Pendred syndrome is caused by mutations in a putative sulphate transporter gene (*PDS*)

Lorraine A. Everett[1], Benjamin Glaser[2], John C. Beck[3], Jacquelyn R. Idol[1], Andreas Buchs[2], Ma'ayan Heyman[2], Faiad Adawi[4], Elizur Hazani[5], Elias Nassir[5], Andreas D. Baxevanis[1], Val C. Sheffield[3] & Eric D. Green[1]

Pendred syndrome is a recessively inherited disorder with the hallmark features of congenital deafness and thyroid goitre. By some estimates, the disorder may account for upwards of 10% of hereditary deafness. Previous genetic linkage studies localized the gene to a broad interval on human chromosome 7q22–31.1. Using a positional cloning strategy, we have identified the gene (*PDS*) mutated in Pendred syndrome and found three apparently deleterious mutations, each segregating with the disease in the respective families in which they occur. *PDS* produces a transcript of approximately 5 kb that was found to be expressed at significant levels only in the thyroid. The predicted protein, pendrin, is closely related to a number of known sulphate transporters. These studies provide compelling evidence that defects in pendrin cause Pendred syndrome thereby launching a new area of investigation into thyroid physiology, the pathogenesis of congenital deafness and the role of altered sulphate transport in human disease.

Pendred syndrome is an autosomal-recessive disorder characterized by congenital sensorineural hearing loss combined with goitre[1–3] [OMIM entry 274600 (http://www3.ncbi.nlm.nih.gov/Omim)]. The incidence of the disease is estimated at 7.5–10 in 100,000 (refs 2,3). Pendred syndrome may account for as many as 10% of the cases of hereditary deafness[3–6], making it the most common form of syndromic deafness. Although typically present at birth, the deafness can be variable in its expression[7], sometimes appearing during childhood, and is most often associated with a Mondini cochlear defect[8–10], in which the basal cochlear turn is retained while the apical turns form a common cavity.

Even more variable in its presentation is the thyroid disease—particularly with respect to the extent of goitre, being apparently absent in some affected individuals. Striking examples of intrafamilial variability in the thyroid manifestations of Pendred syndrome have also been noted[2,7]. The goitre is multinodular and can grow to a large size, with many patients ultimately requiring surgery for tracheal compression. In general, affected individuals are clinically euthyroid and have normal serum T_3 and T_4 levels, although thyroid-stimulating hormone levels are often in the upper end of the normal range and hypothyroidism can develop. Pendred-syndrome patients typically have a raised serum thyroglobulin level as well as a positive perchlorate discharge test (an inappropriate discharge of unincorporated iodide from the thyroid after perchlorate challenge)[2,3]. The latter suggests that the underlying defect interferes with the binding of iodine to thyroglobulin, consistent with the finding of defective iodine organification in thyroid tissue obtained from a Pendred-syndrome patient[11].

The nature of the association between the thyroid and cochlear defects in Pendred syndrome is obscure. Although deafness has been associated with hypothyroidism in both human disease and animal studies[12–20], patients with Pendred syndrome generally display normal thyroid function, particularly early in life. Thus, the identification of the gene that is defective in Pendred syndrome should provide insight about the pathophysiology of this disorder as well as normal thyroid function and cochlear development.

In 1996, two groups reported linkage of Pendred syndrome to the same broad region on human chromosome 7q22–31.1 (refs 11,21). Together, these initial studies indicated that the gene resides within an approximately 5.4-cM interval[22,23]. Subsequent studies[24,25] further reduced the size of the critical region to about 1.7 cM. Genetic linkage of Pendred syndrome to this region of chromosome 7 has been reported in more than twenty families of different ethnic origins[3,11,21,24,25], strongly suggesting locus homogeneity for this disorder. By a combination of genetic analysis of additional families and high-resolution physical mapping coupled with systematic DNA sequencing, we have identified the gene responsible for Pendred syndrome. Here we report the details of our positional cloning effort, the general features of the Pendred syndrome gene (*PDS*) and the disease-causing mutations identified to date.

Genetic refinement of the critical region

We sought to use a positional cloning strategy to identify the gene mutated in Pendred syndrome. Previous genetic mapping studies[11,21,24,25] collectively indicated that the Pendred-syndrome gene resided within an approximately 1.7-cM interval defined by the markers *D7S501* and *D7S692*. In an attempt to reduce the size of this critical region, we performed genetic analysis on three additional large families of Arabic descent (families 3–5 in Fig. 1). The consanguineous nature of these families as well as those we reported previously (families 1 and 2; ref. 11) allows detection of ancestral recombination events by homozygosity mapping. All of the affected individuals in family 3 are associated with an ancestral recombination between markers *D7S2459* and *D7S2456*, as

[1]*Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.* [2]*Department of Endocrinology and Metabolism, Hadassah University Hospital, Ein Karem, Jerusalem, Israel 91120.* [3]*Howard Hughes Medical Institute and Department of Pediatrics, Division of Medical Genetics, University of Iowa, Iowa City, Iowa 52242, USA.* [4]*Department of Endocrinology, Rivka Zeev Hospital, Zafad, Israel 16950.* [5]*Department of Internal Medicine, Western Galilee-Nahariya Hospital, Nahariya, Israel 22100. Correspondence should be addressed to E.D.G. (e-mail: egreen@nhgri.nih.gov), B.G. (e-mail: beng@cc.huji.ac.il) or V.C.S. (e-mail: val-sheffield@uiowa.edu).*

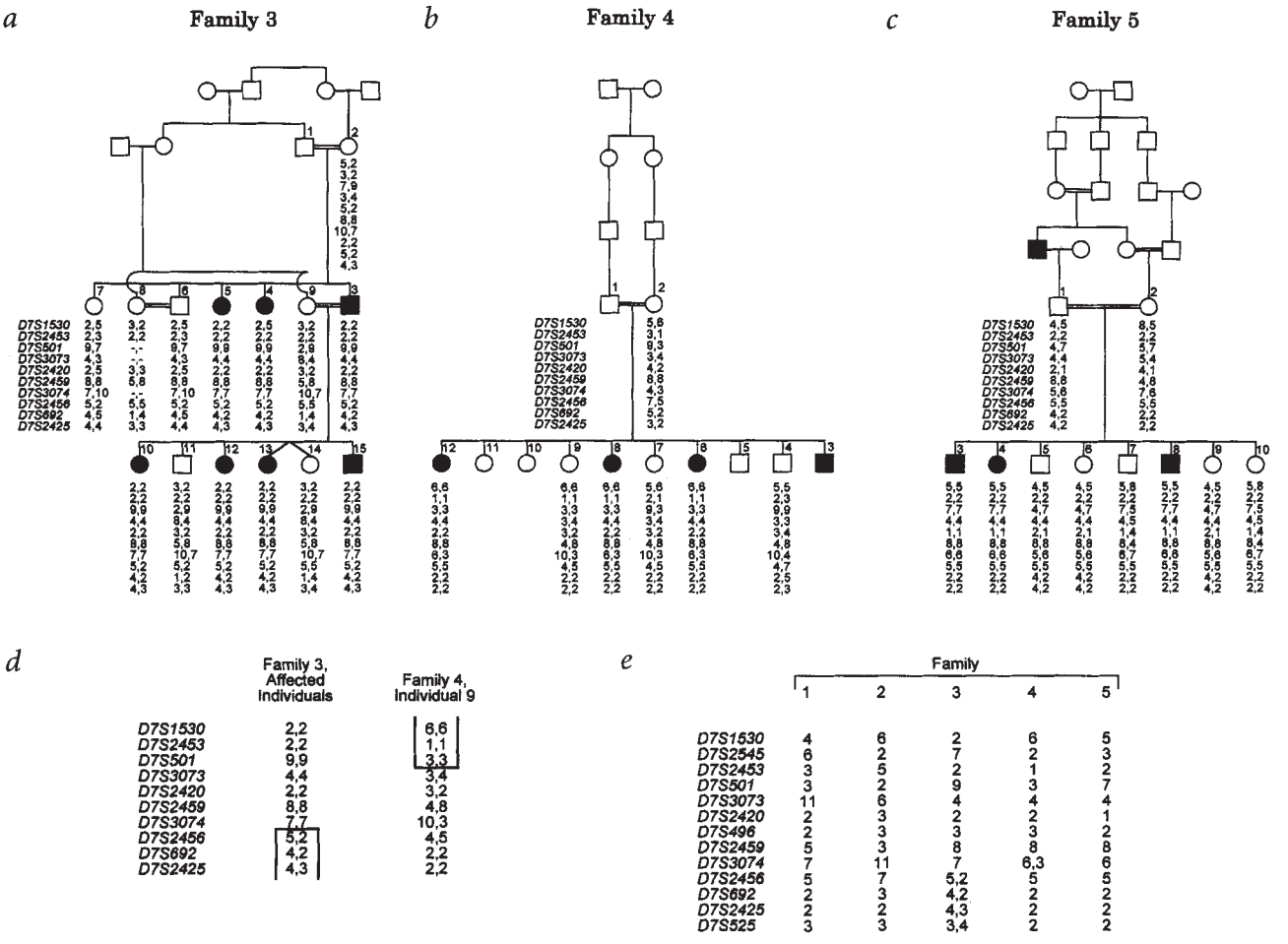

Fig. 1 Three newly ascertained Pendred-syndrome families. Pedigrees and genotype data are provided for families 3 (***a***), 4 (***b***) and 5 (***c***). Individuals affected (solid symbols) and unaffected (open symbols) with Pendred syndrome are indicated. Below each individual is the corresponding genotype data generated for the indicated genetic markers [listed in physical mapping order (Fig. 2) from the centromeric (top) to telomeric (bottom) ends of the genomic interval containing the Pendred-syndrome critical region]. ***d***, The positions of key recombination events in the families 3 and 4 that defined the critical region used for the described positional cloning effort are indicated. **e**, A summary of the disease-associated haplotypes in families 1–5 is provided. Some of the genetic mapping data shown for families 3–5 were generated after the identification of the *PDS* gene. Note that the allele numbering used in the figure differs from that initially reported in Sheffield *et al.*[11] because of the identification of additional alleles in families 3–5.

evidenced by the lack of homozygosity at *D7S2456* (Fig. 1). These data thus defined a new telomeric boundary (*D7S2456*) for the Pendred-syndrome critical region. In addition, the recombination detected in the unaffected individual 9 in family 4 confirms the data reported by others[21,25] and defines *D7S501* as the centromeric boundary for the critical region (Fig. 1).

To refine further the telomeric boundary, we developed new genetic markers between *D7S2459* and *D7S2456*. A new marker derived from genomic sequence data, *D7S3074*, was homozygous in all affected individuals in family 3, indicating that the recombination occurred between *D7S2456* and *D7S3074* (Fig. 1). All individuals in family 4 are heterozygous at *D7S3074*, yet affected individuals in this family are homozygous for all of the flanking markers. The most likely explanation for this apparent discrepancy is that there has been an ancestral mutation within the polymorphism itself that has created the different alleles of the marker in this family. Similar efforts at refining the centromeric boundary included the development of a new genetic marker, *D7S3073*, between *D7S501* and *D7S2420*. The unaffected recombinant individual 9 in family 4 is heterozygous at this locus, thus localizing the recombination between *D7S501* and *D7S3073* (Fig. 1).

These results indicate that, based on recombinants (Fig. 1*d*), the refined Pendred-syndrome critical region is defined by the flanking markers *D7S501* and *D7S2456*, a genetic distance estimated at about 1.1 cM[22,23]. These genetic mapping data guided the positional cloning effort described below.

Physical mapping of the critical region

We previously reported the construction of a complete yeast artificial chromosome (YAC)–based physical map of human chromosome 7 (refs 26–30). The greater Pendred-syndrome critical region is contained within a single YAC contig [N (sWSS6)[29]; see http://www.nhgri.nih.gov/DIR/GTB/CHR7]. Using this contig map and its associated sequence-tagged sites (STSs) as a starting framework, we developed a fully contiguous bacterial artificial chromosome (BAC) contig containing the key flanking markers, *D7S501* (sWSS1170) and *D7S2456* (sWSS2328). Initially, we isolated BACs containing the 38 STSs from the starting YAC-based map[29]. The resulting STS-content data allowed the assembly of six BAC contigs. STSs were then developed from the insert ends of strategically selected clones and used to isolate additional BACs. The process was repeated several times, eventually closing

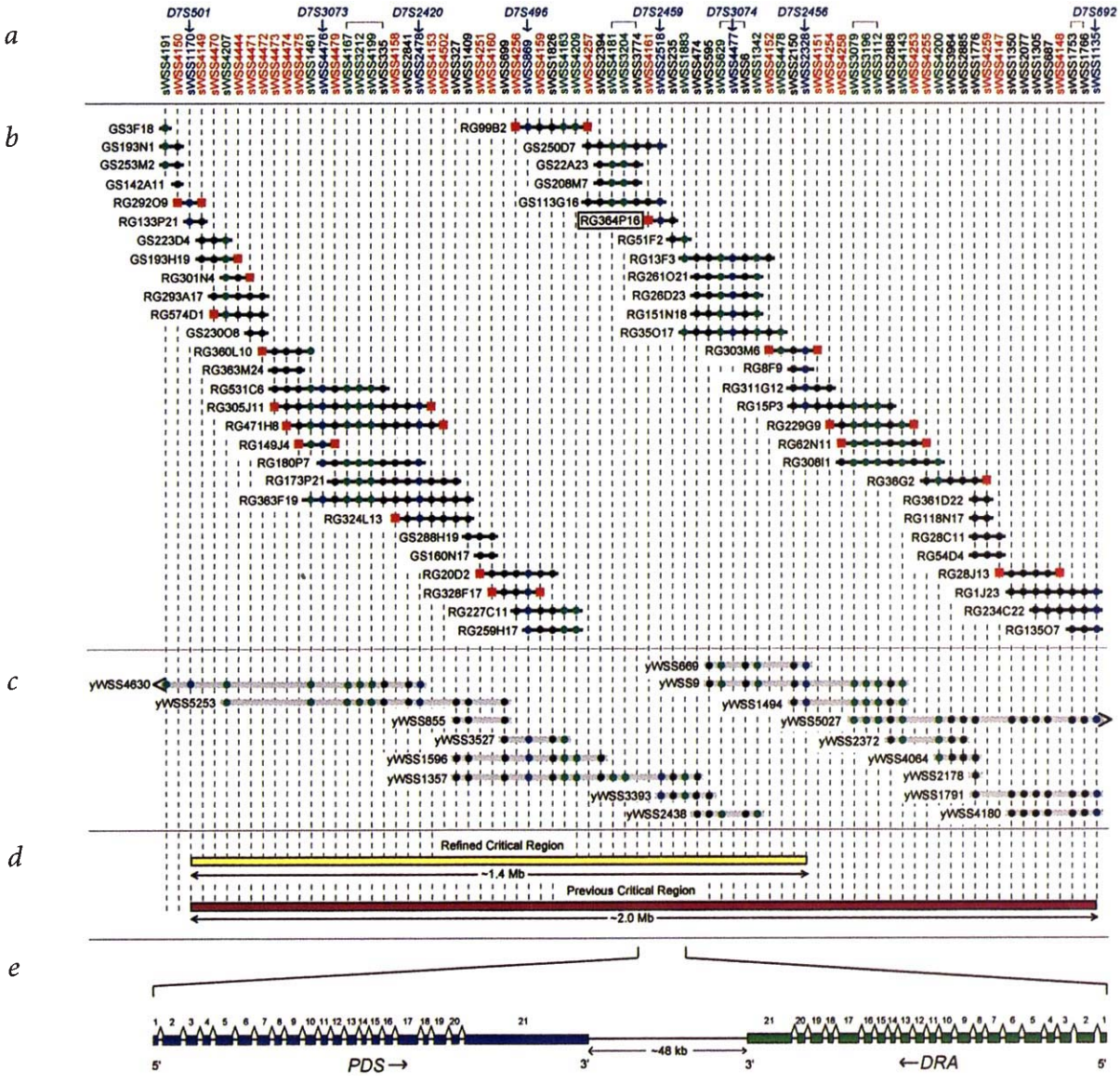

Fig. 2 High-resolution physical map of the Pendred-syndrome critical region. A BAC-based STS-content map of the Pendred-syndrome critical region is depicted (orientated with the centromeric end leftwards and 7q telomeric end rightwards). ***a***, The relative positions of the mapped STSs are indicated, with detailed information about all of the STSs available in GenBank and GDB. The isolated BACs (***b***) and a representative set of YACs (***c***) are shown as horizontal lines. The 'RG' or 'GS' before each BAC name indicates the library of origin (Research Genetics and Genome Systems, respectively). A coloured circle or square indicates that the STS is present in that clone. The sources of STSs were BAC insert ends (red), genetic markers (blue), ESTs or known genes (green) and miscellaneous other sequences (black). The absence of a circle or square at the expected position of an STS within a YAC indicates that the clone was not tested for that marker. The map is displayed with equal spacing between STSs. Markers with brackets above their names could not be uniquely ordered on the basis of the STS content of the BACs. The absence of a bracket above STSs that appear unresolved on the BAC map were, in fact, resolved on the complete YAC map [contig N (sWSS6; ref. 29; see http://www.nhgri.nih.gov/DIR/GTB/CHR7]. ***d***, The boundaries of the previous and refined Pendred-syndrome critical region are indicated. **e**, The general structure and approximate positions of the *PDS* and *DRA* genes are schematically illustrated (without indicating actual intron sizes). Note the similar intron–exon structure of the two genes. A box is placed around the name of BAC RG364P16, as this clone was found to contain most of the *DRA* gene as well as the 3′ portion of the *PDS* gene.

all of the gaps and providing complete BAC-based continuity across the region (Fig. 2). In addition, virtually all of the depicted overlaps between clones were confirmed by restriction enzyme-based fingerprint analysis[31] (M. Marra and J. McPherson, unpublished data). STSs developed from various genes and expressed-sequence tags (ESTs) were also mapped to the isolated BACs. The total interval spanned by the contig is about 2.0 Mb, based on both the measured sizes of the BAC inserts and the long-range restriction map of the region established previously[32]. The contig contains a total of 78 STSs (38 initial STSs and 40 newly developed STSs); therefore, the map contains an STS, on average, approximately every 26 kb. The refined Pendred-syndrome critical region (between *D7S501* and *D7S2456*) is estimated to be about 1.4 Mb in size. As part of a broader collaborative effort to sequence human chromosome 7, BACs (Fig. 2) were sent to the Washington University Genome Sequencing Center for systematic sequencing of the region.

Evaluation of candidate genes in the critical region

Several strategies were used to identify genes within the Pendred-syndrome critical region. Initially, a number of known genes (laminin β1 chain, GenBank M61916; *DRA*, GenBank L02785;

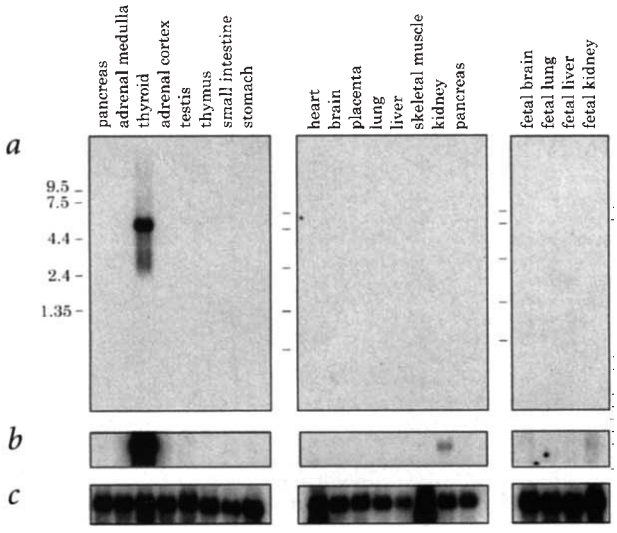

Fig. 3 Expression profile of *PDS*. Northern blots containing polyA+ RNA (2 μg) from the indicated tissues were hybridized sequentially with a *PDS*– (***a***,***b***) and then human β-actin–specific (***c***) probe. After autoradiography for seven hours (***a***), an intense signal of approximately 5 kb was seen with thyroid RNA only. After autoradiography for 100 hours (***b***), weak (but discrete) signals of approximately 5 kb could also be seen with adult and fetal kidney RNA as well as a diffuse signal with fetal brain RNA. ***c***, After hybridization with the β-actin probe, autoradiography was performed for 0.5 hour.

DLD, GenBank L13748; cAMP-dependent protein kinase subunit RII-β, GenBank M31158.) and an EST (WI-6690) were localized within this region during construction of the YAC-based physical map. Next, we rigorously reviewed the genome-wide 'transcript map'[33] to identify ESTs mapping within the general vicinity of the critical region. Specifically, ESTs within the interval defined by the genetic markers *D7S658* and *D7S648* were analysed and then tested for their presence within a BAC contig (Fig. 2). This effort resulted in the identification of five additional candidate genes/ESTs within the Pendred-syndrome critical region. Finally, sequence data from BACs within the critical region (Fig. 2) were routinely scrutinized for the presence of putative genes. Like other groups engaged in large-scale human genomic sequencing within the Human Genome Project, the Washington University Genome Sequencing Center has a policy of making 'unfinished' sequences available electronically to all investigators (http://genome.wustl.edu/gsc). We routinely analysed new sequences from the BACs for the presence of putative genes with a variety of computational tools (see Methods).

Our general approach for assessing the possible role of a gene in Pendred syndrome entailed a standard series of steps, typically including full-length cDNA isolation, northern analysis and direct sequencing of PCR-amplified segments from affected individuals. All (or most) of the open reading frames (ORFs) of seven genes (three known genes and four characterized ESTs) were screened in this fashion. No amino acid–altering changes were encountered. The eighth gene to be analysed was ultimately found to be mutated in Pendred syndrome and therefore corresponds to the Pendred-syndrome gene (*PDS*). A detailed description of *PDS* and its mutations are provided below.

Isolation and characterization of *PDS*

GRAIL analysis[34,35] (http://compbio.ornl.gov) of the partially assembled sequence of BAC RG364P16 (Fig. 2) revealed the presence of several putative exons. These DNA segments were then compared with all sequences in GenBank with BLASTX[36,37], which revealed strong homology to a portion of the *DRA* ('down-regulated in adenoma') gene[38,39]. Of note, the protein encoded by *DRA* functions as a sulphate transporter[40] and is defective in congenital chloride diarrhoea syndrome[41]. *DRA* resides close to *PDS* within the critical region (Fig. 2). With primers specific for the GRAIL-predicted, *DRA*-related sequences, RT-PCR as well as 5′– and 3′–RACE were used to amplify the *PDS* cDNA (see Methods). Northern-blot analysis with a probe derived from the 3′ end of the *PDS* cDNA revealed high expression of an approximately 5-kb mRNA in adult thyroid, markedly lower expression in adult and fetal kidney as well as fetal brain (detectable only after prolonged autoradiography) and no expression in various other tissues (Fig. 3).

Sequence analysis of the isolated *PDS* cDNA revealed a predicted ORF of 2,343 bp. The sequence around the putative ATG translation start site (AGGTCATGGCA) contains the critical –3 purine and +4 guanine residues of a Kozak consensus sequence[42,43]. An in-frame stop codon is present 72 bases upstream of this predicted start site. There is a large (about 2.3 kb) 3′ untranslated region, with a polyadenylation signal (AATAAA) starting at position 4877. The general structure and intron–exon boundaries of *PDS* were deduced by a combination of the direct comparison of the cDNA sequence with the available genomic sequence of BAC RG364P16 and, when genomic sequence was unavailable, by the sequencing of adjacent BACs using cDNA-specific primers (see Methods). The ORF of *PDS* is distributed across 20 exons (starting in exon 2); these range in size from 55 to 231 bp (excluding exon 21), based on the established intron–exon boundaries (Table 1). It is worth noting that the established sequence of the coding region within exons 9–21 perfectly matches the corresponding sequence within BAC RG364P16. The availability of the genomic sequence for *DRA* (within BACs RG364P16 and RG013F03) allowed *DRA* and *PDS* to be compared in greater detail. The genes encode the same number of exons (21; Fig. 2), with corresponding exons being remarkably similar in size. From the data available for BAC RG364P16 (containing the 3′ portion of *PDS*), however, it is evident that corresponding introns are often quite different in size.

Computational analysis of *PDS* and its encoded protein

PDS is predicted to encode a 780-amino-acid (86-kD) protein, which we have named pendrin. BLAST[36] and PSI-BLAST[44] searches of the public sequence databases, using human pendrin as the query sequence, reveal statistically significant homology to thirteen other proteins, with most known to function as sulphate transporters. These sequences are found across a large taxonomic span—including animals, plants (GenBank X96431, P53392, D89631) and yeast (GenBank P38359)—although the two closest relatives are the proteins encoded by the human *DRA*[38,39] and *DTD*[45] genes. The proteins within this family are highly hydrophobic; for example, 57% of the amino acids within pendrin are hydrophobic.

A multiple-sequence alignment of pendrin with its animal homologues (Fig. 4) reveals a large number of conserved positions, with absolute similarity being higher at the N-terminal end of the proteins. At positions 113–134 in the alignment, the pendrin homologues conform to a consensus pattern previously established for sulphate transporters (PROSITE PS01130). Although pendrin shows very high similarity to its homologues in this region, it does not conform to the 'sulphate transporter signature' as currently defined. Specifically, variance with the signature occurs at position 113 (V instead of P) and position 126 (T instead of any one of LIVMFY). There is, however, a proline at position 112, perhaps indicating a single-residue insertion within the signature region. The fact that pendrin is highly simi-

Table 1 • *PDS* coding region: intron–exon structure and PCR/sequencing primers

a Intron–exon organization of *PDS*

Exon/ Intron	Exon length (bp)	Starting position in cDNA	Acceptor splice site[a]	Donor splice site[a]	Intron length (bp)[b]
1	221	1		AGAGACAG/gtgagttcgccctgaa	ND
2	167	222	gtcctctggctcgcag/GTCATGGC	TGCTGCAG/gtagcggccgcgcggg	ND
3	140	389	tcttgctttttgacag/TTGTTCAA	GCTGCAAG/gtaagatgttggcaga	ND
4	111	529	ttgcatgtgctttcag/GGATGGCA	CTCAGTTG/gtaattataagtatat	ND
5	185	640	gacttttttttccccag/GACCTTTT	TTATACAG/gtaatgaacttacaag	596
6	165	825	cttttccttatcgtag/TTGATATT	TTATCTAT/gtaagtgttgcttctt	ND
7	153	990	tattttctttttatag/ACGCTGGT	TAATTGTG/gtaagtagaatatgta	100
8	83	1143	ttttgttttatttcag/ACGATAAT	CCAAGGGG/gtgagtgtggtgttcc	ND
9	148	1226	cttaaaaactcactag/GTTTTTGC	GGAACCAG/gtatgggtgccctttt	923
10	114	1374	ccagtctcttccttag/GAATTCAT	AGACACAG/gtaggaacaacagcct	4165
11	78	1488	ggtgtgtgtcttccag/GTTGCTGG	TGCAGAAG/gtataaccctgcttct	140
12	96	1566	ctgtctctcttggcag/TCGGTCTT	TTGATGCT/gtaagtcacctaccac	1216
13	107	1663	ctatttttttccctag/GTTATCTG	GTTCAGTT/gtgagtaacgtaaaac	2002
14	70	1769	atatttttttcttctag/TCCTTCTT	ACAAAAAC/gtaagtacctttgtga	1843
15	93	1839	tttcttttaatgccag/ATTGAAGA	AGTCCACA/gtaagtaatttttatcc	925
16	96	1932	acatttatttccaaag/GTTGGATT	CAACAAAG/gtgagatgacatcttt	630
17	231	2028	tgttttcttcgtttag/AATGGCAT	TGCGGGTG/gtaaggttctggtttt	2273
18	55	2259	ccttttttatttttag/ATTGTCAA	ACTTCAAG/gtaaatacatatatct	5668
19	146	2314	tgtctttcttttgaag/ATTATGTG	TAGAAACG/gtaaatattcaaccct	2339
20	84	2460	tgttctctttttcaag/ATCACTCT	AGGATGAG/gtatgatcatttttctt	2800
21	2356	2544	ccccttgcttccacag/GCTATGCG		NA

[a]The exonic and intronic sequences are indicated in upper-case and lower-case letters, respectively. [b]ND: not determined; NA: not applicable.

b Primers for PCR amplification and sequencing of *PDS* exons

Exon	Sense primer[a]	Antisense primer[a]	PCR product size (bp)
1	TTCCTCTTCTCCTCCCCATGTG	TGTTTCCGAGCAGCCTGTTC	586
2	ACCCGAAGGTTCTCAGGTGCC	CGAGACTGATGGAGCCACCCTC	593
3	ACAGTTCTTGGCAAAAGCATGG	GAAGGGTAAGCAACCATCTGTCAC	411
4	CCATTGTAAGTTGAGGACTTTCTG	CCAACCTAATAGAGGTATAATGCAC	270
5	CCTATGCAGACACATTGAACATTTG	TGAGCCTTAATAAGTGGGGTCTTG	442
6	AATTTTAGAGTGGTGGAGGAAGGG	ATTGTTTCTGGAATGAACAGTGACC	405
7	CATGGTTTTTCATGTGGGAAGATTC	ACACAAAATCCCAGTCCC	636
8	AATGGCAGTAGCAATTATCG	CAAATGGCTTGACGTTTATCTACACAC	
9	CAGATATAGCATTTGATGAGATGGGG	CCCCTTCTTTAGCTGACACC	1317
10	GGCATGGGAGTTTTCATTC	GCCATTCCTCGACTTGTTCTCTG	
11	GGGGAGACAGGGAAGTATGAAGTG	AAGGCTGTGTTATTCAGGG	1821
12	CCCTTCACTACTCTGCTACCAG	GCACAAGTGAGACTTAGGGG	
13	TTTTATAGGTAGTTATCACATGATGG	TGGTATAAGGAAGCTCAGAGTGTGTTG	
14	CAAAACACCAGAATGATGGGCTC	CGTAAAATGGAGCTGCTGAAACTTC	245
15	AGTTGAGTGCTGCTACCCAGCTC	GCCTATTCCTGATTGGACC	2195
16	TCTTTTTTGGCAGGATAGC	GCTCAGTTGTTCTTTGATACG	
17	CCAAGGAACAGTGTGTAGGTC	CCCATGTATTTGCCCTGTTGC	
18	GCTGGATGTTGCCATCTCTTGAG	CCACAGTCCCAGATAGGAGAAAGG	335
19	GTTGCAGTCAGCAATGATGCC	CTGATGAAAAAACTGAGGCTC	2828
20	CAGTTCACCTTTCAATGTGC	TGCTATGAAGCCATCTTTTCTGC	
21	GCAACAGTGAGTGAGATTCAGTCTCC	TCCAGGTCAGTCAAAAACTG	859
		GCCTGGTTCTGTAGCTTTTAGGTAAAC	

[a]Within each grouping of exons (e.g., 11–13), the underlined primers were used for PCR amplification of the genomic segment containing those exons, whereas all of the primers were then used for direct DNA sequencing of the coding regions contained within the amplified segment. The size of the resulting PCR product is indicated in the far-right column.

lar to homologues that contain the sulphate transporter signature, particularly within the region in question, argues for a redefinition of the signature and is the subject of further study.

As the pendrin homologues are all transmembrane proteins, prediction of putative transmembrane segments in pendrin was performed with PHDhtm, a neural network method with an expected per-residue accuracy of approximately 95% (ref. 46). A total of eleven transmembrane segments were predicted by this method (positions indicated in Fig. 4; see also Fig. 5). The N-terminal leader (before the first transmembrane segment) is positioned inside the membrane (Fig. 5) and does not contain any discernible signal sequence. The sequences of the homologues (Fig. 4) were also analysed with PHDhtm to assess the consistency of the predictions across the alignment. All but one of the sequences were predicted to have eleven transmembrane segments (in the case of the mouse sulphate transporter protein, the predic-

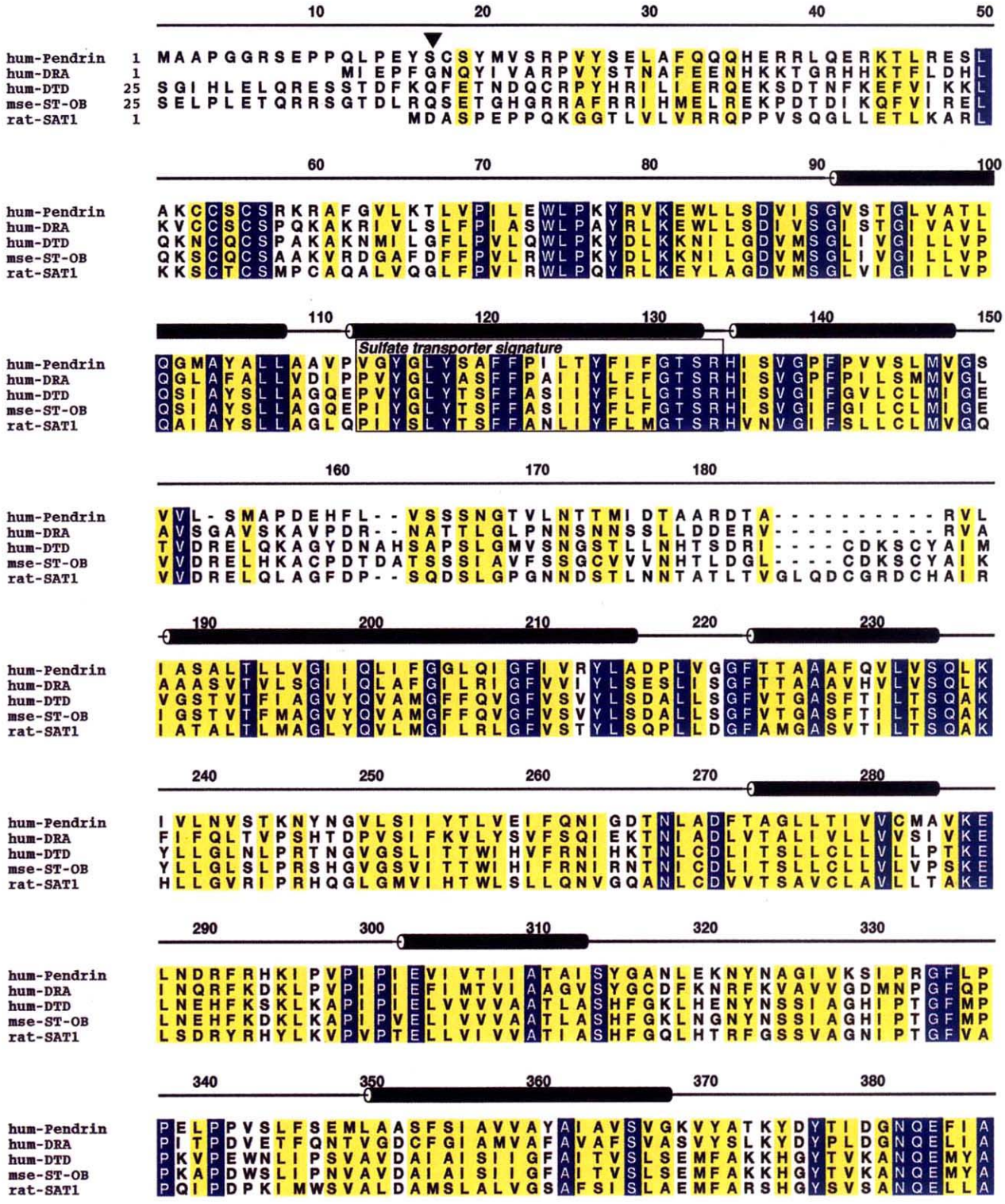

Fig. 4 Multiple-sequence alignment of selected proteins with significant sequence homology to human pendrin. The amino-acid sequence of human pendrin (hum-Pendrin) is aligned relative to the sequences of the closely related human DRA protein (hum-DRA), human diastrophic dysplasia protein (hum-DTD), mouse sulphate transporter protein (mse-ST-OB) and rat sulphate anion transporter 1 (rat-SAT1). The percent amino-acid sequence identity to human pendrin (as determined by pairwise sequence alignment against the human pendrin sequence) for these four proteins is 45%, 32%, 31% and 29%, respectively. Note that the percent identity of human pendrin to plant, yeast and fungal homologues ranged from 16% to 22% (data not shown). Positions exhibiting absolute identity among the five proteins are shown with a blue background; those with conservative substitutions are shown with a yellow background. The positions of the eleven transmembrane segments predicted for pendrin by PHDhtm[46] are shown as dark cylinders; the N-terminal leader is predicted to be inside the membrane. Potential serine O-glycosylation sites[72] are indicated by inverted arrowheads. The region within the proteins (positions 113–134 in the alignment) predicted to contain a 'sulphate-transporter signature' (PROSITE PS01130) is indicated by a box. ALSCRIPT[70] was used to format the alignment.

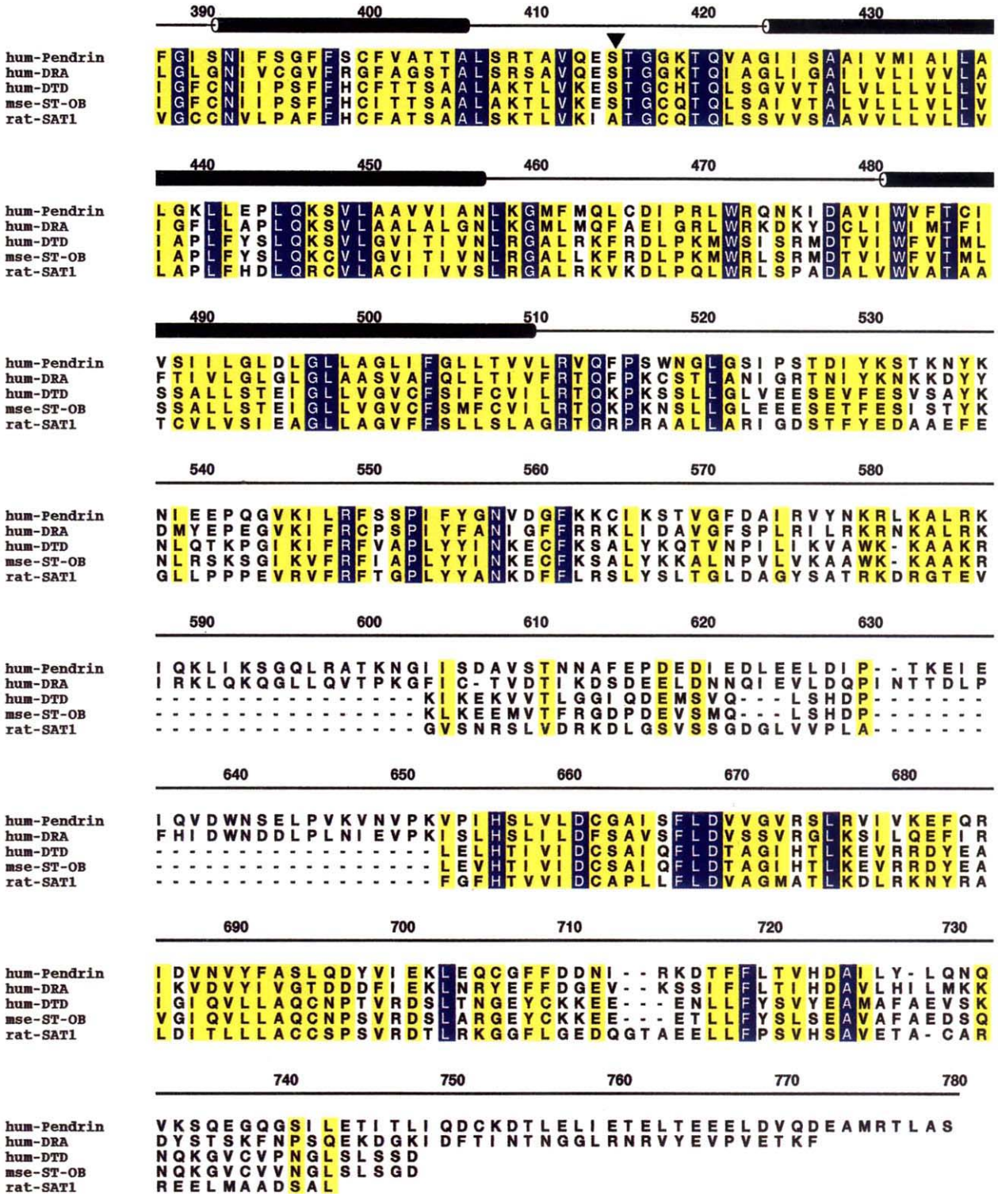

Fig. 4 (continued)

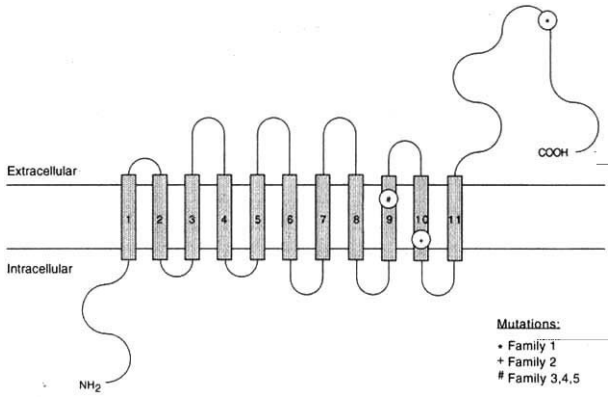

Fig. 5 Model of human pendrin. A schematic representation of the human pendrin protein is depicted based on its predicted amino-acid sequence (Fig. 4). Note the eleven putative transmembrane segments (with the residue numbers of each segment indicated in Fig. 4). The approximate positions of alterations caused by known Pendred-syndrome mutations (Table 2, Fig. 6) are indicated.

tion combined transmembrane segments 2 and 3 into a single one). With minor variation, the transmembrane segments in pendrin and its homologues occur at roughly the same positions. The only major variation encountered was for human *DRA*, whose transmembrane segment 10 is shifted three residues towards the N-terminus and is fourteen residues shorter than that of pendrin.

PDS mutations in Pendred-syndrome families

PCR assays specific for exon-containing fragments across the *PDS* gene were developed (Table 1) and used to amplify the DNA from representative Pendred-syndrome patients, obligate carriers and non-carriers (normal controls). The resulting products were subjected to direct DNA sequencing. Because of the consanguineous nature of the families, we expected to see homozygous mutations in affected individuals and heterozygosity in obligate carriers. However, we also expected several independent mutations, as distinct haplotypes were observed among the families studied (Fig. 1; ref. 11).

Three distinct *PDS* mutations were encountered among the Pendred-syndrome kindreds examined (Fig. 6, Table 2). The mutation in family 1 is a T→G substitution (missense mutation) that changes a predicted Phe to Cys at a highly conserved position (residue 667) near the carboxy terminus of the protein (Figs 4,5). This mutation was detected in both the genomic DNA and thyroid RNA sample available from individual IV-2 of this family[11]. The mutation in family 2 is a single base-pair (G) deletion at position 1565, which is predicted to result either in a frameshift and premature truncation of the protein (Fig. 5) or in aberrant splicing of the mRNA or both. The deleted G is either part of or immediately adjacent to a splice donor site; thus, the precise effect of the mutation cannot be predicted with certainty. The mutation in families 3, 4 and 5 is a single base-pair (T) deletion at position 1421, which is predicted to result in a frameshift and premature truncation of the protein (Fig. 5). In each of the families, affected individuals and obligate carriers are homozygous and heterozygous for the corresponding mutation, respectively, whereas non-carriers are homozygous for the normal sequence. On the basis of the inheritance pattern and the expected highly deleterious effects on pendrin, it seems reasonable to conclude that these mutations are responsible for Pendred syndrome in the corresponding families. In addition, none of these three mutations were detected in 38 random individuals of Arabic descent (76 chromosomes).

At the same time that *PDS* was isolated and implicated in Pendred syndrome, additional genetic mapping data for families 3–5 were generated (Fig. 1*e*). Analysis of the combined data set suggests the presence of regions of a shared haplotype among the affected individuals (markers *D7S3073* to *D7S2459* in families 3 and 4, markers *D7S2459* to *D7S525* in families 4 and 5). Note that all affected individuals in families 3–5 share the same disease-associated allele for *D7S2459*, which is an intragenic marker located about 900 bp from the common mutation in these families. Although these families are not known to be related, our findings suggest that a single founder mutation is responsible for the disease in families 3–5. It is of interest that these families live in northern Israel, whereas families 1 and 2, who have different mutations (Table 2), live in the Jerusalem area. Finally, the new, combined haplotype data (Fig. 1*e*) suggest an even more delimited critical region (between *D7S496* and *D7S3074*) than that defined by recombinants alone (Figs. 1*d*,2). As *PDS* resides between *D7S496* and *D7S3074*, these results provide additional evidence that *PDS* is the Pendred-syndrome gene.

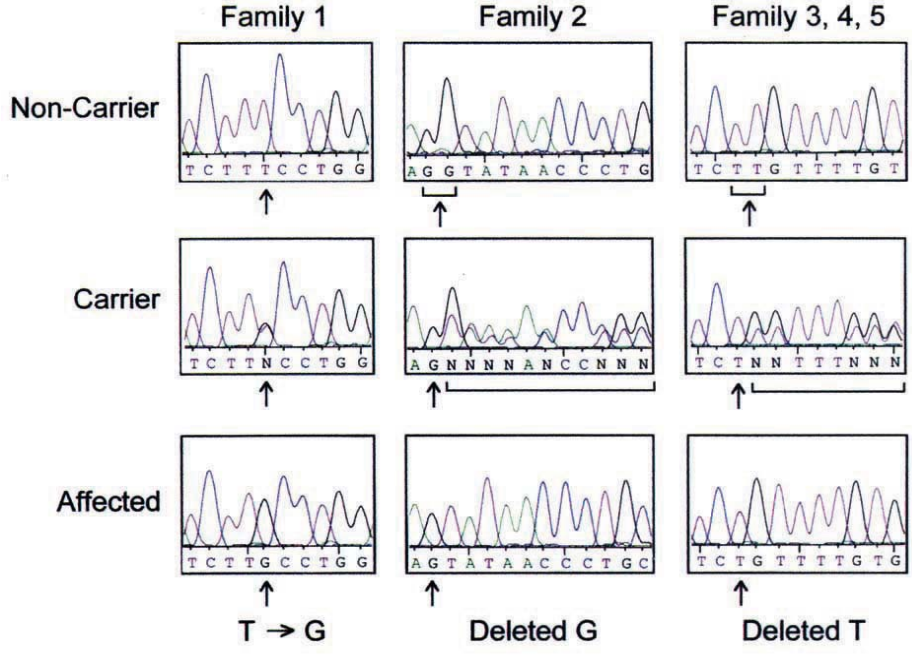

Fig. 6 *PDS* mutations in Pendred-syndrome families. DNA sequence data for the regions immediately surrounding the *PDS* mutations in the indicated Pendred-syndrome families are shown. In each case, representative data generated with the DNA from non-carriers, obligate carriers and affected individuals are depicted. Arrows indicate the positions of the mutations. Note that with families 2–5, the sequence data from the carriers becomes 'mixed' distal to the mutations (indicated by brackets)—a consequence of a single base-pair deletion on one of the two alleles.

Table 2 • *PDS* mutations in Pendred-syndrome families

Family no.[a]	Nucleotide change[b]	Predicted effect
1	2224T→G	Missense F667C
2	1565delG or 1565 + 1delG	Frameshift leading to an altered amino-acid sequence from codon 446, followed by stop at codon 453 Aberrant splicing of mRNA
3,4,5	1421delT	Frameshift leading to an altered amino-acid sequence from codon 400, followed by a stop at codon 430

[a]Families 1 and 2 were reported previously[11]. Families 3–5 are shown in Fig. 1. [b]Mutations are designated according to the suggested nomenclature of Beaudet and Tsui[74].

Discussion

Unravelling the genetic bases of hereditary hearing loss, with an incidence of more than 1 in 2,000 births, represents an important and challenging area of ongoing molecular genetics research. Although a number of genes have been implicated in congenital deafness[47–52], the list of loci responsible for syndromic and particularly non-syndromic deafness, each accounting for a very small percentage of the total, continues to grow[53,54] (see also http://hgins.uia.ac.be/dnalab/hhh). Pendred syndrome, first described just over 100 years ago[1], represents the most common form of syndromic deafness and may account for a significant fraction of all cases of hereditary hearing loss. The obscure constellation of symptoms—abnormal cochlear development, sensorineural hearing loss and diffuse thyroid enlargement (that is, goitre)—has intrigued investigators for years. In this paper, we describe the identification of a gene, *PDS*, that encodes a new putative sulphate transporter (pendrin) and provide evidence that mutations in *PDS* underlie Pendred syndrome.

Our efforts to identify the Pendred-syndrome gene were aided by several key factors. First, the availability of several inbred Pendred-syndrome kindreds greatly facilitated our genetic analyses, allowing homozygosity mapping to be used for delimiting the critical region. Second, our broader efforts at constructing a detailed YAC-based[29] and, more recently, higher-resolution BAC-based physical map of human chromosome 7 provided an invaluable framework for generating additional genetic markers, for mapping known genes and available ESTs, for organizing the evaluation of candidate genes within the critical region and for facilitating the sequencing of the interval. Third, the recent launching of systematic human genomic sequencing as part of the Human Genome Project[55,56] was particularly opportune, as it quickly produced valuable sequence data from the Pendred-syndrome critical region that directly facilitated the identification of *PDS*.

To our knowledge, this is among the first examples whereby the positional cloning of a human disease gene was facilitated by GRAIL analysis[34,35] of primary genomic sequence data. Specifically, the GRAIL-predicted exons focused our attention to the *PDS*-containing portion of the Pendred-syndrome critical region, quickly resulting in our detection of *DRA*-related sequences as well as the single thyroid *PDS*-specific EST in GenBank (EST99218). The presence of only two ESTs (EST99218 and H23053) from *PDS* in GenBank is surprising in light of its strikingly high expression in thyroid (Fig. 3); however, this may reflect the paucity of ESTs generated from thyroid: roughly 2,200 from the thyroid of any organism among the more than 1.2 million entries in dbEST. It is also worth pointing out that GRAIL correctly predicts the presence of seven of the thirteen *PDS* exons for which genomic sequence is thus far available.

From a 'genome evolution' perspective, it is interesting to note that the closely related genes *PDS* and *DRA* are arranged in opposite orientations to one another on chromosome 7, with their 3′ ends separated by only about 48 kb (Fig. 2). This arrangement is suggestive of a distant gene duplication event. Despite the sequence similarity and close physical proximity in the genome, the expression profiles of these two genes are markedly different. In adult tissues, *DRA* is expressed at high levels in intestine and prostate[41], whereas *PDS* was found to be expressed at high levels only in thyroid (Fig. 3). Insights about the regulatory elements controlling the expression of these genes may become available when additional genomic sequence from this region of chromosome 7 is generated.

PDS mutations were identified that precisely co-segregate with Pendred-syndrome alleles. The mutations encountered to date include a missense mutation leading to a non-conservative Phe→Cys substitution and two single base-pair deletions, one causing a translational frameshift and the other causing either a translational frameshift or an mRNA splicing defect (Table 2). It seems convincing that such homozygous mutations cause disease. Some investigators have speculated about the possible presence of two genes accounting for deafness and goitre, respectively, in Pendred syndrome. However, the identification of independent mutations in the same gene on different haplotypes (in conjunction with our failure to identify any other significant genetic changes in candidate genes mapping within the critical region) strongly supports the notion that Pendred syndrome is caused by mutations in a single gene (*PDS*).

Pendrin is a highly conserved protein that, on the basis of strong homologies with proteins of known function in evolutionarily diverse organisms, acts as a sulphate transporter. The two most closely related proteins, those encoded by *DRA* and *DTD*, have both been implicated in human diseases within the past few years (congenital chloride diarrhoea[41] and diastrophic dysplasia[45], respectively). It is thus clear that the transport of sulphate across cell membranes is critical for a variety of physiological processes (for example, modification of proteins and proteoglycans, maintenance of a proper osmotic environment) and that alterations in this function can lead to significant morbidity. We acknowledge the need for formal proof that pendrin is a sulphate transporter, which should be feasible with studies using either viable thyrocytes[11] or experimentally created heterologous systems, such as expressing human *PDS* in *Xenopus* oocytes.

Implicating loss-of-function mutations in *PDS* as the direct cause of thyroid disease in Pendred syndrome is relatively straightforward, as the gene is heavily expressed in thyroid, a tissue known to produce sulphated proteins. In particular, thyroglobulin, the major secretory product of thyroid follicular cells, is sulphated on both Tyr residues and Asn-linked oligosaccharides[57,58]. It is interesting to note that sulphation of thyroglobulin is ubiquitous in organisms ranging from humans to lower eukaryotes[58], suggesting that this form of post-translational modification serves an intrinsic role in thyroglobulin function. Some researchers even suggest that thyroglobulin sulphation can influence the subsequent iodination of the protein[58]. It thus seems reasonable to speculate that altered sulphation of thyroglobulin secondary to a defect in sulphate transport produces the thyroid disease seen in Pendred syndrome. We also recognize, however, that pendrin defects may alter thyroid function by other mechanisms (for instance, perturbation of ion transport or balance).

The role of pendrin in cochlear development is less obvious

and potentially more intriguing. One possibility is that the cochlear malformation seen in Pendred syndrome occurs (at least in part) secondary to thyroid dysfunction at a key developmental stage. A number of studies have implicated an important role for thyroid hormones in inner-ear development, based on both clinical data from deaf patients and animal models of hypothyroidism[13–17,53]. For example, the β-thyroid hormone receptor is expressed early in development within the portion of the embryonic inner ear that gives rise to the cochlea[13]; absence of this receptor in humans[18,19] or mice[20] causes sensorineural deafness. Admittedly, in the latter situations, the cochlear malformations are not the same as those seen in Pendred syndrome. Alternatively (and perhaps more likely), pendrin may play a more direct role in cochlear development and auditory function. In preliminary studies, we detected the presence of *PDS* sequences in a fetal cochlear cDNA library[59] by PCR analysis (data not shown). We are cautious in interpreting these results, as they provide little insight about the level of *PDS* expression and reflect only a particular (relatively late) developmental stage. Thus, the examination of *PDS* expression in the cochlea requires more careful studies. Use of a more tractable developmental system, such as the mouse, will clearly aid in establishing the role of pendrin in the developing cochlea as well as revealing how *PDS*-specific mutations lead to deafness.

With the identification of the Pendred-syndrome gene, the general role that *PDS* plays in congenital deafness can be more precisely established. Roughly half of childhood deafness is estimated to have a genetic origin; however, the locus is poorly defined for a large fraction of these cases. Some have inferred that Pendred syndrome accounts for upwards of 10% of inherited deafness[3–6]. Variable expressivity of the syndrome's phenotypic features, particularly the thyroid manifestations, has led some to propose that there is significant under-ascertainment of Pendred syndrome among deaf patients[3], with goitre often being missed or absent. We speculate that Pendred syndrome may be more common than previously thought among the deaf population. For example, another recessive locus for deafness (designated DFNB4) has been mapped to the same general region on chromosome 7q[60]; it is likely that these affected individuals actually have Pendred syndrome rather than mutations in another gene. It seems prudent to screen a broader deaf population for *PDS* mutations so as to establish more accurately the true prevalence of Pendred syndrome. With the findings reported here, the molecular tools for performing such studies are now in hand.

Methods

Family DNA samples. Blood samples and clinical data were obtained from members of five multiplex Arabic families thought to be unrelated to each other (Fig. 1). Clinical and pedigree data on families 1 and 2 were reported previously[11]. Family 5 is part of a Bedouin tribe known to include a large number of individuals with Pendred syndrome. For the purposes of this study, only a single nuclear family was included from this tribe. Serum thyroxine (T_4), tri-iodothyronine (T_3) and thyroid-stimulating hormone (TSH) levels were measured with an automated chemiluminescence system (Chiron Diagnostics). Thyroglobulin levels were determined with the DELFIA time-resolved fluoro-immunoassay kit (Wallac). DNA was extracted from peripheral lymphocytes with the Puregene kit (Gentra Systems).

With the exception of individual 15 in family 3, the diagnosis of Pendred syndrome was based on the presence of both sensorineural hearing loss and goitre. T_4 and T_3 levels were normal in all affected and unaffected individuals. TSH levels were normal in all unaffected individuals and were normal or slightly elevated in all Pendred-syndrome patients. Thyroglobulin levels were normal in all unaffected individuals but were grossly elevated (90–596 ng/ml; normal, <60 ng/ml) in all but three affected individuals. Individuals 4 in family 3, 8 in family 4 and 4 in family 5 had normal thyroglobulin levels, but all three had previously undergone thyroidectomy [for goitre or papillary thyroid carcinoma (individual 8 in family 4)]. No goitre was palpable in individual 15 in family 3; however, this patient is only nine years old and has elevated thyroglobulin levels (90 ng/ml). Because the goitre in Pendred syndrome is frequently detectable only after puberty, this patient is considered to be probably affected. Note that individual 9 in family 4 has been examined to confirm her unaffected status; she has normal hearing, a thyroid size within the normal range for the geographical region (where goitre is endemic) and normal serum T_4, T_3 and TSH. Thyroglobulin levels could not be assayed because of the presence of anti-thyroglobulin antibodies.

Genotyping. PCR amplification of short tandem repeat polymorphisms (STRPs) was performed with 20 ng of genomic DNA in 8.4-μl reactions containing 1.25 μl of PCR buffer [100 mM Tris-HCl (pH 8.8), 500 mM KCl, 15 mM $MgCl_2$, 0.01% gelatin (w/v)], 200 μM each of dATP, dCTP, dGTP and dTTP, 2.5 pmol of each primer and 0.25 unit of *Taq* polymerase. Samples were subjected to 35 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s. Amplification products were electrophoresed on 6% polyacrylamide gels containing 7.7 M urea at 60 W for approximately 2 h. Gels were silver stained according to the method of Bassem *et al.*[61]. Genotyping data was scored independently by two observers. Genetic markers *D7S3073* (sWSS4476) and *D7S3074* (sWSS4477) were developed from partially finished sequences of BACs RG363F19 and RG013F03, respectively (see Fig. 2), that were available on the Washington University Genome Sequencing Center Web site (http://genome.wustl.edu/gsc).

BAC isolation and development of BAC insert end-specific STSs. BACs were isolated from the Research Genetics and Genome Systems human BAC libraries by PCR-based screening. After isolation, individual BACs were tested for STSs (Fig. 2) to establish their STS content. As BAC contig assembly proceeded, clones at the ends of nascent contigs were selected for the generation of insert end-specific STSs. The choice of clones was also aided by the measurement of BAC insert sizes by pulsed-field gel electrophoresis. Specifically, 50 ng of purified BAC DNA was digested with *Not*I (New England Biolabs) at 37 °C for 1 h and then electrophoresed in a 1% agarose gel with a CHEF DRII apparatus (Bio-Rad; 120 V, 5–12 s ramped switching times for 14 h). The resulting size data often allowed accurate prediction about which clone(s) extended most into a particular gap. BAC DNA was routinely prepared with an Autogen 740 automated nucleic acid system (Integrated Separation Systems) and treated with additional RNAse (Ambion). When used as a sequencing template, the BAC DNA was then concentrated to 200 ng/μl with a Microcon-100 column (Amicon).

For deriving BAC insert end-specific sequences, fluorescent DNA sequencing reactions were performed with the –40M13F and –28M13R energy-transfer dye primers (Amersham) and 5 μg of purified BAC DNA by a Catalyst 800 LabStation [Perkin Elmer/Applied Biosystems; cycling conditions were 95 °C for 120 s, then 20 cycles of 95 °C for 2 s, 55 °C (for –40M13F) or 50 °C (for –28M1R) for 20 s and 72 °C for 60 s]. Products were analysed on an ABI 373A or 377 automated fluorescent sequencer (Perkin Elmer/Applied Biosystems). The resulting sequences were scanned for repetitive elements with RepeatMasker (A.F.A. Smit and P. Green, unpublished data; http://ftp.genome.washington.edu/cgi-bin/RepeatMasker), and primers suitable for PCR were designed with the MacVector program (Kodak). The relevant information about all of the STSs reported in this study has been submitted to GenBank and the Genome Database (GDB).

Gene identification. Additional gene sequences not previously localized on the YAC-based physical map of chromosome 7 (ref. 29) were detected as follows. EST clusters mapping within the broad region between markers *D7S658* and *D7S648* were identified on the 'human transcript map' Web site reported by Schuler *et al.*[33] (http://www.ncbi.nlm.nih.gov/SCIENCE96), and the corresponding sequences were downloaded. Since this Web site does not contain all of the sequences from each cluster, we also downloaded any related ESTs from the NCBI Unigene site[62] (http://www.ncbi.nlm.nih.gov/UniGene/index.html). All of these EST sequences were aligned to detect overlaps with the DNASTAR program Seqman II (Lasergene). Homology searching of ESTs was performed with the BLAST program[36,37] to identify those matching already mapped STSs. Corresponding PCR assays were then used to map the remaining ESTs on the YAC or BAC contigs. In addition to the mapping of ESTs, partially finished sequence data from relevant BAC clones (available on the

Washington University Genome Sequencing Center ftp server; see http://genome.wustl.edu/gsc) were routinely examined by two programs: PowerBLAST[63], which uses the BLAST search algorithm[36,37], and XGRAIL 1.3 (refs 34,35; http://compbio.ornl.gov). The latter analysis was typically reserved for contiguous sequence stretches of >40 kb in size. Putative genes identified by the analysis of ESTs or genomic sequence data served as a starting point for more complete cDNA isolation and mutation screening, typically using the approaches described below for *PDS*.

Isolation of *PDS*. Primers from each of two GRAIL-predicted, *DRA*-related exons from BAC RG364P16 were designed in opposite orientations and used in a PCR assay that included thyroid cDNA [prepared from thyroid polyA$^+$ RNA (Clontech) with the Advantage RT for PCR kit (Clontech)] and Advantage cDNA polymerase mix (Clontech). The resulting ~1-kb product was sequenced with the two initial PCR primers as well as two subsequent walking primers. 5′– and 3′–RACE[64] was performed with the Marathon cDNA amplification kit and human thyroid Marathon-ready cDNA (Clontech). The resulting ~1- and ~4-kb products, respectively, were sequenced by a primer-walking strategy. The generation of high-accuracy sequence across the *PDS* coding region was ensured by the alignment of redundant sequence reads from primers that were spaced roughly 200 bp apart in both directions and that provided sequence reads from both strands. The sequence of the final exon past the 3′ end of the coding region was verified only by single-pass sequence with primers spaced roughly 500 bp apart, as the complete sequence of this segment of the gene is available in BAC RG364P16. The sequences of the various PCR and sequencing primers mentioned above are available on request.

Elucidation of *PDS* intron–exon structure. Direct comparison of the available genomic sequence of BAC RG364P16 and the *PDS* cDNA sequence allowed the intron–exon boundaries to be determined for exons 8–20. For the remaining 5′ exons, cDNA-specific primers were designed [often with the intron–exon structure of *DRA* used as a guide (Fig. 2); details available on request] and used in dye-terminator sequencing reactions with purified DNA from the adjacent BAC GS113G16 (Fig. 2).

Detection of *PDS* mutations. Genomic DNA from Pendred-syndrome family members (affected individuals, obligate carriers and, when available, non-carriers) and normal controls was used as template for PCR amplification of *PDS* exon-containing DNA segments with the Advantage cDNA polymerase mix (Clontech) and the primers indicated in Table 1 (according to conditions recommended by the manufacturer). Each PCR product contained between 1 and 3 exons. The PCR primers along with a set of additional nested primers (Table 1) were then used to sequence the resulting products, obtaining at least one sequence read from each direction across the entire coding region. In the case of thyroid tissue (available only from individual IV-2 in family 1), RNA was purified with the RNA-plus kit (Epicentre), reverse-transcribed and subjected to PCR amplification with exon-specific primers. The resulting PCR products were then sequenced.

Sequencing of PCR products. After amplification, PCR products were purified on a Microcon-100 column (Amicon). Sequencing reactions were then performed with ~20 ng/kb of purified DNA, 6 pmol of primer and ABI PRISM Rhodamine, dRhodamine or BigDye terminator cycle sequencing kits (Perkin Elmer/Applied Biosystems). The cycle-sequencing conditions consisted of 93 °C for 3 min, followed by 25 cycles of 94 °C for 30 s, 55 °C for 5 s and 60 °C for 4 min. Fluorescent sequencing products were analysed as above.

Sequence analysis. The human pendrin sequence was searched against the GenPept[65], EMBL[66], PIR[67] and Swiss-Prot[68] databases with the BLASTP[36] and PSI-BLAST[44] algorithms. The BLAST search cutoffs used to identify the homologues of human pendrin were a Karlin/Altschul score for two aligned sequence segments of ≥90 and a probability of $\leq 10^{-3}$. Sequences showing only local homology, sequence fragments and putative or theoretical translations were removed from the data set. Both pairwise and multiple-sequence alignments of the resultant data set were performed with MultAlin[69]. The amino-acid alignment (Fig. 4) was performed using sequences from the following GenBank records: human DRA (P40879), human DTD (P50443), mouse sulphate transporter protein (D42049), and rat sulphate anion transporter (P45380); with ALSCRIPT[70] was used to format the final alignment. The human pendrin sequence was analysed for the presence of N-terminal signal sequences with SignalP[71]. Putative O-glycosylation sites were determined with NetOGlyc[72]. The prediction of helical transmembrane segments was performed with PHDhtm[46], while the orientation of the N-terminus of pendrin and its homologues with respect to the membrane was determined by PHDtopology[73].

Northern-blot analysis. A 317-bp segment from the 3′ untranslated region of *PDS* was PCR amplified (using primers 5′–TCGGGAGGTCTCTATGAGCAAG–3′ and 5′–TTGGGTCAGCACTCACTCTAACTG–3′) and purified on a 1.25% agarose gel. The DNA was recovered from the agarose with a Micropure-0.22 separator with gel nebulizer and a Microcon-30 spin column (Amicon), radiolabelled with [^{32}P]αdCTP with 'Ready to Go' DNA labelling beads (Pharmacia Biotech) and separated from unincorporated nucleotides with Sephadex G-50 NICK columns (Pharmacia Biotech). The *PDS*-specific probe was hybridized to three Clontech MTN northern blots (human I, endocrine and fetal II) in Express-Hyb solution (Clontech), followed by washing according to the manufacturer's instructions.

GenBank accession numbers. Human *PDS* cDNA, AF030880; BAC RG364P16 (containing the 3′ end of the *PDS* gene), AC002467.

Acknowledgements

We are grateful to the patients and their families for their participation in this study and the Washington University Genome Sequencing Center for their generation and open dissemination of human genomic sequence data. We also thank G. Bouffard and C. Jamison for informatics support; E. Young and N. Dietrich for technical assistance; C. Morton for providing the fetal cochlear cDNA library; R. Smith, G. Van Camp and C. Baldwin for helpful communications about Pendred syndrome; R. Gamborg for administrative assistance; G.A. Green for editorial assistance; and R. Nussbaum, F. Collins and M. Boguski for general advice and critical review of the manuscript. This work was supported in part by NIH grant R01-HG00457 (to V.C.S.).

Received 7 October; accepted 27 October 1997.

1. Pendred, V. Deaf-mutism and goitre. *Lancet* **ii**, 532 (1896).
2. Reardon, W. & Trembath, R.C. Pendred syndrome. *J. Med. Genet.* **33**, 1037–1040 (1996).
3. Reardon, W. *et al.* Pendred syndrome: 100 years of underascertainment? *Q. J. Med.* **90**, 443–447 (1997).
4. Fraser, G.R. Association of congenital deafness with goitre (Pendred's syndrome): a study of 207 families. *Ann. Hum. Genet.* **28**, 201–249 (1965).
5. Batsakis, J.G. & Nishiyama, R.H. Deafness with sporadic goiter. *Arch. Otolaryngol.* **76**, 401–406 (1962).
6. Illum, P., Kiaer, H.W., Hvidberg-Hansen, J. & Sondergaard, G. Fifteen cases of Pendred's syndrome: congenital deafness and sporadic goiter. *Arch. Otolaryngol.* **96**, 297–304 (1972).
7. Johnsen, T., Sorensen, M.S., Feldt-Rasmussen, U. & Friis, J. The variable intrafamiliar expressivity in Pendred's syndrome. *Clin. Otolaryngol.* **14**, 395–399 (1989).
8. Mondini, C. Anatomia surdi nati sectio: De Boroniensi scientarium et artium instituto atque acedemia commnetarii. *Bononiae* **7**, 419–431 (1791).
9. Johnsen, T., Jorgensen, M.B. & Johnsen, S. Mondini cochlea in Pendred's syndrome; a histological study. *Acta Oto-Laryngol.* **102**, 239–247 (1986).
10. Johnsen, T., Larsen, C., Friis, J. & Hougaard-Jensen, F. Pendred's syndrome: acoustic, vestibular and radiological findings in 17 unrelated patients. *J. Laryngol. Otol.* **101**, 1187–1192 (1987).
11. Sheffield, V.C. *et al.* Pendred syndrome maps to chromosome 7q21–34 and is caused by an intrinsic defect in thyroid iodine organification. *Nature Genet.* **12**, 424–426 (1996).
12. Trotter, W.R. The association of deafness with thyroid dysfunction. *Br. Med. Bull.* **16**, 92–98 (1960).
13. Bradley, D.J., Towle, H.C. & Young, W.S., 3rd. Alpha and beta thyroid hormone receptor (TR) gene expression during auditory neurogenesis: evidence for TR isoform-specific transcriptional regulation in vivo. *Proc. Natl. Acad. Sci. USA* **91**, 439–443 (1994).
14. Van Middlesworth, L. & Norris, C.H. Audiogenic seizures and cochlear damage in rats after perinatal antithyroid treatment. *Endocrinology* **106**, 1686–1690 (1980).

15. O'Malley, B.W., Jr., Li, D. & Turner, D.S. Hearing loss and cochlear abnormalities in the congenital hypothyroid (hyt/hyt) mouse. *Hear. Res.* **88**, 181–189 (1995).
16. Deol, M.S. An experimental approach to the understanding and treatment of hereditary syndromes with congenital deafness and hypothyroidism. *J. Med. Genet.* **10**, 235–242 (1973).
17. Uziel, A., Gabrion, J., Ohresser, M. & Legrand, C. Effects of hypothyroidism on the structural development of the organ of Corti in the rat. *Acta Oto-Laryngol.* **92**, 469–480 (1981).
18. Takeda, K., Sakurai, A., DeGroot, L.J. & Refetoff, S. Recessive inheritance of thyroid hormone resistance caused by complete deletion of the protein-coding region of the thyroid hormone receptor- beta gene. *J. Clin. Endocrinol. Metab.* **74**, 49–55 (1992).
19. Takeda, K., Balzano, S., Sakurai, A., DeGroot, L.J. & Refetoff, S. Screening of nineteen unrelated families with generalized resistance to thyroid hormone for known point mutations in the thyroid hormone receptor beta gene and the detection of a new mutation. *J. Clin. Invest.* **87**, 496–502 (1991).
20. Forrest, D., Erway, L.C., Ng, L., Altschuler, R. & Curran, T. Thyroid hormone receptor is essential for development of auditory function. *Nature Genet.* **13**, 354–357 (1996).
21. Coyle, B. *et al.* Pendred syndrome (goitre and sensorineural hearing loss) maps to chromosome 7 in the region containing the nonsyndromic deafness gene *DFNB4*. *Nature Genet.* **12**, 421–423 (1996).
22. Murray, J.C. *et al.* A comprehensive human linkage map with centimorgan density. Cooperative Human Linkage Center (CHLC). *Science* **265**, 2049–2054 (1994).
23. Sheffield, V.C. *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**, 1837–1844 (1995).
24. Coucke, P. *et al.* The gene for Pendred syndrome is located between D7S501 and D7S692 in a 1.7-cM region on chromosome 7q. *Genomics* **40**, 48–54 (1997).
25. Gausden, E. *et al.* Pendred syndrome: evidence for genetic homogeneity and further refinement of linkage. *J. Med. Genet.* **34**, 126–129 (1997).
26. Green, E.D. *et al.* Systematic generation of sequence-tagged sites for physical mapping of human chromosomes: application to the mapping of human chromosome 7 using yeast artificial chromosomes. *Genomics* **11**, 548–564 (1991).
27. Green, E.D. *et al.* Integration of physical, genetic and cytogenetic maps of human chromosome 7: isolation and analysis of yeast artificial chromosome clones for 117 mapped genetic markers. *Hum. Mol. Genet.* **3**, 489–501 (1994).
28. Green, E.D. *et al.* A human chromosome 7 yeast artificial chromosome (YAC) resource: construction, characterization, and screening. *Genomics* **25**, 170–183 (1995).
29. Bouffard, G.G. *et al.* A physical map of human chromosome 7: an integrated YAC contig map with average STS spacing of 79 kb. *Genome Res.* **7**, 673–692 (1997).
30. Bouffard, G.G. *et al.* A collection of 1814 human chromosome 7-specific STSs. *Genome Res.* **7**, 59–64 (1997).
31. Marra, M.A. *et al.* High throughput fingerprint analysis of large-insert clones. *Genome Res.* (in the press).
32. Hoglund, P. *et al.* Positional candidate genes for congenital chloride diarrhea suggested by high-resolution physical mapping in chromosome region 7q31. *Genome Res.* **6**, 202–210 (1996).
33. Schuler, G.D. *et al.* A gene map of the human genome. *Science* **274**, 540–546 (1996).
34. Uberbacher, E.C. & Mural, R.J. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88**, 11261–11265 (1991).
35. Xu, Y., Mural, R., Shah, M. & Uberbacher, E. Recognizing exons in genomic sequence using GRAIL II. *Genet. Eng.* **16**, 241–253 (1994).
36. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
37. Gish, W. & States, D.J. Identification of protein coding regions by database similarity search. *Nature Genet.* **3**, 266–272 (1993).
38. Schweinfest, C.W., Henderson, K.W., Suster, S., Kondoh, N. & Papas, T.S. Identification of a colon mucosa gene that is down-regulated in colon adenomas and adenocarcinomas. *Proc. Natl. Acad. Sci. USA* **90**, 4166–4170 (1993).
39. Byeon, M.K. *et al.* The down-regulated in adenoma (DRA) gene encodes an intestine-specific membrane glycoprotein. *Oncogene* **12**, 387–396 (1996).
40. Silberg, D.G., Wang, W., Moseley, R.H. & Traber, P.G. The down regulated in adenoma (dra) gene encodes an intestine-specific membrane sulfate transport protein. *J. Biol. Chem.* **270**, 11897–11902 (1995).
41. Hoglund, P. *et al.* Mutations of the down-regulated in adenoma (DRA) gene cause congenital chloride diarrhoea. *Nature Genet.* **14**, 316–319 (1996).
42. Kozak, M. Interpreting cDNA sequences: some insights from studies on translation. *Mamm. Genome* **7**, 563–574 (1996).
43. Kozak, M. Regulation of translation in eukaryotic systems. *Annu. Rev. Cell Biol.* **8**, 197–225 (1992).
44. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
45. Hastbacka, J. *et al.* The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* **78**, 1073–1087 (1994).
46. Rost, B., Casadio, R., Fariselli, P. & Sander, C. Transmembrane helices predicted at 95% accuracy. *Protein Sci.* **4**, 521–533 (1995).
47. Weil, D. *et al.* Defective myosin VIIA gene responsible for Usher syndrome type 1B. *Nature* **374**, 60–61 (1995).
48. Baldwin, C.T., Hoth, C.F., Amos, J.A., da-Silva, E.O. & Milunsky, A. An exonic mutation in the *HuP2* paired domain gene causes Waardenburg's syndrome. *Nature* **355**, 637–638 (1992).
49. Tassabehji, M. *et al.* Waardenburg's syndrome patients have mutations in the human homologue of the *Pax-3* paired box gene. *Nature* **355**, 635–636 (1992).
50. Barker, D.F. *et al.* Identification of mutations in the COL4A5 collagen gene in Alport syndrome. *Science* **248**, 1224–1227 (1990).
51. Liu, X.-Z. *et al.* Mutations in the myosin VIIA gene cause non-syndromic recessive deafness. *Nature Genet.* **16**, 188–190 (1997).
52. Kelsell, D.P. *et al.* Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. *Nature* **387**, 80–83 (1997).
53. Petit, C. Genes responsible for human hereditary deafness: symphony of a thousand. *Nature Genet.* **14**, 385–391 (1996).
54. Van Camp, G., Willems, P.J. & Smith, R.J.H. Nonsyndromic hearing impairment: unparalleled heterogeneity. *Am. J. Hum. Genet.* **60**, 758–764 (1997).
55. Olson, M.V. A time to sequence. *Science* **270**, 394–396 (1995).
56. Boguski, M., Chakravarti, A., Gibbs, R., Green, E. & Myers, R.M. The end of the beginning: the race to begin human genome sequencing. *Genome Res.* **6**, 771–772 (1996).
57. Herzog, V. Secretion of sulfated thyroglobulin. *Eur. J. Cell Biol.* **39**, 399–409 (1986).
58. Baumeister, F.A.M. & Herzog, V. Sulfation of thyroglobulin: a ubiquitous modification in vertebrates. *Cell Tissue Res.* **252**, 349–358 (1988).
59. Robertson, N.G., Khetarpal, U., Gutierrez-Espeleta, G.A., Bieber, F.R. & Morton, C.C. Isolation of novel and known genes from a human fetal cochlear cDNA library using subtractive hybridization and differential screening. *Genomics* **23**, 42–50 (1994)
60. Baldwin, C.T. *et al.* Linkage of congenital, recessive deafness (DFNB4) to chromosome 7q31 and evidence for genetic heterogeneity in the Middle Eastern Druze population. *Hum. Mol. Genet.* **4**, 1637–1642 (1995).
61. Bassam, B.J., Caetano-Anolles, G. & Gresshoff, P.M. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* **196**, 80–83 (1991).
62. Boguski, M.S. & Schuler, G.D. ESTablishing a human transcript map. *Nature Genet.* **10**, 369–371 (1995).
63. Zhang, J. & Madden, T.L. PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Res.* **7**, 649–656 (1997).
64. 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. USA* **85**, 8998–9002 (1988).
65. Benson, D.A., Boguski, M.S., Lipman, D.J. & Ostell, J. GenBank. *Nucleic Acids Res.* **25**, 1–6 (1997).
66. Stoesser, G., Sterk, P., Tuli, M.A., Stoehr, P.J. & Cameron, G.N. The EMBL nucleotide sequence database. *Nucleic Acids Res.* **25**, 7–14 (1997).
67. George, D.G. *et al.* The protein information resource (PIR) and the PIR-international protein sequence database. *Nucleic Acids Res.* **25**, 24–28 (1997).
68. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucleic Acids Res.* **25**, 31–36 (1997).
69. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890 (1988).
70. Barton, G.J. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37–40 (1993).
71. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6 (1997).
72. Hansen, J.E. *et al.* Prediction of O-glycosylation of mammalian proteins: specificity patterns of UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase. *Biochem. J.* **308**, 801–813 (1995).
73. Rost, B., Fariselli, P. & Casadio, R. Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci.* **5**, 1704–1718 (1996).
74. Beaudet, A.L. & Tsui, L.-C. A suggested nomenclature for designating mutations. *Hum. Mutat.* **2**, 245–248 (1993).