# POSITIONAL CLONING OF
# THE USHER SYNDROME TYPE 3
# GENE *(USH3)*


Tarja Joensuu

Department of Medical Genetics, Biomedicum Helsinki
University of Helsinki

Department of Molecular Genetics
The Folkhälsan Institute of Genetics

Department of Ophthalmology
Helsinki University Central Hospital
Finland

Academic dissertation

HELSINKI 2002

**Supervised by**

Professor Albert de la Chapelle, MD, PhD
Comprehensive Cancer Center
The Ohio State University
Columbus, Ohio, USA
and
Department of Molecular Genetics
The Folkhälsan Institute of Genetics
Helsinki, Finland

Eeva-Marja Sankila, MD, PhD
Department of Ophthalmology
Helsinki University Central Hospital
and
Department of Medical Genetics, Biomedicum Helsinki
University of Helsinki/
Department of Molecular Genetics
The Folkhälsan Institute of Genetics
Helsinki, Finland

**Reviewed by**

Docent Mirja Somer, MD, PhD
Department of Clinical Genetics
Helsinki University Central Hospital
Helsinki, Finland

Docent Anu Wartiovaara, MD, PhD
University of Helsinki
Department of Neurology and Programme of Neurosciences
Helsinki, Finland

**Official opponent**

Docent Kalle Simola, MD, PhD
Department of Clinical Genetics
Centre for Laboratory Medicine
Tampere University Hospital
Tampere, Finland

*To all in the family*

# TABLE OF CONTENTS

# LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to by their Roman numerals in the text. In addition, some unpublished data are presented.

I      Joensuu T, Blanco G, Pakarinen L, Sistonen P, Kääriäinen H, Brown S, de la Chapelle A, Sankila E-M. Refined mapping of the Usher syndrome type III locus on chromosome 3, exclusion of candidate genes, and identification of the putative mouse homologous region. *Genomics* 38:255-263. (1996).

II    Joensuu T, Hämäläinen R, Lehesjoki A-E, de la Chapelle A, Sankila E-M. A sequence-ready map of the Usher syndrome type III critical region on chromosome 3q. *Genomics* 63:409-416. (2000).

III   Joensuu T*, Hämäläinen R*, Yuan Bo, Johnson C, Tegelberg S, Gasparini P, Zelante L, Pirvola P, Pakarinen L, Lehesjoki A-E, de la Chapelle A, Sankila E-M. Mutations in a novel gene with transmembrane domains underlie Usher syndrome type 3. *Am J Hum Genet* 69:673-684. (2001).

*These authors contributed equally to this work.

# ABBREVIATIONS

| | | | |
|---|---|---|---|
| **BAC** | bacterial artificial chromosome | **PFGE** | pulsed-field gel electrophoresis |
| **bp** | base pair | **RACE** | rapid amplification of cDNA ends |
| **cDNA** | complementary DNA | **RFLP** | restriction fragment length |
| *CEPH* | Centre d'Etude du Polymorphisme | | polymorphism |
| | Humain | **RNA** | ribonucleic acid |
| **cM** | centiMorgan | **RP** | retinitis pigmentosa |
| **CHM** | choroideremia | **RPE** | retinal pigment epithelium |
| **dbEST** | database of expressed sequence tags | **RS** | X-chromosomal retinoschisis |
| **DGGE** | denaturing gradient gel | **RT-PCR** | reverse transcriptase polymerase |
| | electrophoresis | | chain reaction |
| **DNA** | deoxyribonucleic acid | *sina* | seven in absentia |
| **ERG** | electroretinography | *SIAH* | Homo sapiens seven in absentia |
| **EST** | expressed sequence tag | | (*Drosophila*) homolog |
| **HA** | heteroduplex analysis | **SNP** | single nucleotide polymorphism |
| **HGP** | Human Genome Project | **SSCP** | single strand conformation |
| **HL** | hearing loss | | polymorphism |
| *HOPA* | human OPA-containing gene | **STR** | short tandem repeat |
| **HPLC** | high-performance liquid | **STS** | sequence tagged site |
| | chromatography | θ | theta, recombination fraction |
| **kb** | kilobase | **TR** | thyroid hormone receptor |
| **kD** | kiloDalton | **TRAP** | human thyroid hormone receptor- |
| **LD** | linkage disequilibrium | | associated protein complex |
| **LOD** | logarithm of odds | *UCRP* | Usher critical region pseudogene |
| **Mb** | megabase pair | **UTR** | untranslated region |
| **mRNA** | messenger RNA | **USH** | Usher syndrome |
| *NOPAR* | No OPA-repeat gene | **USH1** | Usher syndrome type 1 |
| **OPA** | opposite paired | **USH2** | Usher syndrome type 2 |
| **ORF** | open reading frame | **USH3** | Usher syndrome type 3 |
| **PAC** | P1-artificial chromosome | **VDA** | variant detector array |
| **PAGE** | polyacrylamide gel electrophoresis | **VNTR** | variable number of tandem repeats |
| **PCR** | polymerase chain reaction | **YAC** | yeast artificial chromosome |

**SUMMARY**

Usher syndrome type 3 (USH3) is an autosomal recessive disorder characterized by progressive hearing loss, severe retinal degeneration, and variably present vestibular dysfunction. USH3 comprises about 40% of the Finnish USH population, and thereby it belongs to the Finnish disease heritage (Pakarinen et al., 1995a). After the preliminary localization of *USH3* to a 5-cM interval on chromosome 3q21-q25 by Sankila et al. in 1995, the molecular analysis, by means of genetic and physical mapping, was initiated. Yeast and bacterial artificial chromosome (YAC and BAC) clones covering the region of interest were placed in contiguous series, and the overlaps between different clones were established on the basis of shared sequence tagged sites (STSs) and expressed sequence tags (ESTs) using primers for amplification. Based on the physical map, the positional candidate genes *PFN2*, encoding human profilin-2, and KIAA0001, were assigned to the critical *USH3* interval. By use of an orthologous EST sequence, a putative mouse homologue, *Pfn2*, was mapped on mouse chromosome 3 suggesting localization for the mouse homologue of *USH3*. The identification of novel polymorphic markers, and the analyses of linkage disequilibrium and haplotypes in 32 USH3 families, eventually led to the refined localization of *USH3* to a 250-kb genomic interval. Once the *USH3* region was cloned in manageable ordered pieces, it was possible to determine its complete DNA sequence.

The 149-kb and 207-kb genomic sequence contigs across the *USH3* region were constructed by shotgun cloning and sequencing, followed by contig assembling, of three BAC clones. Later, as the haplotypes in two USH3 chromosomes indicated a novel centromeric border at marker 107G19CA7, the sequence analysis was focused on the 207-kb genomic contig defined by the new marker. The subsequent computational sequence annotation revealed several human and orthologous mouse and rat ESTs, as well as two previously identified protein encoding genes, KIAA0001 and H963. In addition, two novel genes, *NOPAR* and its splice variant *NOPAR2*, as well as an unprocessed pseudogene, *UCRP*, were constructed by RT-PCR and RACE experiments. All these genes were examined in patients and excluded as *USH3*.

As attempts to find additional transcripts in the 207-kb region did not reveal new information, the analysis was extended towards the next centromeric marker 25B8CA2. A retina-derived EST was identified, and a novel gene encoding a predicted 120 amino acid transmembrane protein with no homology to known proteins was constructed from the extended

133-kb *USH3* region. Three disease causing mutations were identified in the coding region of the *USH3* gene in affected individuals. The major Finnish founder mutation is a Y100X nonsense mutation, which is present in 94% of the USH3 patients analyzed. A second mutation, M44K, was identified in two Finnish families, in which four patients were compound heterozygous for Y100X and a missense mutation M44K. Finally, a homozygous 3-bp deletion that results in the amino acid substitution of one methionine for isoleucine and leucine was detected in an Italian family. One carrier of the Y100X mutation was identified among 200 Finnish unrelated controls.

The identification of the *USH3* gene and the study of the mutated gene product will provide novel insight into the understanding of the biological basis of deafness and blindness. It remains to be elucidated by *in vitro* and *in vivo* functional analyses, how the mutations cause the defects of the sensory cells in the inner ear and the retina. At present, the use of mutation detection will be useful for earlier diagnosis of the disease. Genetic testing will enable presymptomatic and prenatal diagnosis in those families who wish it. Moreover, carrier diagnosis can be offered for individuals whose family members are known to be associated with USH3. It will also permit a more accurate genetic counseling to be offered to families. Eventually, it may allow the development of new specific therapies.

# INTRODUCTION

A remarkable progress has been made in identification of deafness genes during the past years. To date, over 400 syndromes including deafness have been listed, and 77 loci for nonsyndromic deafness have been reported (On line Mendelian Inheritance of Man; Hereditary Hearing Loss Homepage). Already over 90 different mouse genes that are known to affect the inner ear have been identified, and around 20 mouse mutants with inner ear defects provide models for human inherited disorders including hearing loss (Steel and Kros, 2001). Many of the human genes for dominant and recessive forms of nonsyndromic and syndromic hearing loss encode proteins, which localize in specialized regions and cell types of the cochlea, and are involved for example in ion homeostasis, in the development of the auditory and vestibular systems, and in transcriptional events (Gillespie and Walker, 2001; Steel and Kros, 2001).

Usher syndrome (USH) is a recessively inherited disorder, which combines simultaneously both hearing impairment and progressive loss of vision due to retinitis pigmentosa (RP). Based on the clinical features, the syndrome has been divided into three subtypes: USH1, USH2, and USH3 (Davenport and Omenn, 1977). Type 1 (USH1) is characterized by a congenital severe to profound and preverbal deafness, an absent vestibular function, and an early onset of retinal deterioration. Type 2 (USH2) has a milder congenital hearing loss, which deteriorates very slowly, and a later onset of retinal degeneration. Balance functions are normal. The existence of a third distinct clinical type of USH, USH3, with a progressive postverbal hearing loss and variable vestibular function, was not generally accepted (Smith et al., 1994), until Sankila et al. (1995) assigned the *USH3* locus on chromosome 3q21-q25 by linkage. In addition, in the study by Pakarinen et al. (1995a), the USH3 phenotype was shown to be the major USH type in Finland comprising about 40% of USH cases, while it was thought to be the rarest clinical type in other countries (Davenport and Omenn, 1977).

Among USH, at least ten different loci have been detected, and six genes have been identified, for USH1B (Weil et al., 1995), for USH1C (Bitner-Glindzicz et al., 2000; Verpy et al., 2000), for USH1D (Bolz et al., 2001; Bork et al., 2001), for USH1F (Ahmed et al., 2001; Alagramam et al., 2001b), for USH2A (Eudy et al., 1998), and for USH3, which we identified in this study by positional cloning methods. At least three USH genes, *USH1B*, *USH1D*, and

*USH2A*, underlie other isolated auditory or visual sensory defects as well (Bork et al., 2001; Liu et al., 1997b; Liu et al., 1997d; Rivolta et al., 2000).

The mutated *USH* genes cause defects, especially in the organ of Corti, where severe degeneration is observed. Although the specific expression patterns for some of the USH proteins in the cochlea and the retina are known, their exact molecular pathways in these organs remains to be elucidated. In this regard, mouse mutants have given important information in the study of molecular pathogenesis of the disease. Similarities between hereditary deafness in humans and mice suggest that mice are important animal models for human deafness. An example is the characterization of the *myosin VIIA* gene (*MYO7A*), which is involved in mouse *shaker-1* (*sh-1*) deafness and in human USH1B (Gibson et al., 1995; Weil et al., 1995). Based on the studies with the mutant *sh-1* mouse, the actin-mediated organization and function of hair cells and photoreceptors, such as assembly of the stereocilia bundles, endocytosis and membrane recycling, protein transport, neurotransmitter release, and pigment granule distribution in the retinal pigment epithelium (RPE), have been suggested to be affected by a defective MYO7A protein (Hasson, 1999; Hasson et al., 1997b; Liu et al., 1998a; Liu et al., 1999a; Richardson et al., 1997; Self et al., 1998).

# REVIEW OF THE LITERATURE

## 1. The Finnish disease heritage

The Finnish disease heritage is a well-known concept, and was first time published by Norio, Nevanlinna and Perheentupa in 1973. There is evidence of habitation already some 10,000 years ago in Finland, although the present-day population is thought to have expanded from a small number of founder settlers of Baltic Finnic and Germanic origin, who immigrated from the south over the Gulf of Finland some 2,000-2,500 years ago, and populated all the coastal areas of the southern and the western parts of Finland, also called the early settlement region (Nevanlinna, 1972; Norio, 1973). In the 12th century, the number of inhabitants was approximately 50,000, and as late as the 15 th century, the number of total population reached 250,000 (de la Chapelle, 1993; de la Chapelle and Wright, 1998; Nevanlinna, 1972; Peltonen et al., 1999). Although genetic bottlenecks (a remarkable decrease in the size of population followed by growth of the original population), such as famine, war, and diseases, prevented the rapid growth of the population, by the end of the 16th century the population had spread over from the coastlines and the area of South Savo, towards the east, to the central, western and northern parts of Finland. During a century, the new Finland, or late settlement region, was finally settled (de la Chapelle, 1993; de la Chapelle and Wright, 1998). Thereafter, the Finnish population has expanded fast to its present number of 5,100,000.

As a consequence of the special Finnish population history characterized by national and regional isolates with rapid expansion and random sampling of inhabitants from the original population, some disease alleles became enriched and formed the Finnish disease heritage, while certain disease alleles, such as cystic fibrosis and phenylketouria, disappeared almost completely (de la Chapelle, 1993; de la Chapelle and Wright, 1998; Norio et al., 1973; Peltonen et al., 1999). Compared to other populations, most of these diseases in Finland are associated with high gene frequencies and low mutation rate, and they exhibit excessive locus homogeneity (Peltonen et al., 2000). Most diseases show regional clustering, but some of the mutations are enriched all over the country, and were probably carried with an early immigrant to Finland (de la Chapelle, 1993; de la Chapelle and Wright, 1998, Peltonen et al., 1999). However, sometimes multiple mutations

account for the disease of regional subisolates, as is exemplified in X-linked recessive choroideremia, CHM (Sankila et al., 1991; Sankila et al., 1992), and X-linked juvenile retinoschisis, RS (Huopaniemi et al., 1999).

The past success in identifying the disease genes in Finland via the random mapping and further fine mapping by haplotype-based restriction analysis has been assisted by features that population isolate possesses, such as population founding sufficient time ago, large population size, and absence of significant immigration during rapid expansion (de la Chapelle and Wright, 1998). Therefore, the subisolates have more uniform genetic and environmental background, prevalence of certain diseases is high, and the diagnostic and phenotypic criteria are easily standardized based on the high quality of good medical and epidemiological registers (Peltonen et al., 2000). Genealogical records include the accurate church records, which allows tracing of the families and ancestors back to the 16th century.

To date, the Finnish disease heritage comprises some 35 genetic disorders more prevalent in Finland than in other populations, and at least 24 disease genes have been identified, which offer a profound basis for functional studies and understanding of the disease mechanisms (de la Chapelle and Wright, 1998; Kere, 2001; Peltonen et al., 1999; Peltonen et al., 2000; Ridanpää et al., 2001). The list includes its own unique set of autosomal recessive diseases, however, two autosomal dominant (amyloidosis type V; FAF, and tibial muscular dystrophy; TMD) and two X-chromosomal (CHM and RS) patterns of inheritance are also included (de la Chapelle and Wright, 1998).

## 2. Identification of genes

### 2.1. The Human Genome Project (HGP)

Detailed information of the human genome sequence offers advances to increase our understanding of genetics and the nature of genes in health and disease. The effort for complete sequencing of human genome was first formally proposed in 1985-1986, and officially it led to the initiation of the Human Genome Project (HGP) in 1990. Many different sequencing centers and hundreds of scientists around the world shared the forthcoming work. The strategy "map first, sequence later", and the principle that all sequencing data were publicly and freely accessible was adopted in the project (Collins et al., 1998). Sequencing the genomes of several other organisms

was also included in the project, many of which have already been completely sequenced, such as the yeast *Saccharomyces Cerevisae* (Goffeau et al., 1996), the nematode *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998), the fruitfly *Drosophila melanogaster* (Myers et al., 2000), and the plant *Arabidopsis thaliana* (*Arabidopsis* Genomics Initiative (AGI), 2000) genomes.

At the 5-year point, the first major task of the HGP was to generate detailed high-resolution genetic and physical maps of the entire human genome (Lander et al., 2001). The work progressed fast, and in 1994, the HGP had constructed high-resolution maps, which were used for sequencing and assembling the overall draft genome sequence. By the end of 1998, large-insert clone contigs had been assembled, and approximately 180 megabases (Mb) of the human genome had been completed. During the following years, 1999 and 2000, the first complete human chromosome sequences for chromosomes 22 and 21, respectively, were published (Dunham et al., 1999; Hattori et al., 2000). Finally, in February 2001, the first complete draft consisting of about one billion bases of the human genome was published (Lander et al., 2001). At the same time, a private company, Celera Genomics, published another draft sequence incorporating all of the public project's accessible data with the whole genome shotgun data of their own (Venter et al., 2001).

### 2.1.1. The first draft of the human genome

The estimated total size of the human genome is 3.2 gigabases (Gb), of which 2.95 Gb is euchromatic (Baltimore et al., 2001). More than 90% of this region is represented in the sequenced clones of the HGP and Celera, and over 94% can be found in the combined publicly available sequence databases (Lander et al., 2001). More than 1.4 million single nucleotide polymorphisms (SNPs), meaning one SNP per 1.91 kb of genomic sequence, have also been identified (Sachidanandam et al., 2001).

Even with the availability of the first human genome draft sequence, gene identification in the genomic sequence presents with challenges. The most suprising finding of the HGP was that although the human genome is around 30 times larger than the worm and fly genomes, there appear to be only some 30,000-40,000 genes coding for proteins which is only twice as many as in worm or fly (Lander et al., 2001; Venter et al., 2001). It may turn out that these numbers are underestimations, since some genes are expressed in rare tissues or encode small proteins, and

may have been missed in these predictions. On the other hand, some predictions may correspond to pseudogenes or gene fragments, giving a spurious estimate. Other recent estimates of gene number vary from 35,000 to 120,000 genes (Ewing and Green, 2000; Liang et al., 2000; Wright et al., 2001). These numbers are not directly comparable, since the sequence annotations are based on different prediction programs and techniques.

Assuming that the human genome consists of approximately 30,000 genes with an average gene size of about 30 kb and transcript size of 1.4 kb, then only 1.5% of the genome codes for proteins, and various types of repeat sequences form over half of the genome (Lander et al., 2001). The rest of the sequence contains regulatory regions, such as promoters and transcriptional sequences, as well as unknown features (Baltimore, 2001; Birney et al., 2001; Lander et al., 2001; Venter et al., 2001). Genes seem to be more complex than has been thought. The different combinations of exons produced by alternative splicing, which affects more the coding sequence than the 5'- or 3'-untranslated regions (UTRs), are seen for approximately 60% of the genes, and are thought to yield a large number of protein products. Proteins appear to be complicated, and the diversity of the human proteome is emphasized by innovation in the creation of new proteins and their post-translatorial modifications (Lander et al., 2001). After evaluating all the proteins potentially encoded in the genome, the public consortium concluded that although humans do not have more types of protein domains than lower organisms, they use those domains in a new manner, assembling new proteins from old parts. The second phase draft of the human genome, to fill the remaining gaps and provide a more accurate catalogue of genes, is now under way and will hopefully be completed by 2003.

## 2.2. Identification and isolation of genes by positional cloning methods

Positional cloning is a process whereby genes responsible for a disease are identified as a direct result of genetic analysis, without any prior knowledge of the defective gene product, in a multi-step process. First, the gene is localized to a particular region of a chromosome by linkage analysis, utilizing families, in which the disease locus segregates. This preliminary localization is followed by a detailed molecular analysis of the region, including genetic fine mapping: the region of interest is covered in a series of overlapping genomic clones (YACs, BACs, and cosmids), DNA isolation, transcript identification, cDNA cloning, and mutation searching of

candidate genes. The proof for identifying a disease gene is that it carries the mutations that segregate consistently with the disease in the families.

Positional cloning of disease genes has progressed rapidly because of the information provided by the HGP. Although many disease genes have been identified successfully by this method, for example the cystic fibrosis gene, *CFTR* (Rommens et al., 1989), and the diastrophic dysplasia gene, *DTD* (Hästbacka et al., 1994), generally this strategy can be very laborious. The size of the critical region defined by linkage analysis usually remains several megabases containing dozens of genes. Instead, positional cloning using approaches, such as linkage disequilibrium and/or large-scale sequencing of genomic clones, followed by transcript mapping, assists genetic mapping to be taken to a very high resolution, and finally the identification of a disease gene.

The positional candidate gene approach is based on strategies to identify candidate genes by a combination of map position, expression, function, or homology to a relevant gene in a model organism. An obvious positional candidate gene was suggested based on the linkage analysis of an Irish retinitis pigmentosa family mapping the disease gene to chromosome 3q, on the same region next to the human *rhodopsin* (*RHO*) gene (McWilliam et al., 1989). Later, a novel *RP1* mutation was identified in the *RHO* gene (Farrar et al., 1992). Recently, a functional candidate gene approach was used successfully in mapping of the megaloblastic anemia 1 (*MGA1*) locus to the same region with the candidate gene encoding the intrinsic factor (IF)-B$_{12}$ receptor *cubilin*, *CUBN*, and in the identification of the underlying mutations in the Finnish MGA1 families (Aminoff et al., 1999).

The position independent candidate gene approach, which is based solely on the molecular pathology of the disease or the phenotypically related diseases, is rarely very successful. Instead, when biochemical basis of a disease is known exactly, it is possible to characterize the gene product through functional cloning approach. The normal protein sequence is determined by aminoacid sequencing or by purifying the corresponding mRNA by immunoprecipitation using antibodies raised against the gene product, followed by probing cDNA libraries with either a combination of gene-specific oligonucleotides or orthologous cDNA, respectively. This approach was used successfully in the characterization of the corresponding mutated genes encoding the blood clotting factor VIII in hemophilia A patients

(Gitschier et al., 1984), and the enzyme phenylalanine hydroxylase (PAH) causing phenylketonuria (Robson et al., 1982).

### 2.2.1. Linkage and linkage disequilibrium analyses

The common method to search for disease genes is to use linkage analysis to reveal the initial disease locus, followed by haplotype analysis to refine the genetic region of interest (Peltonen et al., 2000). The basic idea for linkage analysis is to trace and measure the cosegregation of disease in a given family with marker loci. If the marker locus and the disease gene are linked, the allele at that marker locus is likely to travel with the disease, and the distance between these loci can be calculated by using information from putative recombination events during meiosis (Ott, 1991; Terwilliger and Ott, 1994). The proportion of recombination, recombination fraction ($\theta$), is the likelihood of recombination between the loci. The genetic distance is generally measured in centiMorgans (cM), where 1 cM is approximately the distance between two loci that on average show 1% of recombination ($\theta = 0.01$). The extent of linkage is measured by formulating a LOD (logarithm of odds) score (Z), which is the logarithm of the likelihood ratio comparing the hypothesis of linkage between two loci at certain recombination fraction with the hypothesis of random recombination ($\theta = 0.5$). The maximum value of the LOD score ($Z_{max}$) gives a measure of the statistical significance of the result; a LOD score of 3 (odds ratio 1000:1) or more implies a significant linkage. The closer the marker is to the disease gene, the greater the extent of cosegregation and the higher the LOD score. The recombination fraction at $Z_{max}$ is the most likely estimate of the genetic distance between the loci. Negative LOD scores of less than -2 exclude linkage for a marker and can be used as a measure for exclusion mapping.

Due to limited number of families and generations in analysis, linkage can be detected over large genetic distances. Most often, these regions remain megabases in size, and are therefore too large for effective cloning of the gene (de la Chapelle and Wright, 1998). Linkage disequilibrium (LD) analysis usually overcomes this problem and can be used as a method to complete fine-scale gene localization (de la Chapelle, 1993).

In linkage disequilibrium analysis, information from meiotic events throughout dozens or hundreds of generations are evaluated by assuming that the disease-causing mutation occurred in an individual as a founder effect, and the expansion of the population has occurred by growth

rather than by immigration (de la Chapelle and Wright, 1998). Moreover, enough time has elapsed since the founding, and the population size of the affected individuals is large enough in the analysis. Generally, LD is referred to as a measure of the degree of non-random association of two markers. If LD exists, the alleles at nearby markers are in association with the disease at the population level. Errors in the calculations of the relationship between LD and marker locus physical distances probably will happened, for example, if the disease gene frequencies in the studied population are estimated wrong or multiple founding mutations exist instead of one (de la Chapelle, 1993; Jorde, 2000).

Linkage disequilibrium analysis utilizes several statistical formulas in calculating allelic association, and the use of multiple markers can enhance the power and accuracy of LD mapping. The strength of the allelic association can be calculated by the use of a nonparametric statistical test, such as a likelihood based program DISMULT, which analyses association for multiple loci simultaneously (Terwilliger, 1995). An assumption of a small number of founders followed by rapid expansion of population in a given isolate led Hästbacka et al. in 1992 to adapt a modified formula, which was originally developed for the study of mutation rates in bacterial cultures (Luria and Delbrück, 1943). This method was used successfully in the localization of the diastrophic dysplasia (*DTD*) gene.

Search for common ancestral haplotype shared by the affected individuals is often done as a part of linkage or LD analysis, and offers an excellent shortcut for disease mapping (Peltonen et al., 2000). In haplotype-based recombination analysis, the common ancestral haplotype is disrupted by multiple historical recombinations, and the conserved region suggests the position of the gene precisely. The size of the chromosomal region showing LD depends directly on the number of generations from the founder (Peltonen et al., 1999). The more extended the conserved region is, the more recently has the mutation occurred in the population. In congenital chloride diarrhea, CLD, the ancestral haplotype was conserved throughout a 12-cM genetic interval, suggesting that a single founder mutation has been subjected in the population only some 15-25 generations ago (Höglund et al., 1995). In contrast, in progressive myoclonus epilepsy (EPM1) (Lehesjoki et al., 1993), the ancestral haplotype consisted of a 5-cM interval, referring to a founder mutation having occurred some 100 generations ago. Fine mapping of the chromosomal regions by linkage disequilibrium has been extremely helpful in refining the localizations of the

above-mentioned as well as many other Finnish diseases, and led to the isolation of their respective genes (de la Chapelle and Wright, 1998; Peltonen et al., 1999).

### 2.2.1.1. Polymorphic markers

Recent advances in high-throughput sequencing and genotyping technologies have enabled geneticists to easily characterize genetic variation at the nucleotide level. In the 1980s, the first generation polymorphic DNA markers were based on restriction fragment length polymorphisms (RFLPs). RFLPs have only two alleles depending on the presence or absence of a specific enzyme restriction site (Botstein et al., 1980). Since their scoring by digestion of DNA followed by Southern blotting and hybridization was laborious and time consuming, the invention of polymerase chain reaction (PCR) technology enabled a more specific and fast amplification of the region of interest. To date, the use of multiallelic and thereby very informative variable number of tandem repeats (VNTRs) is preferred in genotyping.

During the 1990s, RFLPs were mostly replaced by simple tandem repeats (STR), consisting of di-, tri- or tetranucleotides, in the study of monogenic diseases (Livak et al., 1995). STRs are preferred in semi-automated systems of genotyping, as they are evenly and widely distributed across the human genome, and display high levels of allelic variation (Gray et al., 2000). However, recent interest of analyzing complex multifactorial diseases has raised the favour of using single base polymorphisms (SNPs). Dense maps of SNP markers and the discovery of intragenic SNPs have provided excellent tools to be used in linkage and association studies (Kwok, 2001). Of the 1.42 million SNPs identified by the International SNP Map working group, 60,000 were found to be within genes, and the amount of diversity varies within the genome (Sachidanandam et al., 2001). Generally, transitional base changes are more common than transversions, and CpG dinucleotides display the highest mutation rate (Gray et al., 2000). Although SNPs are less informative, they are more stable than STRs due to lower mutation rates. To date, the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/index.html) contains already over 2.5 million reference SNP entries.

### 2.2.2. Physical mapping

The primary goal of physical mapping is to identify and isolate a gene in the genomic clones that have been mapped to a certain chromosomal region defined by recombinants or allelic association

in genetic mapping. The initial physical mapping and DNA isolation of large genomic clones has been greatly advanced by the ability to clone large fragments of DNA in yeast artificial chromosomes (YACs) (Burke et al., 1992). YAC libraries can be screened for selected chromosomally positioned DNA markers, termed sequence tagged sites (STSs), in order to construct regional maps, and to join individual overlapping fragments into contiguous fragment arrays. Overlapping YAC clones covering up to several megabases can then be characterized individually and reassembled into the genome. However, in isolation of new polymorphisms and expressed sequences, YAC clones may be too large or chimeric, and are therefore complemented by contigs of overlapping P1-artificial chromosomes (PACs) or bacterial artificial chromosomes (BACs). These clones are selected for shotgun sequencing, and the genomic sequence is reconstructed by assembly of overlapping clone sequences. Recently, as a part of the HGP sequencing project, the International Human Genome Mapping Consortium reported a large human whole-genome BAC map combined with previous genetic maps (McPherson et al., 2001). They also inserted the new data into the database, which can be found from http://genome.wustl.edu/gsc/human/mapping/index.shtlm. Similar physical maps based on BAC clones have been used in sequencing other large genomes, and the technique is being applied to the mouse genome as well (Bouck et al., 2000).

### 2.2.3. Identification of genes from cloned DNA

Several traditional approaches are available for gene identification from cloned DNA. These include searching for evolutionary conservation of DNA fragments by hybridization to Southern blots of genomic DNA isolated from multiple species (zoo blots) (Claudio et al., 1994), identification of CpG islands as signposts for the 5' end of genes (Larsen et al., 1992), cDNA selection (Lovett et al., 1991; Parimoo et al., 1991), and exon trapping (Duyk et al., 1990). Similarly, conserved DNA fragments can be tested for expression by hybridization to multiple-tissue northern blots of RNA isolated from fetal and adult human tissues (Alwine et al., 1977). If a transcript is identified, the genomic DNA fragment can be subsequently hybridized to specific tissue cDNA libraries (Bonaldo et al., 1994).

The dinucleotide CpG islands are of particular interest because many are associated with 5' ends of genes, and they can be identified by cleavage with restriction enzymes (Larsen et al., 1992), and more recently, by genomic annotation software (see below). The density of CpG

islands varies, and correlates reasonably well with estimates of relative gene density on different chromosomes.

In cDNA selection strategy, an amplified cDNA library is hybridized for example to YAC inserts or cosmid clone sequences (Lovett et al., 1991; Parimoo et al., 1991). Hybridized cDNA inserts are then eluted and amplified by PCR. The process is repeated two or three times before cloning the selected cDNAs. The resulting cDNA sub-libraries are enriched for expressed sequences from the genomic region, but still require thorough characterization because of simultaneous selection of cDNAs with homology to pseudogenes and low copy number repeat sequences in the genomic DNA template.

Exon trapping relies on functional sequences required for RNA splicing (Duyk et al., 1990). Subcloned genomic DNA fragments are transfected into an appropriate eukaryotic cell line and transcribed into RNA. If the genomic fragment contains an exon, it can be spliced properly into mature mRNA after expression in mammalian cells in culture. This method relies on there being only one splice site to be selected, and has the disadvantage of selecting genomic fragments with cryptic splice acceptor sites, thereby trapping false positive exons.

In the past few years, new genes have been identified in many species by randomly sequencing cDNA clones representing RNAs expressed in different tissues to produce expressed sequence tags (ESTs), which have been collected into a single database dbEST (Adams et al., 1991; Boguski et al., 1993). This kind of information obtained from a gene sequence and expression, in combination with its possible chromosomal location, will implicate a candidate gene for a disease that has been mapped to the same region (Ballabio, 1993).

As an alternative to hybridization to cDNA libraries on filters, EST fragments can be used to design primers to obtain complete cDNA sequences using a method such as rapid amplification of cDNA ends (RACE) (Frohman et al., 1988). On a larger scale, long stretches of genomic DNA can be assembled by shotgun sequencing method, and potential exons identified using computer algorithms. As automated sequencing and computer programs become more sensitive and efficient at finding coding sequences, the approach of direct sequencing to find genes should become more popular.

## 2.3. Genome annotation through database searches

In around 1996, database searches or "research *in silico*" became a valuable tool in genomic research because of accumulating sequences of partial gene sequences derived from mRNA of human and other species (Boguski, 1995). To date, the attention is focused more on genomic sequence annotation, for which new computational strategies for data generation are developing rapidly. A "virtual mRNA" that reflects a probable transcript is based on the evidence of various gene structures by different gene-prediction programs. However, detection of genes *in silico* requires advanced bioinformatics technology and large-scale annotation of the genomic sequence.

The first step to characterize a gene is to construct a complete gene model by detection and assembly of individual exons with 5' and 3' extremities of the putative transcript (Claverie, 1997). Computational methods can be divided into homology-based methods and those based on general properties of protein-encoding genes. Homology-based gene prediction software detects sequences that are associated with already known genes, ESTs, and complementary DNAs. In general, it gives more reliable evidence of a putative gene than prediction without knowledge about other similar genes (Birney et al., 2001). The general properties of genes include promoter regions, splice sites, open reading frames, initiation and termination codons, and 3' untranslated regions.

Gene finding should be based on combinations of different prediction programs, and it depends on the quality and quantity of sequence data in use. Despite the vast progress in developing accurate annotation programs, some prediction programs tend to overlook gene numbers by producing false positives. Sequencing errors and contigs with gaps cause disrupted reading frames (Birney et al., 2001). The genes themselves can be separated by long stretches of noncoding DNA, and the total sizes of introns occasionally extend thousands of bases between exons, as is exemplified in the human *dystrophin* (*DMD*) gene spanning over 2.5 Mb of the X-chromosome (Pozzoli et al., 2002). Although, the overall sizes of exons are more constant compared to introns, some genes have a considerable number of exons. For example, the human *titin* (*TTN*) gene comprises 178 exons of over 80-kb coding sequence (Labeit and Kolmerer, 1995). Alternative splicing of exons, which most of the programs do not recognize, further complicates construction of genes. Rarely active or highly tissue specific genes are not always detected in searches of expressed genes (Birney, 2001; Lander et al., 2001). Moreover, approximately 5% of the human genome contains large duplicated genomic segments of size of

over 10 kb within the same chromosome or among different chromosomes resulting in copies of genes or nonfunctional pseudogenes (Lander et al., Venter et al., 2001). All these features of the genome undoubtedly complicate gene identification.

In the past, genomic sequences have been analyzed mostly by hand (Harris, 1997). However, the rapid establishment of databases and the rate of sequencing have raised the need for reliable automated techniques for interpreting long genomic sequences and for multiple comparisons among the analyses. The most effective are the integrated gene prediction methods, which provide accurate and automated predictions. Automated sequence annotation software programs, such as Genotator (Harris, 1997) and NIX/PIX (Nucleotide/Protein Identify X, UK HGMP Resource Center), notably enhance sequence analysis by combining and running series of different gene-finding programs and homology searches simultaneously. Moreover, when several different programs make the same prediction on a given sequence, the confidence in the prediction is increased.

### 2.3.1. Similarity searches

Gene prediction makes use of different sequence similarity searches. Basic local alignment tools, BLAST programs, perform similarity searches against EST, protein, and genome-sequence databases (Altschul et al., 1990, 1997). Other programs, such as Repeat Masker2, screen anonymous genomic sequences against repeat databases to filter so-called "junk" DNA prior to similarity searches (Gish and States, 1993). Usually, similarity search programs do not detect an entire gene, and combinations of other prediction methods are needed. Gapped BLAST allows for insertions and deletions within the analyzed genomic sequence, whereas a new version of BLAST, a position-specific-iterated BLAST (PSI-BLAST), detects similarities by a series of repeated homology searches (Altschul et al., 1997). EST_GENOME (Mott, 1998) analyzes either ESTs or cDNAs against genomic sequences based on a simple model of recognition of splice site dinucleotides. As the human genome project is in a phase of systematic sequencing, different intermediate forms of sequences are available within single-pass sequence of Genome Survey Sequence (GSS) and High Throughput Genomic (HTGS) databases of GenBank.

## 2.3.2. Predicting exons and gene-specific elements

To predict protein-coding exons and genes, several distinct types of programs have been developed, for example Genscan (Burge and Karlin 1997), Grail (Uberbacher and Mural, 1991), Genie (Reese et al., 1997), Fgene (Solovyev and Salamov, 1997), and MZEF (Zhang, 1997). Most often, they search for conserved exon-intron structures, such as acceptor and donor splice sites or specific signals for 5' exons (a TATA-box and/or a translation start site) and 3' exons (a stop codon and/or a polyadenylation signal) (Claverie, 1997). There are certain limitations for all of these programs. Most of them assume that the analyzed genomic sequence contains exactly one complete gene structure. Therefore, most of the predicted genes contain incorrect exon boundaries or exons are missed. Genscan is designed to predict partial genes, multiple genes and single-exon genes in given genomic DNA, and it analyzes the genes occurring on either or both DNA strands simultaneously (Burge and Karlin, 1997, Claverie, 1997). On the other hand, the accuracy of all these programs is reduced significantly when analyzing large genomic contigs of hundreds of kilobases.

To predict full extent noncoding parts of transcripts, determination of noncoding exons, putative functional regulatory elements, promoter-containing regions, and 3' extremities is needed. Unfortunately, most traditional gene prediction programs refer only to protein-coding regions, and thereby do not recognize, for example, the noncoding first exons. Gene promoters are located upstream from the translation start codon, and are frequently associated with CpG islands of various sizes, frequency, and GC-content. The sequence signals that support transcription units are TATA-boxes within the promoter region and polyadenylation signals (AATAAA) downstream from the termination codon ending the transcription (Claverie, 1997). Although many useful programs, for example, MatInspector/TRANSFAC (Heinemeyer et al., 1999) and Neural Network Promoter Input (Knudsen, 1999), have been generated to identify these potential consensus target sequences on genes, however, many genes do not have such elements. Only 70%-80% of promoters contain a TATA box and a polyadenylation signal is missing from over half of the 3' ends of transcripts (Claverie, 1997). Also, the functional non-protein coding RNA genes are difficult to detect with the current prediction programs, since these genes are untranslated, often small, and non-polyadenylated (Lander et al., 2001). They can mainly be located by similarity searches corresponding to known RNA genes.

### 2.3.3. Protein structure prediction

Many proteins in sequence databases have conserved sequence patterns based on which they are categorized. Therefore, if two proteins share significant sequence similarity, they should also have similar structures, and the search with a candidate sequence may discover a family or a superfamily relationship with a known protein. The prediction and comparing the protein structures to analyze the secondary structure, the functional domains, usually provides valuable information, although increased confidence is reached when several methods give a similar prediction (Mount, 2001).

Pattern-Hit Initiated BLAST, PHI-BLAST, (Zhang et al., 1998) combines the input protein pattern search with the local similarity searches. For detailed analysis of proteins' structural motifs, a number of programs are available in databases. SWISS-PROT (Bairoch and Apweiler 1999), the Brookhaven Protein Databank, PDB, PROSITE (Hofmann et al., 1999), and Pfam (Bateman et al., 1999) databases are among the most extensively used databases for protein function, domain structure, motifs, families, and post-translational modification. PSORT (Nakai and Horton, 1999) predicts the presence of protein localization signals, whereas SignaIP predicts the location of signal peptide cleavage sites in proteins (Nielson et al., 1997). A hydrophobic $\alpha$-helix, typically 20-30 residues in length, is a good indicator that the region spans a membrane, and can be predicted with the program Tmpred (Kyte and Doolittle, 1982; Hofmann and Stoffel, 1993).

### 2.3.4. Comparative genomics and model organisms

Large genome sequencing projects of different organisms have made possible to predict the function of similar genes in different organisms. Therefore, sequence comparison of genomes of various species is an excellent technique to discover genes and other important regions. Especially, comparisons between human and mice sequences have become a particularly effective resource for studies of human genes, since mouse and human genes often fall into homologous chromosomal regions that have a conserved gene content (Marra et al., 1999). Moreover, identifying of disease genes in mice is a more practicable method, since multiple backgrosses yield a high density of recombination breakpoints within the genetic interval of interest. To find a putative orthologue (i.e. related gene in other species), the most useful way is to compare protein products between species. By comparing human sequences with sequences

from other species, it has been shown that 40% to 60% of the predicted proteins give some sequence similarity to other species, for example fruitfly and worm (Rubin, 2001). Databases, such as XREFdp (Bassett et al., 1995), can be used to compare and align genomic sequences of different model organisms.

**Table 1.** Selected programs for performing gene and protein predictions

| NAME | SOURCE | REFERENCE |
|------|--------|-----------|
| BLAST, PHI-BLAST, PSI-BLAST | http://www.ncbi.nlm.nih.gov/ BLAST | Altschul et al., 1990, 1997; Zhang et al., 1998 |
| EST_GENOME | http://www.hgmp.mrc.ac.uk/ Registered/Option/est_genome. htlm | Mott 1998 |
| GSS, HTGS | http://www.ncbi.nlm.nih.gov/ | See web |
| FGENE | http://genomic.sanger.ac.uk/gf/ gf.shtlm | Solovyev and Salamov, 1997 |
| GENSCAN | http://genes.mit.edu/GENSCAN. htlm | Burge and Karlin, 1997 |
| GENIE | http://www.cse.ucsc.edu/~dkulp/ cgi-bin/genie | Reese et al., 1997 |
| GRAIL | http://compbio.ornl.gov/ | Überbacher and Mural, 1991 |
| MZEF | http://argon.cshl.org/genefinder/ | Zhang, 1997 |
| NIX and PIX | http://www.hgmp.mrc.ac.uk | See web |
| PDB | http://www.rcsn.org | See web |
| Pfam | http://www.sanger.ac.uk/Pfam | Bateman et al., 1999 |
| PROSITE | http://www.expasy.ch/prosite | Hofmann et al., 1999 |
| PSORT | http://psort.nibb.ac.jp/ | Nakai and Horton, 1999 |
| SWISS-PROT | http://www.expasy.ch/sprot/ | Bairoch and Apweiler, 1999 |
| MatInspector/TRANSFAC | http://www.gsf.de/cgibin/ matsearch.pl | Heinemeyer et al., 1999 |
| Neural Network Promoter II Input | http://www.cbs.dtu.dk/services/ promoter | Knudsen, 1999 |
| Repeat Masker2 | http://www.genome.washington.edu/ UWGC/ | See web |
| SignalP | http://www.cbs.dtu.dk/services/ SignalP/ | Nielsen et al., 1997 |
| Tmpred | http://www.ch.embnet.org/ software/TMPRED_form.html | Hofmann and Stoffel, 1993 |
| XREFdp | http://www.ncbi.nlm.nih.gov/ XREFdb/ | Bassett et al., 1995 |

**2.4. Mutation identification**

Many of the current techniques for mutation identification rely on the finding of conformational changes in either double or single-stranded DNA fragment which are distinguished by electrophoresis based on their differential migration in an electrical field. Since single-base changes are the most common type of mutations in the human genome, most of the techniques detect sequence mismatches in heteroduplexes between allelic DNA fragments or single-stranded DNA which adopt slightly different conformations under nondenaturing conditions (Kwok, 2001). All the mutation identification methods have their limitations and none of them are 100% sensitive. Depending on the technique used, they detect mutations on small, 100-600 bp, fragments. Mostly, they do not reveal the site or the type of the mutation, and the sequence context of a base change has an effect on mutation identification. Moreover, mutations are difficult to detect if they are in the domain of high melting-temperature. Methods, such as single strand conformation polymorphisms analysis (SSCP) (Orita et al., 1989), heteroduplex analysis (HA) (Lichten and Fox, 1983), denaturing gradient gel electrophoresis (DGGE) (Fischer and Lerman, 1979), direct DNA sequencing, and variant detector arrays (VDAs) (Wang et al., 1998) are most commonly used for mutation identification.

In SSCP analysis, the amplified DNA fragments are electrophoresed in a nondenaturing polyacrylamide gel, where they form the secondary structures according to their sequence contex. Mutated fragments are identified as a result of their divergent migration pattern. This method detects typically 70% to 95% of the mutations.

In heteroduplex analysis, the denatured DNA fragments reanneal and form a heteroduplex, which can be detected as a band shift on a gel. To enhance formation of heteroduplex structures, wild type and mutant samples are crossmatched by mixing and letting them reanneal after denaturation. Lately, the most popular detection system for heteroduplexes due to simplicity, low cost, and high rate is to run the samples on a high-performance liquid chromatography (HPLC) column (Gray et al., 2000). The subsequent mutation detection is based on differential retention of the heteroduplexes in the column.

Denaturing gradient gel electrophoresis (DGGE) allows the identification of point mutations, which alter the melting behavior of the DNA fragment analyzed. The analysis is based on the electrophoretic mobility of a double-stranded DNA molecule, as mismatches reduce the mobility of heteroduplexes compared with homoduplexes, through linearly increasing

concentration of denaturing agents, such as formamide and urea. The introduction of a GC-rich domain (GC-clamp) to the DGGE PCR primers increases the mutation detection close to 100% (Myers et al., 1985).

While sequencing techniques, and sample capacity per run in capillary systems, have developed enormously during the past years, direct sequencing is usually the favored choice for mutation detection. When scanning large amounts of samples and especially single nucleotide polymorphisms, a comparable method for sequencing is the recently adapted VDA technology. Mutations are detected by hybridization of a PCR product to a glass chip containing oligonucleotide arrays. The difference in hybridization strength between matched and mismatched nucleotides detects the mutation (Gray et al., 2000).

## 2.5. Analysis of gene expression

To understand gene function, it is important to know about gene expression. Many technologies are now available for the analysis of patterns of gene expression. Northern blot hybridization and reverse transcriptase polymerase chain reaction (RT-PCR) allow the detection of rough resolution expression patterns of RNA from different tissues or cell lines. Cell specific localizations of RNA within tissues are normally obtained by *in situ* hybridization (Wilkinson and Green, 1990). For computational methods, gene indexes, such as UniGene and SAGE (serial analysis of gene expression) (Velculescu et al., 1995), provide information about the tissues, in which different transcripts are expressed. However, in recent years the "chips" or microarray techniques are being used to bring together expression data at enormous capacity (Brown and Botstein, 1999). With these methods it is possible to monitor the differences in expression of individual genes and RNA samples from multiple tissues under different physiological conditions by using thousands of target DNAs or entire genomes. The expression and function of genes can be characterized *in vitro* using cultured cells, but the development of the technology to manipulate mammalian genes in embryonic stem cells (ES) combined with the generation of transgenic mice, has made possible to express genes in mice, and create germline gain-of-function and loss-of-function mutations (Camper et al., 1995).

# 3. The human retina and the phototransduction pathway

The human neural retina (Figure 1. A) comprises three layers of cells. The outer nuclear layer contains two kinds of photoreceptor cells, of which 5% are cones and 95% rods. Cones, which mediate colour vision, are concentrated within the central part of the retina, the fovea, which is the area of highest visual acuity. The fovea is surrounded by the rods, which are responsible for vision in dim light. The outer segment of the photoreceptor is filled with phototransduction proteins-containing discs, which are shedded from the tip of the photoreceptor and constantly renewed. Photoreceptors synapse in the outer plexiform layer with a second, inner nuclear layer of bipolar cells, as well as with horizontal and amacrine cells, which process the visual information. The third ganglion cell layer contains amacrine cells as well as ganglion cells, which send their axons along the inner surface of the retina coming together at the optic disc to make up the optic nerve (Rattner et al., 1999). Adjacent to the neural retina is the retinal pigment epithelium (RPE), which functions on phagocytotic processes of the photoreceptor outer segments and transport of metabolites.

Phototransduction is a G-protein (GTP-binding regulatory protein) -coupled enzymatic cascade to amplify the signal derived from the light-absorbing visual pigment of rods, rhodopsin, which consists of an apoprotein opsin (Rattner et al., 1999). Photoactivation requires a continuous recovery of active rod opsin by regeneration of covalently bound chromophore 11-*cis*-retinal, a vitamin A-derivative, through isomerization cycle in the RPE and in the photoreceptors. Figure 1. (B-C) explains the phototransduction cascade in detail. A similar phototransduction cascade has been also proposed for cone photoreceptors, but other photopigments are involved in the cascade.
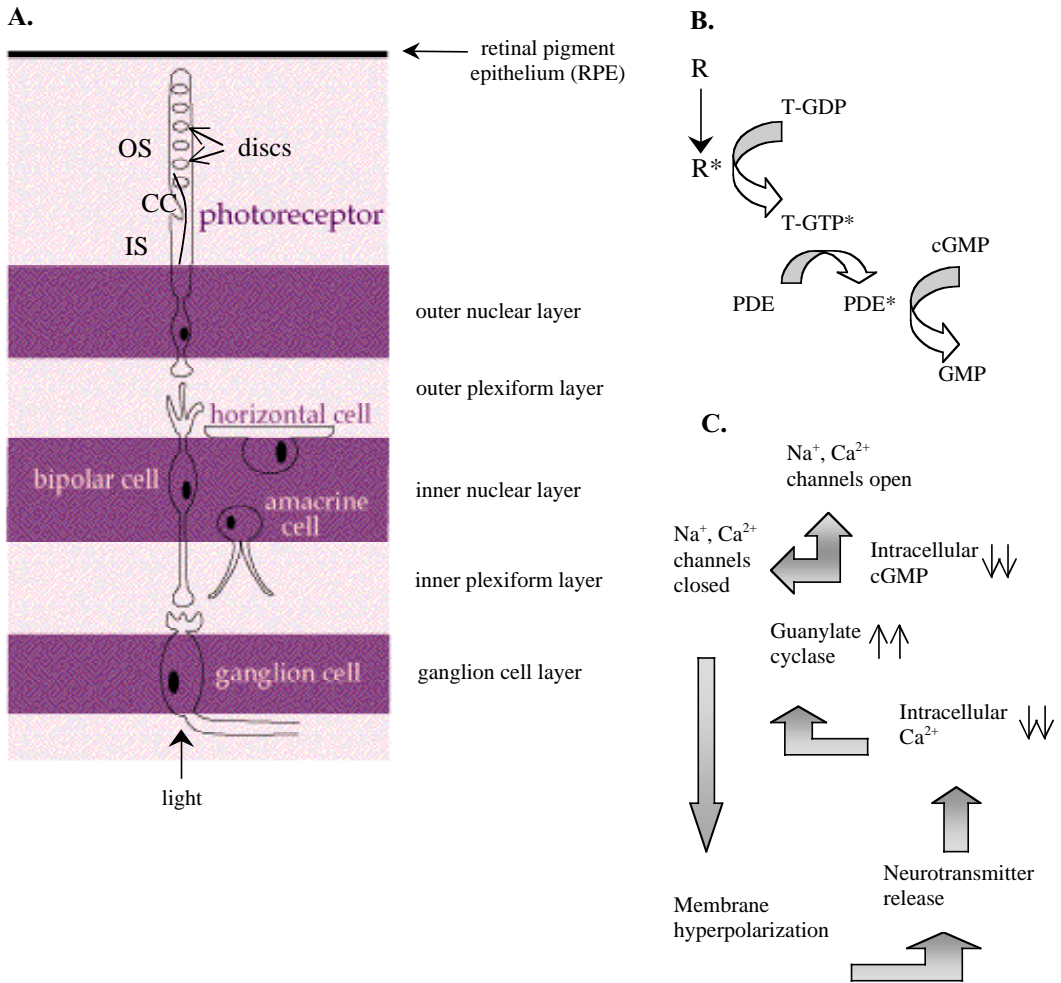
**A.**



retinal pigment epithelium (RPE)

OS   discs

CC   photoreceptor

IS

outer nuclear layer

outer plexiform layer

horizontal cell

bipolar cell

inner nuclear layer

amacrine cell

inner plexiform layer

ganglion cell   ganglion cell layer

light

**B.**

R

R*

T-GDP

T-GTP*

PDE   PDE*

cGMP

GMP

**C.**

$Na^+$, $Ca^{2+}$ channels open

$Na^+$, $Ca^{2+}$ channels closed

Intracellular cGMP

Guanylate cyclase

Intracellular $Ca^{2+}$

Membrane hyperpolarization

Neurotransmitter release

**Figure 1. (A).** Schematic drawing of the retina. The major cell types are shown: retinal pigment epithelium (RPE), photoreceptor, bipolar cells, horizontal cells, amacrine cells, and ganglion cells. A photoreceptor is composed of an outer segment (OS), a connecting cilium (CC), and an inner segment (IS). Adapted from web http://thalamus.wustl.edu. **(B).** Schematic representation of phototransduction pathway. Light enters through the ganglion cell layer, and must penetrate all the other cell layers before it reaches the rod and cone outer segments, containing the visual pigments. Phototransduction begins with the photoactivation induced conformational changes of rhodopsin (R*), which in turn activates the G-protein, transducin. Transducin, bound to GTP (T-GTP), triggers the enzymatic activity of cGMP phosphodiesterase (PDE), which then hydrolyses cGMP to GMP. **(C).** The low level of cGMP leads to closure of the cGMP-gated $Na^+$- and $Ca^{2+}$-channels at the plasma membrane and cell hyperpolarization. Finally, the membrane hyperpolarization leads to the release of the neurotransmitter at the photoreceptor synapse. Within the outer segment of the photoreceptor, drop of intracellular calcium levels activates the guanylate cyclase, which leads to an increase of the cytosolic cGMP concentration and causes the re-opening of the cGMP-gated cation-channels. Reassociation of transducin and phosphodiesterase subunits, as well as inactivation of rhodopsin by phosphorylation and binding of phosphorylated arrestin, completes the visual cycle (Gregory-Evans and Bhattacharaya, 1998).

* Activated.

### 3.1. Retinitis Pigmentosa (RP)

Inherited retinal diseases are a common cause of visual impairment in children and adults which result in a severe loss of vision. Retinitis pigmentosa (RP) is the most frequent form of inherited retinopathy with an incidence of 1 in 3,500 individuals worldwide (Sullivan et al., 1999). The disorder is genetically heterogeneous with approximately 80% of the patients having autosomal recessive forms of RP. Autosomal dominant and X-linked forms account each for approximately 10% of the patients. RP is characterized by night blindness, progressive constriction of the visual field leading to tunnel vision, abnormalities present on electroretinography, and pigmented fundus abnormalities (Roepman et al., 1996).

Advances in molecular genetics have provided new insights into the genes responsible for the pathogenic mechanisms of RP. The genetics of RP is complex, and 79 genes causing retinal disorders have been cloned to date, and additional 53 have been mapped, but not yet identified (RetNet, http://www.sph.uth.tmc.edu/RetNet). RP is the result of mutations in a wide variety of genes encoding proteins located in the photoreceptor cells and in the RPE, and components of phototransduction cascade (Gregory-Evans and Bhattacharaya, 1998). RP affects the sensory cells, the photoreceptors, which gradually deteriorate and die, the rods before the cones. To understand the mechanisms of photoreceptor cell death, invertebrate and vertebrate models carrying mutations in the genes implicated in human retinal diseases have been examined (Travis, 1998). For example, induction of apoptotic pathways leading to degeneration of retinal cells has been suggested to be associated in RP based on the studies of *Drosophila* visual system mutants, in which the mutations in the genes encoding visual cascade components, such as phopholipase C, arrestin and phosphatidylinositol transfer protein, lead to abnormal accumulation of rhodopsin-arrestin complexes in the photoreceptor cell (Alloway et al., 2000). In electron micrographs, the retinas of these mutants appear abnormal with large intercellular holes and vesicular bodies adjacent to photoreceptor cells. A possible mechanism for photoreceptor degeneration in RP, caused by a mutation in the *peripherin-RDS* (retinal degeneration slow)-gene (Farrar et al., 1991), is exemplified in *rds/rds* null mouse mutant, which fails to form outer segments of photoreceptors. Loss of the outer segments and, thereby, loss of the cGMP-gated ion-channels results in toxic levels of oxygen in photoreceptor cells leading to cell death (Travis, 1998).

## 4. The human ear and the mechanisms of normal hearing

The adult mammalian ear is made up of three distinct parts, the external, the middle, and the inner ear (Figure 2). Sound is collected by the external ear comprising the auricle and the external auditory canal, and transmitted to the tympanic membrane in the middle ear. In the middle ear, the three auditory bones, the malleus, the incus, and the stapes, transmit the sound waves to the oval window of the inner ear. The inner ear regulates two sensory systems, the auditory system, which is responsible for hearing, and the vestibular system, which sensors orientation and balance. It consists of two labyrinths, a membranous labyrinth filled with the endolymph in a cavity of bony labyrinth, which contains the vestibule, the cochlea, and the semicircular canals with the sacculus and the utriculus immersed with the perilymph (Petit, 1996; Willems, 2000). The membranous labyrinth consists of canals of the vestibular apparatus and the cochlear duct. The cochlear duct contains the organ of Corti located between the tectorial and the basilar membranes. In response to a movement of the basilar membrane and shearing of the tectorial membrane induced by sound vibrations, the auditory receptors, the hair cell stereocilia, bend. As a result, connecting tip-links between the stereocilia stretch followed by opening of the mechano-gated potassium channels at the stereocilia membrane (Kalatzis and Petit, 1998).

Each hair cell consists of a kinocilium defining the orientation of a well-organized V-shaped array of stereocilia in rows of graded height which has a vital role in hair cell mechano-electrical transduction (Petit, 2001). Each stereocilium consists of an actin-core, which is anchored into an actin cytoskeleton constituting the cuticular plate. The stereocilia are joined to one another by the tip-links, which have been proposed to be the gating springs that regulate opening and closing of the transduction channels located at or near the tip-links, through the action of a motor protein, probably myosin (Gillespie and Corey, 1997; Petit, 1996). These motor proteins move along actin filaments using energy from ATP hydrolysis. Due to the tension of the stereocilia and tip links, a small portion of the channels are open, even in the rest (Steel and Kros, 2001). The mechanic stimulation of the stereocilia leads to an influx of endolymphatic potassium resulting in hair cell depolarization. This leads to the release of neurotransmitter induced by the inflow of calcium ions ($Ca^{2+}$) on the basolateral side into the hair cell, thereby activating the acoustic nerve (Hudspeth, 1997; Willems, 2000).
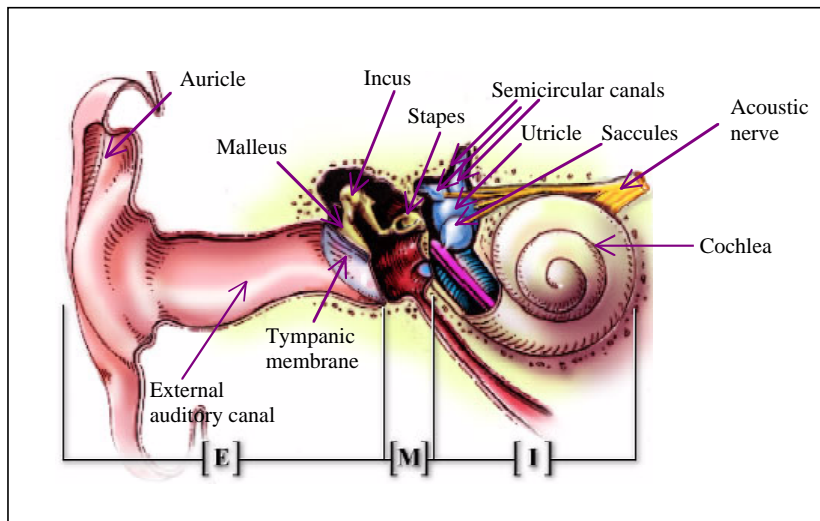
**Figure 2**. Drawing of the external (E), middle (M), and inner (I) ear. Adapted from web http://www.iurc.montp.inserm.fr/cric/audition/.

### 4.1. Classification of hearing loss (HL)

Hearing loss (HL) is the most common sensory disorder in humans caused by genetic and environmental factors, or both (Petit, 1996). Among the environmental factors, it can be caused by meningitis, rubella, perinatal complications, bacterial and viral infections, trauma, and ototoxic drugs, such as aminoglycosides. It is classified according to the age of onset as prelingual or postlingual, and syndromic or nonsydromic hearing loss (Willems, 2000). Most often, the hereditary hearing loss is caused by peripheral auditory defects, which can be conductive (external or middle ear defects), sensorineural (cochlear defects) or mixed (Petit, 1996). Severity of hearing loss is measured in decibels (dB) of the loss of hearing level, and is assorted as mild (27 to 40 dB), moderate (41 to 55 dB), moderate to severe (56 to 70 dB), severe (71 to 90 dB), and profound (>90 dB) (Kalatzis and Petit, 1998).

Hereditary deafness can be monogenic and result from a mutation in a single gene or multifactorial resulting from a combination of mutations in different genes and environmental factors. In developed countries, approximately 60% of deafness have a genetic background, as HL results from a mutation in a single gene (Marazita et al., 1993). Furthermore, approximately 70% of genetic HL is classified as nonsyndromic and the remaining 30% as syndromic hearing

loss. Nonsyndromic HL can be divided according to mode of inheritance to autosomal recessive (77%), autosomal dominant (22%), X-linked (1%) or mitochondrial (less than 1%), and the defect is generally of sensorineural origin (Resendes et al., 2001).

The overall impact of HL is affected by the age of onset. HL at birth or during prelingual period affects approximately 1/1000 infants. Postlingual HL results generally from a combination of additional genetic and environmental causes and accounts for a further 1/1000 children before adulthood (Morton, 1991; Resendes et al., 2001). Moreover, postlingual HL is more frequent than prelingual HL, and affects normal communication in 10% of the population by the age of 65 years (Willems, 2000). Many of the recently identified deafness genes cause progressive HL that begins in adulthood, suggesting that single genes may have an important role in HL in the population as a whole (Steel, 1998).

## 4.2. Hereditary hearing loss is genetically and clinically heterogeneous

Knowledge of the molecular basis of normal auditory function, and the progress in identifying genes involved in hearing loss has improved enormously over the past years. The molecular composition of the different structures in the cochlea, and especially changes in structural proteins have been associated with hearing loss. The mutated genes encode a large diversity of molecules, including molecules of cytoskeletal components, extracellular matrix components, transcription factors, enzymes, ion pumps, and mitochondrial genes (Kalatzis and Petit, 1998; Steel and Kros, 2001).

At present, 26 new genes involved in non-syndromic and over 32 genes for syndromic hearing impairments have been identified, some of the underlying genes being responsible for both forms of HL (Hereditary Hearing Loss Homepage; Resendes et al., 2001). Phenotypic variation from profound congenital to slowly progressing adult-onset HL results from allelic and locus heterogeneity (Adato et al., 1999; Balciuniene et al., 1998; Keats and Berlin, 1999). HL results from allelic heterogeneity, for example, in syndromic deafness USH1B (Weil et al., 1995), in nonsyndromic DFNB2 (Liu et al., 1997d; Weil et al., 1997), and in DFNA11 (Liu et al., 1997b), as well as in the atypical form of Usher syndrome (Liu et al., 1998b), where distinct mutations in the *myosin VIIA* (*MYO7A*) gene have been shown to underlie these diseases. Moreover, mutations in a gene encoding PDS, an anion transporter named pendrin, cause both syndromic and nonsyndromic hearing losses in Pendred syndrome (Everett et al., 1997) and in

DFNB4 (Li et al., 1998), respectively. Furthermore, correlation between different mutations and phenotypes have been observed, for example in the case of *cadherin 23* (*CDH23*), whose mutations may lead to both syndromic (USH1D) and nonsyndromic (DFNB12) hearing loss (Bolz et al., 2001; Bork et al., 2001). Digenic inheritance has been reported in a Swedish family, whose affected members are carriers for mutations in two different genes, *DFNA2* and *DFNA12* (Balciuniene et al., 1998). Also, possible interaction between two different genes, *USH1B* and *USH3*, giving the distinct phenotypes in an USH3 family of Jewish Yemenite origin, has been suggested (Adato et al., 1999).

It is difficult to ascertain whether the mutation type accounts for the expression profile of different phenotypes. Phenotypic variation between individuals with the same disease alleles can be also due to genetic and/or environmental modulating factors, such as modifier genes at another locus, which control the function and the expression of the gene underlying hearing loss (Friedman et al., 2000). The possible role for a modifier gene has been suggested in maternally inherited deafness associated with the A1555G mutation in the mitochondrial 12S ribosomal RNA gene (Bykhovskaya et al., 2000). The aminoglycosides have been shown to trigger HL in individuals with the A1555G mutation, though some patients with the A1555G mutation have a congenital HL in the absence of exposure to aminoglycosides. Through a genomewide screen, a nuclear locus for a possible modifier gene has been found for a mitochondrial DNA disorder (Bykhovskaya et al., 2000). Although there are limited examples of modifier genes in humans, they are expected to be a common cause of phenotypic variability, and are frequent in mice (Nadeau, 2001).

## 5. Usher syndrome (USH)

Von Graefe (1858) and Liebreich (1861) documented the earliest descriptions of the simultaneous occurrence of pigmentary retinopathy with deafness. Thereafter, Usher syndrome (USH), was named after a Scottish ophthalmologist Charles Usher (1914), who was the first to emphasize the recessive hereditary nature of the syndrome and suggested that it constituted a separate entity among other forms of retinitis pigmentosa.

USH comprises a group of autosomal recessive disorders associated with bilateral sensorineural deafness, loss of vision caused by retinitis pigmentosa, and variable vestibular

deficit. It is the most frequent cause of genetically determined combined visual and hearing loss, and accounts for more than 50% of the deaf-blind population (Boughman et al., 1983). Among congenital deaf individuals and RP population, USH accounts for approximately 3%-6% and 12%-20%, respectively (Boughman and Fishman, 1983; Haim, 2002; Rosenfeld et al., 1994; Vernon, 1969).

The prevalence of USH has been estimated at 2.4-3.6/100,000 in Scandinavia (Gröndahl, 1987; Hallgren, 1959; Lindenov, 1945; Nuutila, 1968; Nuutila, 1970) and at 4.4/100,000 in the United States (Boughman et al., 1983). USH is enriched particularly in Northern Sweden and eastern and northern parts of Finland, where the estimated prevalence is 10-21/100,000 (Hallgren, 1959; Nuutila, 1968; Nuutila, 1970). Recent studies suggest that Usher type 2 (USH2) is the most common phenotype of the three subtypes of USH worldwide, accounting for more than half of all the cases (Hope et al., 1997; Rosenberg et al., 1997), whereas USH1B seems to be the most common variety of Usher type 1 (USH1) (Weston et al., 1996). In the United States and Northern Europe, distribution of USH1 and USH2 patients varies between 33%-44% and 56%-67% (Petit, 2001), respectively. In Finland, there are more than 300 USH cases diagnosed, giving an overall prevalence of 6.6/100,000. Of those, 40% have progressive hearing loss suggesting USH3, whereas 12% account for USH2, 34% for USH1, and 14% for an unknown type (Pakarinen et al., 1995a). In Birmingham, UK, in the population over 15 years of age, the prevalence for all USH subtypes was 6.2/100,000 and USH3 was reported to account for 20% of all USH cases (Hope et al., 1997).

## 5.1. Clinical classification of USH

The existence of genetic heterogeneity with at least two distinct forms of USH was raised based on the differences between families in regard to the severity of hearing loss and vestibular findings (Davenport et al., 1978; Davenport and Omenn, 1977; Fishman et al., 1983; Forsius et al., 1971; Hallgren, 1959; Nuutila, 1970). On the basis of phenotypic classification suggested by Davenport and Omenn (1977), and the clinical features published by the Usher Syndrome Consortium (Smith et al., 1994), three different USH subtypes have been defined (Table 2). All USH patients have progressive RP appearing as loss of night vision and restriction of the visual field, which leads eventually to blindness, while the degree of hearing loss and vestibular function differs among the three clinical types. An USH1 patient has a profound congenital

deafness with onset of RP by the first decade, whereas in USH2 the hearing loss is congenital, from mild to severe. The age at diagnosis of RP in USH2 overlaps with USH1. USH1 and USH2 can be distinguished based on vestibular function. USH1 patients lack vestibular responses which causes delay in early motor development, whereas USH2 patients are normal in this regard (Kumar et al., 1984; Möller et al., 1989; Nuutila, 1970). Therefore, children with USH1 are late to sit and often do not walk before the age of 18 months (Möller et al., 1989). USH1 patients communicate with sign language, while USH2 patients have residual hearing and benefit from hearing aids. A third type of Usher syndrome (USH3) was suggested based on the progressive nature of hearing loss, and RP noted at puberty (Davenport and Omenn, 1977), however, these patients were estimated to account only for 2% of USH, and USH3 was not accepted as an established entity (Smith et al., 1994). Some other clinical findings in USH1 patients, such as cilia anomalies with slow sperm motility and bronchiectasis have been reported (Bonneau et al., 1993; Hunter et al., 1986).

**Table 2**. Clinical classification of Usher syndromes

| Type | Hearing impairment | Balance dysfuction | Vision loss from retinitis pigmentosa (RP) |
|------|--------------------|--------------------|--------------------------------------------|
| **USH1** | Profound congenital hearing loss. | Absent vestibular responses cause delay in early motor development. | Night blindness in infancy or first decade of life. Legally blind by early adulthood. |
| **USH2** | Mild to severe congenital hearing loss. | Normal vestibular responses. | Night blindness in first or second decade of life. Legally blind by early adulthood. |
| **USH3** | Normal or mild hearing loss at birth, progressive hearing loss leads to moderate to severe deafness within a few years or several decades. | Normal or decreased vestibular responses. | Night blindness in childhood or teens. Legally blind in early to mid adulthood. Rate of progression variable. |

**5.2. Molecular genetics of USH**

At present, ten different loci account for subsets of USH (the Hereditary Hearing Loss Homepage), of which six genes (*USH1B, USH1C, USH1D, USH1F, USH2A, USH3*) have been cloned. Six loci (USH1A-USH1F) correspond to Usher type 1, (Chaib et al., 1997; Kaplan et al.,

1992; Kimberling et al., 1992; Smith et al., 1992; Wayne et al., 1996; Wayne et al., 1997), and three loci (USH2A-C) for Usher type 2 (Hmani et al., 1999; Kimberling et al., 1990; Lewis et al., 1990; Pieke-Dahl et al., 2000). The first locus for Usher type 3 (USH3) was assigned in Finnish families by Sankila et al. (1995) (Table 3).

**Table 3**. Molecular classification of Usher syndromes

| Locus Name | Location | *Gene* | Protein | References |
|---|---|---|---|---|
| **USH1A** | 14q32 | *unknown* | unknown | Kaplan et al., 1992 |
| **USH1B** | 11q13.5 | *MYO7A* | Myosin VIIA | Kimberling et al., 1992; Smith et al., 1992; Weil et al., 1995 |
| **USH1C** | 11p15.1 | *USH1C* | Harmonin | Smith et al., 1992; Bitner-Glindzicz et al., 2000; Verby et al., 2000 |
| **USH1D** | 10q2 | *CDH23* | Cadherin 23 | Wayne et al., 1996; Bolz et al., 2001; Bork et al., 2001 |
| **USH1E** | 21q21 | *unknown* | unknown | Chaib et al., 1997 |
| **USH1F** | 10q21 | *PCDH15* | Protocadherin 15 | Wayne et al., 1997; Ahmed et al., 2001 |
| **USH2A** | 1q41 | *USH2A* | Usherin | Kimberling et al., 1990; Lewis et al., 1990; Eudy et al., 1998 |
| **USH2B** | 3p23-p24.2 | *unknown* | unknown | Hmani et al., 1999 |
| **USH2C** | 5q14.3-q21.3 | *unknown* | unknown | Pieke-Dahl et al., 2000 |
| **USH3** | 3q21-q25 | *USH3* | USH3 | Sankila et al., 1995; this study |

### 5.2.1. The USH1 genes

Kaplan et al. (1992) first localized the *USH1A* gene to chromosome 14q32 in ten families, of which eight originated from the Poitou-Charentes region in Western France. In the same year, Smith et al. discovered that 8 French Acadian and 11 British USH1 families were linked to two different regions in chromosome 11. USH1B was assigned to 11q14 whereas USH1C localized to 11p14, and was restricted to the French Acadian population in Louisiana. The 11q14 locus was supported in a study by Kimberling et al. (1992), who showed linkage to chromosome 11q in families from the United States, Sweden, Ireland and South Africa.

In 1995, Gibson et al. identified the defective gene (*sh-1*) of a naturally occurring deaf mouse mutant *shaker-1* encoding an unconventional murine myosin VIIA (Myo7a) by positional cloning strategy. Based on the localization of the *sh-1* phenotype to the mouse chromosomal

region homologous to human USH1B, and the role of unconventional myosins in auditory processes (Gillespie et al., 1993), *Myo7a* was suggested to be a likely homologue for USH1B. This led to the cloning of the human homologue, *MYO7A*, which was shown to be responsible for USH1B (Weil et al., 1995). *MYO7A* consists of 48 exons extending over a 100-kb genomic region, and it has several alternatively spliced forms (Chen et al., 1996). To date, at least 81 different mutations spreading throughout the *MYO7A* gene have been detected (Adato et al., 1997; Bharadwaj et al., 2000; Cuevas et al., 1999; Cuevas et al., 1998; Espinos et al., 1998a; Janecke et al., 1999; Lévy et al., 1997; Liu et al., 1997b; Liu et al., 1997c; Liu et al., 1997d; Liu et al., 1998b; Weil et al., 1995; Weston et al., 1996). The gene responsible for USH1B has also been shown to be mutated in two forms of nonsyndromic hearing loss, autosomal recessive deafness, DFNB2 (Liu et al., 1997d; Weil et al., 1997), and autosomal dominant deafness, DFNA11 (Liu et al., 1997b). Moreover, in patients with atypical USH, with symptoms similar to the USH3 phenotype, mutations in the *MYO7A* gene have been detected (Liu et al., 1998b).

Recently, a gene for USH1C was identified as a PDZ domain-containing protein, named harmonin (Bitner-Glindzicz et al., 2000; Verpy et al., 2000). The novel gene was identified using a candidate gene approach by sequencing random cDNA clones from a subtracted mouse library generated from inner ear sensory areas (Verpy et al., 2000). One clone showed sequence homology to a gene, encoding the human PDZ domain containing-protein, which mapped to the USH1C region. In a parallel study by Bitner-Glindzicz et al. (2000), a homozygous 122-kb deletion, which included most of the genomic region of harmonin, was identified by positional cloning in patients with severe hyperinsulism, deafness, enteropathy and renal tubular dysfunction. So far, mutation analyses have indicated altogether six different mutations in patients (Bitner-Glindzicz et al., 2000; Verpy et al., 2000; Zwaenepoel et al., 2001). Pathogenesis of USH1C is still unknown, since no animal model exists for the disease.

Harmonin consists of 28 exons of alternatively spliced transcripts with different expression patterns in various tissues. A functional role for harmonin has been suggested based on the function of other PDZ domain-containing proteins and the subcellular localization of harmonin (Bitner-Glindzicz et al., 2000; Montell, 2000; Verpy et al., 2000). PDZ domains are originally detected in scaffold proteins <u>P</u>SD/SAP90 (postsynaptic density protein), <u>D</u>lg-A (*Drosophila* tumor suppressor), and <u>Z</u>O-1 (tight junction protein) (Montell, 2000; Petit, 2001). PDZ proteins are known to be involved in protein-protein interactions, to organize multiprotein

complexes, for example receptors and ion-channels, at the plasma membrane (Montell, 2000). By immunofluorescence, harmonin has been detected in mouse sensory hair cells, including the cytoplasm and the stereocilia bundle (Verpy et al., 2000). Bitner-Glindzicz et al. (2000) detected low levels of antibody staining in the apical and basal surfaces of the sensory areas in the developing human ear. No expression was seen in neonatal mouse eyes, although low levels of anti-harmonin staining was seen on the outer neuroblastic layer of the retina in the developing human eye (Bitner-Glindzicz et al., 2000). Based on the expression profile in hair cells and the possible role in stabilizing the hair bundle, Verpy et. al (2000) suggested that harmonin and myosin VIIA might be components of the same protein complex. As well, harmonin can be modulating the activity of ion-channels or it can be localizing cytoskeletal and signaling proteins (Montell, 2000; Verpy et al., 2000).

The fourth locus for USH1 (USH1D) overlaps the locus for nonsyndromic autosomal recessive deafness (DFNB12) (Chaib et al., 1996), and was assigned to chromosome 10q2 in a Pakistani pedigree (Wayne et al., 1996). The underlying gene, a novel human cadherin, *CDH23*, was identified by positional candidate approach (Bolz et al., 2001) based on a parallel study determining the gene defect in the orthologous mouse mutant *walzer* (*v*) (Di Palma et al., 2001). Mutations in the orthologous mouse gene, *Cdh23*, cause disorganization of the stereocilia and misposition of the kinocilium, supporting the role of *Cdh23* as a critical component for the hair bundle formation. By RT-PCR analysis from several tissues, including retina, *Cdh23* was found to have a broad expression profile. However, by *in situ* hybridization from neonate cochlea samples, *Cdh23* expression was shown to be restricted to outer and inner hair cells (Di Palma et al., 2001). Bork et al. (2001) showed *CDH23* to be expressed in the human retina by northern blot analyses. Based on the mutation analyses, USH1D and DFNB12 have been shown to be caused by altogether 17 allelic mutations of *CHD23* (von Brederlow et al., 2002; Bolz et al., 2001; Bork et al., 2001). In general, classical cadherins are involved in cell-to-cell adhesions and cell-extracellular interactions. Probably, *CDH23* mediates similar functions. Also, it has been shown that in the presence of a calcium chelator, the hair cell stereocilia tip-links are disrupted. In the absence of calcium, cadherins undergo proteolysis, which indicates that *CDH23* may be associated with these links (Petit, 2001).

The fifth identified locus for USH1, USH1E, was assigned to chromosome 21q21 by Chaib et al. (1997). Using homozygosity mapping in a consanguineous family from Morocco,

USH1E was mapped to a delimited 15-cM interval. The gene underlying USH1E has not been cloned so far.

USH1F was assigned to chromosome 10q21 (Wayne et al., 1997) in a region of conserved synteny to mouse chromosome 10 containing a novel protocadherin gene, *Pcdh15*. The deafness and balance disorder in a mouse mutant, *ames waltzer* (*av*), is caused by a recessive *av* mutation of the *Pcdh15* gene (Alagramam et al., 2001a), and *av* was thereby suggested as a mouse model for USH1F, although no retinopathy had been reported in these mice. Subsequent protein sequence homology searches of overlapping BAC clones in the *USH1F* region indicated the human orthologue, *PCDH15*, which spanned approximately 1.6 Mb of genomic DNA. On a northern blot, a variety of putative *PCDH15* transcripts were identified in several tissues, including retina (Ahmed et al., 2001). By immunohistochemistry, *PCDH15* has been shown to be expressed especially in the synaptic and nerve fiber layers of retina, whereas in the human fetal cochlea, *PCDH15* expression was detected in the hair cells, in the supporting cells, and in the spiral ganglion cells (Alagramam et al., 2001b). In the following mutation analysis of 33 exons of *PCDH15*, four different mutations were indicated in two Pakistani families, in a Hutterite family from Alberta, and Indian families with USH1F (Ahmed et al., 2001, Alagramam et al., 2001b). According to the findings of abnormal supporting cells and stereocilia of both outer and inner hair cells in *av* mice, *PCDH15* has been proposed to have a possible role in maintaining the structure of the stereocilia bundles of the apical surface of hair cells (Alagramam et al., 2001a; Alagramam et al., 2001b).

## 5.2.2. The USH2 genes

In 1990, Kimberling et al. and Lewis et al. mapped the first gene for USH2 on chromosome 1q32-q44. Later, this gene, *USH2A*, comprising 21 exons, was identified by positional cloning from BAC subclones and a retina cDNA library as a putative extracellular matrix protein, usherin, with laminin epidermal growth factor and fibronectin type III motifs (Eudy et al., 1998; Weston et al., 2000). In contrast to the *USH1* genes, immunohistolabelling suggests that usherin has a broad but not ubiquitous tissue distribution, and is a component of basement membranes in the retina, in the lens capsule, in the Bruch's layer between the RPE and the choroid layer, and in the cochlea (Bhattacharya et al., 2002). Thus, the molecular pathogenesis of USH2A may be different from that of USH1 (Petit, 2001).

In the initial mutation analysis, three novel mutations were detected, of which one, 2299delG (named also 2314delG) leads to a premature stop codon within a crucial region in the laminin epidermal growth factor motif of *USH2A* (Eudy et al., 1998). Additional mutation searches of USH2A patients (Adato et al., 2000; Dreyer et al., 2000; Weston et al., 2000, Leroy et al., 2001) and patients with recessive retinitis pigmentosa without hearing loss (Rivolta et al., 2000), as well as USH3-like patients with late onset of progressive deafness and variable vestibular dysfunction (Liu et al., 1999b) indicated altogether 33 distinct disease-causing mutations.

The 2299delG mutation in the *USH2A* gene seems to be extremely common in many populations, in southern and northern Europe, in South America, in South Africa and in China, and may account for a substantial proportion of retinitis pigmentosa in humans (Beneyto et al., 2000; Dreyer et al., 2001; Eudy et al., 1998; Leroy et al., 2001; Liu et al., 1999b; Rivolta et al., 2000; Weston et al., 2000). In the haplotype analyses of 116 unrelated patients with USH2 from 14 countries, Dreyer et al. (2001) found that all 2299delG chromosomes are in complete association with one specific core haplotype. Their data strongly suggest that the widespread geographic distribution of 2299delG is the result of an ancestral mutation that has spread throughout Europe as a result of migration, instead of being a mutational hotspot in *USH2A* (Eudy et al., 1998; Liu et al., 1999b).

Although most USH2 families show linkage to 1q41, evidence for genetic heterogeneity was showed in non-consanguineous families from northern European origin (Pieke Dahl et al., 1993) and in one consanguineous family from Tunisia (Hmani et al., 1999). The locus for USH2B was localized to chromosome 5q (Pieke-Dahl et al., 2000), whereas the Tunisian variant USH2C was assigned to chromosome 3p23 (Hmani et al., 1999). So far, the genes coding *USH2B* and *USH2C* have not been identified.

### 5.2.3. Usher syndrome type 3 (USH3)

Nuutila first described patients with retinitis pigmentosa and progressive sensorineural hearing loss, the familial dystrophia retinae pigmentosa - dysacusis syndrome (DRD), which began during puberty, in 3% of the Finnish USH patients (Nuutila, 1968; Nuutila, 1970). Although several other studies of progressive hearing loss in patients have been published in Finland (Karjalainen et al. 1983, Karjalainen et al., 1989) and elsewhere (Davenport and Omenn, 1977;

Gorlin et al., 1979; Gröndahl and Mjöen, 1986; Karp and Santore, 1983), the third clinically and genetically distinct entity of USH was not generally accepted (Smith et al., 1994) until Sankila et al. (1995) excluded by linkage four other chromosomal sites, to which USH had been previously mapped (Kaplan et al., 1992; Kimberling et al., 1992; Kimberling et al., 1990; Lewis et al., 1990; Smith et al., 1992), and assigned the *USH3* locus to chromosome 3q21-q25 by linkage in ten Finnish multiplex affected families. Twelve out of twenty of the USH parental chromosomes were found to share a common ancestral haplotype comprising alleles at five marker loci, suggesting enrichment of a major mutation in Finland. Multipoint analyses assigned *USH3* within the 5-cM genetic region between the polymorphic DNA markers *D3S1555* and *D3S1279*. Moreover, Pakarinen et al. (1995a) showed that USH3 with progressive hearing loss is the most common form of USH in Finland and covers 40% of all the studied 229 USH cases. These Finnish patients originated mainly from the eastern part of Finland suggesting enrichment of the putative *USH3* mutation. Later, other USH families showing linkage to the *USH3* region have also been reported from the US and Sweden (Kimberling et al., 1995), Spain (Espinos et al., 1998b), Israel (Adato et al., 1999), and Italy (Gasparini et al., 1998).

### 5.2.3.1. Clinical manifestations of USH3

Usher syndrome type 3 (*USH3*; MIM No. 276902) is an autosomal recessively inherited disorder characterized by congenital, childhood or adulthood onset of progressive hearing loss, progressive retinitis pigmentosa and decreased vestibular dysfunction in some patients. The main clinical features of USH3 are summarized on the basis of detailed clinical analysis of 42 Finnish USH3 patients from 23 families (Pakarinen et al. 1995b; Pakarinen et al., 1996; Pakarinen, 1997), whose diagnosis was confirmed by earlier linkage and haplotype analysis (Sankila et al., 1995; see also this study). The clinical features of USH3 were described as follows: hearing impairment is usually postlingual and develops by the age of eight years, some cases of congenital or adulthood onset also occurs. The hearing loss is progressive, sensorineural and cochlear resulting in moderate to severe deafness within a few years or decades. The typical audiometric curve is slightly down sloping or U-shaped. Usually patients benefit from hearing aids and communicate verbally, but later as the hearing loss progresses, the method of communication is sign language in conjugation with lip reading. Cochlear implantation is a suitable treatment for USH3 patients with profound hearing loss. The diagnosis for retinitis pigmentosa is made usually by

electroretinogram (ERG) and typical fundus findings; thin vessels, granular pigmentation in the midperiphery of the retina, clusters of pigment starting from the nasal peripheric retina or rough pigment epithelium. RP is diagnosed at the mean age of 17 years. The symptoms of night-blindness, followed by mid peripheral visual field defects, lead to narrowed visual fields and to slowly progressive tunnel vision at a mean age of 30 years. The progression of RP by the age of 20-30 years results in a severe visual handicap. High degree hypermetropia together with concurrent astigmatism has been found in 40% of the patients. Patients have either normal or decreased vestibular function. Mental retardation or psychiatric illness have been reported in USH patients (Carvill, 2001; Hallgren, 1959; Karjalainen et al., 1983; Nuutila, 1970; Sharp et al., 1994; Waldeck et al., 2001), but they are unlikely to be part of the USH, since non-USH members of the families also had similar mental symptoms. Pakarinen et al. (1995b) reported three USH3 families with lowered intelligence and psychiatric symptoms or schizophrenia-like psychosis, but most USH3 patients had normal intelligence.

Based on the study by Shinkawa and Nadol (1986), the histopathology of the inner ear of an USH3 patient with progressive loss of hearing and vision showed abnormalities in the cochlea, such as loss of hair cells in the basal turn of the organ of Corti, widespread neural degeneration accompanied by severe loss of spiral ganglion cells and degeneration of supporting cells filled with vacuoles of degenerative cytoplasmic proteins.

On clinical grounds, an early and accurate diagnosis, as well as differentiation of USH3 from the other USH types, is difficult. RP does not differ from the other Usher types, and the type of hearing loss and vestibular function are the distinguishing features. Also, the progression of hearing loss is sometimes difficult to detect, since it can be prelingual, slow, or late onset (Pakarinen et al., 1995b; Pakarinen, 1997).

## 5.3. *Shaker-1* as a mouse model for pathogenesis of USH1B
Knowledge of the pathogenesis of heterogeneous USH has been mainly obtained from the anomalies observed in deafness mouse models. Moreover, identification of genes involved in human deafness has been facilitated by first cloning the causative orthologous mouse genes (Alagramam et al., 2001a; Bork et al., 2001; Gibson et al., 1995; Verpy et al., 2000). The expression patterns of the disease gene, and subcellular localization of the encoded protein probably have been studied most extensively in the *sh-1* mouse, which was first described already

in 1929 by Lord and Gates. *Sh-1* suffers from symptoms associated with inner-ear defects: vestibular dysfunction causing hyperactivity, head-tossing and circling, combined with neuroepithelial abnormalities and degeneration of the organ of Corti due to a defective *myosin VIIA* gene (Gibson et al., 1995; Weil et al., 1995). Both the human USH1 and the mouse *sh-1* phenotypes involve sensorineural hearing loss and vestibular dysfunction. In the mouse cochlea and vestibular tissues, *Myo7a* is expressed exclusively very early in sensory hair cell development (el-Amraoui et al., 1996). It colocalizes with cross-links between stereocilia, to the cuticular plate, and between the adherens junction and the cuticular plate, known as pericuticular necklace (el-Amraoui et al., 1996; Hasson et al., 1995; Hasson et al., 1997a; Hasson et al., 1997b; Self et al., 1998). In the retina, *Myo7a* is present in both the RPE and the photoreceptor cells, especially in the connecting cilium joining the inner and outer segments of the photoreceptors (el-Amraoui et al., 1996; Hasson et al., 1995; Liu et al., 1997a; Liu and Willems, 2001). The complete sequence and at least seven different mutations in *sh-1* encoding Myo7a have been determined in detail (Hasson et al., 1997b; Mburu et al., 1997).

### 5.3.1. Myosin VIIA function in cochlea

The exact role of myosin VIIA in sensory hair cell function is still unknown while the study by Self et al. (1998) with three *sh-1* mutants has expedited the understanding of both the development and function of cochlear hair cells. Most of the *sh-1* mutants show normal growth and heights of stereocilia, but the kinocilium is misplaced, and the stereocilia bundles are progressively disorganized leading to degeneration of hair cells (Gibson et al., 1995; Self et al., 1998). The *sh-1* mice exhibit degeneration of the organ of Corti and the afferent neurons during the first weeks after birth (Hasson et al., 1995). According to Hasson et al., the defect lies within the hair cells, and the degeneration of the neurons is secondary.

The location of myosin VIIA along the length of stereocilia and to the lower end, near the bases of stereocilia in hair cells (Hasson et al., 1995; Hasson et al., 1997a; Hasson et al., 1997b; Self et al., 1998), is consistent with the formation of a protein complex, which joins adjacent stereocilia to their neighbours and gives a precise arrangement of the stereocilia. Myosin VIIA has been proposed to anchor these ankle links to the actin cytoskeleton through a complex of ubiquitous transmembrane protein, vezatin, and a component of the ankle link protein, a PDZ domain-containing protein (Küssel-Andermann et al., 2000) (Figure 3). Recently, a gene

encoding a new PDZ domain-containing protein, harmonin, has been discovered to underlie USH1C (Bitner-Glindzicz et al., 2000; Verpy et al., 2000). Interaction of vezatin with myosin VIIA in controlling the tension of stereocilia may be mediated by $Ca^{2+}$-dependent cell-cell adhesion molecules, for example cadherins, which also may be involved in binding the lateral links or tip links of stereocilia (Küssel-Andermann et al., 2000; Müller and Littlewood-Evans, 2001). Interestingly, mutations in *cadherin 23* (*CDH23*) and *protocadherin 15* (*PCDH15*) have been shown to underlie human USH1D, DFNB12, and USH1F (Ahmed et al., 2001; Alagramam et al., 2001b; Bolz et al., 2001; Bork et al., 2001). The exact function of these anchoring proteins is not clear, although it has been suggested that they have an important role in controlling the location of myosin VIIA along the stereocilia either by preventing myosin VIIA from moving up actin filaments or inhibit the enzymatic activity of myosin VIIA (Hasson et al., 1997a).

Mutations in *Myo7a* have been demonstrated to slack the gating springs in resting stereocilia and to close the transduction channels (Richardson et al., 1997). Thus, a possible role for myosin VIIA in adjusting the tension of the tip-link and/or transduction channel complex has been suggested (Steel and Kros, 2001).

Stereocilia rootlets in the hair cell are anchored by the actin-rich structure, cuticular plate. The pericuticular necklace of the hair cell is rich in membrane vesicles and it can serve as a reservoir of components needed in cuticular plate and stereocilia (Hasson et al., 1997a). Self et al. (1998) demonstrated that *sh-1* mutants show irregular bulges of the cuticular plate as they are interrupted by areas of vesicle-rich cytoplasm. A possible role for myosin VIIA at adherens junctions is to stabilize the cuticular plate linked to rootlets of stereocilia through the interactions with vezatin, which also localizes at the junctions between hair cells and supporting cells (Küssel-Andermann et al., 2000). Furthermore, myosin VIIA may be involved in microtubule-bound vesicular traffic or membrane uptake events in the pericuticular necklace supported by the findings that the hair cells of *sh-1* mice have a defect in aminoglycoside uptake (Richardson et al., 1997).
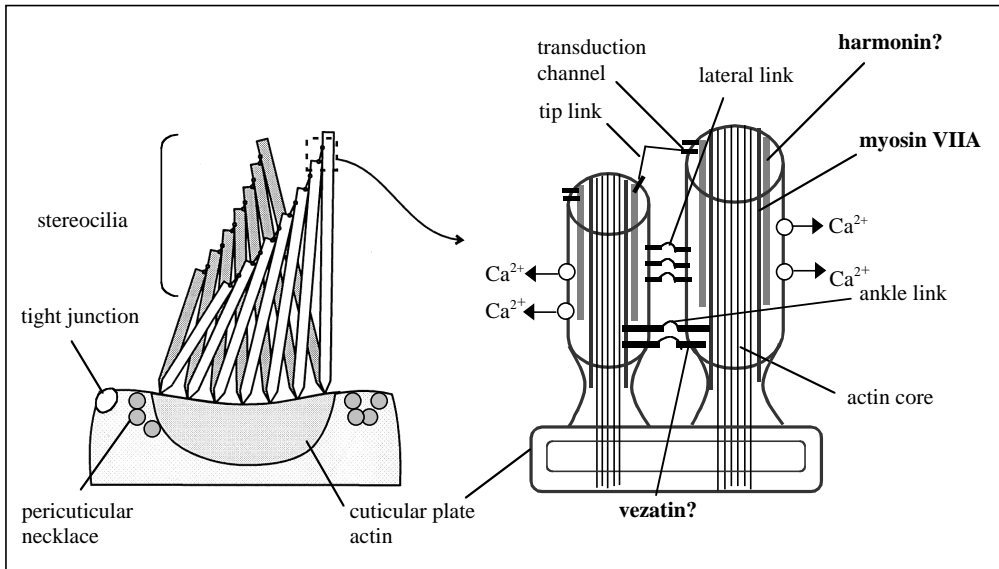
stereocilia

tight junction

pericuticular
necklace

cuticular plate
actin

transduction
channel

tip link

lateral link

harmonin?

myosin VIIA

$Ca^{2+}$

$Ca^{2+}$

$Ca^{2+}$

$Ca^{2+}$

$Ca^{2+}$

ankle link

actin core

vezatin?

**Figure 3**. Schematic picture of apical surface of a hair cell and its stereocilia. Localization of myosin VIIA, harmonin (speculative), and vezatin (speculative) is shown. Source García-Añoveros and Corey (1997), and Steel and Kros (2001).

### 5.3.2. Myosin VIIA function in retina

In retinitis pigmentosa, the sensory cells in the retina eventually die (Rattner et al., 1999). While USH1B patients manifest progressive RP leading to blindness, none of the known *sh-1* allelic mutants appear to develop retinopathy (Gibson et al., 1995; Weil et al., 1995). The reason why retinal degeneration is observed only in humans is still unknown. The mutations in *MYO7A* underlying USH1B may be different from those found in the mutant mouse *Myo7a*, and the genetic background may play an important role in the severity of the disease (Self et al., 1998). Recently, Libby and Steel (2001) showed evidence for a physiological abnormality in the retinas of several different *Myo7a* mutant mice by ERG. As in humans associated with USH1 (Seeliger et al., 2000), mutations in mouse *Myo7a* were found to lead to decreased ERG a-wave- (resulting from the photoreceptors' response to light) and b-wave (resulting from retinal interneurons activated by photoreceptors) amplitudes, suggesting that photoreceptors in mutant mice are not responding properly to light. However, the retinal dysfunction seems to be similar in all USH1 cases, since no statistically significant difference in visual function between USH1 patients with or without pathogenic *MYO7A* mutations was seen in the study by Bharadwaj et al. (2000).

In a study of myosin VIIA expression in developing mouse and human eye, myosin VIIA has been immunolocalized in the apical surface of RPE and the photoreceptor cells (el-Amraoui et al., 1996; Hasson et al., 1995; Liu et al., 1997a). In the first study of retinal phenotype associated with mutant myosin VIIA, Liu et al. (1998a) showed that myosin VIIA is required for correct distribution of the melanosomes in the *sh-1* mouse RPE. However, they concluded that this organelle transport in itself does not seem to be critical for retinal viability, since other mice with melanosome abnormalities do not normally undergo retinal degeneration.

Titus (1999) studied myosin VIIA distribution within apical microvilli of RPE in myosin VIIA knockout amoebae *Dictyostelium discoideum*, and suggested that myosin VIIA has a possible role in phagocytosis of the outer segments of exfoliated photoreceptors. Phagocytosis has not reported in *sh-1* mice, however, the mutant *Dictyostelium discoideum* showed an 80% decrease in the phagocytosis of particles, such as yeast, bacteria or latex beads. The defect in phagocytosis is neither because of altered particle binding nor inability to form actin-filled phagocytic cups, rather it appears to lie in actin-dependent particle internalization, where myosin VIIA is suggested to play an important role (Titus, 1999).

In photoreceptors, myosin VIIA localizes in the connecting cilium joining the inner and outer segments (Liu et al., 1997a; Liu and Williams, 2001). In addition, myosin VIIA has been detected in the synaptic endings of human photoreceptor cells, where it could be involved in the formation and trafficking of the vesicle complexes (el-Amraoui et al., 1996). The appearance of myosin VIIA in the connecting cilium of photoreceptors suggests that it might be involved in the visual cascade by transporting phototransductive proteins, such as opsin, through the connecting cilium into the outer segment of the photoreceptor. This is supported by electron immunomicroscopic studies of *sh-1* mice photoreceptor cells, which showed that the mutant *Myo7a* results in abnormal accumulation of opsin in the connecting cilium (Liu et al., 1999a). However, Liu et al. did not detect a complete blockage of transport from the site of opsin synthesis in the inner segment to the outer segment. Mutant myosin VIIA did not affect either the distribution of other outer segment proteins, indicating that the abnormal accumulation of opsin is not the result of a general defect affecting the transport of all proteins (Liu et al., 1999a). Further work by Liu and Willems (2001) showed that coincident onset of expression of opsin and myosin VIIA in the mouse cilium was consistent with the role of myosin VIIA in the transport of opsin. Abnormal opsin transport over a prolonged period thus might contribute to blindness in USH.

# AIMS OF THE PRESENT STUDY

The purpose of this study was to identify the gene and the disease causing mutation(s) underlying Usher syndrome type 3 (USH3) to provide a basis for understanding the molecular pathogenesis of the disease and to enable exact and even presymptomatic diagnosis of the disease.

This work was initiated when the chromosomal localization of *USH3* had been assigned to chromosome 3q21-q25 (Sankila et al., 1995). Thereafter, the positional cloning strategy was undertaken to:

1. Further delineate the chromosomal region of *USH3* by genetic mapping using an extended number of families with a denser map of markers.
2. Establish a physical map across the critical region and construct a contig of genomic clones providing tools for isolation of gene sequences.
3. Construct a transcript map for the characterization of positional candidate genes.
4. Identify the *USH3* gene among the candidate genes, and characterize the disease causing mutations.

## MATERIALS AND METHODS

### 1. USH3 families and control individuals

In addition to the 10 multiplex USH3 families studied by Sankila et al. (1995), blood samples from 22 Finnish families, of which 4 were multiplex and 18 were simplex, three non-Finnish (1 Italian and 2 Spanish) multiplex families and two patients (1 English simplex patient, and 1 from Belgium with autosomal dominant hearing loss assigned to the same chromosomal region) were included in this study. All the Finnish USH3 patients were diagnosed and ascertained by L. Pakarinen, and the clinical findings have been described elsewhere (Pakarinen et al., 1995b; Pakarinen et al., 1996; Pakarinen, 1997; Sankila et al., 1995). Clinical and genealogical data of the Italian and Spanish families have been reported elsewhere (Espinos et al., 1998b; Gasparini et al., 1998).

Control DNA samples from 200 anonymous blood donors from Eastern (high prevalence of USH3), Western and Southern Finland (low prevalence of USH3) were provided by the Finnish Red Cross Blood Transfusion Service. Moreover, a total of 51 grandparents of Centre d'Etude du Polymorphisme Humain (CEPH) families were also studied as control.

### 2. Genealogical studies

Genealogical studies of the 32 Finnish USH3 families were performed to determine the parental and grandparental birthplaces of the patients. In addition, some families were traced back in search for common ancestors.

### 3. Genotyping and linkage analyses

Genomic DNA was extracted from leukocytes contained in 10-20 ml of peripheral venous blood using standard methods. The families were typed with a total of nine polymorphic microsatellite markers from Genéthon (Dib et al., 1996), Cooperative Human Linkage Center (Sheffield et al., 1995), and the published polymorphism for the type 1 angiotensin II receptor gene (*AGTR1*) (Davies et al., 1994). Polymorphic markers were amplified in a PCR-reaction of a total volume of

10 μl with 30 ng of genomic DNA. The amplification was carried out during 30 cycles of 94$^{O}$C 45 sec, and 55$^{O}$C 45 sec, followed by a final extension at 72$^{O}$C for 4 min. The alleles were analyzed in 6% polyacrylamide, 7 M urea denaturing gels, and visualized by silver staining (Bassam et al., 1991) or $^{32}$P-dATP labelled autoradiography.

Pairwise linkage analysis was carried out using the MLINK program of the LINKAGE program package (Lathrop et al., 1984). The allele frequencies were calculated from 43 normal parental chromosomes. The USH3 disease frequency was set to 0.001, and complete penetrance and sex average recombinations were assumed. The multipoint linkage analysis was carried out using the program VITESSE (O'Connell and Weeks, 1995). The intermarker distances used in the VITESSE and the DISMULT analyses (see below) were estimated based on the published genetic data (Dib et al., 1996; Sankila et al., 1995), and the physical and haplotype mapping results obtained in this study.

## 4. Linkage disequilibrium and haplotype analyses

Association analysis for multiple loci simultaneously was carried out using the program DISMULT version 2.1 (Terwilliger, 1995), and the confidence interval was calculated as $Z_{max}$ (±)3 lod units. The significance of the linkage disequilibrium was measured by the single locus association analysis program DISLAMB (Terwilliger, 1995), in which the analysis is based on the parameter $\lambda$ that is equal to the proportion of increase of a certain allele in disease chromosomes, relative to its population frequency. The confidence intervals for $\lambda$ were calculated as 1-lod units. In all allelic association calculations, the parental non-USH3 associated alleles were used as controls.

For the estimation of genetic distances between the close markers and *USH3*, Luria-Delbrück-based calculation was used (de la Chapelle, 1993; Hästbacka et al., 1992; Lehesjoki et al., 1993). The number of generations since founding (g) was assessed to 10-200, the mutation rate ($\mu$) of the *USH3* locus was set to 1 x 10$^{-6}$, and the *USH3* gene frequency (q) was estimated to be 0.003. Allelic association was calculated according to the formula $p_{excess}= (p_{affected}-p_{normal})/(1-p_{normal})$, where p equals the marker allele frequency. The recombination fraction between *USH3* and the marker locus, the theta value ($\theta$), was calculated using the formula $p_{excess} = (1-\mu gq^{-1})(1-$

θ)$^g$, where $(1-\mu gq^{-1})$ denotes the proportion of USH3 chromosomes carrying the same ancestral mutation.

Haplotypes of the USH3 families were constructed manually, assuming the minimum number of recombinations in each family.

## 5. Isolation of genomic clones

A panel of mega yeast artificial chromosome (YAC) clones positive for the originally linked markers across the *USH3* region (Sankila et al., 1995) were provided by the Centre d'Etude du Polymorphisme Humain (CEPH) in Paris. The clones were cultured, total yeast DNA was extracted and stored in agarose beads.

The ESTs and STSs (see below) present on the YACs were used for screening of the bacterial artificial chromosome (BAC) library (Research Genetics Inc., Huntsville, AL) by PCR. The positive colonies were cultured overnight in Luria Broth media containing the appropriate antibiotic, and DNA was extracted using Qiagen's standard protocol for Plasmid Midiprep (tip 100) kit. *NotI*-digested BAC DNA was subjected to pulsed-field electrophoretic separation to determine the insert sizes in the clones, and the separated fragments were visualized in UV-light. The sizes of individual inserts were determined relative to a high molecular weight λ ladder (New England BioLabs).

## 6. EST and STS mapping

The YAC and BAC clones were screened for the presence or absence of ESTs and STSs obtained from the Chromosome 3 Database (The University of Texas Health Science Center at San Antonio; http://mars.uthscsa.edu; (Naylor et al., 1996)), the Genome Data Base (Cuticchia et al., 1993), the database at the Whitehead Institute for Biomedical Research/MIT Center for Genome Research (http://www-genome.wi.mit.edu), and the National Center for Biotechnology (http://www.ncbi.nlm.nih.gov/SCIENCE96/ and /genemap98/) by PCR amplification using the conditions described earlier. The PCR primers used in the physical mapping were designed using the program Primer3 (www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi).

New STSs were generated from the BAC-ends by sequencing with T7 and SP6 primers. The sequences obtained were screened against a library of repetitive elements using the program Repeat Masker2 at the University of Washington Genome Center (http://www.genome.washington.edu/UWGC/). Primer pairs were designed from the end-sequences, except for when the sequence contained repetitive elements. The STSs generated from the BAC-ends were used to further screen clones from the BAC libraries and to determine the orientation of BACs in the contigs by PCR.

## 7. Generation and mapping of novel polymorphic markers

Novel polymorphic $(CA)_n$-repeat markers were isolated by subcloning BAC clones into the *pUC18* vector (Pharmacia, Biotech). BAC DNA was digested with *RsaI*, ligated to *SmaI*-cut *pUC18* vector, transformed into *E. Coli* SURE competent cells (Stratagene), and plated onto selective agar plates. Putative polymorphisms were screened with $[\gamma\text{-}^{32}P]$dATP radiolabelled $(GT)_{16}$ oligonucleotide probes. Double positive clones were selected, and DNA was extracted using Qiagen's Plasmid Miniprep Kit following the manufacturer's protocol. Positive clones were sequenced using T7 and SP6 primers. Primers flanking an identified $(CA)_n$-repeat were designed and used in genotyping by PCR. The allele sizes and frequencies for the novel markers were calculated from 48 non-USH3 associated chromosomes.

Through genomic sequencing of a positional candidate gene, *NOPAR*, (see below), the novel single nucleotide polymorphisms, TSNP1 and RSNP2, were identified. Intronic forward and reverse primers were designed and used to genotype the USH3 families by PCR. The alleles were detected either by direct sequencing or by SSCP analysis in 0.7 x MDE gel for 18 h, 5 W at room temperature, and they were visualized by silver staining.

## 8. Genetic mapping of the *USH3* mouse homologue

The mouse EST sequence fragment r75057, homologous to the 3' UTR of human *profilin-2* (*PFN2*) (EMBL accession no. R75057), was used in linkage analysis of the European Collaborative Interspecific Backcross (EUCIB) mapping resource comprising 982 backcross progeny (Breen et al., 1994). DNA samples were provided by the MRC HGMP Resource Centre

(Hinxton, Cambridge). For PCR reactions, 2 ng of mouse genomic DNA was used as a template for a 25-$\mu$l PCR reaction. The cycling conditions were 94$^O$C for 4 min, followed by 35 cycles of: 93$^O$C 3 sec, 53$^O$C 10 sec and 72$^O$C 15 sec. For SSCP analysis, PCR samples were electrophoresed overnight at 4$^O$C on an 8% non-denaturing polyacrylamide gel, and the alleles were visualized using silver staining. The linkage analysis was computed using the *MBx* database supporting EUCIB (Breen et al., 1994).

## 9. Large-scale sequencing and contig assembly

The complete genomic sequence of the *USH3* region was constructed by shotgun cloning of the BAC clones. Shotgun libraries were prepared and plasmid DNA was extracted by Seqwright (Houston, TX) and GATC GmbH Shotgun Library Construction services (Konstanz, Germany). The individual BAC subclones were sequenced from both directions using a PE9700 Thermal Cycler (Perkin Elmer).

To assemble the genomic contigs, all sequence readings were edited and filtered using the GASP software (http://www.genome.washington.edu/UWGC/analysistools) in a UNIX-based system. The genomic 149-kb contig was assembled based on the fragment order on the physical map using the program Sequencer 3, whereas the 207-kb contig was assembled using Phil Green's Phred, Phrap and Consed sequence assembly programs (http://www.phrap.org). To find additional BACs or P1-derived artificial clones mapping in the same region, the contig sequences were compared with High-Throughput Genome Sequence (HTGS) and Genomic Survey Sequence (GSS) databases at NCBI. The orientation of the contigs from centromere to telomere was determined based on STS mapping.

## 10. Database analyses

The assembled genomic sequence contigs were screened for repetitive elements, and introduced to BLAST (http://www.ncbi.nlm.nih.gov/BLAST) programs against different databases, as well as to HINT (Human Index for Nonredundant Transcripts) (http://pandora.med.ohio-state.edu/HINT), to UNIGENE (http://www.ncbi.nlm.nih.gov) and to TIGR's Human Gene Index

(http://www.tigr.org/tdb/hgi). To compare orthologous mouse genomic sequences, homology searches were performed using Celera Assembled and Annotated Mouse Genome (http://publication.celera.com/). The gene and exon predictions were carried out using programs GRAIL (Uberbacher and Mural, 1991) and GENSCAN (Burge and Karlin, 1997), as well as automated sequence annotation softwares Genotator (Harris, 1997) and NIX (http://www.hgmp.mrc.ac.uk).

## 11. Isolation and mutation analysis of the positional candidate genes

Positional candidate genes were isolated as full-length cDNAs with rapid amplification of cDNA ends (RACE-PCR) and RT-PCR using the Marathon Ready Human Retina cDNA library and human placenta or retina RNA preparations. Primers for the mutation analyses of the novel genes were designed and used in the PCR amplification. The PCR fragments were analyzed by SSCP analysis or by direct sequencing of all exonic regions with adjacent splice sites from patients' genomic DNA and when possible, from total and poly(A) RNA from EBV-transformed lymphoblasts extracted from two USH3 patients, two carriers, and two normal individuals. The SSCP alleles were separated in 0.7 x MDE gel for 18 h, 5 W at room temperature and visualized by silver staining. The purified amplification products (PCR purification Kit, Qiagen) were sequenced using ABI 377-sequencer (Perkin Elmer).

## 12. Isolation and characterization of the *USH3* gene

The predicted exons and the EST 35f2 were combined by PCR amplification from human retina cDNA and Marathon Ready Human Retina cDNA library (Clontech) using the interexon, 5'- and 3'-UTR primers. The exon-intron boundaries were established comparing the *USH3* cDNA with the genomic sequence obtained from databases. For the mutation analysis, the genomic fragments encompassing *USH3* exons were amplified with intronic primers. Mutations were screened by direct sequencing or by SSCP according to the protocol described earlier.

Predictions for regulatory elements and promoters at the 5'-flanking genomic sequences of *USH3* were performed using MatInspector/TRANSFAC and Neural Network Promoter Input programs at Search Launcher (http://searchlauncher.bcm.tmc.edu/). Protein predictions were

carried out using softwares PIX (http://www.hgmp.mrc.ac.uk) and LASERGENE (DNASTAR Ltd, UK), and programs Tmpred (http://www.ch.embnet.org/software/TMPRED_form.html), SignalP (http://www.cbs.dtu.dk/services/SignalP/) and PSORTII (http://psort.nibb.ac.jp/).

## 13. Northern hybridizations

For northern blots, 2 µg of poly(A)RNA or 10 µg of total RNA were extracted from lymphoblasts of USH3 patients, carriers, and controls, and the samples were electrophoresed on a 1% denaturing formaldehyde-agarose gel, followed by transfer to a nylon membrane. The membrane and the commercial human or mouse multiple tissue northern blots were hybridized with [$\alpha^{32}$P]dCTP random-labelled cDNAs derived from different positional candidate genes as probes according to manufacturer's protocols. The blots were hybridized with an β-actin probe as a control.

## 14. In situ expression studies

A 262-bp PCR fragment homologous to *USH3* was amplified from Marathon Ready Mouse Brain cDNA library (Clontech) and cloned into a pGEM-T Easy vector (Promega) according to the manufacturer's instructions. The synthesis of cRNA probes and *in situ* hybridization were done as described by Wilkinson and Green (1990). The 4% paraformaldehyde –fixed paraffin sections from an embryonic day 16 mouse and from dissected adult mouse eyes and inner ears were hybridized with *in vitro* transcribed ^{35}S-labelled antisense and control-sense cRNA probes.

**RESULTS AND DISCUSSION**

## 1. Genetic refinement of the *USH3* locus by linkage analyses (I)

The efforts towards the positional cloning of the *USH3* gene were started based on the initial assignment of *USH3* to a 5-cM interval between the markers *D3S1555* and *D3S1279* by linkage analyses in ten USH3 families (Sankila et al., 1995). The panel of families genotyped was enlarged from 10 to 32 families with nine polymorphic DNA markers across the critical region. Two-point linkage analysis with *USH3* and each of the markers showed the highest lod score of 9.84 at *D3S1299* ($\theta$ = 0.0) between markers *D3S1555* and *D3S3625*. In a six-point linkage analysis between *USH3* and five linked marker loci *D3S1555*, *D3S1308*, *D3S2401*, *D3S1299*, and *D3S3625*, a maximum six-point lod score of 13.89 was obtained at *D3S1299* assigning *USH3* to a 4-cM region between *D3S1555* and a novel marker *D3S3625*.

## 2. Further refinement of the *USH3* locus to a 1-cM region by linkage dis-equilibrium (I, II)

To further refine the localization of the *USH3* gene, linkage disequilibrium analysis was performed utilizing the program DISMULT with nine markers at intermarker distances (in kilobase pairs) of cen-*D3S3661*-(500)-*D3S1555*-(1000)-*D3S1308*-(1000)-*D3S2401*-(1000)-*D3S1299*-(1000)-*D3S3625*-(1000)-*D3S1594*-(500)-*D3S1279*-(2000)-*D3S1746*-tel. Significant linkage disequilibrium was detected with all the markers analyzed, and the maximum likelihood estimate ($Z_{max}$ = 57.02) for the location of *USH3* was obtained between markers *D3S3625* and *D3S1594* at 0.4 cM proximal to *D3S1594*. As the confidence interval was calculated as $Z_{max}$ ($\pm$3) lod units, the *USH3* genomic region was shown to span ~1.3 cM around the loci *D3S3625* and *D3S1594*.

For quantitative estimations of genetic distances between *USH3* and five close marker loci *D3S2401*, *D3S1299*, *D3S3625*, *D3S1594*, and *D3S1279*, calculations based on Luria-Delbrück formula were performed. The mutation age was estimated based on an observation of a conserved ancestral disease-associated haplotype spanning ~6.5 cM around the *USH3* locus (see below), and

of geographical distribution of the grandparental birthplaces in the middle parts of Finland. Assuming 20 generations (g) of expansion of the *USH3* mutation, the *USH3* gene frequency (q) of 0.003, and a mutation rate ($\mu$) of 1 x $10^{-6}$, the Luria-Delbrück-based analysis suggested that *USH3* lies within less than 0.5 cM from the loci *D3S3625* and *D3S1594*. The calculated $P_{excess}$ values for markers *D3S1299*, *D3S3625*, and *D3S1594* were as 0.83, 0.94, and 0.94, respectively. Finally, consistent with the above results, the single locus association program DISLAMB (Terwilliger, 1995) with three novel microsatellites, 25B8CA1, 25B8CA2, and 107G19CA7, (see below), detected the strongest allelic association at the locus 107G19CA7 ($\lambda$-value 0.973 with p-value < $10^{-15}$) placing *USH3* between the markers *D3S1299* and *D3S3625*.

## 3. Haplotypes suggest only one ancestral *USH3* mutation in Finland (I)

The haplotype analysis by Sankila et al. (1995) suggested a major ancestral founding USH3 haplotype in Finland. Of the 20 parental haplotypes studied, 12 chromosomes carried the same alleles at five marker loci. However, in one family, the disease haplotype of the mother differed from the major USH3 haplotype suggesting another mutation segregating in the family. Based on these results, the disease-associated haplotypes were constructed in the 20 original and 57 additional USH3 chromosomes with nine marker loci. The ancestral haplotype was present in 37 of 77 USH3 bearing chromosomes spanning ~6.5 cM around the *USH3* locus. The other haplotypes suggested historical crossovers placing *USH3* between the markers *D3S1299* and *D3S3625*, none of the haplotypes being present in non-USH3 chromosomes. With this denser map of markers, the dissimilar haplotype previously observed in one of the families (Sankila et al., 1995), was found identical to the ancestral haplotype at markers *D3S3625* and *D3S1594*, and was thus thought to represent the major mutation. A similar haplotype was observed in another novel family as well. According to historical recombinations suggested by haplotypes, the location of the Finnish *USH3* mutation could be narrowed to an approximately 1-cM interval between the markers *D3S1299* and *D3S3625*.

# 4. Construction of the physical map across the critical *USH3* region

## 4.1. Novel STSs and ESTs allow joining of the genomic clones into contiguous contigs and refining of the critical region (I, II)

In order to map new polymorphisms and to identify positional candidate genes for USH3, we constructed a genomic YAC contig across the region between the markers *D3S1308* and *D3S1279*. Altogether 20 partly overlapping CEPH megaYACs covering approximately an 8.5-cM genomic region were ordered based on the known genetic order of the polymorphic markers used in the linkage studies. Seven new polymorphic and four non-polymorphic STSs, as well as three ESTs, carboxypeptidase A3 (*CPA3*), angiotensin II receptor gene (*AGTR1*) (Davies et al., 1994), and *D3S1319E*, previously assigned to the proximity of the *USH3* locus defined by cytogenetic breakpoints of somatic cell hybrids (Naylor et al., 1996), were localized within the contig. We assigned *CPA3* centromeric to the marker *D3S1299*, and it was therefore excluded as *USH3* positionally. *AGTR1* was excluded by segregation analysis ($Z_{max}$ = 3.12; $\theta$ = 0.08), in accordance with its physical localization between *D3S1308* and *D3S1555* in the contig. Although we localized *D3S1319E* centromeric to the marker *D3S1299*, based on the homology to *profilin-2* (*PFN2*), and the regulatory role of profilins in actin polymerization (Theriot and Mitchison, 1993), it was considered to be a good functional candidate for USH3 (see below). Two ESTs, six genes, and two STSs previously mapped to chromosome 3q (Naylor et al., 1996), were excluded from the YAC contig positionally: *D3S1324E*, *D3S1492E*, cellular retinol-binding protein 1 (*CRBP1*), ribophorin I (*RPN1*), membrane metalloendopeptidase (*MME*), phosphatidylinositol 3-kinase (*PIK3CA*) (Volinia et al., 1994), neuronally expressed *EPH*-related tyrosine kinase (*NET*) (Tang et al., 1995), myosin light chain kinase (*MLCK*) (Potier et al., 1995), *D3S1111*, and *D3S1978*.

To further refine the minimum critical *USH3* interval, and to isolate additional candidate genes, we set out to construct a physical map consisting of BAC clones between the markers *D3S1299* and *D3S1279*. We screened the Research Genetics Human BAC library using the STSs previously mapped to the YAC contig or sequenced from the ends of the BAC clones. Altogether, 21 BAC clones and 28 novel STSs generated from the ends of those clones were placed on the physical map. Two BAC clones, 282N2 and 280J1, were found to be rearranged or chimeric. As a result, a BAC contig with one gap close to the marker *D3S1299* at the centromeric end of the

critical region was constructed. Finally, at least a five times clone coverage was obtained between a new centromeric flanking marker 107G19CA7 (see below) and the telomeric marker *D3S3625*.

Initially, two BAC clones (25B8 and 545J16) were selected for screening novel CA-repeats by hybridizing with the oligonucleotide $(GT)_{16}$ as a probe. As a result, two polymorphic markers, 25B8CA1 and 25B8CA2, were obtained from the BAC clone 25B8. After identifying several uninformative CA-repeats, and isolating of additional BAC clones, the third microsatellite marker, 107G19CA7, was found in the clone 107G19. The physical order of the new markers between the flanking markers *D3S1299* and *D3S3625*, cen-25B8CA2-25B8CA1-107G19CA7-tel, was confirmed by STS content mapping of the BACs and supported by the results of the haplotype analysis. Two USH3 chromosomes bearing ancestral alleles solely at two markers (see previous results) revealed a different allele at 107G19CA7, which was thus thought to represent a novel centromeric boundary of the *USH3* locus allowing further narrowing of the critical interval (Table 4). Based on the sizes of the BACs determined by pulsed-field electrophoresis, the *USH3* critical region was approximately 250 kb bordered by the markers 107G19CA7 and *D3S3625* (Figure 5.).

**Table 4**. Haplotypes and $p_{excess}$ values associated with the *USH3* locus. The conserved haplotype is boxed.

| Marker cen | $p_{excess}$ | No. of chrs with founder allele USH3/ctrl | Haplotypes (cen→tel) |
|---|---|---|---|
| D3S3661 | 0.79 | 71/25 | 2 2 2 2 2 2 1 2 2 2 2 2 4 2 2 2 1 1 2 2 2 2 2 2 2 2 1 1 2 2 2 2 |
| D3S1555 | 0.71 | 64/17 | 3 3 3 3 3 3 3 2 1 1 1 3 2 2 3 1 1 1 3 3 3 3 3 3 3 6 3 1 1 1 3 3 |
| D3S1308 | 0.69 | 66/22 | 2 2 2 2 2 2 2 2 2 2 1 3 3 4 3 1 4 2 2 2 3 2 2 2 2 2 2 1 1 2 2 |
| D3S2401 | 0.78 | 69/19 | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 1 6 8 1 1 10 1 7 7 5 5 1 1 |
| D3S1299 | 0.83 | 69/11 | 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 5 5 5 2 6 6 2 3 3 5 5 6 6 |
| 25B8CA2 | 0.80 | 69/8 | 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 6 6 6 6 6 6 3 5 |
| 25B8CA1 | 0.95 | 76/8 | 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 6 6 5 5 |
| 107G19CA7 | 0.97 | 77/1 | 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 3 3 6 6 |
| D3S3625 | 0.94 | 76/17 | 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 6 6 |
| D3S1594 | 0.94 | 75/9 | 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 8 8 |
| D3S1279 | 0.87 | 72/16 | 4 4 4 4 4 11 4 4 4 4 6 4 4 4 4 4 4 4 4 4 4 4 4 4 4 11 3 2 2 |
| D3S1746 tel | 0.72 | 59/5 | 6 8 7 1 5 6 6 6 6 7 5 6 6 6 6 8 6 5 6 7 6 6 5 6 6 5 6 6 2 2 6 6 |
| *No. of chrs: USH3/ctrl* | | *79/48* | 36/0 1/0 5/0 1/0 3/0 1/0 1/0 1/0 1/0 1/0 1/0 2/0 1/0 2/0 1/0 1/0 1/0 2/0 1/0 1/0 1/0 1/0 1/0 2/0 1/0 2/0 1/0 1/0 1/0 2/0 1/0 |

**4.2. ESTs provide probes to localize the positional *USH3* candidate genes (I, II)**

A nucleotide BLAST search (BLASTn) (Altschul et al., 1990) with the EST *D3S1319E* revealed homologies in opposite directions of the sequence, for the *ADP-ribosylation factor 1* gene and *profilin-2* (*PFN2*), suggesting that it represents a chimeric cDNA clone containing sequences from two different genes. The sequence for *PFN2* (GenBank accession no. L10678) had been published (Honore et al., 1993), but the gene had not been localized previously. The PCR primers that we used in the physical mapping of *D3S1319E*, were derived from the *PFN2* sequences suggesting that we had localized the gene for profilin-2 to chromosome 3q. Based on the possible interactions of profilins with actins, *PFN2* was studied as a functional candidate for USH3. A 633 bp of the *PFN2* cDNA was amplified by RT-PCR from EBV-transformed lymphoblast-derived mRNA from two USH3 patients and healthy individuals. Direct sequencing of the PCR fragments did not reveal any changes. Later, it has been suggested that the PFN2 protein has two different isoforms, PFN2a and PFN2b, being involved in different biochemical events, such as neural tube development and regulation of synaptic vesicle traffic (Lambrechts et al., 2000).

Altogether seven out of fourteen ESTs derived from the Transcript Map of the Human Genome and the Gene Map 98 of the Human Genome could be further mapped by PCR to the BAC clones within the interval *D3S1299-D3S1279.* All seven ESTs, *TIGRA002B48*, *WI-11937*, *WI-11588*, *SHGC-133*, *WI-17533*, *KIAA0001*, and *D3S3882* were subjected to BLAST searches, and four of them, *WI-11588*, *SHGC-133, KIAA0001,* and *D3S3882*, were found to represent two separate genes, whose mRNA sequences were available in GenBank. These genes were considered as potential candidate genes for *USH3* and were further studied in patients.

According to BLAST searches, the ESTs *WI-11588* and *SHGC-133* were found to represent the *Homo sapiens seven-in-absentia* (*Drosophila*) *homolog 2* gene, *SIAH2* (GenBank accession no. NM_005067), which has been characterized in detail by Hu et al. (1997). *SIAH2* was an intriguing functional candidate for USH3, as the orthologous gene *seven-in-absentia* (*sina*) has been shown to perform an important role in the photoreceptor development of *Drosophila melanogaster;* loss-of-function mutations in *sina* cause the R7 photoreceptor cell to adopt a cone cell fate (Carthew and Rubin, 1990). In addition, the mouse homologue, *Siah2,* has been mapped to the middle of mouse chromosome 3 (Holloway et al., 1997), close to the same chromosomal region, on which we assigned the mouse *Pfn2* gene (see below). *SIAH2* containing a 972-bp coding region plus 100 bp of the 5'- and 317 bp of the 3'-UTR was sequenced in two

USH3 patients, two carriers and two controls from PCR amplified genomic DNA, but no alterations were found.

The ESTs *KIAA0001* and *D3S3882* were found to be 5'-end sequences of the previously described *KIAA0001* mRNA (GenBank accession no. NM_014879) (Nomura et al., 1994). The possible function of the *KIAA0001* gene is unknown, although BLAST searches suggest that the *KIAA0001* mRNA sequence represents a probable ortholog of the rat *VTR (ventral tegmentum receptor) 15-20* gene. The VTR 15-20 protein is a possible cell-surface orphan G-protein-coupled receptor expressed throughout the mammalian nervous system and in peripheral tissues including the spleen. The possible role in neuroimmune function has been suggested, although no evidence for ligand binding activity was found so far (Charlton et al., 1997). For mutation analysis of the *KIAA0001* gene in USH3 patients, PCR amplification and subsequent sequencing were performed, but no relevant sequence alterations could be detected in the 1016-bp coding region from patients' genomic DNA, while mutations in the so far uncharacterized regulatory regions could not be excluded. Northern blot analysis, using a 1010-bp amplified product from the *KIAA0001* coding region as a probe, detected a 2.4-kb transcript highly expressed in heart and placenta, and less in brain, lung, liver, skeletal muscle, kidney, and pancreas.

Based on our physical map, the ESTs *TIGRA002B48* and *WI-11937*, were excluded by location as candidates for *USH3*. An EST *WI-17533*, a positional candidate for *USH3*, was thought to represent the 3'-UTR of a putative gene with no known homology. Several attempts to isolate the full-length cDNA of the novel gene by RT-PCR and RACE PCR were performed without success. Later, the EST was shown to reside slightly outside of the critical 250-kb region and was not considered for further analysis.


### 4.3. Mouse homologue of *profilin-2* (I)

In the localization of a putative mouse homologue for *USH3*, the EUICB mapping resource was used to map an orthologous EST (r75057) sequence fragment to the *PFN2* 3'-UTR (GenBank accession no. R75057). The EST r75057 was mapped to an approximately 5-cM region of mouse chromosome 3 between the markers *D3Mit66* and *D3Mit22* by linkage and haplotype analyses. This region was known to contain two other loci, *Mme* and *Si-s*, mapping to human chromosome 3q as well, and was thereby consistent with the localization of *PFN2* in our physical map. Thus,

we defined a putative mouse *USH3* chromosomal region as a conserved segment defined by *Pfn2*, *Mme*, and *Si-s*.

# 5. From physical map to complete DNA sequence of the *USH3* region (unpublished, III)

Initially, a BAC clone 25B8 was chosen as a target for shotgun sequencing (unpublished data). The gaps between the genomic contigs were filled and partly connected by additional sequencing of plasmid DNA by primer walking from the contig end sequences and by PCR across the gaps. As a result, the genomic sequence of 148.7-kb consisting of seven contiguous sequence contigs of sizes of 0.5 kb to 57.6 kb was assembled (unpublished data). Orientation from centromere to telomere was confirmed by the marker order obtained from our physical map. Later, as a novel polymorphism 107G19CA7 was obtained from BAC 107G19, and the critical *USH3* region between the markers 107G19CA7 and *D3S3625* was suggested, we focused on studying the partly overlapping BAC clones 355L4 and 545J16, which were subcloned and assembled by methods used earlier. To extend the distal sequences, the database-derived BAC clone RP11-385G14 (GenBank accession no. AC011103) sequence was aligned with genomic contigs of the subcloned BAC 545J16. Eventually, a final sequence contig of 207 kb was assembled.

## 5.1. Full 207-kb DNA sequence allows identification of additional candidate genes

A BLASTn search of the 207-kb contig identified 30 distinct or partly overlapping human ESTs and 11 orthologous mouse or rat ESTs. The region further contained a previously identified and excluded gene, *KIAA0001,* as well as a human platelet activating receptor homolog, H963 (GenBank accession no. AF002986). A protein similarity search (BLASTx) revealed a high sequence similarity to a human OPA-containing gene, HOPA (GenBank accession no. AF132033) (Philibert et al., 1999), also called trinucleotide repeat-containing 11 gene, TNRC11, or thyroid hormone receptor associated protein, TRAP230 (GenBank accession no. NM_005120) (Ito et al., 1999). In addition, the complete sequence annotation, which was carried out using the sequence annotation tools Genotator and NIX, predicted several exons and gene specific elements on the 207-kb sequence contig. Exons predicted by at least two different programs were chosen

for further analysis. After the *in silico* analysis of the genomic region, construction of the predicted genes was undertaken by use of a combination of RT-PCR, RACE experiments, and direct PCR of cDNA libraries. In this way, we were able to isolate full-length cDNAs of two novel genes, *UCRP* (Usher critical region pseudogene) (GenBank accession no. AF388367) and *NOPAR* (no OPA-repeat) (GenBank accession no. AF388364), as well as its alternatively spliced transcript, *NOPAR2* (GenBank accession no. AF388365).

### 5.1.1. Platelet activating receptor homolog, H963

As a result of BLASTn search, a cluster of four ESTs (GenBank accession nos. AA353758, AV653286, AV653266, and BE843843) was found to represent the 5' and 3' ends of the previously described human platelet activating receptor homolog mRNA, H963. Significant alignments were also obtained with several mouse and rat ESTs indicating an orthologous rodent gene for H963. Based on its position in the critical region and the possible role as a novel G-protein-coupled receptor (Jacobs et al., 1997), H963 was selected for further analyses. Since the entire 1,272-bp cDNA sequence of H963 was available in GenBank, we were able to determine the genomic structure of the gene by aligning the published sequence with the genomic 207-kb sequence. The H963 gene spans an area of 4.9 kb and is made up of three exons of which the first two are 5' untranslated exons. We sequenced the entire coding region of H963 in USH3 patients and controls without revealing any alterations in sequence.

### 5.1.2. Usher critical region pseudogene, *UCRP*

Two novel ESTs (GenBank accession nos. R79610 and R79611) were found to represent the same cDNA clone (IMAGE no. 146349) and partly overlap another cDNA clone (IMAGE no. 60197) with two additional ESTs (GenBank accession nos. T40442 and T39157). The latter was contaminated by phage sequences and was not further analyzed. The full length 2,081 bp cDNA of the novel gene, Usher critical region pseudogene (*UCRP*), was identified by RT-PCR from human placenta cDNA by bridging the predicted exons in the vicinity of the EST R79611, and by RACE PCR from Human Marathon Ready Retina cDNA library. Although the ESTs R79610 and R79611 were derived from the 5' and 3' end of the same cDNA IMAGE clone, we were not able to amplify the EST R79610 from neither placenta cDNA nor retina cDNA library by RT-PCR, suggesting a genomic contamination in the clone. We further examined the expression pattern of

*UCRP* by hybridizing multitissue northern blots with different *UCRP* cDNA fragments. A single transcript of approximately 2.0 kb was expressed in all the tissues examined (unpublished data). Sequence alignments revealed that the *UCRP* cDNA consists of five exons spanning over 70-kb genomic DNA. However, the ORF is disrupted and thus the gene is probably an unprocessed pseudogene. Despite this assumption, the entire gene was examined by direct sequencing from USH3 patient and control DNA samples. The analyses did not reveal any differences between the samples. Furthermore, the expression pattern of USH3 patients' lymphoblast derived total RNA was normal compared to controls examined by northern blot (unpublished data).

### 5.1.3. No OPA-repeat gene, *NOPAR*

The BLASTx similarities of the 207-kb deduced amino acid sequence to the amino acid sequence of HOPA/TRAP230 prompted further analysis of a putative homologous gene in the *USH3* region. A chimeric cDNA clone (IMAGE no. 1565326) representing an EST AA931311 showed homology to HOPA, and was thought to contain part of the 3'-UTR region of the novel putative gene. Finally, two alternatively spliced transcripts, *NOPAR* and *NOPAR2*, were amplified by RT-PCR from human placenta cDNA and Marathon Ready Human Retina cDNA library by bridging the predicted exons until the 3'-UTR sequences contained in the IMAGE cDNA clone. To find out whether some sequences at the 5'-UTR were still unidentified, primer pairs were placed randomly upstream from the first identified ATG codon in the genomic sequence of the gene, and PCR was performed to the predicted exon 2. This resulted in the refining of two complete cDNAs with ORFs of 2,163 bp and 2,268 bp encoding putative polypeptide isoforms of 721 and 756 amino acids. *NOPAR2* comprised an additional exon 13b between exons 13 and 14. By comparing the genomic contig sequence with the cDNA sequences, the exon-intron boundaries were confirmed and shown to follow the GT-AG splicing rule. Exon 1 contains the first in-frame ATG codon in accordance with the Kozak consensus motif (Kozak, 1987), while the first stop codon TAA is in exon 14. No splice acceptor was found at the 5' end of exon 1, suggesting this exon to be the first. An in-frame polyadenylation signal was found 158 bp downstream from the first termination codon and was confirmed by RT-PCR from exon 13 to the end of the signal. The *NOPAR* gene spans an area of 120 kb determined by aligning the cDNA sequence with the genomic 207-kb sequence.

Promoter analysis using Neural Network Promoter Input programs predicted three possible promoters for the gene. Using various coding and 5'-UTR sequences as probes in northern analysis, abundant levels of different transcripts of sizes of 1.0 kb to 7.5 kb, and a widespread expression pattern were detected in all 16 adult tissues tested (unpublished data). It is highly likely that these different fragments represent differentially spliced or related transcripts found in human tissues. No changes in expression level or in the sizes of the transcripts compared to controls were detected, as the expression pattern of USH3 patients' lymphoblast derived total RNA or poly(A) RNA was analyzed.

NOPAR shows 61% amino acid similarity to the amino terminus of HOPA, however, it is only 1/3 the length of HOPA and does not contain a 3'OPA-repeat. HOPA contains two overlapping ligand-dependent nuclear hormone receptor recognition motifs near the amino terminus, and a highly glutamine-rich C-terminal region that results from the CAG-trinucleotide repeats, namely an OPA (opposite paired)-motif (Philibert et al., 1999). OPA-repeats have multiple possible functions depending on the location of the repeat in the gene, such as mediating protein-protein interactions (Stott et al., 1995), or regulating gene expression (Feng et al., 1995; Otten and Tapscott, 1995). The *HOPA* gene has been described by Nagase et al. (1996), and assigned to chromosome Xq13 (Philibert et al., 1998). It has also been cloned by Ito et al. (1999), as a 230-kD subunit of thyroid hormone receptor associated protein complex (TRAP) having a widespread expression pattern. The exact role of *HOPA/TRAP230* is still unknown, but it has a possible role in thyroid hormone receptor function (Ito et al., 1999). An association with mental retardation has been suggested (Philibert et al., 1998), as well. Thyroid hormone receptor-β (TR-β) has been shown to have an essential role in the development of auditory and visual systems (Ng et al., 2001; Rüsch et al., 1998). Congenital thyroid disorders impair hearing, and hypothyroidism in mice and rats causes deformities in the organ of Corti. Rüsch et al. (1998) identified a defect in inner hair cells (IHC) of the organ Corti in TR$^{\beta-/-}$ mice indicating that a TR-mediated transcriptional pathway is required for the physiological differentiation of IHCs by the control of genes encoding ion channels. In embryonic retina, TR is expressed on the outer nuclear layer, being required for the development of green cone photoreceptors. Mutations in the gene may be associated with human cone disorders (Ng et al., 2001).

Genomic sequencing of *NOPAR* and its alternatively spliced exon 13b indicated no disease causing mutations in patients' DNA, while two novel single nucleotide polymorphisms,

RSNP2 and TSNP1, were detected within its introns. Thus RSNP2 and TSNP1 were localized between markers 107G19CA7 and D3S3625; TSNP1 was telomeric to RSNP2. A haplotype analysis revealed conservation of the segment RSNP2-TSNP1 in all of the 48 USH3 chromosomes studied. This supported the previous haplotype analysis placing *USH3* telomeric to 107G19CA7. However, a recombination event in the same USH3 family, which was previously used to refine the site of the *USH3* mutation telomeric to 107G19CA7, excluded the RSNP2-TSNP1-*D3S3625*-tel segment as the site of the causative mutation, restricting the disease locus to a 60-kb region between markers 107G19CA7 and RSNP2 (Figure 4).
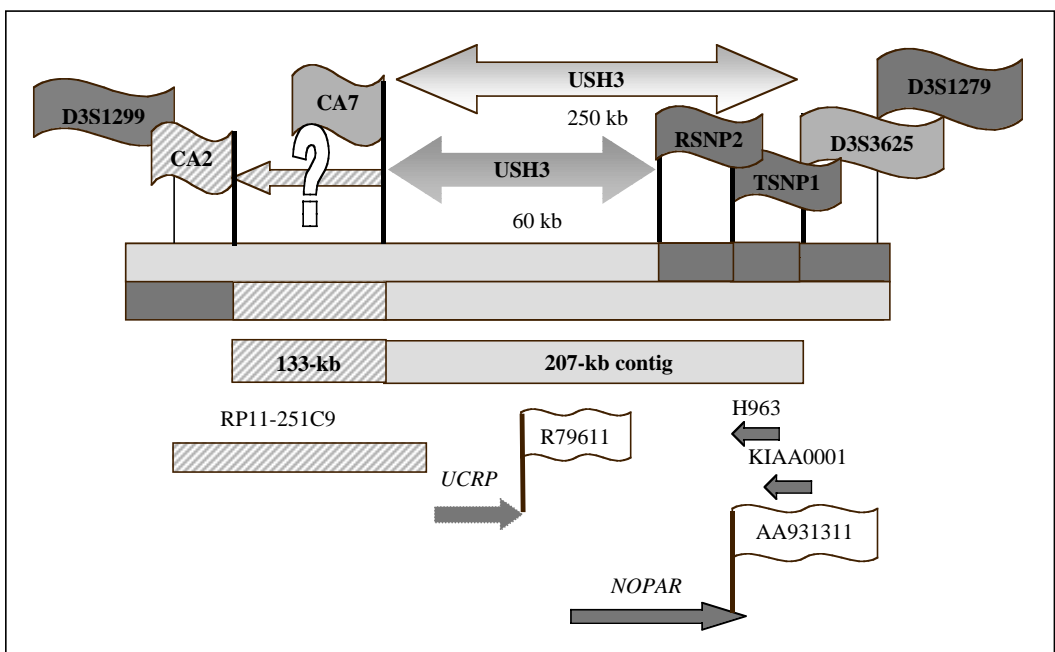


**Figure 4**. Refined mapping of the *USH3* locus between markers 107G19CA7 and RSNP2. The isolated genes (arrows) *UCRP* (Usher critical region pseudogene), *NOPAR* (No OPA-repeat gene), H963, and KIAA0001, as well as ESTs (white flags) are shown below the 207-kb contig. The question mark indicates the possibility for the *USH3* mutation to lie in the extended 133-kb interval between markers 25B8CA2 and 107G19CA7 (see results). Database-derived BAC clone RP11-251C9 is indicated with a striped box. Polymorphic markers are indicated with grey flags.

# 6. Identification of the *USH3* gene (III)

As no additional transcripts in the remaining 60-kb critical region between markers 107G19CA7 and RSNP2 were found, and the centromeric border of the critical interval at 107G19CA7 was suggested based on an assumed historical recombination observed in two disease associated chromosomes only, we considered a possibility of these two haplotypes representing another, non-founder mutation, and the common short haplotype would be a result of common alleles present in this chromosome by chance. This prompted us to extend the sequence analysis centromeric to 107G19CA7, towards 25B8CA2, which was conserved in 69/79 USH3 chromosomes, thus suggesting a more reliable historical recombination breakpoint (Table 4).

Similarity searches using the two assembled 148.7-kb and 207-kb contig sequences against HTGS and GSS databases at NCBI revealed a BAC clone RP11-251C9 (GenBank accession no. AC020636) sequences between markers 25B8CA2 and 107G19CA7. A new 133-kb contig was constructed by aligning RP11-251C9 sequences with the contigs (Figure 5), and a subsequent BLASTn analysis revealed 11 novel ESTs without any homology to known cDNAs or genes. Since one EST, 35f2 (GenBank accession no. W27577), was found to be retina-derived based on the database information, and the sequence annotation software NIX placed several exons in the vicinity of this EST (Figure 6), we performed interexon PCR in order to obtain a full-length cDNA. RT-PCR from human retina-derived total RNA preparations detected two different *USH3* transcripts. The obtained *USH3* cDNA is 1,444 bp in length, with a 5'-UTR of at least 392 bp and a 692-bp 3'-UTR (Figure 7A). The 360-bp open reading frame encodes a putative polypeptide of 120 amino acids with a calculated molecular mass of 13.4 kD, and comprises four exons. A splice variant, *USH3* gene_b, comprises an additional 87-bp exon (1b) with an inframe stop codon between exons 1 and 2. The first in-frame ATG codon in exon 1 conforms to the Kozak motif (Kozak, 1987). Computational promoter analysis revealed five possible promoters for the gene. The putative promoter of *USH3* lacks TATA boxes, but has GC and CAAT boxes, which are usually strong determinants of promoter efficiency. Three alternative polyadenylation signals were found at 156 bp, 764 bp, and 3,568 bp downstream from the termination codon, consistent with the northern blot analysis indicating three transcripts of different sizes.
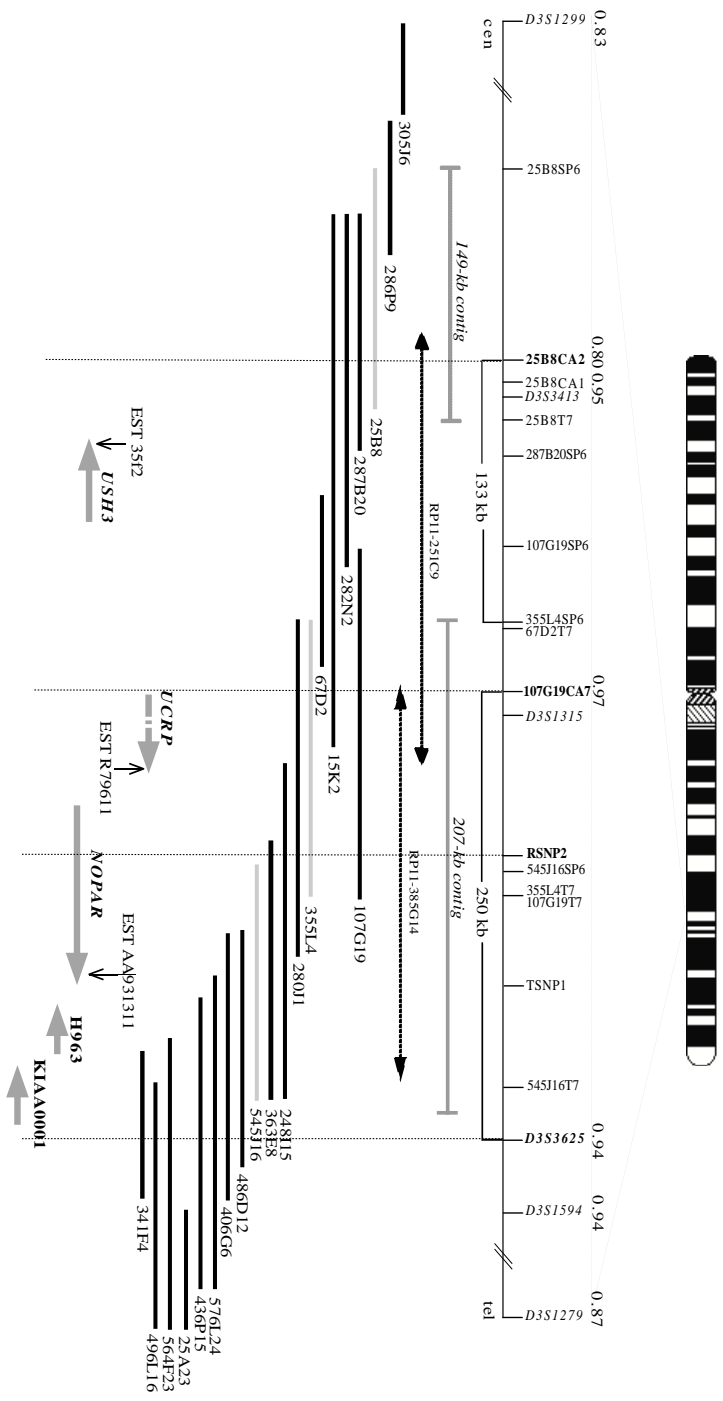
**Figure 5.** The physical and transcript map with overlapping BAC clones between markers *D3S1299* and *D3S1279* at 3q21-q25. Vertical lines indicate polymorphic markers and STSs. $P_{excess}$ values for the markers are indicated above the horizontal line. Sequenced BACs are indicated with a grey line, and database derived BAC clones are indicated with arrows. Horizontal and vertical arrows below the BAC clones show five genes and the corresponding EST sequences, respectively.
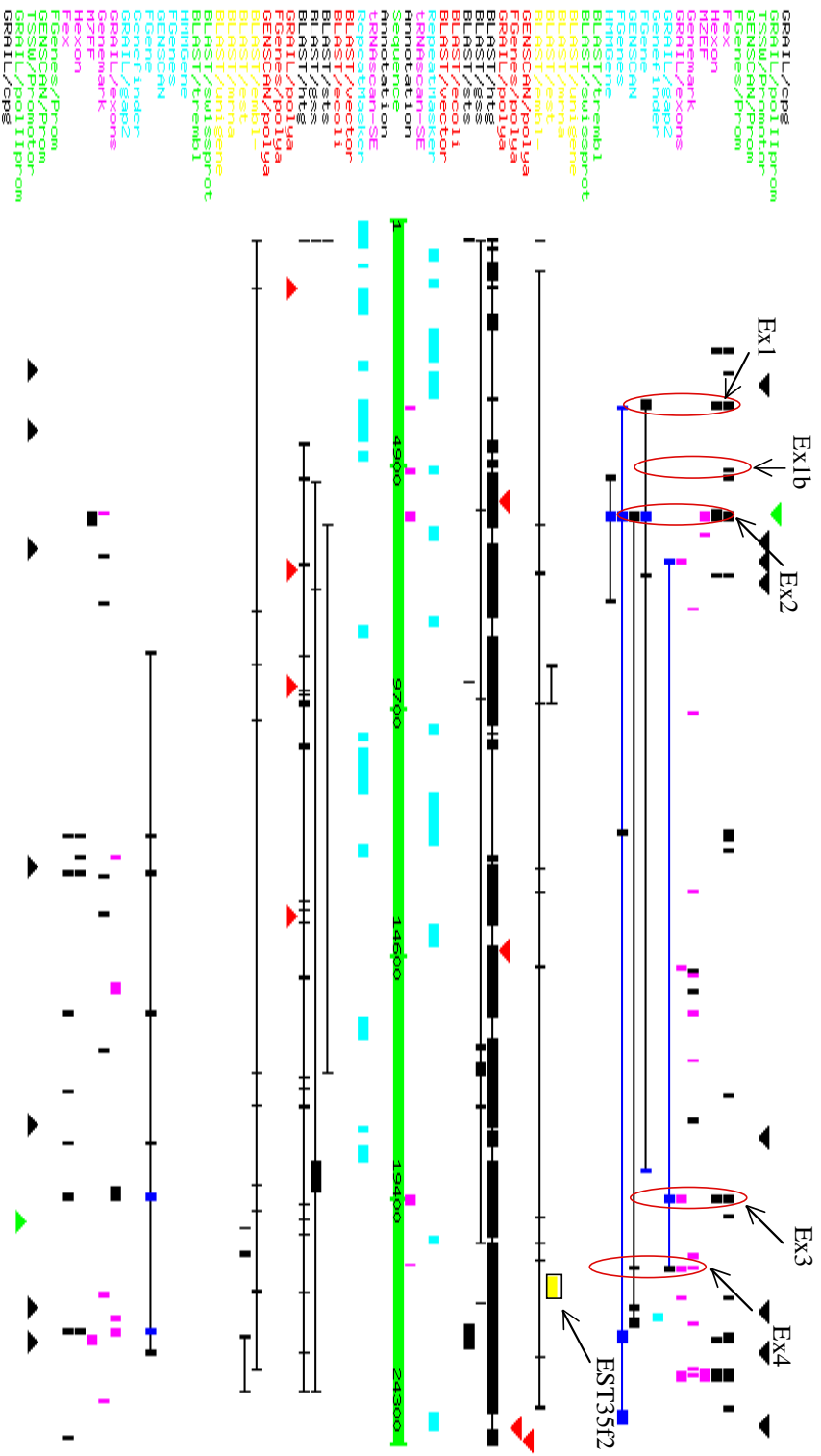
70

**Figure 6.** A graphical overview of the Nucleotide Identify X (NIX) analysis of the *USH3* gene genomic region. NIX runs different programs, which are presented on the left side of the graphics, and displays their results side by side. Repeat masked sequences, BLAST homologies, and predicted exons are indicated with bars and lines connecting them. Poly(A) signals and putative promoter regions are indicated with triangles. The size of the region in nucleotides is illustrated by the bar in the middle. The predicted *USH3* exons and homology to an EST 35f2 sequence are indicated with circles and arrows.

71

### 6.1. *USH3* mutation analyses

The *USH3* gene was screened for mutations by PCR-amplification of the genomic fragments in patients' DNA, followed by SSCP analysis and/or sequencing. We identified three causative mutations in Finnish and Italian USH3 patients in exons 2 and 3. The Finnish founder mutation, $Fin_{major}$ (Y100X), is a nonsense mutation c.300T>G at codon 100 in exon 3, and was found in homozygous form in 52 USH3 patients being responsible for 94% of the USH3 patients studied. All 36 parents analyzed and 26 out of 45 unaffected siblings were heterozygous for the mutation. In two families, one parent was affected and homozygous for the mutation. The second Finnish mutation, $Fin_{minor}$ (M44K), is a c.131T>A transversion at codon 44 in exon 2 resulting in an amino acid substitution of lysine for methionine. It was found in a heterozygous form in four patients in two Finnish families. These patients have inherited the common ancestral haplotype from one parent, and a partly distinctive haplotype from the other parent, being compound heterozygous for both mutations. No homozygous patient for $Fin_{minor}$ has been found so far. Both Finnish mutations originate from the same region in the late settlement area (new Finland) where USH3 is known to be prevalent (Figure 8), strongly supporting the hypothesis that the mutations are introduced into the population after the 1500s (Norio et al., 1973; de la Chapelle, 1993). Furthermore, the ancestral haplotype shares a long, over 6.5-cM genetic interval which strongly indicates the founding of the $Fin_{major}$ of about 20 generations ago. All the analyzed Finnish USH3 patients had a mutation in both alleles. Among 200 anonymous blood donors, one control from the high prevalence region was heterozygous for $Fin_{major}$, giving a carrier frequency of 1:200 in Finland. In addition to these Finnish mutations, a 3-bp deletion, c.231-233delATT, leading to the substitution of one methionine for isoleucine and leucine at codons 77 and 78 in exon 3, was identified in an Italian family. These results strongly suggest that the observed gene defects underlie USH3. So far, no mutations were detected in the other non-Finnish families and patients included in the study.

All the *USH3* mutations alter the coding sequence of the gene, and there are possible consequences for the *USH3* gene expression for these mutations. The founder mutation producing the premature stop codon could trigger the nonsense-mediated mRNA decay (NMD), in which mRNA carrying a premature stop codon is rapidly degraded *in vivo* (Hentze and Kulozik, 1999), thus preventing the synthesis of the corresponding protein. Alternatively, the mutation could produce a truncated polypeptide. How the twenty-one aminoacid truncation affects the function

of the protein is difficult to predict. Most likely, it affects the stability and the structure of the protein, or its possible interactions and assembly with other proteins. The second Finnish *USH3* mutation results in a substitution of a charged polar amino acid lysine for a nonpolar amino acid methionine which most likely affects the secondary structure of the protein. Probably, the Italian mutation, which occurs in the predicted transmembrane domain, leads to abnormal folding of the USH3 protein.

### 6.2. USH3 protein prediction and expression analyses

The *USH3* open reading frame predicts a protein with no homology to any known genes so far. Prediction of membrane topology reveals that the USH3 protein contains two potential transmembrane helices, separating one extracellular loop, at amino acid positions 25 to 41 (inside to outside helix) and 63 to 79 (outside to inside helix), N-terminus being inside (Figure 7B). The N-terminal residues of the USH3 protein do not exhibit characteristics of a signal peptide, instead a weak endoplasmic reticulum-like (ER) signal for ER retention exists in the C-terminus. Northern blot analysis of a variety of human tissues utilizing the *USH3* cDNA as a probe detected transcripts of approximately 4.5 kb, 1.5 kb, and 1.0 kb. Three signals could correspond to transcripts with alternative splicing in the 3'-UTR region recognized by different poly(A) sites or differentially spliced exons. From the tissues studied, only human lymphoblasts did not show a signal, as confirmed both by RT-PCR and northern blotting. *In situ* hybridization, using *in vitro* transcribed ^{35}S-labelled antisense and control-sense mouse *USH3* cRNA probes detected a low-level, ubiquitous expression in mouse retinal and cochlear tissues.

**A.**

ATG    2,061 bp    13,380 bp    1,192 bp    STOP    156 bp    764 bp    3,568 bp

| 5'UTR | E1 | | E2 | | E3 | | E4 | 3'UTR |

392 bp   25 bp   180 bp   137 bp   21 bp

E1b
87 bp

Poly(A)'   Poly(A)''   Poly(A)'''

**B.**

NH2 — COOH

Italian mutation

TM1    TM2

$Fin_{major}$

$Fin_{minor}$
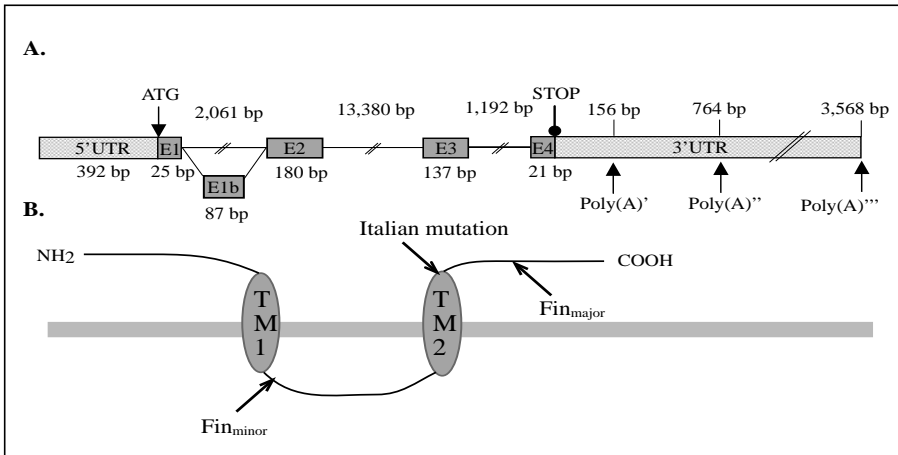
**Figure 7**. Predicted schematic structure of the *USH3* gene and protein. **A.** The four exons are depicted by numbered boxes. The sizes of the exons are shown below the boxes and those of the three separating introns above the lines. The translation initiation site, and the first stop codon are indicated. Three polyadenylation signals and their predicted locations downstream of the termination codon are indicated by arrows. The alternatively spliced exon 1b (E1b) is indicated. **B.** The USH3 protein is predicted to contain two transmembrane helices. The locations of $Fin_{major}$ (Y100X), $Fin_{minor}$ (M44K), and Italian c.231-233delATT mutations in the protein are indicated with an arrow.

| D3S1299 | 6 | 2 | | 5 | 2 |
| 25B8CA2 | 3 | 5 | | 6 | 3 |
| 25B8CA1 | 5 | 5 | | 6 | 6 |
| USH3 | maj | wt | | min | wt |
| 107G19CA7 | 6 | 5 | | 3 | 5 |
| RSNP2 | C | C | | C | G |
| TSNP1 | T | T | | T | C |
| D3S3625 | 2 | 3 | | 2 | 8 |
| D3S1594 | 11 | 3 | | 11 | 3 |
| D3S1279 | 4 | 5 | | 3 | 5 |

I   1   2

II   1   2   3   4

| D3S1299 | 6 | 2 | 6 | 5 | 6 | 2 | 6 | 5 |
| 25B8CA2 | 3 | 3 | 3 | 6 | 3 | 5 | 3 | 6 |
| 25B8CA1 | 5 | 6 | 5 | 6 | 5 | 5 | 5 | 6 |
| USH3 | maj | wt | maj | min | maj | wt | maj | min |
| 107G19CA7 | 6 | 5 | 6 | 3 | 6 | 5 | 6 | 3 |
| RSNP2 | C | C | C | C | C | C | C | C |
| TSNP1 | T | T | T | T | T | T | T | T |
| D3S3625 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 |
| D3S1594 | 11 | 11 | 11 | 11 | 11 | 3 | 11 | 11 |
| D3S1279 | 4 | 4 | 4 | 3 | 4 | 5 | 4 | 3 |

$Fin_{minor}$
$Fin_{minor}$

**Figure 8**. A pedigree of a Finnish USH3 family segregating a paternal $Fin_{major}$ (c.300T>G) mutation and a maternal $Fin_{minor}$ (c.131T>A) mutation. The chromosome assumed to carry the disease allele is typed in bold and boxed. A recombination in the maternal chromosome of individual II-1 excludes the segment RSNP2-*D3S1279* as the site of the *USH3* mutation. The map of Finland shows the parental birthplaces of USH3 patients. The birthplaces of the two mothers carrying the $Fin_{minor}$ mutation are indicated with circles.

### 6.4. A mouse homologue for *USH3* (III, unpublished)

Alignment of the *USH3* gene ORF with Celera Assembled and Annotated Mouse Genome databases indicated several genomic fragments highly homologous to the exons 2 and 3 of the gene. The mouse Marathon Ready Retina cDNA was PCR-amplified with primers homologous to the *USH3* coding region, producing a fragment of 262 bp. Subsequent sequencing indicated that this fragment is highly conserved at nucleotide level between mouse and human (87% identity) (unpublished results). The efforts to identify the full-length cDNA by RT-PCR or RACE PCR have not yet been successful. The mouse multi-tissue northern blot hybridization with the 262-bp fragment detects a 2.0-kb transcript in all tissues examined (heart, brain, spleen, lung, liver, skeletal muscle, and testis). In addition, a weak 1.0-kb transcript in all tissues examined as well as a 4.0-kb transcript in heart and liver were detected (unpublished data) (Figure 9).
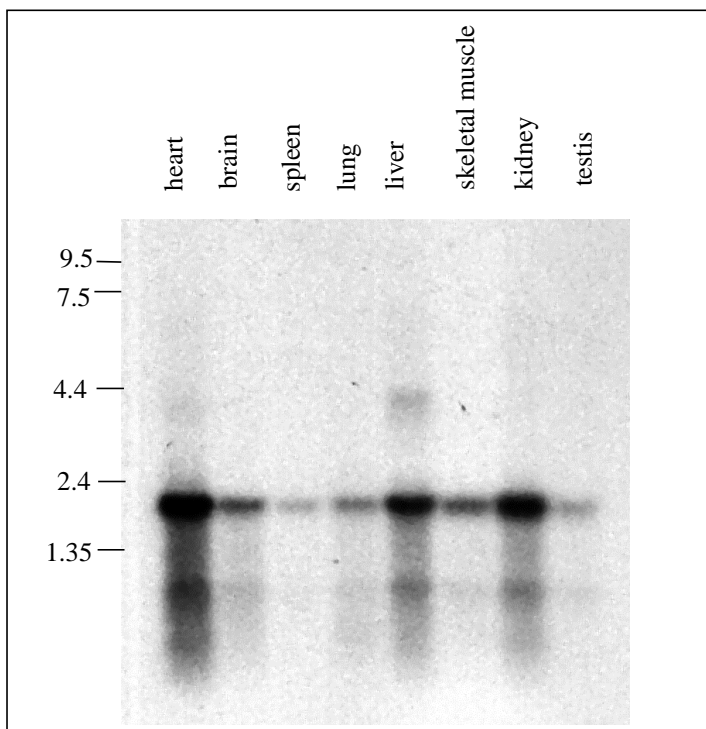


**Figure 9**. Mouse *ush3* gene expression. Northern analysis of a mouse multiple-tissue blot was performed using a probe homologous to the coding region of the human *USH3* gene. Two signals of approximately 2.0 kb and 1.0 kb are detected in all tissues tested. In addition, a weak signal of 4.0 kb is detected in heart and liver mRNA.

# CONCLUSIONS AND FUTURE PROSPECTS

This study describes the positional cloning of the *USH3* gene which was initiated after the assignment of the *USH3* locus to chromosome 3q21-q25 by genome-wide linkage mapping in Finnish USH3 families representing progressive hearing loss, retinitis pigmentosa, and variable vestibular dysfunction (Sankila et al., 1995). The molecular analyses, including linkage, linkage disequilibrium, and physical mapping, aided us to refine the *USH3* locus from 5-cM to a 250-kb region. Thereafter, the genomic contigs were constructed by shotgun sequencing, and mutation analyses of several *in silico* positional candidate genes, as well as an additional sequence analysis of an extended 133-kb genomic region, were needed until the candidate for USH3 was identified. Identification of three different *USH3* mutations in Finnish and Italian patients provided evidence for this particular gene to underlie the disease. Based on the mutation results and the earlier clinical data (Pakarinen et al., 1995b; Pakarinen et al., 1996; Pakarinen 1997), no correlation between different *USH3* mutations and patients' phenotypes were observed. Effects of putative modifier genes at different loci could explain the phenotype differences among the USH3 patients.

Using traditional experimental methods in the transcript isolation is usually time-consuming and laborious. The release of the whole genome sequence combined with the sequence annotation data by the Human Sequencing Consortium allows novel approaches to gene identification, further facilitating the process of positional cloning by "virtual cloning". In this study, the progress from genetic localization of *USH3*, defined by linkage disequilibrium and haplotype analyses, to the identification of the disease gene was accelerated significantly with the resources of the large-scale genomic sequencing data, which we had produced, and the human genome working draft. Several gene prediction programs were used, which enabled us to analyze the genomic contigs through the similarity searches against databases of well-characterized genes, ESTs, and proteins, as well as to predict genes in a given raw DNA sequence by defining their splicing patterns. Systematic PCR amplification of retina cDNA using primers spanning genes of interest, and sequencing of the PCR products confirmed the discovery of a novel gene.

The *USH3* gene is ubiquitously expressed, but probably only the retina and the cochlea are affected by the inherited mutations. *USH3* being a novel gene with no homology to any known genes or proteins provides limited information about the possible molecular pathways of

the disease, and further studies in the wider contex, defining the specific function and the localization of the USH3 protein, are needed. Functional tests *in vitro*, such as different antibody labelling and detection systems using cultured cells or cell extracts, can show how mutations in the *USH3* gene affect its expression and function. While cultured transfected cells give information about the gene expression and subcellular localization, only tests *in vivo* using animal models in gene targeting, such as mice homozygous for a targeted deletion of the *USH3* gene or natural animal models of the disease, should lead to a better understanding of the functional consequences of the defective *USH3*. A yeast two-hybrid screen could provide some clues about the molecular interactions between different proteins and identify other genes that may be involved in the disease. The resulting information of the biological processes involved in the normal and pathogenic states can be finally used to design new gene therapies.

Individuals with USH3 have a progressive loss of hearing and vision. The severity of the hearing loss and the degree of its progression are variable complicating the confirmation of the diagnosis. Therefore, knowledge about the mutated gene provides improvements in the accuracy of presymptomatic diagnosis, especially in children with late onset hearing loss, and in carrier identification. It is also important in genetic counseling of the families, and might prove helpful for example in making educational choices.

# ACKNOWLEDGEMENTS

those stress-releasing discussions. Particularly, I want to thank Maria Aminoff-Backlund for her friendship and making my life easier in Ohio by lending her "check-book" whenever needed.

I want to acknowledge Docent Maaret Ridanpää for her valuable advice in lab. I am extremely grateful to her and Timo Tuomi for their friendship and giving such a good home to Ossi during my stay in Ohio. I thank Docent Tiina Alitalo for her kindness and introducing me to "the population genetics team" at Folkhälsan at M97.

Aila Riikonen and Minna Maunula are thanked for all the help in so many practical matters between earth and heaven.

I acknowledge Monica Virtaneva for revising the language of my thesis.

I want to thank especially Jaana Kari for being my dear friend for so many years.

I am deeply grateful to my mother for her never-failing interest and supporting attitude towards my studies. My dear sister, Leila, is thanked for her support and encouragement, especially during the last two years.

Thanks to Ossi and Onni for their companion and taking me out for exercise; they reminded me to live the so-called normal life also. Especially, Ossi is thanked for the help in lab and sharing many exciting pipetting hours by hunting the Falcon-tubes from the trashes at M97.

Finally, with all my heart I owe my warmest thanks to Pepelu for his constant love and understanding. His continuous support and knowledge in any fields of this study are gratefully acknowledged. I thank him for his endless patience during all these tiring weeks and spoiled weekends. I also thank him for his good sense of humor in learning the Finnish language.

Helsinki, April 2002

# REFERENCES

The *Arabidopsis* Genome Initiative (AGI). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*. **408:** 796-815. (2000).

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., et al. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*. **252:** 1651-6. (1991).

Adato, A., Kalinski, H., Weil, D., Chaib, H., Korostishevsky, M., Bonne-Tamir, B. Possible interaction between USH1B and USH3 gene products as implied by apparent digenic deafness inheritance. *Am J Hum Genet*. **65:** 261-5. (1999).

Adato, A., Weil, D., Kalinski, H., Pel-Or, Y., Ayadi, H., Petit, C., Korostishevsky, M., Bonne-Tamir, B. Mutation profile of all 49 exons of the human myosin VIIA gene, and haplotype analysis, in Usher 1B families from diverse origins. *Am J Hum Genet*. **61**: 813-21. (1997).

Adato, A., Weston, M.D., Berry, A., Kimberling, W.J., Bonne-Tamir, A. Three novel mutations and twelve polymorphisms identified in the USH2A gene in Israeli USH2 families. *Hum Mutat*. **15:** 388. (2000).

Ahmed, Z.M., Riazuddin, S., Bernstein, S.L., Ahmed, Z., Khan, S., Griffith, A.J., Morell, R.J., Friedman, T.B., Wilcox, E.R. Mutations of the protocadherin gene PCDH15 cause Usher syndrome type 1F. *Am J Hum Genet*. **69:** 25-34. (2001).

Alagramam, K.N., Murcia, C.L., Kwon, H.Y., Pawlowski, K.S., Wright, C.G., Woychik, R.P. The mouse Ames waltzer hearing-loss mutant is caused by mutation of Pcdh15, a novel protocadherin gene. *Nat Genet*. **27:** 99-102. (2001a).

Alagramam, K.N., Yuan, H., Kuehn, M.H., Murcia, C.L., Wayne, S., Srisailpathy, C.R., Lowry, R.B., Knaus, R., Van Laer, L., Bernier, F.P., Schwartz, S., Lee, C., Morton, C.C., Mullins, R.F., Ramesh, A., Van Camp, G., Hagemen, G.S., Woychik, R.P., Smith, R.J. Mutations in the novel protocadherin PCDH15 cause Usher syndrome type 1F. *Hum Mol Genet*. **10:** 1709-18. (2001b).

Alloway, P.G., Howard, L., Dolph, P.J. The formation of stable rhodpsin-arrestin complexes induces apoptosis and photoreceptor cell degeneration. *Neuron*. **28:** 129-138. (2000).

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. Basic local alignment search tool. *J Mol Biol*. **215:** 403-10. (1990).

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25:** 3389-402. (1997).

Alwine, J.C., Kemp, D.J., Stark, G.R. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A*. **74:** 5350-4. (1977).

Aminoff, M., Carter, J.E., Chadwick, R.B., Johnson, C., Grasbeck, R., Abdelaal, M.A., Broch, H., Jenner, L.B., Verroust, P.J., Moestrup, S.K., de la Chapelle, A., Krahe, R. Mutations in CUBN, encoding the intrinsic factor-vitamin B12 receptor, cubilin, cause hereditary megaloblastic anaemia 1. *Nat Genet*. **21:** 309-13. (1999).

Bairoch, A., Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res*. **27:** 49-54. (1999).

Balciuniene, J., Dahl, N., Borg, E., Samuelsson, E., Koisti, M.J., Pettersson, U., Jazin, E.E. Evidence for digenic inheritance of nonsyndromic hereditary hearing loss in a Swedish family. *Am J Hum Genet*. **63:** 786-93. (1998).

Ballabio, A. The rise and fall of positional cloning? *Nat Genet*. **3:** 277-9. (1993).

Baltimore, D. Our genome unveiled. *Nature*. **409:** 814-6. (2001).

Bassam, B.J., Caetano-Anolles, G., Gresshoff, P.M. Fast and sensitive silver staining of DNA in polyacrylamide gels [published erratum appears in Anal Biochem 1991 Oct;198(1):217]. *Anal Biochem*. **196:** 80-3. (1991).

Bassett, D.E., Jr., Boguski, M.S., Spencer, F., Reeves, R., Goebl, M., Hieter, P. Comparative genomics, genome cross-referencing and XREFdb. *Trends Genet*. **11:** 372-3. (1995).

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Finn, R.D., Sonnhammer, E.L. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res*. **27:** 260-2. (1999).

Beneyto, M.M., Cuevas, J.M., Millan, J.M., Espinos, C., Mateu, E., Gonzalez-Cabo, P., Baiget, M., Domenech, M., Bernal, S., Ayuso, C., Garcia-Sandoval, B., Trujillo, M.J., Borrego, S., Antinolo, G., Carballo, M., Najera, C. Prevalence of 2314delG mutation in Spanish patients with usher syndrome type II (USH2). *Ophthalmic Genet*. **21:** 123-8. (2000).

Bharadwaj, A.K., Kasztejna, J.P., Huq, S., Berson, E.L., Dryja, T.P. Evaluation of the myosin VIIA gene and visual function in patients with Usher syndrome type I. *Exp Eye Res*. **71:** 173-81. (2000).

Bhattacharya, G., Miller, C., Kimberling, W.J., Jablonski, M. Localization and expression of usherin: a novel basement protein defective in people with Usher's syndrome type IIa. *Hearing Res*. **163:** 1-11. (2002).

Birney, E., Bateman, A., Clamp, M.E., Hubbard, T.J. Mining the draft human genome. *Nature*. **409:** 827-8. (2001).

Bitner-Glindzicz, M., Lindley, K.J., Rutland, P., Blaydon, D., Smith, V.V., Milla, P.J., Hussain, K., Furth-Lavi, J., Cosgrove, K.E., Shepherd, R.M., Barnes, P.D., O'Brien, R.E., Farndon, P.A., Sowden, J., Liu, X.Z., Scanlan, M.J., Malcolm, S., Dunne, M.J., Aynsley-Green, A., Glaser, B. A recessive contiguous gene deletion causing infantile hyperinsulinism, enteropathy and deafness identifies the Usher type 1C gene. *Nat Genet*. **26:** 56-60. (2000).

Boguski, M.S. The turning point in genome research. *Trends Biochem Sci*. **20:** 295-6. (1995).

Boguski, M.S., Lowe, T.M., Tolstoshev, C.M. dbEST--database for "expressed sequence tags". *Nat Genet*. **4:** 332-3. (1993).

Bolz, H., von Brederlow, B., Ramirez, A., Bryda, E.C., Kutsche, K., Nothwang, H.G.,Seeliger, M., del, C.S.C.M., Vila, M.C., Molina, O.P., Gal, A., Kubisch, C. Mutation of CDH23, encoding a new member of the cadherin gene family, causes Usher syndrome type 1D. *Nat Genet*. **27:** 108-12. (2001).

Bonaldo, M.F., Yu, M.T., Jelenc, P., Brown, S., Su, L., Lawton, L., Deaven, L., Efstratiadis, A., Warburton, D., Soares, M.B. Selection of cDNAs using chromosome-specific genomic clones: application to human chromosome 13. *Hum Mol Genet*. **3:** 1663-73. (1994).

Bonneau, D., Raymond, F., Kremer, C., Klossek, J.M., Kaplan, J., Patte, F. Usher syndrome type I associated with bronchiectasis and immotile nasal cilia in two brothers. *J Med Genet*. **30:** 253-4. (1993).

Bork, J.M., Peters, L.M., Riazuddin, S., Bernstein, S.L., Ahmed, Z.M., Ness, S.L., Polomeno, R., Ramesh, A., Schloss, M., Srisailpathy, C.R., Wayne, S., Bellman, S., Desmukh, D., Ahmed, Z., Khan, S.N., Kaloustian, V.M., Li, X.C., Lalwani, A., Bitner-Glindzicz, M., Nance, W.E., Liu, X.Z., Wistow, G., Smith, R.J., Griffith, A.J., Wilcox, E.R., Friedman, T.B., Morell, R.J. Usher syndrome 1D and nonsyndromic autosomal recessive deafness DFNB12 are caused by allelic mutations of the novel cadherin-like gene CDH23. *Am J Hum Genet*. **68:** 26-37. (2001).

Botstein, D., White, R.L., Skolnick, M., Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*. **32:** 314-31. (1980).

Bouck, J.B., Metzker, M.L., Gibbs, R.A. Shotgun sample sequence comparisons between mouse and human genomes. *Nat Genet*. **25:** 31-3. (2000).

Boughman, J.A., Fishman, G.A. A genetic analysis of retinitis pigmentosa. *Br J Ophthalmol*. **67:** 449-54. (1983).

Boughman, J.A., Vernon, M., Shaver, K.A. Usher syndrome: definition and estimate of prevalence from two high-risk populations. *J Chronic Dis*. **36:** 595-603. (1983).

von Brederlow, B., Bolz, H., Janecke, A., La, O., Cabrera, A., Rudolph, G., Lorenz, B., Schwinger, E., Gal, A. Identification and in vitro expression of novel CDH23 mutations of patients with Usher syndrome type 1D. *Hum Mut*. **19:** 268-73. (2002).

Brown, P.O., Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nat Genet*. **21:** 33-7. (1999).

Burge, C., Karlin, S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*. **268:** 78-94. (1997).

Burke, D.T., Carle, G.F., Olson, M.V. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Biotechnology*. **24:** 172-8. (1992).

Bykhovskaya, Y., Estivill, X., Taylor, K., Hang, T., Hamon, M., Casano, R.A., Yang, H., Rotter, J.I., Shohat, M., Fischel-Ghodsian, N. Candidate locus for a nuclear modifier gene for maternally inherited deafness. *Am J Hum Genet*. **66:** 1905-10. (2000).

Camper, S.A., Saunders, T.L., Kendall, S.K., Keri, R.A., Seasholtz, A.F., Gordon, D.F., Birkmeier, T.S., Keegan, C.E., Karolyi, I.J., Roller, M.L., et al. Implementing transgenic and embryonic stem cell technology to study gene expression, cell-cell interactions and gene function. *Biol Reprod*. **52:** 246-57. (1995).

The C. *elegans* Sequencing Consortium. Genome sequence of the nematode C. *elegans*: a platform for investigating biology. *Science*. **282:** 2012-8. (1998).

Carthew, R.W., Rubin, G.M. Seven in absentia, a gene required for specification of R7 cell fate in the Drosophila eye. *Cell*. **63:** 561-77. (1990).

Carvill, S. Sensory impaiments, intellectual diability and psychiatry. *J Intel Disability Res*. **45:** 467-83. (2001).

Chaib, H., Kaplan, J., Gerber, S., Vincent, C., Ayadi, H., Slim, R., Munnich, A., Weissenbach, J., Petit, C. A newly identified locus for Usher syndrome type I, USH1E, maps to chromosome 21q21. *Hum Mol Genet*. **6:** 27-31. (1997).

Chaib, H., Place, C., Salem, N., Chardenoux, S., Vincent, C., Weissenbach, J., El-Zir, E., Loiselet, J., Petit, C. A gene responsible for a sensorineural nonsyndromic recessive deafness maps to chromosome 2p22-23. *Hum Mol Genet*. **5:** 155-8. (1996).

Charlton, M.E., Williams, A.S., Fogliano, M., Sweetnam, P.M., Duman, R.S. The isolation and characterization of a novel G protein-coupled receptor regulated by immunologic challenge. *Brain Res*. **764:** 141-8. (1997).

Chen, Z.Y., Hasson, T., Kelley, P.M., Schwender, B.J., Schwartz, M.F., Ramakrishnan, M., Kimberling, W.J., Mooseker, M.S., Corey, D.P. Molecular cloning and domain structure of human myosin-VIIa, the gene product defective in Usher syndrome 1B. *Genomics*. **36:** 440-8. (1996).

Claudio, J.O., Marineau, C., Rouleau, G.A. The mouse homologue of the neurofibromatosis type 2 gene is highly conserved. *Hum Mol Genet*. **3:** 185-90. (1994).

Claverie, J.M. Computational methods for the identification of genes in vertebrate genomic sequences. *Hum Mol Genet*. **6:** 1735-44. (1997).

Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L. New goals for the U.S. Human Genome Project: 1998-2003. *Science*. **282:** 682-9. (1998).

Cuevas, J.M., Espinos, C., Millan, J.M., Sanchez, F., Trujillo, M.J., Ayuso, C., Beneyto, M., Najera, C. Identification of three novel mutations in the MYO7A gene. *Hum Mutat*. **14:** 181. (1999).

Cuevas, J.M., Espinos, C., Millan, J.M., Sanchez, F., Trujillo, M.J., Garcia-Sandoval, B., Ayuso, C., Najera, C., Beneyto, M. Detection of a novel Cys628STOP mutation of the myosin VIIA gene in Usher syndrome type Ib. *Mol Cell Probes*. **12:** 417-20. (1998).

Cuticchia, A.J., Fasman, K.H., Kingsbury, D.T., Robbins, R.J., Pearson, P.L. The GDB human genome data base anno 1993. *Nucleic Acids Res*. **21:** 3003-6. (1993).

Davenport, S.L.H., Omenn, G.S. The heterogenity of Usher syndrome. (Abstract). *Vth Int. Conf. on Birth defects*, Montreal, 8/1977.

Davenport, S.L., O'Nuallain, S., Omenn, G.S., Wilkus, R.J. Usher syndrome in four hard-of-hearing siblings. *Pediatrics*. **62:** 578-83. (1978).

Davies, E., Bonnardeaux, A., Lathrop, G.M., Corvol, P., Clauser, E., Soubrier, F. Angiotensin II (type-1) receptor locus: CA repeat polymorphism and genetic mapping. *Hum Mol Genet*. **3:** 838. (1994).

de la Chapelle, A. Disease gene mapping in isolated human populations: the example of Finland. *J Med Genet*. **30:** 857-65. (1993).

de la Chapelle, A., Wright, F.A. Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc Natl Acad Sci U S A*. **95:** 12416-23. (1998).

Di Palma, F., Holme, R.H., Bryda, E.C., Belyantseva, I.A., Pellegrino, R., Kachar, B., Steel, K.P., Noben-Trauth, K. Mutations in Cdh23, encoding a new type of cadherin, cause stereocilia disorganization in waltzer, the mouse model for Usher syndrome type 1D. *Nat Genet*. **27:** 103-7. (2001).

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc,S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J., Weissenbach, J. A comprehensive genetic map of the human genome based on 5,264 microsatellites [see comments]. *Nature*. **380:** 152-4. (1996).

Dreyer, B., Tranebjaerg, L., Brox, V., Rosenberg, T., Möller, C., Beneyto, M., Weston, M.D., Kimberling, W.J., Nilssen, O. A common ancestral origin of the frequent and widespread 2299delG USH2A mutation. *Am J Hum Genet*. **69:** 228-34. (2001).

Dreyer, B., Tranebjaerg, L., Rosenberg, T., Weston, M.D., Kimberling, W.J., Nilssen, O. Identification of novel USH2A mutations: implications for the structure of USH2A protein. *Eur J Hum Genet*. **8:** 500-6. (2000).

Dunham, I., Shimizu, N., Roe, B.A., Chissoe, S., Hunt, A.R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Smink, L.J., Ainscough, R., Almeida, J.P., Babbage, A., Bagguley, C., Bailey, J., Barlow, K., Bates, K.N., Beasley, O., Bird, C.P., Blakey, S., Bridgeman, A.M., Buck, D., Burgess, J., Burrill, W.D., O'Brien, K.P., et al. The DNA sequence of human chromosome 22. *Nature*. **402:** 489-95. (1999).

Duyk, G.M., Kim, S.W., Myers, R.M., Cox, D.R. Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc Natl Acad Sci U S A*. **87:** 8995-9. (1990).

el-Amraoui, A., Sahly, I., Picaud, S., Sahel, J., Abitbol, M., Petit, C. Human Usher 1B/mouse shaker-1: the retinal phenotype discrepancy explained by the presence/absence of myosin VIIA in the photoreceptor cells. *Hum Mol Genet*. **5:** 1171-8. (1996).

Espinos, C., Millan, J.M., Sanchez, F., Beneyto, M., Najera, C. Ala397Asp mutation of myosin VIIA gene segregating in a Spanish family with type-Ib Usher syndrome. *Hum Genet*. **102:** 691-4. (1998a).

Espinos, C., Najera, C., Millan, J.M., Ayuso, C., Baiget, M., Perez-Garrigues, H., Rodrigo, O., Vilela, C., Beneyto, M. Linkage analysis in Usher syndrome type I (USH1) families from Spain. *J Med Genet*. **35:** 391-8. (1998b).

Eudy, J.D., Weston, M.D., Yao, S., Hoover, D.M., Rehm, H.L., Ma-Edmonds, M., Yan,D., Ahmad, I., Cheng, J.J., Ayuso, C., Cremers, C., Davenport, S., Möller, C., Talmadge, C.B., Beisel, K.W., Tamayo, M., Morton, C.C., Swaroop, A., Kimberling, W.J., Sumegi, J. Mutation of a gene encoding a protein with extracellular matrix motifs in Usher syndrome type IIa. *Science*. **280:** 1753-7. (1998).

Everett, L.A., Glaser, B., Beck, J.C., Idol, J.R., Buchs, A., Heyman, M., Adawi, F., Hazani, E., Nassir, E., Baxevanis, A.D., Sheffield, V.C., Green, E.D. Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS). *Nat Genet*. **17:** 411-22. (1997).

Ewing, B., Green, P. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet*. **25**: 232-4. (2000).

Farrar, G.J., Findlay, J.B., Kumar-Singh, R., Kenna, P., Humphries, M.M., Sharpe, E., Humphries, P. Autosomal dominant retinitis pigmentosa: a novel mutation in the rhodopsin gene in the original 3q linked family. *Hum Mol Genet*. **1**: 769-71. (1992).

Farrar, G.J., Kenna, P., Jordan, S.A., Kumar-Singh, R., Humphries, M.M., Sharp, E.M., Sheils, D.M., Humphries, P. A three-base-pair deletion in the peripherin-RDS gene in one form of retinitis pigmentosa. *Nature*. **354:** 478-80. (1991).

Feng, Y., Zhang, F., Lokey, L.K., Chastain, J.L., Lakkis, L., Eberhart, D., Warren, S.T. Translational suppression by trinucleotide repeat expansion at FMR1. *Science*. **268:** 731-4. (1995).

Fischer, S.G., Lerman, L.S. Length-independent separation of DNA restriction fragments in two- dimensional gel electrophoresis. *Cell*. **16:** 191-200. (1979).

Fishman, G.A., Kumar, A., Joseph, M.E., Torok, N., Anderson, R.J. Usher's syndrome. Ophthalmic and neuro-otologic findings suggesting genetic heterogeneity. *Arch Ophthalmol*. **101:** 1367-74. (1983).

Forsius, H., Eriksson, A., Nuutila, A., Vainio-Mattila, B., Krause, U. A genetic study of three rare retinal disorders: dystrophia retinae dysacusis syndrome, X-chromosomal retinoschisis and grouped pigments of the retina. *Birth Defects Orig Artic Ser*. **7:** 83-98. (1971).

Friedman, T., Battey, J., Kachar, B., Riazuddin, S., Noben-Trauth, K., Griffith, A., Wilcox, E. Modifier genes of hereditary hearing loss. *Curr Opin Neurobiol*. **10:** 487-93. (2000).

Frohman, M.A., Dush, M.K., Martin, G.R. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A*. **85:** 8998-9002. (1988).

Gasparini, P., De Fazio, A., Croce, A.I., Stanziale, P., Zelante, L. Usher syndrome type III (USH3) linked to chromosome 3q in an Italian family. *J Med Genet*. **35:** 666-7. (1998).

Gibson, F., Walsh, J., Mburu, P., Varela, A., Brown, K.A., Antonio, M., Beisel, K.W., Steel, K.P., Brown, S.D. A type VII myosin encoded by the mouse deafness gene shaker-1. *Nature*. **374:** 62-4. (1995).

Gillespie, P.G., Corey, D.P. Myosin and adaptation by hair cells. *Neuron*. **19:** 955-8. (1997).

Gillespie, P.G., Wagner, M.C., Hudspeth, A.J. Identification of a 120 kd hair-bundle myosin located near stereociliary tips. *Neuron*. **11:** 581-94. (1993).

Gillespie, P.G., Walker, R.G. Molecular basis of mechanosensory transduction. *Nature*. **413:** 194-202. (2001).

Gish, W., States, D.J. Identification of protein coding regions by database similarity search. *Nat Genet*. **3:** 266-72. (1993).

Gitschier, J., Wood, W.I., Goralka, T.M., Wion, K.L., Chen, E.Y., Eaton, D.H., Vehar, G.A., Capon, D.J., Lawn, R.M. Characterization of the human factor VIII gene. *Nature*. **312:** 326-30. (1984).

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G. Life with 6000 genes. *Science*. **274:** 546, 563-7. (1996).

Gorlin, R.J., Tilsner, T.J., Feinstein, S., Duvall, A.J., 3rd. Usher's syndrome type III. *Arch Otolaryngol*. **105:** 353-4. (1979).

von Grafe, A. Exceptionelles verhalten des gesichtsfeldes bei pigmententartung der netzhaut. *Graefes Arch Klin Exp Ophtalmol* **4**: 250-253. (1858).

Gray, I.C., Campbell, D.A., Spurr, N.K. Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet*. **9:** 2403-8. (2000).

Gregory-Evans, K., Bhattacharya, S.S. Genetic blindness: current concepts in the pathogenesis of human outer retinal dystrophies. *Trends in Genet*. **14**: 103-109. (1998).

Gröndahl, J. Estimation of prognosis and prevalence of retinitis pigmentosa and Usher syndrome in Norway. *Clin Genet*. **31**: 255-64. (1987).

Gröndahl, J, and Mjöen, S. Usher syndrome in four Norwegian counties. *Clin Genet*. **30**: 14-28. (1986).

Haim, M. Epidemiology of retinitis pigmentosa in Denmark. *Acta Opthalmol Scan Suppl*. **80**: 1-34. (2002).

Hallgren, B. Retinitis pigmentosa combined with congenital deafness; with vestibulo-cerebellar ataxia and mental abnormality in a proportion of cases. A clinical and genetico-statistical study. *Acta Psychiatr Scand*. (Suppl.) **138**: 5-101. (1959).

Harris, N.L. Genotator: a workbench for sequence annotation. *Genome Res*. **7**: 754-62. (1997).

Hasson, T. Molecular motors: sensing a function for myosin-VIIa. *Curr Biol*. **9**: R838-41. (1999).

Hasson, T., Gillespie, P.G., Garcia, J.A., MacDonald, R.B., Zhao, Y., Yee, A.G., Mooseker, M.S., Corey, D.P. Unconventional myosins in inner-ear sensory epithelia. *J Cell Biol*. **137**: 1287-307. (1997a).

Hasson, T., Heintzelman, M.B., Santos-Sacchi, J., Corey, D.P., Mooseker, M.S. Expression in cochlea and retina of myosin VIIa, the gene product defective in Usher syndrome type 1B. *Proc Natl Acad Sci U S A*. **92**: 9815-9. (1995).

Hasson, T., Walsh, J., Cable, J., Mooseker, M.S., Brown, S.D., Steel, K.P. Effects of shaker-1 mutations on myosin-VIIa protein and mRNA expression. *Cell Motil Cytoskeleton*. **37**: 127-38. (1997b).

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Schibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, T., Nagamine, K., Mitsuyama, S., Antonarakis, S.E., Minoshima, S., Shimizu, N., Nordsiek, G., Hornischer, K., Brant, P., Scharfe, M., Schon, O., Desario, A., Reichelt, J., Kauer, G., Blocker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesselmann, L., Dagand, E., Haaf, T., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R., Yaspo, M.L. The DNA sequence of human chromosome 21. *Nature*. **405**: 311-9. (2000).

Heinemeyer, T., Chen, X., Karas, H., Kel, A.E., Kel, O.V., Liebich, I., Meinhardt, T., Reuter, I., Schacherer, F., Wingender, E. Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. *Nucleic Acids Res*. **27**: 318-22. (1999).

Hentze, M.W., Kulozik, A.E. A perfect message: RNA surveillance and nonsense-mediated decay. *Cell*. **96**: 307-10. (1999).

Hereditary Hearing Loss Homepage. Van Gamp, G., Smith, R.J.G. http://dnalab-www.uia.ac.be/dnalab/hhh/.

Hmani, M., Ghorbel, A., Boulila-Elgaied, A., Zina, Z.B., Kammoun, W., Drira, M., Chaabouni, M., Petit, C., Ayadi, H. A novel locus for Usher syndrome type II, USH2B, maps to chromosome 3 at p23-24.2. *Eur J Hum Genet*. **7**: 363-7. (1999).

Hofmann, K., Bucher, P., Falquet, L., Bairoch, A. The PROSITE database, its status in 1999. *Nucleic Acids Res*. **27**: 215-9. (1999).

Hofmann, K., Stoffel, W. TMbase - A database of membrane spanning proteins segments. *Biol Chem Hoppe-Seyler* **374**: 166. (1993).

Holloway, A.J., Della, N.G., Fletcher, C.F., Largespada, D.A., Copeland, N.G., Jenkins, N.A., Bowtell, D.D. Chromosomal mapping of five highly conserved murine homologues of the Drosophila RING finger gene seven-in-absentia. *Genomics*. **41**: 160-8. (1997).

Honore, B., Madsen, P., Andersen, A.H., Leffers, H. Cloning and expression of a novel human profilin variant, profilin II. *FEBS Lett*. **330**: 151-5. (1993).

Hope, C.I., Bundey, S., Proops, D., Fielder, A.R. Usher syndrome in the city of Birmingham--prevalence and clinical classification. *Br J Ophthalmol*. **81**: 46-53. (1997).

Hudspeth, A.J. How hearing happens. *Neuron*. **19**: 947-50. (1997).

Hunter, D.G., Fishman, G.A., Mehta, R.S., Kretzer, F.L. Abnormal sperm and photoreceptor axonemes in Usher's syndrome. *Arch Ophthalmol*. **104**: 385-9. (1986).

Huopaniemi, L., Rantala, A., Forsius, H., Somer, M., de la Chapelle, A., Alitalo, T. Three widespread founder mutations contribute to high incidence of X- linked juvenile retinoschisis in Finland. *Eur J Hum Genet*. **7**: 368-76. (1999).

Hästbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A., Lander, E. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet*. **2**: 204-11. (1992).

Hästbacka, J., de la Chapelle, A., Mahtani, M.M., Clines, G., Reeve-Daly, M.P., Daly, M., Hamilton, B.A., Kusumi, K., Trivedi, B., Weaver, A., et al. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell.* **78:** 1073-87. (1994).

Höglund, P., Sistonen, P., Norio, R., Holmberg, C., Dimberg, A., Gustavson, K.H., de la Chapelle, A., Kere, J. Fine mapping of the congenital chloride diarrhea gene by linkage disequilibrium. *Am J Hum Genet.* **57:** 95-102. (1995).

Ito, M., Yuan, C.X., Malik, S., Gu, W., Fondell, J.D., Yamamura, S., Fu, Z.Y., Zhang, X., Qin, J., Roeder, R.G. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol Cell.* **3:** 361-70. (1999).

Jacobs, K.A., Collins-Racie, L.A., Colbert, M., Duckett, M., Golden-Fleet, M., Kelleher, K., Kriz, R., LaVallie, E.R., Merberg, D., Spaulding, V., Stover, J., Williamson, M.J., McCoy, J.M. A genetic selection for isolating cDNAs encoding secreted proteins. *Gene.* **198:** 289-96. (1997).

Janecke, A.R., Meins, M., Sadeghi, M., Grundmann, K., Apfelstedt-Sylla, E., Zrenner, E., Rosenberg, T., Gal, A. Twelve novel myosin VIIA mutations in 34 patients with Usher syndrome type I: confirmation of genetic heterogeneity. *Hum Mutat.* **13:** 133-40. (1999).

Jorde, L.B. Linkage disequilibrium and the search for complex disease genes. *Genome Res.* **10:** 1435-44. (2000).

Kalatzis, V., Petit, C. The fundamental and medical impacts of recent progress in research on hereditary hearing loss. *Hum Mol Genet.* **7:** 1589-97. (1998).

Kaplan, J., Gerber, S., Bonneau, D., Rozet, J.M., Delrieu, O., Briard, M.L., Dollfus, H., Ghazi, I., Dufier, J.L., Frezal, J., et al. A gene for Usher syndrome type I (USH1A) maps to chromosome 14q. *Genomics.* **14:** 979-87. (1992).

Karjalainen, S., Pakarinen, L., Teräsvirta, M., Kääriäinen, H., Vartiainen, E. Progressive hearing loss in Usher's syndrome. *Ann Otol Rhinol Laryngol.* **98:** 863-6. (1989).

Karjalainen, S., Teräsvirta, M., Karja, J., Kääriäinen, H. An unusual otological manifestation of Usher's syndrome in four siblings. *Clin Genet.* **24:** 273-9. (1983).

Karp, A., Santore, F. Retinitis pigmentosa and progressive hearing loss. *J Speech Hear Disord.* **48:** 308-14. (1983).

Keats, B.J., Berlin, C.I. Genomics and hearing impairment. *Genome Res.* **9:** 7-16. (1999).

Kere, J. Human population genetics: Lessons from Finland. *Annu Rev Genomics Hum Genet.* **2:** 103-28. (2001).

Kimberling, W.J., Möller, C.G., Davenport, S., Priluck, I.A., Beighton, P.H., Greenberg, J., Reardon, W., Weston, M.D., Kenyon, J.B., Grunkemeyer, J.A., et al. Linkage of Usher syndrome type I gene (USH1B) to the long arm of chromosome 11. *Genomics.* **14:** 988-94. (1992).

Kimberling, W.J., Weston, M.D., Möller, C., Davenport, S.L., Shugart, Y.Y., Priluck, I.A., Martini, A., Milani, M., Smith, R.J. Localization of Usher syndrome type II to chromosome 1q. *Genomics.* **7:** 245-9. (1990).

Kimberling, W.J., Weston, M.D., Möller, C., van Aarem, A., Cremers, C.W., Sumegi, J., Ing, P.S., Connolly, C., Martini, A., Milani, M., et al. Gene mapping of Usher syndrome type IIa: localization of the gene to a 2.1-cM segment on chromosome 1q41. *Am J Hum Genet.* **56:** 216-23. (1995).

Knudsen, S. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics.* **15:** 356-61. (1999).

Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15:** 8125-48. (1987).

Kumar, A., Fishman, G., Torok, N. Vestibular and auditory function in Usher's syndrome. *Ann Otol Rhinol Laryngol.* **93:** 600-8. (1984).

Küssel-Andermann, P., El-Amraoui, A., Safieddine, S., Nouaille, S., Perfettini, I., Lecuit, M., Cossart, P., Wolfrum, U., Petit, C. Vezatin, a novel transmembrane protein, bridges myosin VIIA to the cadherin-catenins complex. *Embo J.* **19:** 6020-9. (2000).

Kwok, P.Y. Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet.* **2:** 235-58. (2001).

Kyte, J., Doolittle, RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* **157**: 105-132. (1982).

Labeit, S., Kolmerer, B. Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science.* **270**: 293-96. (1995).

Lambrechts, A., Braun, A., Jonckheere, V., Aszodi, A., Lanier, L.M., Robbens, J., Van Colen, I., Vandekerckhove, J., Fassler, R., Ampe, C. Profilin II is alternatively spliced, resulting in profilin isoforms that are differentially expressed and have distinct biochemical properties. *Mol Cell Biol.* **20:** 8209-19. (2000).

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C.,

Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. Initial sequencing and analysis of the human genome. *Nature*. **409:** 860-921. (2001).

Larsen, F., Gundersen, G., Lopez, R., Prydz, H. CpG islands as gene markers in the human genome. *Genomics*. **13:** 1095-107. (1992).

Lathrop, G.M., Lalouel, J.M., Julier, C., Ott, J. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci U S A*. **81:** 3443-6. (1984).

Lehesjoki, A.E., Koskiniemi, M., Norio, R., Tirrito, S., Sistonen, P., Lander, E., de la Chapelle, A. Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. *Hum Mol Genet*. **2:** 1229-34. (1993).

Leroy, B. P., Aragon-Martin, J.A., Weston, M. D., Bessant, D. A. R., Willis, C., Webster, A. R., Bird, A. C., Kimberling, W. J., Payne, A. M., Bhattachararya, S.S. Spectrum of mutations in *USH2A* in British patients with Usher syndrome type II. *Exp Eye Res*. **72:** 503-09. (2001).

Lévy, G., Levi-Acobas, F., Blanchard, S., Gerber, S., Larget-Piet, D., Chenal, V., Liu, X.Z., Newton, V., Steel, K.P., Brown, S.D., Munnich, A., Kaplan, J., Petit, C., Weil, D. Myosin VIIA gene: heterogeneity of the mutations responsible for Usher syndrome type IB. *Hum Mol Genet*. **6:** 111-6. (1997).

Lewis, R.A., Otterud, B., Stauffer, D., Lalouel, J.M., Leppert, M. Mapping recessive ophthalmic diseases: linkage of the locus for Usher syndrome type II to a DNA marker on chromosome 1q. *Genomics*. **7:** 250-6. (1990).

Li, X.C., Everett, L.A., Lalwani, A.K., Desmukh, D., Friedman, T.B., Green, E.D., Wilcox, E.R. A mutation in PDS causes non-syndromic recessive deafness. *Nat Genet*. **18:** 215-7. (1998).

Liang, F., Ingeborg, H., Pertea, G., Karamycheva, S., Salzberg, S.L., Quackenbush, J. Gene Index of the human genome estimates approximately 120,000 genes. *Nat Genet*. **25:** 239-40.

Lichten, M.J., Fox, M.S. Detection of non-homology-containing heteroduplex molecules. *Nucleic Acids Res*. **11:** 3959-71. (1983).

Liebreich, R. Abkunft aus ehen unter blutsverwandten als grund von retinitis pigmentosa. *Dtsch Klin*. **13:** 53. (1861).

Lindenov, H. The aetiology of deaf mutism with special reference to heredity. Thesis (väitöskirja). Cobenhagen. Opera ex Domo University of Hafniensis no. 8. (1945).

Liu, X., Ondek, B., Williams, D.S. Mutant myosin VIIa causes defective melanosome distribution in the RPE of shaker-1 mice. *Nat Genet*. **19:** 117-8. (1998a).

Liu, X., Udovichenko, I.P., Brown, S.D., Steel, K.P., Williams, D.S. Myosin VIIa participates in opsin transport through the photoreceptor cilium. *J Neurosci*. **19:** 6267-74. (1999a).

Liu, X., Vansant, G., Udovichenko, I.P., Wolfrum, U., Williams, D.S. Myosin VIIa, the product of the Usher 1B syndrome gene, is concentrated in the connecting cilia of photoreceptor cells. *Cell Motil Cytoskeleton*. **37:** 240-52. (1997a).

Liu, X., Williams, D.S. Coincident onset of expression of myosin VIIa and opsin in the cilium of the developing photoreceptor cell. *Exp Eye Res*. **72:** 351-5. (2001).

Liu, X.-Z., Walsh, J., Tamagawa, Y., Kitamura, K., Nishizawa, M., Steel, K.P., Brown, S.D.M. Autosomal dominant non-dyndromic deafness caused by a mutation in the myosin VIIA gene. *Nat Genet*. **17:** 268-9. (1997b).

Liu, X.Z., Hope, C., Liang, C.Y., Zou, J.M., Xu, L.R., Cole, T., Mueller, R.F., Bundey, S., Nance, W., Steel, K.P., Brown, S.D. A Mutation (2314delG) in the Usher Syndrome Type IIA Gene: High Prevalence and Phenotypic Variation. *Am J Hum Genet*. **64:** 1221-25. (1999b).

Liu, X.Z., Hope, C., Walsh, J., Newton, V., Ke, X.M., Liang, C.Y., Xu, L.R., Zhou, J.M., Trump, D., Steel, K.P., Bundey, S., Brown, S.D. Mutations in the myosin VIIA gene cause a wide phenotypic spectrum, including atypical Usher syndrome. *Am J Hum Genet*. **63:** 909-12. (1998b).

Liu, X.Z., Newton, V.E., Steel, K.P., Brown, S.D. Identification of a new mutation of the myosin VII head region in Usher syndrome type 1. *Hum Mutat*. **10:** 168-70. (1997c).

Liu, X.Z., Walsh, J., Mburu, P., Kendrick-Jones, J., Cope, M.J., Steel, K.P., Brown, S.D. Mutations in the myosin VIIA gene cause non-syndromic recessive deafness. *Nat Genet*. **16:** 188-90. (1997d).

Livak, K.J., Marmaro, J., Todd, J.A. Towards fully automated genome-wide polymorphism screening. *Nat Genet*. **9:** 341-2. (1995).

Lord, EM., Gates, W.H. Shaker, a new mutation of the house mouse (Mus musculus). *Am Naturalist*. **63:** 435-42. (1929).

Lovett, M., Kere, J., Hinton, L.M. Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc Natl Acad Sci U S A*. **88:** 9628-32. (1991).

Luria, S.E., Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*. **28:** 491-511. (1943).

Marazita, M.L., Ploughman, L.M., Rawlings, B., Remington, E., Arnos, K.S., Nance, W.E. Genetic epidemiological studies of early-onset deafness in the U.S. school-age population. *Am J Med Genet*. **46:** 486-91. (1993).

Marra, M., Hillier, L., Kucaba, T., Allen, M., Barstead, R., Beck, C., Blistain, A., Bonaldo, M., Bowers, Y., Bowles, L., Cardenas, M., Chamberlain, A., Chappell, J., Clifton, S., Favello, A., Geisel, S., Gibbons, M., Harvey, N., Hill, F., Jackson, Y., Kohn, S., Lennon, G., Mardis, E., Martin, J., Waterston, R., et al. An encyclopedia of mouse genes. *Nat Genet*. **21:** 191-4. (1999).

Mburu, P., Liu, X.Z., Walsh, J., Saw, D., Jr., Cope, M.J., Gibson, F., Kendrick-Jones, J., Steel, K.P., Brown, S.D. Mutation analysis of the mouse myosin VIIA deafness gene. *Genes Funct*. **1:** 191-203. (1997).

McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J.,Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K., Fulton, R., Kucaba, T.A., Wagner-McPherson, C., Barbazuk, W.B., Gregory, S.G., Humphray, S.J., French, L., Evans, R.S., Bethel, G., Whittaker, A., Holden, J.L., McCann, O.T., Dunham, A., Soderlund, C., Scott, C.E., Bentley, D.R., Schuler, G., Chen, H.C., Jang, W., Green, E.D., Idol, J.R., Maduro, V.V., Montgomery, K.T., Lee, E., Miller, A., Emerling, S., Kucherlapati, Gibbs, R., Scherer, S., Gorrell, J.H., Sodergren, E., Clerc-Blankenburg, K., Tabor, P., Naylor, S., Garcia, D., de Jong, P.J., Catanese, J.J., Nowak, N., Osoegawa, K., Qin, S., Rowen, L., Madan, A., Dors, M., Hood, L., Trask, B., Friedman, C., Massa, H., Cheung, V.G., Kirsch, I.R., Reid, T., Yonescu, R., Weissenbach, J., Bruls, T., Heilig, R., Branscomb, E., Olsen, A., Doggett, N., Cheng, J.F., Hawkins, T., Myers, R.M., Shang, J., Ramirez, L., Schmutz, J., Velasquez, O., Dixon, K., Stone, N.E., Cox, D.R., Haussler, D., Kent, W.J., Furey, T., Rogic, S., Kennedy, S., Jones, S., Rosenthal, A., Wen, G., Schilhabel, M., Gloeckner, G., Nyakatura, G., Siebert, R., Schlegelberger, B., Korenberg, J., Chen, X.N., Fujiyama, A., Hattori, M., Toyoda, A., Yada, T., Park, H.S., Sakaki, Y., Shimizu, N., Asakawa, S., et al. A physical map of the human genome. *Nature*. **409:** 934-41. (2001).

McWilliam, P., Farrar, G.J., Kenna, P., Bradley, D.G., Humphries, M.M., Sharp, E.M., McConnell, D.J., Lawler, M., Sheils, D., Ryan, C., et al. Autosomal dominant retinitis pigmentosa (ADRP): localization of an ADRP gene to the long arm of chromosome 3. *Genomics*. **5:** 619-22. (1989).

Montell, C. A PDZ protein ushers in new links. *Nat Genet*. **26:** 6-7. (2000).

Morton, N.E. Genetic epidemiology of hearing impairment. *Ann N Y Acad Sci*. **630:** 16-31. (1991).

Mott, R. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Comput Appl Biosci*. **13:** 477-8. (1998).

Mount, D. Bioinformatics: sequence and genome analysis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. (2001).

Müller, U., Littlewood-Evans, A. Mechanisms that regulate mechanosensory hair cell differentiation. *Trends Cell Biol*. **11:** 334-42. (2001).

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A., Anson, E.L., Bolanos, R.A., Chou, H.H., Jordan, C.M., Halpern, A.L., Lonardi, S., Beasley, E.M., Brandon, R.C., Chen, L., Dunn, P.J., Lai, Z., Liang, Y., Nusskern, D.R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G.M., Adams, M.D., Venter, J.C. A whole-genome assembly of Drosophila. *Science*. **287:** 2196-204. (2000).

Myers, R.M., Fischer, S.G., Lerman, L.S., Maniatis, T. Nearly all single base substitutions in DNA fragments joined to a GC- clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Res*. **13:** 3131-45. (1985).

Möller, C.G., Kimberling, W.J., Davenport, S.L., Priluck, I., White, V., Biscone-Halterman, K., Odkvist, L.M., Brookhouser, P.E., Lund, G., Grissom, T.J. Usher syndrome: an otoneurologic study. *Laryngoscope*. **99:** 73-9. (1989).

Nadeau, J.H. Modifier genes in mice and humans. *Nat Rev Genet*. **2:** 165-74. (2001).

Nagase, T., Seki, N., Ishikawa, K., Tanaka, A., Nomura, N. Prediction of the coding sequences of unidentified human genes. V. The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res* **3:** 17-24. (1996).

Nakai, K., Horton, P. PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. *Trends Biochem Sci*. **24:** 34-6. (1999).

Naylor, S.L., Moore, S., Garcia, D., Xiang, X., Xin, X., Mohrer, M., Reus, B., Linn, R., Stanton, V., O'Connell, P., Leach, R.J. Mapping 638 STSs to regions of human chromosome 3. *Cytogenet Cell Genet*. **72:** 90-4. (1996).

Nevanlinna, H.R. The Finnish population structure. A genetic and genealogical study. *Hereditas* **71**: 195-236. (1972).

Ng, L., Hurley, J.B., Dierks, B., Srinivas, M., Salto, C., Vennstrom, B., Reh, T.A., Forrest, D. A thyroid hormone receptor that is required for the development of green cone photoreceptors. *Nat Genet*. **27:** 94-8. (2001).

Nielson, H., Engelbrecht J., Brunak, S., von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their celavage sites. *Protein eng*. **10**: 1-6. (1997).

Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayasi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K., Tabata, S. Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1 [published erratum appears in DNA Res 1995 Aug 31;2(4):211]. *DNA Res*. **1:** 27-35. (1994).

Norio, R., Nevanlinna, H.R., Perheentupa, J. Hereditary diseases in Finland; rare flora in rare soul. *Ann Clin Res*. **5:** 109-41. (1973).

Nuutila, A. Neuropsychiatric and genetic aspects of the dystrophia retinae pigmentosa - dysacusis syndrome. Thesis (väitöskirja), University of Helsinki. (1968).

Nuutila, A. Dystrophia retinae pigmentosa--dysacusis syndrome (DRD): a study of the Usher- or Hallgren syndrome. *J Genet Hum*. **18:** 57-88. (1970).

O'Connell, J.R., Weeks, D.E. The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance [see comments]. *Nat Genet*. **11:** 402-8. (1995).

OMIM, Online Mendelian Inheritance in Man. http://www.ncbi.nlm.nih.gov/omim/.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., Sekiya, T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A*. **86:** 2766-70. (1989).

Ott, J. Analysis of Human Genetic Linakge. Baltimore: John Hopkins University Press, Baltimore. (1991).

Otten, A.D., Tapscott, S.J. Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc Natl Acad Sci U S A*. **92:** 5465-9. (1995).

Pakarinen, L. Usher syndrome type III (USH3). Thesis (väitöskirja), University of Tampere, no. 574. (1997).

Pakarinen, L., Karjalainen, S., Simola, K.O., Laippala, P., Kaitalo, H. Usher's syndrome type 3 in Finland. *Laryngoscope*. **105:** 613-7. (1995a).

Pakarinen, L., Sankila, E-M., Tuppurainen, K., Karjalainen S., Kääriäinen, H. (1995b). Usher syndrome type III (USH3): the clinical manifestations in 42 patients. *Scand J Long Phon*. **20:** 141-9.

Pakarinen, L., Tuppurainen, K., Laippala, P., Mäntyjärvi, M., Puhakka, H. The ophthalmological course of Usher syndrome type III. *Int Ophthalmol*. **19:** 307-11. (1996).

Parimoo, S., Patanjali, S.R., Shukla, H., Chaplin, D.D., Weissman, S.M. cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc Natl Acad Sci U S A*. **88:** 9623-7. (1991).

Peltonen, L., Jalanko, A., Varilo, T. Molecular genetics of the Finnish disease heritage. *Hum Mol Genet*. **8:** 1913-23. (1999).

Peltonen, L., Palotie, A., Lange, K. Use of population isolates for mapping complex traits. *Nat Rev Genet*. **1:** 182-90. (2000).

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

Petit, C. Usher Synrome: From Genetics to Pathogenesis. *Annu Rev Genomics Hum Genet*. **2:** 271-97. (2001).

Philibert, R.A., King, B.H., Winfield, S., Cook, E.H., Lee, Y.H., Stubblefield, B., Damschroder-Williams, P., Dea, C., Palotie, A., Tengstrom, C., Martin, B.M., Ginns, E.I. Association of an X-chromosome dodecamer insertional variant allele with mental retardation. *Mol Psychiatry*. **3:** 303-9. (1998).

Philibert, R.A., Winfield, S.L., Damschroder-Williams, P., Tengstrom, C., Martin, B.M., Ginns, E.I. The genomic structure and developmental expression patterns of the human OPA-containing gene (HOPA). *Hum Genet*. **105:** 174-8. (1999).

Pieke Dahl, S., Kimberling, W.J., Gorin, M.B., Weston, M.D., Furman, J.M., Pikus, A., Möller, C. Genetic heterogeneity of Usher syndrome type II. *J Med Genet*. **30:** 843-8. (1993).

Pieke-Dahl, S., Möller, C.G., Kelley, P.M., Astuto, L.M., Cremers, C.W., Gorin, M.B., Kimberling, W.J. Genetic heterogeneity of Usher syndrome type II: localisation to chromosome 5q. *J Med Genet*. **37:** 256-62. (2000).

Potier, M.C., Chelot, E., Pekarsky, Y., Gardiner, K., Rossier, J., Turnell, W.G. The human myosin light chain kinase (MLCK) from hippocampus: cloning, sequencing, expression, and localization to 3qcen-q21. *Genomics*. **29:** 562-70. (1995).

Pozzoli, U., Sironi, M., Cagliani, R., Comi, G.P., Bardoni, A., Bresolin, N. Comparative analysis of the human dystrophin and utrophin gene structures. *Genetics*. **160:** 793-98. (2002).

Rattner, A., Sun, H., Nathans, J. Molecular genetics of human retinal disease. *Annu Rev Genet*. **33:** 89-131. (1999).

Reese, M.G., Eeckman, F.H., Kulp, D., Haussler, D. Improved splice site detection in Genie. *J Comput Biol*. **4:** 311-23. (1997).

Resendes, B.L., Williamson, R.E., Morton, C.C. At the speed of sound: gene discovery in the auditory system. *Am J Hum Genet*. **69:** 923-35. (2001).

Richardson, G.P., Forge, A., Kros, C.J., Fleming, J., Brown, S.D., Steel, K.P. Myosin VIIA is required for aminoglycoside accumulation in cochlear hair cells. *J Neurosci*. **17:** 9506-19. (1997).

Ridanpää, M., van Eenennaam, H., Pelin, K., Chadwick, R., Johnson, C., Yuan, B., vanVenrooij, W., Pruijn, G., Salmela, R., Rockas, S., Mäkitie, O., Kaitila, I., de la Chapelle, A. Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. *Cell*. **104:** 195-203. (2001).

Rivolta, C., Sweklo, E.A., Berson, E.L., Dryja, T.P. Missense mutation in the USH2A gene: association with recessive retinitis pigmentosa without hearing loss. *Am J Hum Genet*. **66:** 1975-8. (2000).

Robson, K.J., Chandra, T., MacGillivray, R.T., Woo, S.L. Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA. *Proc Natl Acad Sci U S A*. **79:** 4701-5. (1982).

Roepman, R., van Duijnhoven, G., Rosenberg, T., Pinckers, A.J., Bleeker-Wagemakers, L.M., Bergen, A.A., Post, J., Beck, A., Reinhardt, R., Ropers, H.H., Cremers, F.P., Berger, W. Positional cloning of the gene for X-linked retinitis pigmentosa 3: homology with the guanine-nucleotide-exchange factor RCC1. *Hum Mol Genet*. **5:** 1035-41. (1996).

Rommens, J.M., Iannuzzi, M.C., Kerem, B., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*. **245:** 1059-65. (1989).

Rosenberg, T., Haim, M., Hauch, A.M., Parving, A. The prevalence of Usher syndrome and other retinal dystrophy-hearing impairment associations. *Clin Genet*. **51:** 314-21. (1997).

Rosenfeld, P.J., McKusick, V.A, Amberger, J., Dryja, T.P. Recent advances in the gene map of inherited eye disorders: Primary hereditary diseases of the retina, choroid, and vitreous. *J Med Genet*. **31:** 903-15. (1994).

Rubin, G.M. The draft sequences. Comparing species. *Nature*. **409:** 820-1. (2001).

Rüsch, A., Erway, L.C., Oliver, D., Vennstrom, B., Forrest, D. Thyroid hormone receptor beta-dependent expression of a potassium conductance in inner hair cells at the onset of hearing. *Proc Natl Acad Sci U S A*. **95:** 15758-62. (1998).

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., Atshuler, D. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. **409:** 928-33. (2001).

Sankila, E.M., Pakarinen, L., Kääriäinen, H., Aittomäki, K., Karjalainen, S., Sistonen, P., de la Chapelle, A. Assignment of an Usher syndrome type III (USH3) gene to chromosome 3q. *Hum Mol Genet*. **4:** 93-8. (1995).

Sankila, E.M., Sistonen, P., Cremers, F., de la Chapelle, A. Choroideremia: linkage analysis with physically mapped close DNA- markers. *Hum Genet*. **87:** 348-52. (1991).

Sankila, E.M., Tolvanen, R., van den Hurk, J.A., Cremers, F.P., de la Chapelle, A. Aberrant splicing of the CHM gene is a significant cause of choroideremia. *Nat Genet*. **1:** 109-13. (1992).

Seeliger, M.W., Apfelstedt-Sylla, E., Jaissle, G.B. ERG implicit time separates Usher syndrome I and II (Arvo abstract). *Invest Opthalmol Vis Sci*. **40:** 1020-24. (2000).

Self, T., Mahony, M., Fleming, J., Walsh, J., Brown, S.D., Steel, K.P. Shaker-1 mutations reveal roles for myosin VIIA in both development and function of cochlear hair cells. *Development*. **125:** 557-66. (1998).

Sharp, C.W., Muir, W.J., Blackwood, D.H., Walker, M., Gosden, C., St Clair, D.M. Schizophrenia and mental retardation associated in a pedigree with retinitis pigmentosa and sensorineural deafness. *Am J Med Genet*. **54:** 354-60. (1994).

Sheffield, V.C., Weber, J.L., Buetow, K.H., Murray, J.C., Even, D.A., Wiles, K., Gastier, J.M., Pulido, J.C., Yandava, C., Sunden, S.L., et al. A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. *Hum Mol Genet*. **4:** 1837-44. (1995).

Shinkawa, H., Nadol, J.B. Histopathology of the inner ear in Usher's syndrome as observed by light and electron microscopy. *Ann Otol Rhinol Laryngol*. **95:** 313-8. (1986).

Smith, R.J., Berlin, C.I., Hejtmancik, J.F., Keats, B.J., Kimberling, W.J., Lewis, R.A., Möller, C.G., Pelias, M.Z., Tranebjaerg, L. Clinical diagnosis of the Usher syndromes. Usher Syndrome Consortium. *Am J Med Genet*. **50:** 32-8. (1994).

Smith, R.J., Lee, E.C., Kimberling, W.J., Daiger, S.P., Pelias, M.Z., Keats, B.J., Jay, M.,Bird, A., Reardon, W., Guest, M., et al. Localization of two genes for Usher syndrome type I to chromosome 11. *Genomics*. **14:** 995-1002. (1992).

Solovyev, V., Salamov, A. The Gene-Finder computer tools for analysis of human and model organisms genome sequences. *Proc Int Conf Intell Syst Mol Biol*. **5:** 294-302. (1997).

Steel, K.P. A new era in the genetics of deafness. *N Engl J Med*. **339:** 1545-7. (1998).

Steel, K.P., Kros, C.J. A genetic approach to understanding auditory function. *Nat Genet*. **27:** 143-9. (2001).

Stott, K., Blackburn, J.M., Butler, P.J., Perutz, M. Incorporation of glutamine repeats makes protein oligomerize: implications for neurodegenerative diseases. *Proc Natl Acad Sci U S A*. **92:** 6509-13. (1995).

Sullivan, L.S., Heckenlively, J.R., Bowne, S.J., Zuo, J., Hide, W.A., Gal, A., Denton, M., Inglehearn, C.F., Blanton, S.H., Daiger, S.P. Mutations in a novel retina-specific gene cause autosomal dominant retinitis pigmentosa. *Nat Genet*. **22:** 255-9. (1999).

Tang, X.X., Biegel, J.A., Nycum, L.M., Yoshioka, A., Brodeur, G.M., Pleasure, D.E., Ikegaki, N. cDNA cloning, molecular characterization, and chromosomal localization of NET (EPHT2), a human EPH-related receptor protein-tyrosine kinase gene preferentially expressed in brain. *Genomics*. **29:** 426-37. (1995).

Terwilliger, J.D. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am J Hum Genet*. **56:** 777-87. (1995).

Terwilliger, JD., Ott, J. Handbook of Human Genetic linkage. Baltimore: John Hopkins Press. (1994).

Theriot, J.A., Mitchison, T.J. The three faces of profilin. *Cell*. **75:** 835-8. (1993).

Titus, M.A. A class VII unconventional myosin is required for phagocytosis. *Curr Biol*. **9:** 1297-303. (1999).

Travis, G.H. Mechanisms of cell death in the inherited retinal degenerations. *Am J Hum Genet*. **62:** 503-8. (1998).

Usher, CH. On the inheritance of retinitis pigmentosa, with notes of cases. *Roy Lond Opthalmol Hosp Rev*. **19:** 130-6. (1914).

Überbacher, E.C., Mural, R.J. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc Natl Acad Sci U S A*. **88:** 11261-5. (1991).

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W. Serial analysis of gene expression. *Science*. **270:** 484-7. (1995).

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. The sequence of the human genome. *Science*. **291:** 1304-51. (2001).

Vernon, M. Usher's Syndrome-Deafness and progressive blindness: clinical case, prevention, theory and literature survey. *J Chronic Dis*. **22:** 133-51. (1969).

Verpy, E., Leibovici, M., Zwaenepoel, I., Liu, X.Z., Gal, A., Salem, N., Mansour, A., Blanchard, S., Kobayashi, I., Keats, B.J., Slim, R., Petit, C. A defect in harmonin, a PDZ domain-containing protein expressed in the inner ear sensory hair cells, underlies Usher syndrome type 1C. *Nat Genet*. **26:** 51-5. (2000).

Volinia, S., Hiles, I., Ormondroyd, E., Nizetic, D., Antonacci, R., Rocchi, M., Waterfield, M.D. Molecular cloning, cDNA sequence, and chromosomal localization of the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene. *Genomics.* **24:** 472-7. (1994).

Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lander, E.S., et al. Large-scale identification, mapping, and genotyping of single- nucleotide polymorphisms in the human genome. *Science.* **280:** 1077-82. (1998).

Wayne, S., Der Kaloustian, V.M., Schloss, M., Polomeno, R., Scott, D.A., Hejtmancik, J.F., Sheffield, V.C., Smith, R.J. Localization of the Usher syndrome type ID gene (Ush1D) to chromosome 10. *Hum Mol Genet.* **5:** 1689-92. (1996).

Wayne S., Lowry R. B., McLeod D. R. Knaus R., Farr C., Smith R. J. H. et al. Localization of the Usher syndrome type 1F (Ush1F) to chromosome 10. *Am J Hum Gen.* **61**: A300. (1997).

Weil, D., Blanchard, S., Kaplan, J., Guilford, P., Gibson, F., Walsh, J., Mburu, P., Varela, A., Levilliers, J., Weston, M.D., et al. Defective myosin VIIA gene responsible for Usher syndrome type 1B. *Nature.* **374:** 60-1. (1995).

Weil, D., Kussel, P., Blanchard, S., Levy, G., Levi-Acobas, F., Drira, M., Ayadi, H., Petit, C. The autosomal recessive isolated deafness, DFNB2, and the Usher 1B syndrome are allelic defects of the myosin-VIIA gene. *Nat Genet.* **16:** 191-3. (1997).

Weston, M.D., Eudy, J.D., Fujita, S., Yao, S., Usami, S., Cremers, C., Greenberg, J., Ramesar, R., Martini, A., Möller, C., Smith, R.J., Sumegi, J., Kimberling, W.J., Greenburg, J. Genomic structure and identification of novel mutations in usherin, the gene responsible for Usher syndrome type IIa. *Am J Hum Genet.* **66:** 1199-210. (2000).

Weston, M.D., Kelley, P.M., Overbeck, L.D., Wagenaar, M., Orten, D.J., Hasson, T., Chen, Z.Y., Corey, D., Mooseker, M., Sumegi, J., Cremers, C., Möller, C., Jacobson, S.G., Gorin, M.B., Kimberling, W.J. Myosin VIIA mutation screening in 189 Usher syndrome type 1 patients. *Am J Hum Genet.* **59:** 1074-83. (1996).

Willems, P.J. Genetic causes of hearing loss. *N Engl J Med.* **342:** 1101-9. (2000).

Wright, F.A., Lemon, W.J., Zhao, W.D., Sears, R., Zhuo, D., Wang, J-P., Yang H-Y., Baer, T., Stredney D., Spitzner J., Stutz, A., Krahe, R., Yuan, B. A draft annotation and overview of the human genome. *Genome Biol.* **2:** 0025.1.-0025.18. (2001).

Zhang, M.Q. Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proc Natl Acad Sci U S A.* **94:** 565-8. (1997).

Zhang, Z., Schaffer, A.A., Miller, W., Madden, T.L., Lipman, D.J., Koonin, E.V., Altschul, S.F. Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res.* **26:** 3986-90. (1998).

Zwaenepoel, I., Verpy, E., Blanchard, S., Meins, M., Apfelstedt-Sylla, E., Gal, A., Petit, C. Identification of three novel mutations in the USH1C gene and detection of thirty-one polymorphisms used for haplotype analysis. *Hum Mutat.* **17:** 34-41. (2001).