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy;
5. Institute of Internal Medicine & Geriatrics, Haemostasis Research Center, Catholic University School of Medicine, Rome, Italy;
6. Institute of Experimental Neurology and Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy

Abstract

Background:

Next-generation sequencing is currently the technology of choice for gene/mutation discovery in genetically-heterogeneous disorders, such as inherited sensorineural hearing loss (HL).

Methods:

A combination of molecular genetics (whole-exome sequencing, WES, and Sanger sequencing), bioinformatics, and biochemical analyses were used to identify and functionally characterize novel causative mutations underlying sensorineural HL.

Results:

WES of a single Italian proband affected by recessive HL identified a novel missense mutation within the *PRPS1* gene (p.A113S), which segregates with nonsyndromic, post-lingual, bilateral, progressive deafness in the proband's family. Defects in this gene, which codes for the phosphoribosyl pyrophosphate synthetase 1 (PRS-I) enzyme, determine either X-linked syndromic conditions associated with hearing impairment (such as Arts syndrome and Charcot-Marie-Tooth neuropathy type X-5, CMTX5), or nonsyndromic HL (*DFNX1 locus*). A subsequent screening of the entire *PRPS1* gene in 16 unrelated probands from X-linked deaf families led to the discovery of two additional missense mutations (p.M115V and p.V309F) segregating with hearing impairment, and associated with peripheral neuropathy. All three mutations result in a marked reduction (>60%) of the PRS-I activity in the patients' erythrocytes compared to controls, with the p.M115V and p.V309F variants affecting more severely the enzyme function.

Conclusions:

Our data significantly expand the current spectrum of pathogenic mutations in *PRPS1* and highlight their recurrence, suggesting that this gene may represent a major *locus* for X-linked deafness. Moreover, our results highlight the importance of both functional studies and proper clinical evaluation for the identification of concomitant neurological symptoms.

Introduction

Nonsyndromic sensorineural hearing loss (NSHL) is the most common sensory disorder in humans, affecting about 278 million people worldwide, with severe consequences on public health. It is estimated that about 60-70% of cases are due to genetic factors [1], of which about 5% are transmitted as X-linked traits. Overall, inherited hearing loss (HL) is characterized by a high genetic heterogeneity, with about 70 causal genes identified so far (http://www.hereditaryhearingloss.org), and many still to be discovered.

The extremely high number of NSHL-causing genes, as well as the low frequency of most known pathogenic mutations, leave the vast majority of patients with no definitive genetic diagnosis, principally because the genetic screening of all known deafness genes by traditional mutation analysis methods is unfeasible. Instead, whole-exome sequencing (WES) currently represents one of the most efficient strategy

to search for rare variants underlying Mendelian diseases, and has already been applied to the discovery of novel genes/mutations responsible for autosomal (*DFNB82 locus*) and X-linked (*DFNX4 locus*) recessive NSHL [2-4]. Before the identification of *SMPX* as the causative gene at the *DFNX4 locus*, only two other genes, *POU3F4* (*DFNX2 locus*, OMIM #304400) and *PRPS1* (*DFNX1 locus*, OMIM #304500), were clearly associated with X-linked forms of deafness [5, 6].

PRPS1 codes for phosphoribosylpyrophosphate synthetase 1 (PRS-I), a member of an evolutionarily conserved family of enzymes, which catalyzes the synthesis of phosphoribosyl pyrophosphate (PRPP) from ATP and ribose-5-phosphate, and is essential for the *de-novo* synthesis of nucleotides in both eukaryotes and prokaryotes [7]. The active PRS-I enzyme is a hexamer composed of three homodimers disposed in a propeller-like shape, each containing a catalytic and two allosteric regulatory sites [8, 9].

The *DFNX1 locus* is associated with phenotypically heterogeneous NSHL. In general, hearing impairment in male patients with *PRPS1* mutations is bilateral, moderate to profound, and can be pre- or post-lingual, progressive or non-progressive. Female carriers may also be affected by unilateral or bilateral hearing impairment. So far, five *DFNX1* families were described, four with post-lingual progressive NSHL, and one with congenital profound deafness [6]. Four missense mutations in *PRPS1* (NM_002764.3: p.D65N, p.A87T, p.I290T, and p.G306R) were identified in these families as disease-causing. All of them result in a reduction of about half of PRS-I activity, as shown by *in-vitro* enzymatic assays on patients' erythrocytes and cultured fibroblasts. None of these mutations are predicted to cause a major structural change in the PRS-I protein [6].

Mutations in *PRPS1* might also result in a spectrum of syndromic conditions, including PRS-I superactivity (OMIM #300661; seven mutations so far identified) [10], Charcot-Marie Tooth neuropathy type X-5 (CMTX5 or Rosenberg-Chutorian syndrome, OMIM #311070; two known mutations) [11], and Arts syndrome (OMIM #301835; two mutations reported to date) [12]. In general, *PRPS1* gain-of-function mutations determine a loss of feedback regulation and hence enzyme superactivity, which in turn causes excessive uric acid production and gout. Conversely, loss-of-function mutations are responsible for NSHL, CMTX5, or Arts syndrome. In this case, the severity of the phenotype seems to correlate with the residual functionality of the enzyme, with milder mutations resulting only in HL, and more severe genetic defects causing also peripheral neuropathy (CMTX5), or central nervous system involvement and increased liability to infections (Arts syndrome). Taken together, these findings suggest that the four *PRPS1*-related phenotypes are part of the same disease spectrum [13], thus stressing the importance of the functional characterization of pathogenic variants as well as the interpretation of genotype/phenotype correlations. In particular, CMTX5, which is characterized by a triad of symptoms: early-onset bilateral profound HL, peripheral neuropathy, and optic neuropathy, might be under-recognized by physicians and thus under-diagnosed [14].

Here, we used a WES approach to search for novel genetic determinants of NSHL in a genetically-undiagnosed Italian family affected by post-lingual profound deafness, leading to the identification of a novel missense mutation in *PRPS1*. Two additional novel mutations in *PRPS1* were discovered by

extending the screening of the entire gene to 16 families affected by X-linked sensorineural HL. *PRPS1* mutations were functionally characterized to assess their impact on the enzyme activity, and detailed neurological evaluation was performed in order to refine genotype/phenotype correlations.

Materials and Methods

Patients

This study was approved by local Ethical Committees and was performed according to the Declaration of Helsinki and to the Italian legislation on sensible data recording. Signed informed consent was obtained from all participants and from parents of subjects younger than 18 years.

Genomic DNA extraction from peripheral blood was performed using a semi-automatic DNA extractor (Fujifilm Europe GmBH, Düsseldorf, Germany), whereas DNA from buccal swabs from young children was purified according to the QIAamp DNA Mini Kit extraction protocol (Qiagen, Hilden, Germany). DNA samples were quantified on a Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), and stored at -20°C. All recruited patients did not carry mutations within gap-junction proteins connexin 26 and 30 (*GJB2*, *GJB6*), and the mitochondrial 12S rRNA (*MTRNR1*) genes. Genetic screening of *GJB2*, *GJB6*, and *MTRNR1* was performed as described [15].

Audiological evaluation

A total of 17 genetically-undiagnosed NSHL families with a likely X-linked recessive inheritance pattern were recruited from the Genetic Service and the Audiology Unit of the Ospedale Maggiore Policlinico of Milan. Clinical history ruled out environmental factors as cause of the HL and physical examination did not reveal any evidence of dysmorphic features.

All patients underwent ear, nose and throat, and audiological examinations. For adults and collaborative children, hearing levels were determined by pure-tone audiometry, in accordance with International Standard Organization (ISI 8253-1-3) protocols. The hearing impairment diagnosis of young children (<1 year) was obtained through the auditory brainstem responses and the observation of their behaviours, whereas in older children a behavioural audiometry was performed. In addition, evoked otoacoustic emissions and tympanometry with acoustic reflex thresholds were evaluated. Average thresholds in the range of 21-40 dB were defined as mild, 41-70 dB as moderate, 71-95 dB as severe, and >95 dB as profound HL.

A total of 123 individuals with no familiarity for HL and instrumentally-verified normal auditory function (mean age at withdrawal 32±9) were also collected as controls.

Neurological evaluation

Patients underwent a standard neurological examination, including evaluation of muscle weakness and disability, gait, sensory loss, deep tendon reflexes, coordination, presence of pyramidal or cerebellar signs, cranial nerve involvement, as well as the presence of bone abnormalities, joint retractions, and skin lesions.

Patients were evaluated with electromyography (EMG) and motor and sensory electroneurography, and routine laboratory tests including muscle enzymes, transaminases, metabolites, electrolytes, and screening for disimmune/inflammatory disorders.

Targeted DNA capture and exome sequencing

One NSHL family (Family 1), with a clear recessive inheritance pattern and two affected individuals (Figure 1A) was selected for WES. Genomic capture and sequencing were performed on a single proband (II-2).

In brief, 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. WES yielded 4.8 Mb raw sequence data with about 97.5% coverage of the targeted exome with a mean coverage ≥50X (Supplementary Figure 1).

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 [16], 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 (http://www.broadinstitute.org/gatk/). 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 (corresponding to 99% accuracy of a base call). The called variants were annotated according to their position and predicted functional impact on the protein (Supplementary Tables 1 and 2). 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 (ftp://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, including data on 5,400 exomes of

Caucasian and African American origins). SIFT (http://sift.jcvi.org), Condel (http://bg.upf.edu/condel/home) and MutationTaster (http://www.mutationtaster.org/) were used to predict whether an amino-acid substitution affects protein function. Rare and potentially deleterious variants mapping within known NSHL *loci* were extracted and prioritized for validation. Sanger sequencing was used to confirm the presence of the identified potential disease-causing variants and to verify their segregation with the phenotype within Family 1.

PRPS1 mutational screening

All seven exons and exon-intron boundaries of *PRPS1* were PCR amplified using sets of primers designed on the basis of the known genomic sequence of the gene (GenBank accession number NM_002764). PCRs were performed on 10-20 ng of genomic DNA in a 25-μL reaction volume, following standard procedures. Primer sequences, as well as the specific PCR conditions for each primer couple are available on request.

Direct sequencing of amplified fragments was performed on both strands by means of the BigDye Terminator Cycle Sequencing Ready Reaction Kit v.1.1 and an automated ABI-3130XL DNA sequencer (Applied Biosystem, Foster City, CA, USA). The Variant Reporter software (Applied Biosystems) was used for mutation detection.

Molecular modeling

The effects of the p.A113S, p.M115V, and p.V309F mutations on the structure of PRS-I were examined by using the coordinates of the 3D-model of the homohexamer kindly provided by Dr S.B. Nabuurs and Dr A. de Brouwer [12], as well as the coordinates of the crystal structure of the human protein (PDB entry 2H06; http://www.pdb.org) [9]. The analysis was performed using the FoldX plugin for the Yet Another Scientific Artificial Reality Application (YASARA) program. This algorithm allowed us to evaluate the effect of mutations on the stability of a single subunit and on the interaction energy of the subunits in the complex [17-19]. Figures were produced using PyMOL (DeLano Scientific, San Carlos, CA, USA; http://www.pymol.org).

PRS-1 activity assay

Venous blood samples were collected in EDTA tubes from five wild-type control subjects (one healthy sibling of an affected patient and four unrelated controls), five affected males, as well as four female carriers of *PRPS1* mutations.

Immediately after phlebotomy, peripheral blood mononuclear cells (PBMCs) and erythrocytes were separated by centrifugation on a Lympholyte Cell separation media (Cederlane Laboratories Limited, Hornby, Ontario, Canada) gradient. Red cells were then washed three times in phosphate buffered saline, packed by centrifugation for 10 min at 2,000 rpm, and stored at -20°C. On the day of the assay, each sample was thawed and treated with activated charcoal (3 mg/mL) for 15 min at 4°C. After centrifugation, the supernatant was assayed for PRS-1 activity, essentially as described by Torres and colleagues [20].

Briefly, 100 μL of the charcoal-treated hemolysate were incubated for 60 min at 37°C with 1.9 mL of the pH 7.4 reaction buffer (50 mM Tris-HCl, 5 mM MgCl$_2$, 1 mM EDTA, 0.4 mM DTT, 0.5 mM ATP, 0.35 mM ribose-5-phosphate, 32 mM NaPi, and 0.25 mM diadenosine pentaphosphate, Ap5A). At five time points (0, 15, 30, 45, and 60 min), a 0.3 mL aliquot of each reaction was stopped by adding 30 μL of 100 mM EDTA, and centrifuged in Amicon Ultra Ultracel 10K Membrane filter (Millipore, Billerica, MA, USA) at 4,000 x g for 30 min at 4°C to remove proteins. Control reactions without ribose-5-phosphate were set up for each time-point using the hemolysate of a healthy control. ATP, ADP, and AMP in the filtrate were separated by high-performance liquid chromatography (HPLC).

The hemoglobin content for each hemolysate was determined by the Drabkin method [21], and used for normalization.

HPLC separation

Samples were appropriately diluted and injected into an HPLC C$_{18}$ column (ReproSil-Pur C18-AQ RP-HPLC; Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) and eluted at room temperature with 50 mM K(Na)H$_2$P0$_4$ (pH 6.0) at a flow rate of 1.0 mL/min, following a slightly modified method by Torres and colleagues [20]. The HPLC apparatus was a double-pump Model 2080 HPLC system equipped with a Model 2070 UV spectrophotometric and a FP-2020 fluorescence detector (Jasco, Tokyo, Japan). Absorbance was measured at 254 nm, and different concentrations of AMP and ATP (100-3.25 μM) were injected as standard. PRS-I activity was expressed as concentration of AMP generated per time unit and per mg of hemoglobin in the hemolysate.

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) mutation, 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 mutation under analysis (*PRPS1*_ex3cDNA_F: 5'-GTGGTTGTGGCGAAATCAAT-3', and *PRPS1*_ex3cDNA_R: 5'-ACCAGCATCAGGTGAGACAA-3' for the NM_002764.3:c.343A>G mutation; *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 [22].

X-inactivation analysis

Skewing of X-inactivation in female carriers of the p.M115V or p.V309F mutation 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 [23]. 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 [23, 24], 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 1 μg 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 [25]. 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/).

Results

Exome sequencing identifies a novel PRPS1 mutation segregating with NSHL

We sequenced the exome of one individual (II2) from an Italian family (Family 1) affected by post-lingual bilateral profound NSHL, with a likely recessive pattern of inheritance (autosomal or X-linked). The patient is a profoundly deaf 55-years-old man with an affected brother, one normal-hearing brother and one normal-hearing child (Figure 1A); the proband's mother is moderately deaf. Both the patient and his brother had normal speech development, and received bilateral cochlear implants.

Approximately 2 Gb of polished sequence data mapping to the exome-targeted region was generated as paired-end, 90-bp reads. After discarding reads that had duplicated start sites, nearly 97% of the intended target was covered with an average depth of 58X, with more than 80% of target covered >10X (Supplementary Figure 1). To identify potentially pathogenic mutations among called variants, we focused on rare non-synonymous (NS) variants, splice-site mutations (SS), and short coding insertions or deletions (indels; I). Under a recessive model, we then looked for homozygous/hemizygous or compound heterozygous variations, and prioritized those variants located in known deafness genes (Supplementary Table 1). This analysis revealed as the most promising candidate a novel missense mutation within exon 3 of *PRPS1,* a gene already known to be associated with X-linked syndromic and nonsyndromic HL [13]. The variant was a NM_002764.3:c.337G>T nucleotide transversion, which results in the substitution of the Alanine at position 113 with a Serine (NP_002755.1:p.A113S). This mutation segregates with the NSHL phenotype in the patient's family, as confirmed by Sanger sequencing, being present in the hemizygous state in both affected males (II2 and II4) and in the heterozygous state in the I2 obligate female carrier (Figure 1). Moreover, the variant was absent in a cohort of 123 Italian normal-hearing controls.

PRPS1 mutations are a frequent cause of X-linked deafness

To better determine the incidence of *PRPS1* mutations in X-linked sensorineural deafness, we screened by Sanger sequencing all seven *PRPS1* exons, including intron-exon boundaries, in 16 additional unrelated male patients with familiarity for HL, without reported neurological symptoms, and a likely X-linked inheritance pattern.

Among these, we identified two additional novel missense variants. One was a NM_002764.3:c.343A>G transition in *PRPS1* exon 3, leading to the NP_002755.1:p.M115V amino-acid change, which was found in a 12-years-old Italian patient affected by post-lingual progressive bilateral hearing impairment. The proband has a normal-hearing brother, two affected cousins and two affected uncles (Figure 2A). He presents a down-sloping audiometric profile, with moderate HL at low and middle frequencies and a severe hearing impairment at higher frequencies. The proband's mother, obligate female carrier for the mutation, suffers from a mild form of deafness. The mutation segregates with HL in Family 2, being present in all affected individuals, and absent in all analyzed normal-hearing relatives (Figure 2A).

Finally, the third identified mutation was a NM_002764.3:c.925G>T transversion in *PRPS1* exon 7, which results in the substitution of the Valine at position 309 with a Phenylalanine (NP_002755.1:p.V309F). This variant was found in a 14-years-old proband from a Peruvian family (Family 3; Figure 2B) affected by post-lingual bilateral progressive deafness, characterized by a audiometric profile very similar to that of the Family 2 proband, with higher frequencies being more affected than low and middle ones. The mutation segregates with X-linked HL in the family, being present in an affected uncle from the proband's mother side, who has profound HL.

The p.M115V and p.V309F mutations are associated with unrecognized peripheral neuropathy

As *PRPS1* mutations may cause peripheral neuropathy, all *PRPS1*-mutated patients underwent neurological evaluation at Clinical Centre for hereditary neuropathies (Ospedale San Raffaele, Milano). Among the three families, patients in Family 2 (p.M115V) and Family 3 (p.V309F) showed clinical or subclinical signs of peripheral neuropathy. Main neurological characteristics are described in Table 1. As expected, males and females showed significant differences at neurological examination: in females we only observed subclinical signs of peripheral neuropathy (pes cavus, reduced or absent deep tendon reflexes, chronic denervation in distal muscles of lower limbs), whereas males displayed more evident findings of neuropathy. Even in males, the neuropathy was mildly symptomatic, did not cause weakness or evident motor deficits (except for III2, Family 2), whereas prevailed sensory signs and symptoms, such as Romberg positivity, absent deep tendon reflexes, paresthesias and cramps. Almost all patients, except the carrier female from family 3, showed chronic denervation at EMG evaluation, whereas only males showed mild/moderate axonal neuropathy. In Family 2, we observed reduction of both motor and sensory nerve amplitudes, whereas in family 3 only motor nerves were affected.

The newly identified PRPS1 mutations significantly reduce the enzyme activity in vivo

The open reading frame of human *PRPS1* consists of 957 nucleotides coding for 318 amino acids (GenBank accession number NM_002764). All three novel mutations affect amino-acid residues that are highly evolutionarily conserved, from zebrafish to humans (data not shown).
PRS-I is thought to be physiologically functional as a hexamer, which consists of three homodimers arranged in a propeller-like shape, each with an active site and two regulatory allosteric sites. The active site comprises binding sites for both ATP and ribose-5-phosphate, is largely located at the interface of two subunits within one homodimer (dimer interface), and extends to the interface between homodimers (trimer interface). The ATP-binding site is composed of three main structural elements, namely the flexible loop (residues Phe92-Ser108) and the PPi (pyrophosphate)-binding loop (residues Asp171-Gly174) at the dimer interface, and the flag region (residues Val30-Ile44 of an adjacent subunit) at the trimer interface; the ribose-5-phosphate binding site (residues Asp220-Thr228) face the trimer interface too [Liu *et al.*, 2010]. Allosteric site I (residues Gln135, Asp143, Asn144, Ser308-Phe313 of one subunit, Lys100-Arg104 of the second subunit within the homodimer, residue Ser47, Arg49, Ala 80 of the third subunit) is located at the trimer interface, and allosteric site II (residues Ser132, Gln135, Asn144 and

Tyr146 of one subunit and Lys100, Asp101 and –lys102 of the second subunit) is at the dimer interface [de Brouwer *et al.*, 2010]. To predict the effects of the identified missense variants on the structure/function of PRS-I, molecular models of the protein carrying the p.A113S, the p.M115V, or the p.V309F change were hence built on the crystal structure of human PRS-I.

The p.A113S mutation is located in an alpha-helix participating to the trimer interface (Figure 3). In particular, Ala113 is not contributing to the trimer interface, but is pointing towards the tightly packed hydrophobic core of the N-terminal domain of the corresponding subunit. Changing Ala113 to a more polar Serine side chain will destabilize the surrounding environment, affecting ATP-binding through the direct dislocation of the flexible-loop (Figure 3, panels A and B).

Met115 is located in the same alpha-helix, and is pointing towards the trimer interface. Two Methionine 115 side chains from two different PRS-I dimers participate in a hydrophobic interaction at the interface. The p.M115V substitution is thus predicted to disturb the packing, likely destabilizing the ATP-binding site and the allosteric site I, since both consist of amino acids of different PRS-I molecules coming together at the trimer interface (Figure 3, panels C and D).

Val309 is part of the allosteric site I and is located at the trimer interface near the center of the hexamer. The p.V309F is predicted to disturb the hexameric structure and thus the allosteric site I function as well as the hexameric assembly (Figure 3, panels E and F).

To assess the consequences of these mutations on protein function, we measured PRS-I enzymatic activity in erythrocytes from five affected males (II2 and II4 from Family 1, carrying the p.A113S substitution; III2 and III4 from Family 2 carrying the p.M115V mutation; III1 from Family 3, carrying the p.V309F change), four female carriers, three with the p.M115V substitution (II2, II4, and II6 from Family 2) and one with the p.V309F mutation (II2, Family 3), as well as one unaffected male (III1 from Family 2) and four unrelated healthy controls (CTR1-4; two males, two females). The assay is based on HPLC measurement of AMP, which is produced in equimolar amounts with PRPP. The hemolysates were incubated for 60 minutes at 37°C with saturating amounts of substrates and an inhibitor of adenylate kinase (Ap5A) to prevent conversion of AMP to ADP. Aliquots of each reaction were terminated at five different times (0, 15, 30, 45, and 60 min), and proteins removed by filtration. The amount of ATP and AMP in the eluates was evaluated by HPLC separation.

Statistically significant differences in PRS-I activity in erythrocytes of affected males, female carriers, and wild-type controls were observed (ANOVA $p < 2*10^{-5}$, at time point 60 min) (Figure 4). In particular, a significant reduction was evident in affected males, although the impairment in the enzymatic activity seemed less severe in patients carrying the p.A113S mutation than in the probands with the p.M115V or p.V309F substitutions. Indeed, PRS-I activity in patients II2 and II4 from Family 1 (p.A113S/-) was 5.29 and 7.33 nmoles/mg/hr, respectively, whereas in the Family 2 III2 and III4 patients (p.M115V/-) the enzyme activity was reduced to 2.11 and 1.06 nmoles/mg/hr and in the Family 3 proband (III1, p.V309F/-) to 0.84.

PRPS1 allelic expression is unbalanced in the female carrier of the p.V309F mutation

The p.M115V heterozygous female (II2, Family 2) suffers from a mild form of deafness whereas the p.V309F II2 carrier from Family 3 shows normal hearing (Figure 2). However, both these women showed a decreased PRS-I activity similar to that observed in patients hemizygous for the p.A113S substitution. Starting from this evidence, we investigated the possibility of an unbalanced expression of the wild-type (wt) and the mutant alleles. First, the relative expression of the two alleles was assessed by comparing the sequence of *PRPS1* mRNA at the mutant position, amplified by RT-PCR from PBMC RNA, with that obtained from genomic DNA. The results showed a preferential expression of the mutant *PRPS1* allele compared to the wild-type one in the c.343A>G carrier of Family 2 (60% mutant *vs* 40% wt allele) (Supplementary Figure 2A); conversely, an almost-exclusive expression of the wild-type *PRPS1* allele in comparison to the mutant one in the c.925G>T carrier of Family 3 was observed (85% wt *vs* 15% mutant allele) (Supplementary Figure 2B). However, while the unbalanced allelic expression observed in the individual from Family 2 is not due to a skewed inactivation of the X-chromosome, as determined by methylation assays, in the II2 carrier from Family 3 a mild skewed inactivation (≥75%) of the mutant X-chromosome is evident (Supplementary Figure 2).

PRPS1, PRPS2, PPAT, and HGPRT expression levels in PRPS1 mutation carriers

To verify whether other modifier genes may contribute to the phenotypic variability among the three analyzed families, we measured expression levels of *PRPS2*, PRPP amidotransferase (*PPAT*) and hypoxanthine-guanine phosphoribosyltransferase (*HGPRT*), all known to be involved in the purine metabolism pathway [14]. To this aim, we performed semi-quantitative real-time RT-PCR assays on RNA extracted from PBMCs of the five deaf patients (two with the p.A113S substitution, two with the p.M115V mutation and one with the p.V309F variation), four female carriers (three with the p.M115V amino-acid change and one with the p.V309F substitution), and six controls (three males, three females). 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 (Supplementary Figure 3).

Discussion

The extremely high genetic heterogeneity of sensorineural HL makes the screening of all candidate genes by traditional methods unfeasible. In the present study, WES of a single proband allowed the molecular diagnosis of the genetic defect underlying post-lingual hearing impairment in a recessive NSHL family. In particular, a novel missense variant, p.A113S, was found in the *PRPS1* gene on chromosome X, and confirmed by segregation analysis as the likely cause of deafness in this family. The subsequent mutational

screening of all *PRPS1* exons by Sanger sequencing in 16 unrelated male probands with a likely X-linked deafness, identified two additional missense variants, p.M115V and p.V309F, segregating with early-onset HL. All of the three newly identified variations were not detected in 123 audiologically-tested normal-hearing controls, as well as in an in-house database collecting all variants derived from WES of a large cohort of Italian individuals (about 3,500). These results confirm the usefulness of next-generation sequencing (NGS) for the rapid and cost-effective genetic screening of inherited sensorineural HL, which might help to increase the success rate in the molecular diagnosis of patients. Our data also highlight the recurrence of *PRPS1* mutations in X-linked deafness, suggesting that this gene should be prioritized in genetic screenings and included in gene panels for NGS-based targeted resequencing. More importantly, we show that molecular diagnosis could guide additional clinical evaluations to reveal subtle or subclinical neuropathy that might otherwise be overlooked. In particular, two of the families carrying novel variants (Families 2 and 3) were initially diagnosed as having NSHL, but following the identification of *PRPS1* mutations, they underwent neurological evaluation, and displayed signs and symptoms of peripheral neuropathy, at various severity (Table 1). In this frame, our data significantly increase the mutational spectrum of *PRPS1*-associated NSHL/CMTX5 phenotypes, since only six missense mutations, four responsible for NSHL (p.D65N, p.A87T, p.I290T, p.G306R), and two for CMTX5 (p.E43D, p.N114S), have been described in the literature so far [11, 6]. We now report, and characterize by both *in-silico* structural analysis and *in-vitro* enzymatic assays, three novel pathogenic mutations (p.A113S, p.M115V, p.V309F), thus contributing in better understanding the impact of different variants on the enzyme function, as well as genotype-phenotype correlations.

In general, molecular-modeling indicates that mutations associated with milder phenotypes are predicted to have a minor impact on PRS-I structure. Indeed, all previously reported NSHL-causing mutations are predicted to either disturb the local stability of PRS-I or moderately affect interactions at the trimer interface; conversely, CMTX5-causing mutations are likely to affect both the ATP-binding pocket and the allosteric site I [13]. The here-reported missense variants are expected to exert a mild destabilization of the enzyme structure, although the p.M115V and p.V309F mutations are predicted to have a more severe effect than p.A113S. *In-vitro* enzymatic assays, performed on erythrocytes of compliant patients and controls, substantially confirmed these *in-silico* predictions. All analyzed mutations resulted in a marked reduction (>60%) of PRS-I activity in the patients, p.M115V and p.V309F leading to a much more severe decrease (to about 6% of normal enzyme activity) than p.A113S (about 30% residual activity). These results well correlate with clinical data, with p.A113S being responsible of the sole NSHL, and the p.M115V and p.V309F mutations resulting in the presence of peripheral neuropathy (Table 2). It is interesting to note that the p.M115V mutation found in Family 2 involves the same residue that was found mutated (p.M115T) in a previously described CMTX5 kindred [11], suggesting that Met115 might be particularly critical for enzyme function (Supplementary Figure 4).

Inter- and intra-familial phenotypic variability of disease expression has been reported for Arts syndrome and PRS-I superactivity [14]. Here we show a wide, and almost continuous, spectrum of neurological manifestations also in CMTX5, which might range from a subclinical peripheral neuropathy, with prevalent axonal motor neuropathy (Family 3), to an axonal sensory-motor neuropathy (Family 2) (Table

1). Attenuation/exacerbation of the clinical manifestations may be due to several biological factors, such as: i) functional redundancy and partial compensation by other PRS isoforms (especially *PRPS2*, which is widely expressed): ii) differences in the expression levels of the rate-limiting enzyme, PRPP amidotransferase (PPAT); iii) presence of additional modifier *loci*; or, in females, iv) skewed X-chromosome inactivation. Indeed, the partial skewed inactivation of the mutant X-chromosome could be one of the reasons underlying the phenotypic variability observed in HL in female carriers of the here-identified mutations. In fact, the p.M115V heterozygous female (II2, Family 2) was mildly deaf and did not show skewed inactivation of the mutant X-chromosome, whereas the normal-hearing p.V309F II2 carrier from Family 3 showed a mild skewed inactivation of the mutant *PRPS1* allele (Supplementary Figure 2). We also attempted to verify whether genes involved in the purine metabolism pathway could contribute to the phenotypic variability among the three analyzed families by a compensatory-based mechanism. To this aim, we measured *PRPS2, PPAT,* and *HGPRT* expression levels in heterozygous female carriers of the identified mutations, in affected males, and in healthy controls. We observed increased expression only for the *HGPRT* transcript in heterozygous female carriers in comparison to healthy controls (almost 2-fold increase, p=0.01; Supplementary Figure 3). These data should however be considered with caution, since *HGPRT* itself is a X-linked gene and the number of available heterozygous female carriers is probably not sufficient to reach a robust statistical significance.

Finally, the description and characterization of novel *PRPS1* mutations has important consequences not only for diagnosis, but also for therapeutic and prevention purposes. Indeed, despite the central role of PRS-I enzyme in *de-novo* purine synthesis, purine nucleotides can be produced also by an alternative pathway, which uses S-adenosylmethionine (SAM). This alternative route involves the conversion of SAM into S-adenosylhomocysteine, which in turn is hydrolysed into adenosine and l-homocysteine by *S-*adenosylhomocysteine hydrolase. Purine nucleotide recycling will eventually convert the SAM-derived adenosine nucleotides into guanosine nucleotides [13]. Importantly, SAM appears to be able to cross the intestinal wall, the blood-brain barrier, and possibly the blood-labyrinth barrier, and has already been used as treatment for many different pathologies, including depression, neurological deficits, liver disease and, more recently, Arts syndrome [26, 14]. Dietary supplementation with SAM in Arts syndrome patients, in particular, alleviated their symptoms and stabilized the progression of hearing impairment [13]. Hence, SAM supplementation in the diet might represent a promising therapy to postpone or slow down the onset and progression of neurological and audiological symptoms in patients with *PRPS1* mutations as well as mildly affected carrier females.

Conflict of interest

All authors declare to have no conflict of interest.

References

1 Raviv D, Dror AA, Avraham KB. Hearing loss: a common disorder caused by many rare alleles. *Ann N Y Acad Sci* 2010;1214:168-79.

2 Walsh T, Shahin H, Elkan-Miller T, Lee MK, Thornton AM, Roeb W, Abu Rayyan A, Loulus S, Avraham KB, King MC, Kanaan M. Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *Am J Hum Genet* 2010;87:90-4.

3 Schraders M, Haas SA, Weegerink NJ, Oostrik J, Hu H, Hoefsloot LH, Kannan S, Huygen PL, Pennings RJ, Admiraal RJ, Kalscheuer VM, Kunst HP, Kremer H. Next-generation sequencing identifies mutations of SMPX, which encodes the small muscle protein, X-linked, as a cause of progressive hearing impairment. *Am J Hum Genet* 2011;88:628-34.

4 Huebner AK, Gandia M, Frommolt P, Maak A, Wicklein EM, Thiele H, Altmüller J, Wagner F, Viñuela A, Aguirre LA, Moreno F, Maier H, Rau I, Giesselmann S, Nürnberg G, Gal A, Nürnberg P, Hübner CA, del Castillo I, Kurth I. Nonsense mutations in SMPX, encoding a protein responsive to physical force, result in X-chromosomal hearing loss. *Am J Hum Genet* 2011;88:621-7.

5 de Kok YJ, van der Maarel SM, Bitner-Glindzicz M, Huber I, Monaco AP, Malcolm S, Pembrey ME, Ropers HH, Cremers FP. Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4. *Science* 1995;267:685-8.

6 Liu X, Han D, Li J, Han B, Ouyang X, Cheng J, Li X, Jin Z, Wang Y,Bitner-Glindzicz M, Kong X, Xu H, Kantardzhieva A, Eavey RD, Seidman CE, Seidman JG, Du LL, Chen ZY, Dai P, Teng M, Yan D, Yuan H. Loss-of-function mutations inthe PRPS1 gene cause a type of nonsyndromic X-linked sensorineural deafness,DFN2. *Am J Hum Genet* 2010;86:65-71.

7 Becker MA. Phosphoribosylpyrophosphate synthetase and the regulation of phosphoribosylpyrophosphate production in human cells. *Prog Nucleic Acid Res Mol Biol* 2001;69:115-48.

8 Eriksen TA, Kadziola A, Bentsen AK, Harlow KW, Larsen S. Structural basis for the function of Bacillus subtilis phosphoribosyl-pyrophosphate synthetase. *Nat Struct Biol* 2000;7:303-8.

9 Li S, Lu Y, Peng B, Ding J. Crystal structure of human phosphoribosylpyrophosphate synthetase 1 reveals a novel allosteric site. *Biochem J* 2007;401:39-47.

10 Sperling O, Eilam G, Sara-Persky-Brosh, De Vries A. Accelerated erythrocyte 5-phosphoribosyl-1-pyrophosphate synthesis. A familial abnormality associated with excessive uric acid production and gout. *Biochem Med* 1972;6:310-6.

11 Kim HJ, Sohn KM, Shy ME, Krajewski KM, Hwang M, Park JH, Jang SY, Won HH, Choi BO, Hong SH, Kim BJ, Suh YL, Ki CS, Lee SY, Kim SH, Kim JW. Mutations in PRPS1, which encodes the phosphoribosyl pyrophosphate synthetase enzyme critical for nucleotide biosynthesis, cause hereditary peripheral neuropathy with hearing loss and optic neuropathy (cmtx5). *Am J Hum Genet* 2007;81:552-8.

12 de Brouwer AP, Williams KL, Duley JA, van Kuilenburg AB, Nabuurs SB, Egmont-Petersen M, Lugtenberg D, Zoetekouw L, Banning MJ, Roeffen M, Hamel BC, Weaving L, Ouvrier RA, Donald JA, Wevers RA, Christodoulou J, van Bokhoven H. Arts syndrome is caused by loss-of-function mutations in PRPS1. *Am J Hum Genet* 2007;81:507-18.

13 de Brouwer AP, van Bokhoven H, Nabuurs SB, Arts WF, Christodoulou J, Duley J. PRPS1 mutations: four distinct syndromes and potential treatment. *Am J Hum Genet* 2010;86:506-18.

14 Liu XZ, Xie D, Yuan HJ, de Brouwer AP, Christodoulou J, Yan D. Hearing loss and PRPS1 mutations: Wide spectrum of phenotypes and potential therapy. *Int J Audiol* 2013;52:23-8.

15 Primignani P, Trotta L, Castorina P, Lalatta F, Sironi F, Radaelli C, Degiorgio D, Curcio C, Travi M, Ambrosetti U, Cesarani A, Garavelli L, Formigoni P, Milani D, Murri A, Cuda D, Coviello DA. Analysis of the GJB2 and GJB6 genes in Italian patients with nonsyndromic hearing loss: frequencies, novel mutations, genotypes, and degree of hearing loss. *Genet Test Mol Biomarkers* 2009;13:209-17.

16 Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;25:1966-7.

17 Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res* 2005;33:W382-8.

18 Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 2002;320:369-87.

19 Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. *Proteins* 2002;47:393-402.

20 Torres RJ, Mateos FA, Puig JG, Becker MA. A simplified method for the determination of phosphoribosylpyrophosphate synthetase activity in hemolysates. *Clin Chim Acta* 1994;224:55-63.

21 Balasubramaniam P, Malathi A. Comparative study of hemoglobin estimated by Drabkin's and Sahli's methods. *J Postgrad Med* 1992;38:8-9.

22 Castaman G, Platè M, Giacomelli SH, Rodeghiero F, Duga S. Alterations of mRNA processing and stability as a pathogenic mechanism in von Willebrand factor quantitative deficiencies. *J Thromb Haemost* 2010;8:2736-42.

23 Coene KL, Roepman R, Doherty D, Afroze B, Kroes HY, Letteboer SJ, Ngu LH, Budny B, van Wijk E, Gorden NT, Azhimi M, Thauvin-Robinet C, Veltman JA, Boink M, Kleefstra T, Cremers FP, van

Bokhoven H, de Brouwer AP. OFD1 is mutated in X-linked Joubert syndrome and interacts with LCA5-encoded lebercilin. *Am J Hum Genet* 2009;85:465-81.

24 Bodega B, Bione S, Dalprà L, Toniolo D, Ornaghi F, Vegetti W, Ginelli E, Marozzi A. Influence of intermediate and uninterrupted FMR1 CGG expansions in premature ovarian failure manifestation. *Hum Reprod* 2006;21:952-7.

25 Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;3:RESEARCH0034.

26 Bottiglieri T. S-Adenosyl-L-methionine (SAMe): from the bench to the bedside--molecular basis of a pleiotrophic molecule. *Am J Clin Nutr* 2002;76:1151S-7S.
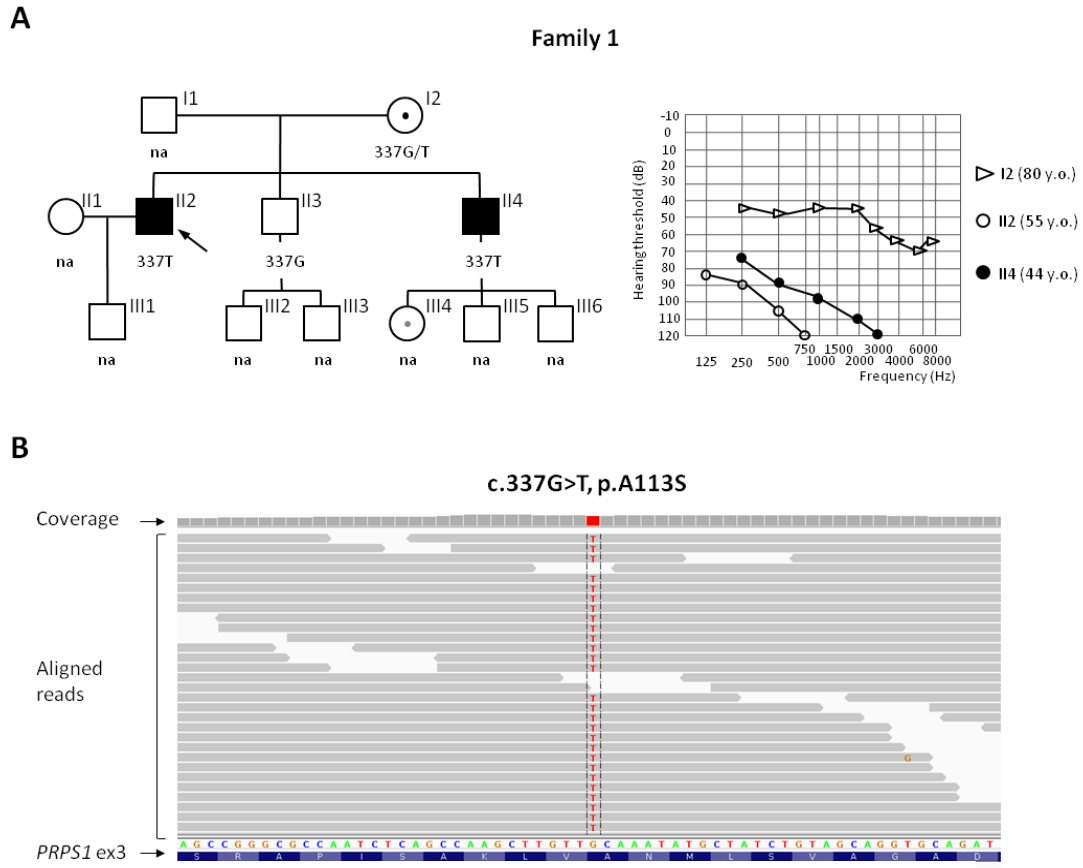
Figure 1. Identification of the novel p.A113S *PRPS1* mutation, segregating with NSHL, by whole-exome sequencing.

A) Pedigree of Family 1 and audiograms of both affected males as well as one carrier female (average HL for the right and left ears are shown). This family, which showed a clear recessive inheritance pattern of the disease, was selected for targeted capture and exome sequencing (the analyzed subject is pointed by an arrow). The genotype of each individual is indicated below the corresponding symbols. na: not analyzed. B) IGV (Integrative Genome Viewer) screenshot showing sequencing reads that support the mutant allele (the T nucleotide is reported) identified in the II2 proband. The identified G>T transversion maps within *PRPS1* exon 3 and results in the A113S amino-acid change. The mean coverage of the region here shown is 31.
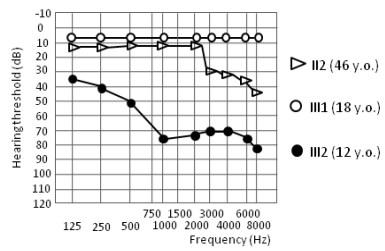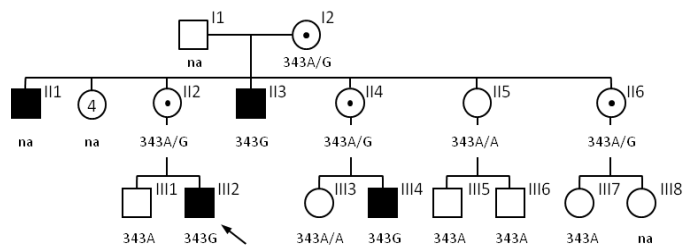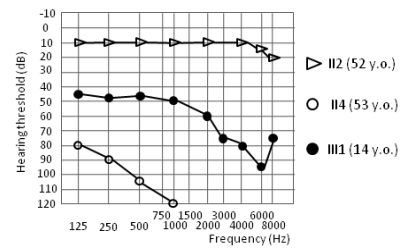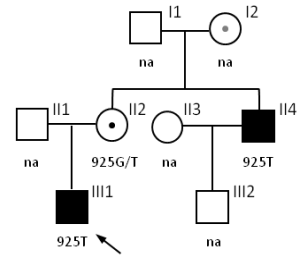
Figure 2. Identification of novel missense mutations (p.M115V, p.V309F) in *PRPS1* in two additional families with X-linked deafness.

Pedigree of Families 2 (A) and 3 (B) and corresponding audiograms, showing the average HL for the right and left ears. The genotype of each individual is indicated below the corresponding symbols. The probands are pointed by an arrow. na: not analyzed.
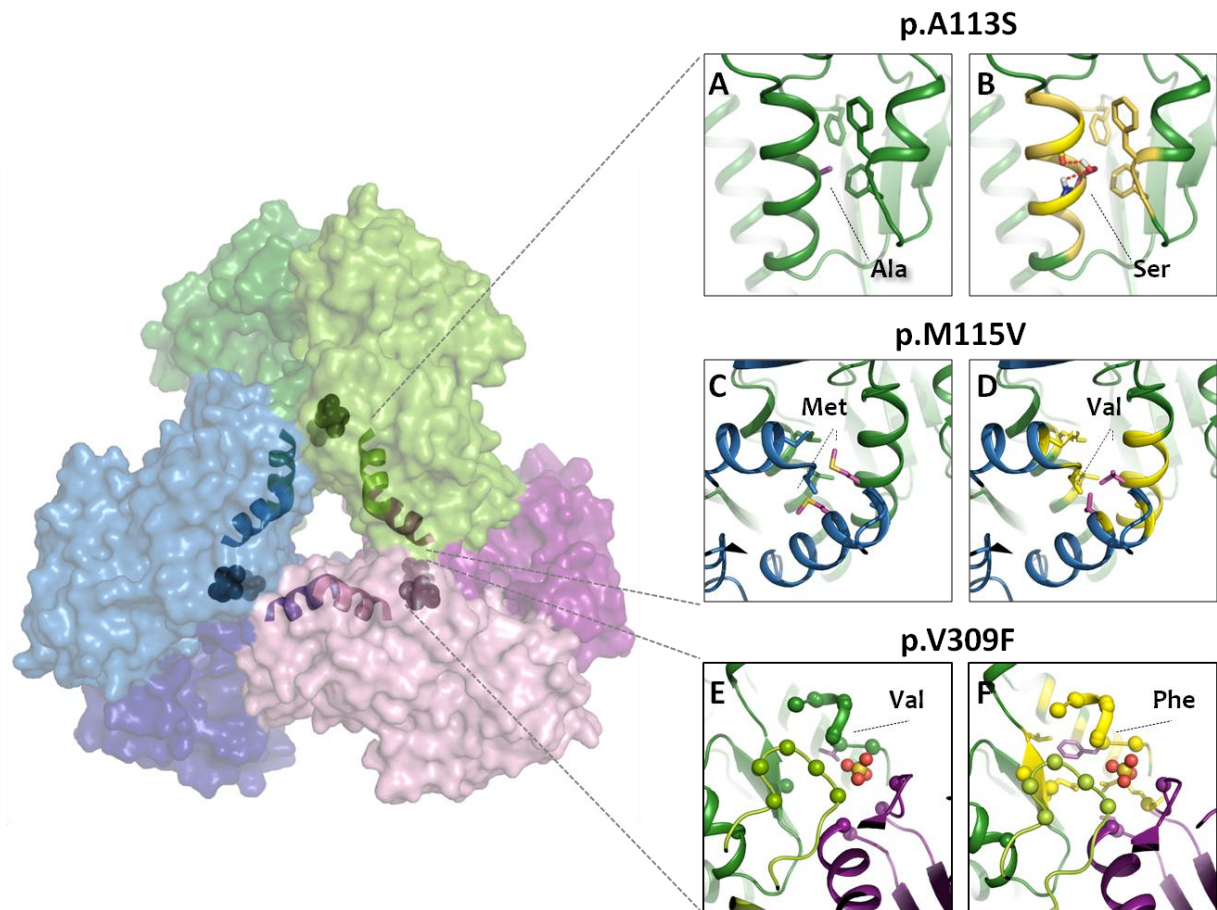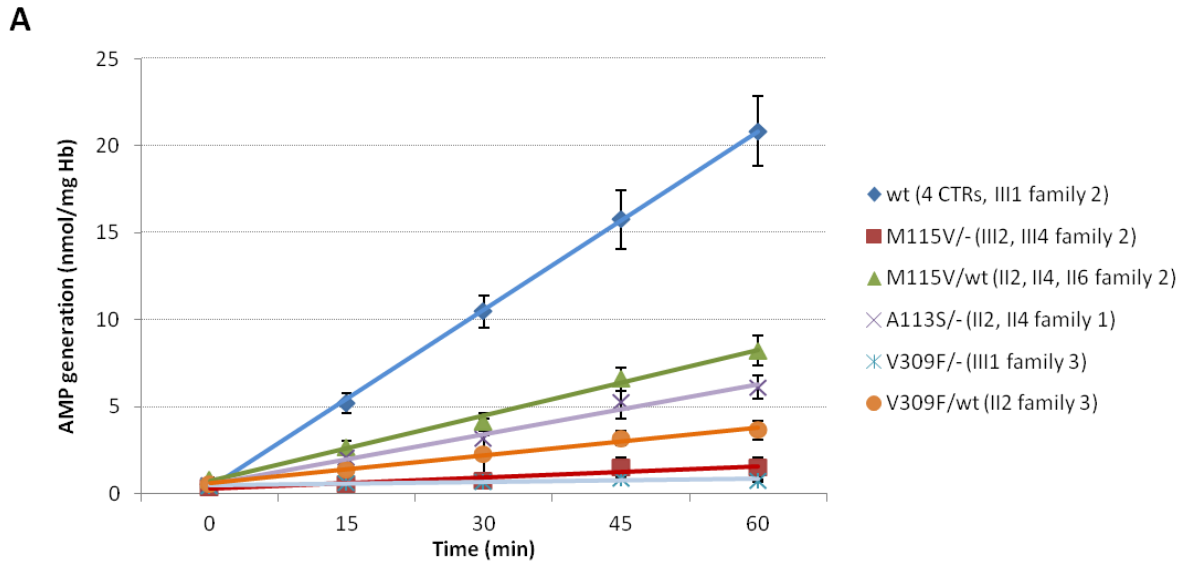
Figure 3. Structural analysis of the newly identified PRS-I mutations.
On the left: Molecular surface of the three-dimensional structure of the PRS-I enzyme in its physiologic assembly, with the six different subunits indicated by separated colors. The 2 monomers forming a dimer are highlighted with different shades of the same color. The alpha-helices hosting p.A113S and p.M115V mutations are shown as ribbons. Residue Val309 is shown by spheres. A-B) Structural analysis of the p.A113S mutation. Left: Ala113 is completely buried in the hydrophobic core of the corresponding subunit, interacting with Phe92 (belonging to the flexible-loop), Phe137, and Phe138. Right: Ser113 differs from Alanine in that one of the methylenic hydrogens is replaced by a hydroxyl group, through which this amino acid makes a new hydrogen bond. The larger and polar (hydrophilic) side chain of Ser113 makes this region to rearrange. C-D) Structural analysis of the p.M115V mutation. Left: Met115 is completely buried at the trimer interface, making hydrophobic interactions with the surrounding amino acids. Right: Val115 is a beta-branched residue, introducing a completely different steric hindrance, making this region to rearrange. E-F) Structural analysis of the p.V309F mutation. Left: Val309 is part of the allosteric site I and is located at the trimer interface near the center of the hexamer. Right: Phe309 is bulkier than Val and surrounding residues are predicted to move to avoid Van der Waals crashes. Residues of allosteric site I are indicated with spheres. A SO_4^{2-} ion present in the crystal structure is also shown. In all panels, the mutated residue is shown by stick: C atoms are colored magenta, O red, S yellow and H white. Residues predicted by FoldX to be perturbed are colored in yellow.

A

B

PRS-I Activity Assay in Erythrocytes from Family 1, 2 and 3					
Individual	Sex	Genotype	Age at testing	Activity (±SD) (nmoli/mg/hr)	Percentage
Affected Males					
II2 family 1	M	A113S/-	54	5.29 (±1.38)	25.4
II4 family 1	M	A113S/-	43	7.33 (±0.37)	35.2
III2 family 2	M	M115V/-	12	2.11 (±1.32)	10.2
III4 family 2	M	M115V/-	9	1.06 (±0.19)	5.1
II1 family 3	M	V309F/-	14	0.84 (±0.1)	4.05
Female Carrier					
II2 family 2	F	M115V/wt	46	7.52 (±1.62)	36.2
II4 family 2	F	M115V/wt	49	7.22 (±0.11)	34.7
II6 family 2	F	M115V/wt	60	9.93 (±0.46)	47.7
II2 family 3	F	V309F/wt	52	3.8 (±0.64)	18.3
Unaffected Male					
III1 family 2	M	wt/-	18	22.10 (±0.12)	106.2
Unrelated Controls					
CTR-1	M	wt/-	46	15.43 (±3.01)	74.2
CTR-2	F	wt/wt	44	17.02 (±4.26)	81.8
CTR-3	F	wt/wt	29	23.03 (±3.96)	110.7
CTR-4	M	wt/-	23	26.44 (±4.94)	127.1

Figure 4. Functional characterization of the identified *PRPS1* mutations.
A) PRS-I activity in erythrocytes was evaluated by measuring the accumulation of AMP by HPLC in five affected males, four female carriers, as well as five healthy control individuals (wt). The diagram shows the correlation between ribose-5-phosphate-dependent AMP generation (y-axis) and the reaction incubation time (x-axis). Error bars represent the standard error of the mean (SEM) for genotypes represented by >1 individual (wt, n=5; M115V/-, n=2; M115V/wt, n=3; A113S/-, n=2), whereas they indicate the standard deviation (SD) for genotypes represented by a single individual. B) Summary of the functional data obtained from PRPP synthetase activity assay. The enzyme activity is expressed as nmoles of AMP per mg of Hb per hr. SD was calculated from three independent experiments performed in triplicate. The percentage was calculated as the ratio between the individual PRS-I activity and the mean of the PRS-I activity of the unaffected male and wt controls.

Table 1. Clinical and electrophysiological features from nerve conduction studies

Patient		Age (y) /sex	Onset (y)	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 / M		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

Legend: DTR, deep tendon reflexes; EMG, needle electromyography; Pt, patient; 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 Materials

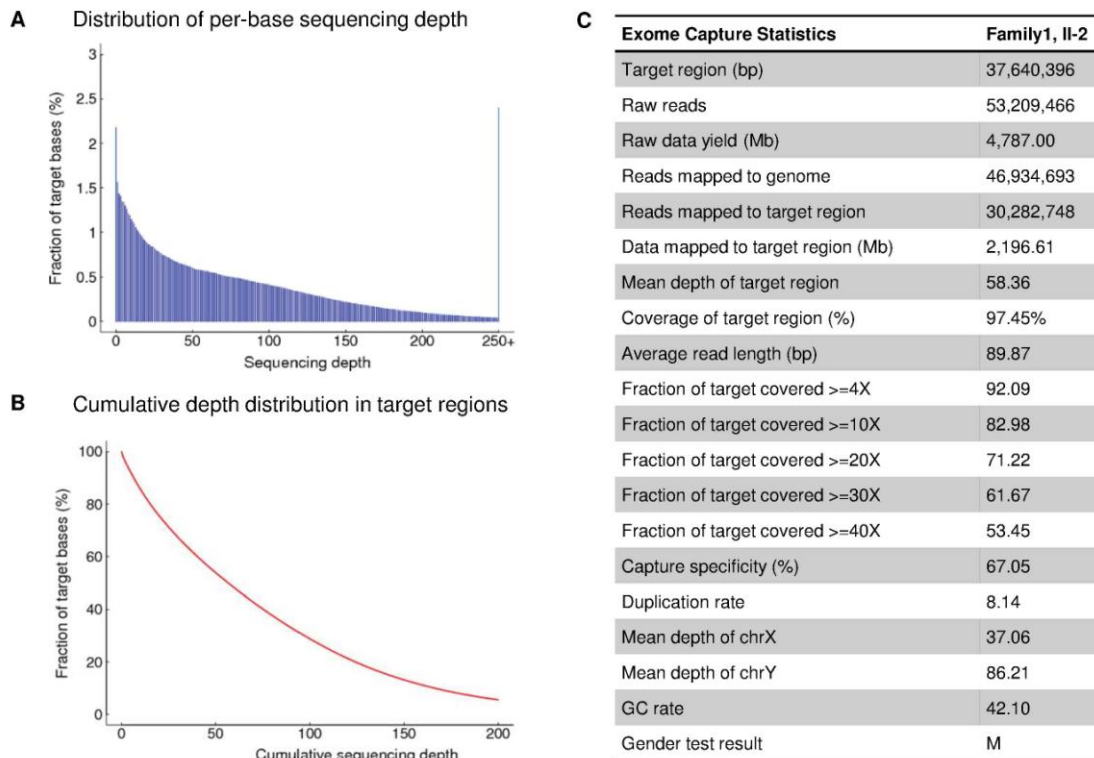**Supplementary Figure 1: Overview of exome data production.**

Supplementary Table 1: Identification of candidate pathogenic variants.

Supplementary Figure 2: Allelic-specific expression in p.M115V and p.V309F carriers.

Supplementary Figure 3: *PRPS1*, *PRPS2*, *PPAT*, and *HGPRT* expression levels by qPCR.

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

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. Identification of candidate pathogenic variants.

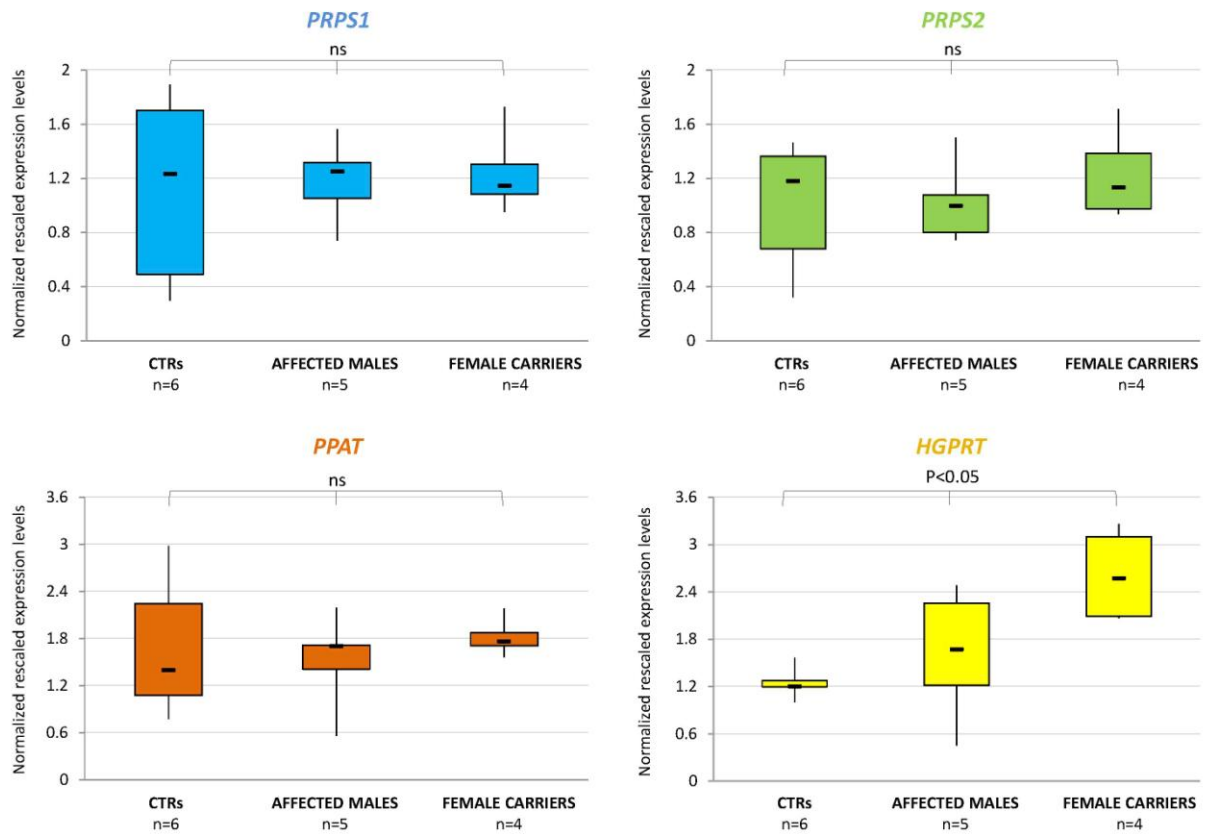| Filter type	Nr of variants (genes)
NS/SS/I	6,855 (4,289)
Not in dbSNP130/1000Genomes	638 (551)
Not in ESP5400 exomes	281 (238)
Hom/compound Het	50 (26)
Within known NSHL genes	**1 (1)**

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/).

A c.343A>G (p.M115V) **B** c.925G>T (p.V309F)
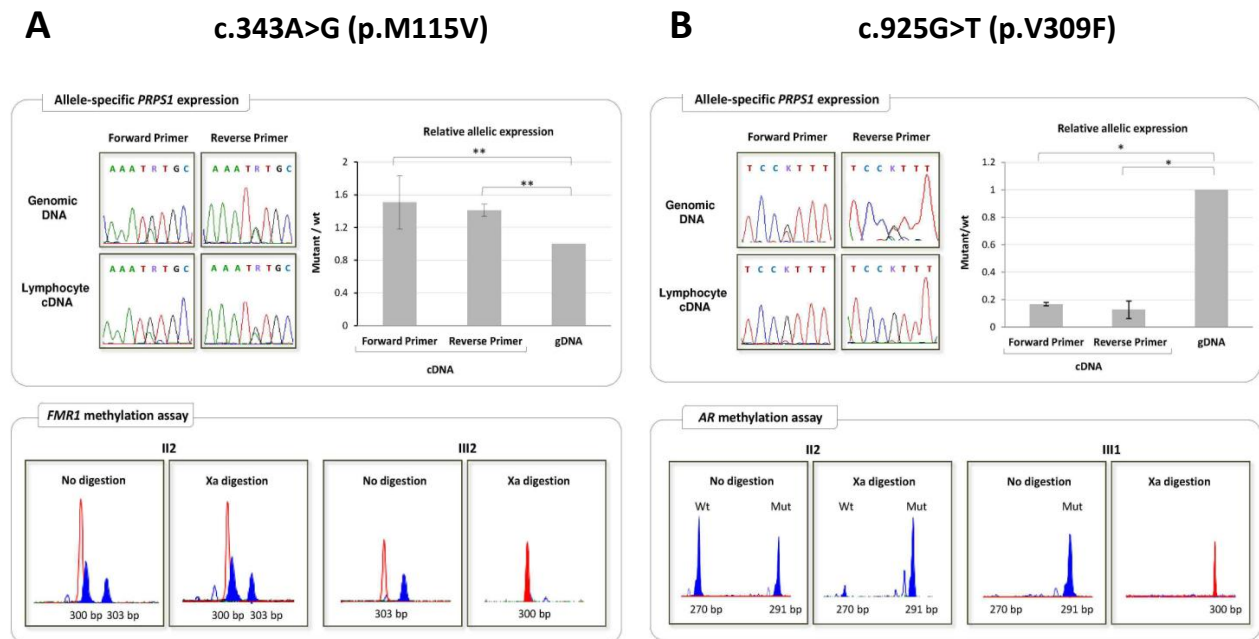
Supplementary Figure 2. *PRPS1* **allelic-specific expression in c.343A>G (p.M115V) and c.925G>T (p.V309F) carriers.**
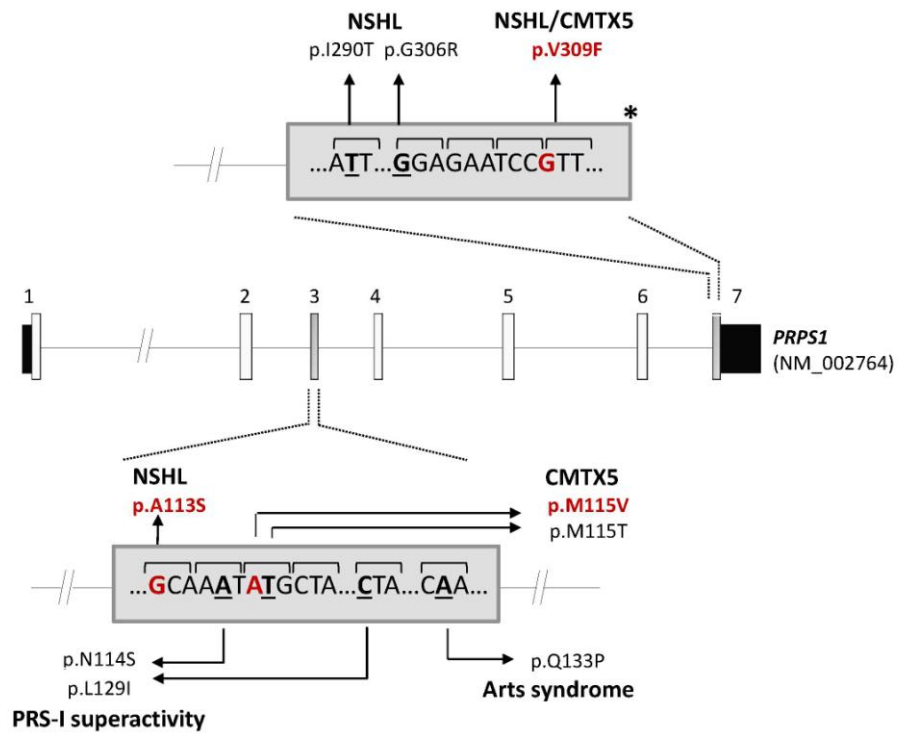
Upper panels. On the left: DNA sequence electropherograms showing the region surrounding the variant position, PCR-amplified from genomic DNA or from lymphocyte cDNA of either the heterozygous individual carrying the c.343A>G mutation (A) or of the female carrier of the c.925G>T transversion (B). On 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. Peak Scan 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. *PRPS1, PRPS2, PPAT,* and *HGPRT* expression levels by qPCR.
PRPS1, PRPS2, PPAT, and *HGPRT* expression levels were measured 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 mutation), four female heterozygotes (three for the p.M115V, one for the p.V309F variant), and six wild-type controls (CTRs). 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.

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 mutations (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.