JARO
Journal of the Association for Research in Otolaryngology

Gene Discovery in the Auditory System: Characterization of Additional Cochlear-Expressed Sequences

Barbara L. Resendes,[1,3] Nahid G. Robertson,[1] Joseph D. Szustakowski,[4] Robert J. Resendes,[5] Zhiping Weng,[4] and Cynthia C. Morton [1,3]

[1]Departments of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, MA 02115, USA
[2]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA
[3]Harvard Medical School, Boston, MA 02115, USA
[4]Department of Biomedical Engineering, Boston University, Boston, MA 02115, USA
[5]Sun Microsystems, Inc., Burlington, MA 01803, USA

ABSTRACT

To identify genes involved in hearing, 8494 expressed sequence tags (ESTs) were generated from a human fetal cochlear cDNA library in two distinct sequencing projects. Analysis of the first set of 4304 ESTs revealed clones representing 517 known human genes, 41 mammalian genes not previously detected in human tissues, 487 ESTs from other human tissues, and 541 cochlear-specific ESTs (http://hearing.bwh.harvard.edu). We now report results of a DNA sequence similarity (BLAST) analysis of an additional 4190 cochlear ESTs and a comparison to the first set. Among the 4190 new cochlear ESTs, 959 known human genes were identified; 594 were found only among the new ESTs and 365 were found among ESTs from both sequencing projects. *COL1A2* was the most abundant transcript among both sets of ESTs, followed in order by *COL3A1, SPARC, EEF1A1,* and *TPTI.* An additional 22 human homologs of known nonhuman mammalian genes and 1595 clusters of ESTs, of which 333 are cochlear-specific, were identified among the new cochlear ESTs. Map positions were determined for 373 of the new cochlear ESTs and revealed 318 additional loci. Forty-nine of the mapped ESTs are located within the genetic interval of 23 deafness loci. Reanalysis of unassigned ESTs from the prior study revealed 338 additional known human genes. The total number of known human genes identified from 8494 cochlear ESTs is 1449 and is represented by 4040 ESTs. Among the known human genes are 14 deafness-associated genes, including *GJB2* (connexin 26) and *KVLQT1.* The total number of nonhuman mammalian genes identified is 43 and is represented by 58 ESTs. The total number of ESTs without sequence similarity to known genes is 4055. Of these, 778 also do not have sequence similarity to any other ESTs, are categorized into 700 clusters, and may represent genes uniquely or preferentially expressed in the cochlea. Identification of additional known genes, ESTs, and cochlear-specific ESTs provides new candidate genes for both syndromic and nonsyndromic deafness disorders.

Keywords: ESTs, genes, cochlea, cochlear-expressed genes

INTRODUCTION

Hearing loss is the most frequent sensory defect in humans. The prevalence of severe to profound bilateral congenital hearing loss is estimated at 1 in 1000 births (Gorlin et al. 1995). About 50% of congenital deafness is thought to be due to environmental factors, such as acoustic trauma, ototoxicity (e.g., aminoglycoside antibiotics), viral infections (e.g., rubella), or bacterial infections (e.g., bacterial meningitis). The

remaining 50% is attributed to genetic causes and is categorized as syndromic or nonsyndromic hearing loss. Approximately 77% of hereditary deafness is estimated to be inherited in an autosomal recessive mode, 22% is autosomal dominant, 1% is X-linked, and less than 1% segregates through the maternal lineage via the mitochondria (Morton 1991). Over 400 syndromes are recognized in which hearing loss is among the clinical findings and over 60 loci have been mapped for nonsyndromic hearing loss (Gorlin et al. 1995; Van Camp and Smith 2001). At least 46 genes known to cause either syndromic or nonsyndromic deafness have been identified to date (Steel and Kros 2001; Van Camp and Smith 2001).

Traditional methods for identification of genes involved in disease, such as genetic linkage analysis, have a less than optimal use in gene discovery efforts for hearing disorders, mainly because of the complex genetic nature of deafness. Although more than 60 disease loci for nonsyndromic deafness are known, the number of disease loci is expected to be considerably larger due to the unparalleled degree of genetic heterogeneity that characterizes hereditary deafness (Van Camp et al. 1997). A complementary method to genetic linkage analysis for gene identification is one that utilizes tissue-specific cDNA libraries (Hedrick et al. 1984; Jones and Reed 1989; Gurish et al. 1992). To this end, we constructed a human fetal cochlear cDNA library and have generated over 8494 ESTs in two sequencing projects from the cochlear cDNA clones (Robertson et al. 1994; Skvorak et al. 1999). Several auditory genes, namely *ATQ1, COCH,* and *OTOR*, of which the latter two are novel cochlear genes, have been identified using our human fetal cochlear cDNA library (Robertson et al. 1997, 2000; Skvorak et al. 1997). *COCH* was further shown to be responsible for a sensorineural deafness and vestibular disorder, DFNA9 (Robertson et al. 1998).

This article describes the production and analysis of the second set of 4190 cochlear EST sequences and compares these results with that of the first set of 4304 cochlear ESTs.

METHODS

Generation of EST sequences from the human fetal cochlear cDNA library

A human fetal (16–22 weeks developmental age) cochlear cDNA library was constructed by Robertson et al. (1994) in accordance with guidelines established by the Human Research Committee at the Brigham and Women's Hospital. Briefly, cDNAs were reverse transcribed from oligo(dT)-primed poly(A)+ RNAs and were directionally cloned into the Uni-ZAP XR vector (Stratagene, La Jolla, CA). The first set of 4304 ESTs were analyzed previously (Skvorak et al. 1999) and have GenBank accession numbers starting with either of the letters H or N. For generation of the second set of 4190 ESTs, a second aliquot of the human fetal cochlear cDNA library was subjected to mass *in vivo* excision according to the manufacturer's protocol to remove phage sequences. Gridding of the clones was performed at Lawrence Livermore National Laboratory. Sequencing of the 5′ ends of the inserts using a T7 primer (5′ TAATACGACTCACTATAGGG 3′) and generation of the ESTs, as described previously (Skvorak et al. 1999), was performed at the NIH Intramural Sequencing Center. GenBank accession numbers for the second cochlear ESTs start with the letters AW.

EST nucleotide sequence analysis

To obtain the data described here, the nucleotide sequence of each cochlear EST was used to conduct sequence similarity (BLAST) searches against various databases (Altschul et al. 1997). EST sequences were compared with the nucleotide sequences in GenBank primate database (release 117) and with sequences in the EST, STS, and nonhuman mammalian databases (release 116) to determine if the ESTs shared nucleotide similarity with known sequences, as previously described (Skvorak et al. 1999). Briefly, a cochlear EST sequence was considered to be a significant match to a nucleotide sequence when $E < 1E - 30$, the aligned region was >50 nucleotides, and the nucleotide identify was >85% over the aligned region, or $E < 1E - 20$, the aligned region was >50 nucleotides, and the nucleotide identity was >90% over the aligned region. The E value is the Expect value; it describes the number of hits one can expect by chance alone when searching a database of a particular size (Altschul et al. 1997). The majority of cochlear EST sequences that identified human genes were >95% identical to the nucleotide sequence. If a BLAST search revealed a significant match to more than one hit, only the highest scoring hit was included in the data set. Any cochlear EST sequences that did not have sequence similarity to any known human or nonhuman genes or EST sequences from other cDNA libraries were considered cochlear-specific, with the caveat that this categorization is dependent on the diversity and completeness of transcripts deposited in the public databases at the time of analysis.

Comparison of ESTs from both sequencing projects

Redundancy among cochlear ESTs from the two sequencing projects was determined by comparing the

nucleotide sequence of each new cochlear EST with the nucleotide sequence of the cochlear ESTs from project 1 and identifying those ESTs with nucleotide sequence similarity, as determined by the same criteria as above for BLAST analysis. Of note, ESTs from the second sequencing project were generated from the 5′ ends of the cDNA clones, in contrast to the first sequencing effort, in which one-third were generated from the 5′ ends of the cDNA inserts and two-thirds from the 3′ ends.

Assignment of ESTs to chromosomal map positions

A cochlear EST was considered mapped when it had nucleotide sequence similarity to an STS according to the criteria described above (E values, identity) and the STS had a chromosomal assignment. In the majority of cases, the EST sequence matched the STS sequence with <95% identity. To determine the chromosomal assignment, marker interval, and genetic interval for a particular STS, the following websites were used: Whitehead Institute for Biomedical Research/MIT Center for Genome Research (http://www-genome.wi.mit.edu/), the Stanford Human Genome Center (http://www-shgc.stanford.edu/), NCBI's Query STS Sequence Database (http://www2.ncbi.nlm.nih.gov/dbST/dbsts_query.html), and NCBI's UniGene database (http://www.ncbi.nlm.nih.gov/UniGene/Hs.Home.html). Some genetic intervals may be unknown if fine mapping information was unavailable for a particular genetic marker. Marker intervals for deafness loci were obtained from the Hereditary Hearing Loss home page (Van Camp and Smith 2001) and their genetic intervals were determined by searching the STS database at the Whitehead Institute/MIT's website as listed above.

RESULTS

Production of a second set of human cochlear EST sequences

As a result of the second round of sequencing of clones from the Morton human fetal cochlear cDNA library (see Methods), an additional 4190 human cochlear ESTs were deposited in GenBank. Only 186 (4.5%) of these cochlear EST sequences were found not to be useful for further analysis, either because of insufficient sequence length (i.e., less than 100 bases, $n = 21$), repetitive sequence content (i.e., Alu or L1 elements, $n = 97$), or homology to yeast genes (>95%

similarly over entire EST sequence, $n = 68$) (Table 1). Of note, the 68 cochlear ESTs with >95% similarity to yeast genes have been deleted from GenBank. In sum, 186 ESTs were omitted from the present analysis; 4004 (95.5%) new cochlear ESTs were included in this analysis and represent the denominator for the statistical analyses.

Identification of known genes in the cochlear library

Of the 4004 new cochlear ESTs, 2050 (51%) were assigned to 959 known human genes (Table 2). 365 (38%) of these genes, represented by 1279 (32%) of the new cochlear ESTs, had been identified among the ESTs from project 1 and thus represent genes identified by both sequencing efforts (Fig. 1). Examples of genes found in both analyses are collagen type I alpha 2 (COL1A2), collagen type III alpha 1 (COL3A1), elongation factor 1 alpha (EEF1A1), and SPARC (Table 3). These four genes were also some of the most frequently detected genes among the ESTs, reflecting their abundant expression in the human fetal cochlear library (Table 3). Genes detected less frequently and found in common among the two analyses included the following genes found once among ESTs from both projects: calpactin 1 light chain, ribosomal protein L31, and dynein light chain 1. Of the 959 known human genes identified in the second project, 594 (62%) genes, represented by 771 (19%) ESTs, were identified specifically among the new cochlear ESTs and were not found among the previously analyzed cochlear ESTs (Fig. 1). 476 of these newly detected genes were found only once and 118 were found more than once; the most abundant newly detected genes were initiation factor 4B ($n = 10$), ribosomal protein S24 ($n = 6$), MHC protein homologous to chicken B complex protein ($n = 6$), and cyclin I ($n = 5$). A complete list of the known human genes found among the new cochlear ESTs can be accessed via our website (http://hearing.bwh.harvard.edu).

Among the 1954 cochlear ESTs not representing a known human gene, 28 ESTs had homology to 22 nonhuman mammalian genes, with nucleotide sequence identity ranging from 83% to 96% (Table 4). Among these ESTs are genes encoding membrane proteins, extracellular matrix proteins, and trafficking proteins. These cochlear ESTs had homology to genes in species such as cow, mouse, and rat and may represent the human homologs. None of these 22 nonhuman mammalian genes were identified during project 1, although 41 nonhuman mammalian genes were identified in that effort (20 for which the human homolog has since been identified).

TABLE 1

Summary of cochlear ESTs from projects 1 and 2

	Project 1	Project 2	Total
Number of ESTs produced	4304	4190	8494
5′ reads	1207	4190	5397 (63.5%)
3′ reads	3097	0	3097 (36.5%)
ESTs eliminated from analysis	155 (3.6%)	186 (4.5%)	341 (4.0%)
ESTs used for analysis	4149 (96.4%)	4004 (95.5%)	8153 (96.0%)

TABLE 2

Categorization of cochlear ESTs

	Project 1	Project 2	Total
Human genes identified (number of ESTs)	855 (1990)	959 (2050)	1449[a] (4040) (49.6% of ESTs)
Human homologs of nonhuman genes identified (number of ESTs)	21 (30)	22 (28)	43 (58) (0.7% of ESTs)
ESTs with sequence similarity to no human or nonhuman genes	2129	1926	4055 ESTs (49.7% of ESTs)
Total ESTs	4149	4004	8153

[a]Value accounts for genes detected in both projects.
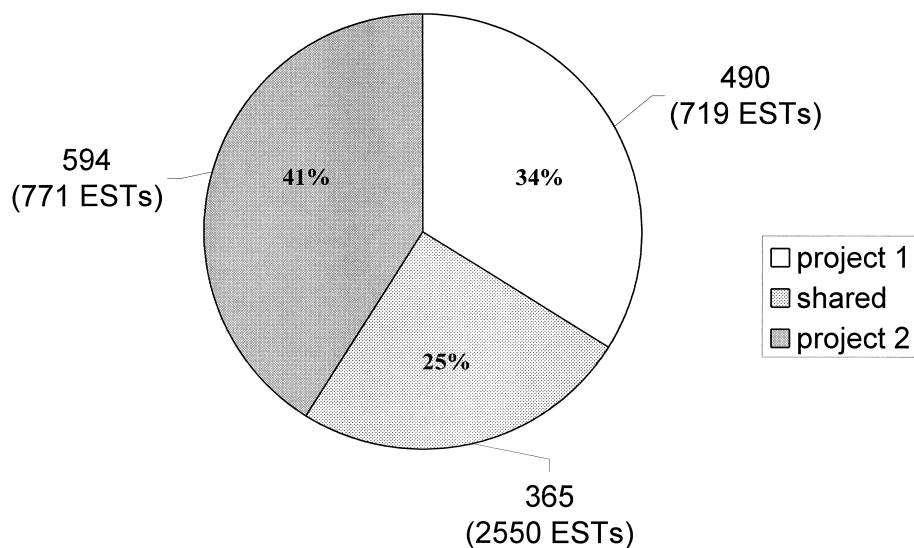
FIG. 1. Distribution of 1449 human genes identified among 8494 cochlear ESTs. Of the 1449 human genes identified by cochlear ESTs, 490 of these genes were identified during project 1, 594 were identified during project 2, and 365 of the genes were identified by both projects.

BLAST analysis of cochlear ESTs against the EST database

1926 (48%) of the project 2 cochlear ESTs did not have significant sequence similarity to any known gene (Table 5). Of these, 1568 ESTs represent 1262 clusters or genes that have sequence similarity to ESTs from other tissue-derived libraries. The remaining 358 ESTs, which are categorized into 333 clusters, did not match any other ESTs in GenBank and may be unique to the cochlear library, suggesting they may represent genes specifically or preferentially expressed in the cochlea. Of the 333 cochlear-specific clusters, 327 are new and 6 had already been identified by project 1. Combining results from projects 1 (updated) and 2, a total of 4055

cochlear ESTs, representing 2966 clusters, do not have nucleotide sequence similarity to any known gene in GenBank. Of these, 778 ESTs are unique ESTs in Gen-Bank, are grouped by nucleotide sequence similarity into 700 clusters, and may be considered cochlear-specific (Table 5).

Chromosomal map position of cochlear ESTs for positional candidate deafness genes

Nucleotide sequence similarity (BLAST) analysis was performed against the STS database to determine the chromosomal map positions of the 1926 project 2 cochlear ESTs that were not assigned as either a human

TABLE 3

Twenty most abundant genes found in the human fetal cochlear cDNA library

			Frequency		
No.	GenBank ID	Gene description	Project 1	Project 2	Total
1	K01078	Collagen type I alpha 2 (COL1A2)	86	99	185
2	X14420	Collagen type III alpha 1 (COL3A1)	38	41	79
3	X16869	Elongation factor 1 alpha 1 (EEF1A1)	30	37	67
4	J03040	SPARC	16	47	63
5	M32798	Collagen type I alpha 1 (COL1A1)	10	26	36
6	M25246	Vimentin	11	21	32
7	L13806	Translationally controlled tumor protein (TCTP)	13	16	29
8	U30521	P311 HUM (3.1)	9	12	21
9	J03801	Lysozyme	6	14	20
10	M58458	Ribosomal protein S4, X-linked	6	13	19
11	J04765	Osteopontin	10	8	18
12	U68105	Poly(A) binding protein	9	8	17
13	X04098	Actin, gamma	6	9	15
13	M54927	Proteolipid protein 1	10	4	14
15	X06747	Heterogeneous nuclear ribonucleoprotein hnRNP A1	7	7	14
16	M23613	Nucleophosmin	9	3	12
17	M92934	Connective tissue growth factor	8	4	12
18	X07036	Guanine nucleotide binding protein G(S), alpha subunit	6	6	12
19	M11353	Histone H3 class A (H3.3)	8	3	11
20	D45887	Calmodulin 2 (phosphorylase kinase, delta)	7	4	11

TABLE 4

Human cochlear ESTs with homology to nonhuman genes

Cochlear accession number	Cluster size[a]	Accession number	Species	Gene name[b]	Nucleotide sequence identity[c]
AW020703	1	M13095	Rat	0-44	263/297 (88%)
AW023063	1	AF190797	Mouse	Actin-related protein 11	262/299 (87%)
AW021288	1	AF223953	Mouse	ARL-6 interacting protein-1	240/262 (91%)
AW020027	2	L24753	Cow	BTLF3 lactoferrin	362/420 (86%)
AW022825	2	U14908	Rat	Cytochrome-450 11-beta hydroxylase/ aldosterone synthase (CY11B2)	223/263 (84%)
AW021213	3	AF079765	Mouse	Enhancer of polycomb (Epc1)	283/295 (95%)
AW022101	1	AF001532	Mouse	Fn54	128/137 (93%)
AW021163	1	AF033663	Mouse	Gene-trap line CT 143 PR4 protein kinase homolog (cb143)	468/504 (92%)
AW023727	1	U34259	Mouse	Golgi 4-transmembrane spanning transporter MT	356/391 (91%)
AW020184	1	L07925	Rat	Guanine nucleotide dissociation stimulator	79/82 (96%)
AW023859	2	U02026	Mouse	Insulin-like growth factor binding protein 5	118/126 (93%)
AW021127	1	AF074020	Mouse	Integral membrane protein 2A (Itm2a)	197/235 (83%)
AW023587	1	U47024	Mouse	Maternal–embryonic 3 (Mem3)	237/274 (86%)
AW023273	1	D78572	Mouse	Membrane glycoprotein	180/188 (95%)
AW020586	1	D32249	Rat	Neurodegeneration associated protein 1	169/193 (87%)
AW022367	1	D45913	Mouse	NLRR-1 leucine-rich-repeat protein	318/360 (88%)
AW023554	1	AF100421	Rat	p80	147/166 (88%)
AW023626	1	AB020504	Rat	PMF31	208/233 (89%)
AW022120	1	U55057	Mouse	Receptor protein tyrosine phosphatase-lambda	222/259 (85%)
AW023212	1	D78188	Mouse	SCID complementing gene 2	180/210 (85%)
AW020966	1	AB024935	Mouse	Sid3177	199/219 (90%)
AW022731	2	U13176	Rat	Ubiquitin conjugating enzyme (E217kB)	213/224 (95%)

[a]Number of cochlear ESTs in gene cluster that have sequence similarity to the particular gene sequence.
[b]Gene name of nucleotide sequence to which the cochlear EST sequence most closely matches.
[c]Length of sequence similarity of cochlear EST sequence to the gene sequence and percentage of sequence identity in the aligned region.

TABLE 5

Categorization of cochlear ESTs that have sequence similarity to no known human or nonhuman genes

| | Number of ESTs[a] | | |
Categorization	Project 1	Project 2	Total
Cochlear ESTs with sequence similarity to ESTs produced from other tissues	1709 (1186 clusters)	1568 (1262 clusters)	3277 (2268 clusters)[b]
Cochlear-specific ESTs	420 (373 clusters)	358 (333 clusters)	778 (700 clusters)[b]
Total number of ESTs	2129	1926	4055

[a]Value in parentheses represents number of clusters into which ESTs with sequence similarity have been grouped.
[b]Sum reflects clusters in common.

or a nonhuman gene; 404 cochlear ESTs were found to have significant sequence similarity to an STS marker. Of the 404 cochlear ESTs assigned to an STS marker, 12 were cochlear-specific. For those STS markers for which information was available (see Methods), the chromosome and genetic interval were then determined. The chromosomal assignment was found for 373 ESTs (7 cochlear-specific) and represented 318 loci (3 loci for cochlear-specific ESTs) (see http://hearing.bwh.harvard.edu for a complete listing). The remaining 31 ESTs (5 cochlear-specific) representing 28 loci (3 loci for cochlear-specific ESTs) could not be assigned to a chromosome at this time because they had nucleotide sequence similarity to STS markers that are currently unassigned. Of the 373 ESTs assigned to a chromosome, the genetic interval (in centiMorgans) was known for 231 that represent 194 loci.

Every chromosome, except the Y chromosome, had at least one additional locus assigned as a result of newly assigned cochlear ESTs. Chromosomes 1 (39 ESTs representing 36 loci) and 2 (37 ESTs representing 30 loci) had the most ESTs assigned to them, followed by chromosome 11 (30 ESTs representing 22 loci), chromosome 3 (23 ESTs representing 19 loci), chromosome 10 (23 ESTs representing 12 loci), chromosome 12 (22 ESTs representing 19 loci), and chromosome 5 (22 ESTs representing 17 loci). Each of the remaining chromosomes had less than 20 ESTs assigned. Of the cochlear ESTs assigned to a chromosome, 7 ESTs (representing 3 loci) were cochlear-specific and were assigned to chromosomes 9, 12, and 17. These three loci are identified by the following ESTs: AW023375 (chromosome 9), AW020502 (chromosome 12), and AW020266 (chromosome 17).

New cochlear ESTs map to deafness loci

The map positions of 49 project 2 cochlear ESTs, representing 28 distinct loci, fall within the genetic intervals of 23 syndromic and nonsyndromic deafness loci

TABLE 6

New cochlear ESTs that map to deafness loci

Deafness locus	Map location	Cochlear EST accession numbers	Cluster size
DFNA7	1q21-23	AW023221	1
BOR2	1q31	AW021744	1
DFNA16	2q24	AW022164	1
		AW022528	1
USH2B	3p23-24	AW020214	1
DFNA18	3q22	AW021801	1
DFNB15	3q21-q25	AW020808	2
		AW020885	2
		AW020363	1
		AW022579	2
USH2C	5q14.3-q21	AW022128	1
DFNB14	7q31	AW022543	1
DFNB13	7q34-36	AW021896	1
DFNB7, DFNB11	9q13-q21	AW023065	2
USH1F	10	AW023248	1
		AW020745	1
DFNB20	11q25-qter	AW020110	9
DFNB5	14q12	AW021286	1
		AW020024	1
USH1A	14q32	AW021421	1
DFNB16	15q21-22	AW021503	1
DFNA4	19q13	AW020688	1
USH1E	21q	AW022229	3
		AW023417	1

(Table 6). For example, two new cochlear ESTs, AW022164 and AW022528, map to chromosome 2 within the genetic interval for the nonsyndromic deafness disorder DFNA16 (171.9–182.5 cM). The cochlear EST AW023065 maps to chromosome 9 within the genetic interval of DFNB7 and DFNB11. (Note: AW023065 is assigned to a cluster consisting of 2 overlapping cochlear ESTs. Because the library was not normalized, the cluster size reflects the relative level of expression of a particular gene.) Another EST, AW023248, maps to chromosome 10 within the genetic interval for the syndromic deaf/blind disorder USH1F (60.4–77.2 cM).

TABLE 7

Deafness-associated genes found in the human fetal cochlear cDNA library

	Frequency			
Gene symbol (protein)	Project 1	Project 2	Total	Hearing disorder
COL1A2 (collagen type I alpha 2)	86	98	184	Osteogenesis imperfecta
PMP22 (peripheral myelin protein 22)	5	7	12	Charcot-Marie-Tooth disease type 1A
COL11A1 (collagen type XI alpha 1)	6	6	12	Stickler syndrome (STL3)
COCH (cochlin)	4	6	10	DFNA9
GJB2 (connexin 26)	5	3	8	DFNB1, DFNA3
COL2A1 (collagen type II alpha 1)	2	6	8	Stickler syndrome (STL1)
SOX9	2	1	3	Camptomelic dysplasia
EDNRB (endothelin-B receptor)	2	1	3	Waardenburg syndrome type IV
KVLQT1	2	0	2	Jervell and Lange–Nielsen (JLNS1)
EYA1A (eyes absent 1A)	2	0	2	Branchio-oto-renal syndrome
COL4A5 (collagen type IV alpha 5)	0	1	1	Alport syndrome
COL11A2 (collagen type XI alpha 2)	1	0	1	DFNA13, Stickler syndrome (STL2)
GJB6 (connexin 30)	1	0	1	DFNA3
SOX10	0	1	1	Waardenburg syndrome type IV

Identification of genes known to be involved in deafness in the cochlear library

Currently, at least 46 genes have been shown to cause deafness (Steel and Kros 2001; Van Camp and Smith 2001). Table 7 lists the 14 genes detected among the ESTs in the human fetal cochlear library whose mutant alleles are involved in human hearing loss. For example, the following genes and their corresponding disorders are among those deafness genes found in the cochlear library: COCH (DFNA9), COL2A1 (Stickler syndrome type 1), COL4A5 (Alport syndrome), COL11A1 (Stickler syndrome type 3), EDNRB (Waardenburg syndrome type IV), and GJB2 (DFNA3, DFNB1).

DISCUSSION

This study was undertaken to identify additional genes important for hearing and deafness. A tissue-specific approach, using total RNA extracted from 16–22 week human fetal cochlea, was our source for expressed auditory genes. Nucleotide sequence similarity (BLAST) analysis was performed with various GenBank databases to determine which known human and novel human homologs of nonhuman mammalian genes were present among the cochlear ESTs. The EST database was also included in the analysis for identification of corresponding ESTs from other tissue-derived libraries. Any cochlear EST that did not have nucleotide sequence similarity to a gene or other EST was considered to be "cochlear-specific." The chromosomal map position was determined for those ESTs that did not have significant sequence similarity to any known gene to establish whether a particular EST was located within the genetic interval of a deafness locus.

It is of interest to know how the cochlear ESTs compare with other tissue-specific ESTs, particularly with respect to the identification of novel sequences. The UniGene database at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) consists of GenBank sequences that have been partitioned into a nonredundant set of gene clusters. Each UniGene cluster contains sequences, both of well-characterized genes as well as of novel ESTs, that represent a unique gene. Currently 358 libraries representing various tissues have been used to produce EST sequences. The number of sequences contributed by each library ranges from 7 to 85,989. According to the UniGene library reports that are updated monthly and available at the NCBI website, as of April 13, 2001, the Morton Fetal Cochlear cDNA Library is ranked 87 of 358 libraries for contribution of sequences, 69 for contribution of clusters (genes), and 72 for contribution of nucleotide sequences that make up novel UniGene entries.

A total of 8494 human cochlear ESTs were generated from our human fetal cochlear cDNA library, of which 5393 (63.5%) were 5′ reads and 3097 (36.5%) were 3′ reads. 8153 (96%) of these ESTs were used for BLAST analyses. About 50% of these ESTs ($n = 4040$) had sequence similarity to a total of 1449 known human genes, of which 365 were identified by ESTs from both sequencing projects; 490 genes were identified uniquely by ESTs from project 1, and 594 were identified only by ESTs from project 2. A total of 43 human homologs of nonhuman mammalian genes have also been identified. Of the remaining 4055 ESTs, 3277 have sequence similarity to other ESTs and repre-

TABLE 8

Tissue expression profile of nonsyndromic deafness genes based on EST sequence similarity

Gene	Deafness	UniGene ID	No. ESTs	cDNA sources[a]
COCH	DFNA9	Hs.21016	59	Brain, CNS, ear, eye, kidney, lung, pooled, prostate, testis, colon, eye, head–neck, lung
DFNA5	DFNA5	Hs.13530	59	Bone, brain, foreskin, germ cell, head and neck, heart, kidney, muscle, pancreas, placenta, pooled, uterus, breast, head–neck, lung, muscle, nervous-normal, pancreas, prostate-tumor, uterus, uterus-tumor
DIAPH1	DFNA1	Hs.26584	76	Blood, brain, colon, esophagus, foreskin, germ cell, head and neck, heart, kidney, lung, lymph, muscle, ovary, pancreas, placenta, pooled, prostate, stomach, testis, tissue culture, tonsil, uterus, breast, breast-normal, cervix, colon, head-neck, lung, nervous-tumor, pancreas, placenta, prostate-tumor, skin, stomach, tongue, uterus
GJB2	DFNB1, DFNA3	Hs.5566	320	Blood, brain, breast, colon, ear, esophagus, eye, foreskin, gall bladder, germ cell, heart, kidney, liver, lung, lymph, muscle, ovary, pancreas, placenta, prostate, spleen, stomach, testis, tonsil, uterus, whole embryo, brain, breast, breast-normal, cervix, colon, colon-normal, eye, head–neck, kidney, leiomios, lung, lymph, nervous-normal, nervous-tumor, ovary, pancreas, placenta, pnet, skin, stomach, thymus, pooled, uterus
GJB3	DFNA2	Hs.98485	26	Germ cell, heart, ovary, pancreas, prostate, stomach, uterus, lung, ovary, stomach
GJB6	DFNA3	Hs.248213	4	Head–neck
KCNQ4	DFNA2	Hs.241376	0[b]	No information
OTOF	DFNB9	Hs.91608	7	Brain
TECTA	DFNA8, DFNA12, DFNB21	Hs.248162	2	Brain, testis
TMPRSS3	DFNB10	Hs.283805	13	Colon, kidney, lung, uterus, whole embryo, ovary, pancreas

[a]Taken from UniGene (NCBI).
[b]Manual BLAST against dbEST revealed 7 ESTs derived from brain tumor cDNA clones (E values $< 1 \times 10^{-27}$).

sent 2266 clusters or genes. The remaining 778 ESTs have sequence similarity to no known genes nor ESTs and are considered to be "cochlear-specific."

Identification of cochlear ESTs with significant nucleotide sequence similarity only to known mammalian genes not found previously in human tissue is important because these cochlear ESTs may represent the respective human homologs. The nonhuman genes may be characterized already, facilitating both the cloning and functional assessment of the human homologs, and thus aiding the process of understanding their role in human hearing. Among the group of 43 nonhuman mammalian genes are genes that encode several transmembrane proteins of interest. Transmembrane proteins, such as the connexins, have been shown to be important for the hearing process, specifically for maintenance of the potassium-rich endolymph by actively recycling potassium ions from within hair cells to supporting cells, stria vascularis, and back to the endolymph.

Identification of cochlear ESTs that map to deafness loci provide another way to identify genes important for hearing, especially those involved in hearing impairment. All cochlear ESTs mapped on human chromosomes, including those from project 1, that do not have nucleotide sequence similarity to a known gene sequence are presented on our website (http://hearing.bwh.harvard.edu); also included are the known syndromic and nonsyndromic deafness loci. A total of 788 loci are represented by cochlear ESTs, 318 of which were identified by the present study. Of the project 2 cochlear ESTs that have nucleotide sequence similarity to no known gene, 373 have been mapped in the human and at least 4 (representing 2 loci) and as many as 39 (representing 36 loci) are found on every chromosome, excluding the Y chromosome. Forty-nine of these mapped project 2 ESTs represent 28 distinct genetic loci that are located within the genetic intervals of 23 deafness loci. Including the mapped ESTs from project 1, a total of 120 cochlear

ESTs map within the genetic interval of 34 deafness loci and represent 99 positional candidate genes for these deafness disorders.

None of 49 cochlear ESTs that map within the genetic interval of various deafness loci are cochlear-specific (i.e., they match ESTs produced from other tissue-specific cDNA libraries). This is not a surprising finding as all of the genes responsible for nonsyndromic deafness identified to date are also not "cochlear-specific" (Table 8). Of the nonsyndromic deafness genes identified, only *COCH, GJB6, KCNQ4, OTOF,* and *TECTA* have nucleotide sequence similarity to relatively few ESTs. Most deafness genes discovered to date are expressed in many different tissues/organs, as determined by the tissue expression profile of the ESTs to which they have nucleotide sequence similarity (Table 8). Of note, although all of these nonsyndromic deafness genes are expressed in the inner ear, only *COCH, GJB2,* and *GJB6* have been identified among the 8494 cochlear ESTs, reflecting their high expression levels in the cochlea and/or the incomplete gene expression profile of the cochlear transcripts.

In summary, the two sequencing projects have resulted in identification of 1449 known human genes expressed in the cochlea; 59% were identified by the first project and 41% by the second project (Fig. 1). The finding of an additional 594 known genes among ESTs from project 2 indicates that the sequencing effort was clearly valuable to ascertain additional transcripts present in the human fetal cochlear library. Of note, the finding of additional genes during project 2 is not because new genes have been identified and deposited in GenBank since our initial analysis of project 1 ESTs, since we updated the data by reanalyzing the unassigned ESTs from project 1. A more likely explanation for this finding is the great complexity and diversity of cochlear genes. The identification of additional genes in project 2 suggests that future sequencing efforts would likely lead to identification of yet additional known and novel human genes expressed in the cochlea.

ACKNOWLEDGMENTS

REFERENCES

ALTSCHUL SF, MADDEN TL, SCHÄFFER AA, ZHANG J, ZHANG Z, MILLER W, LIPMAN DJ. Gapped BLAST and PSI–BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402, 1997.

GORLIN RJ, TORIELLO HV, COHEN MM. Hereditary hearing loss and its syndromes. Oxford University Press Oxford, 1995.

GURISH MF, BELL AF, SMITH TJ, DUCHARME LA, WANG RK, WEIS JH. Expression of murine beta 7, alpha 4, and beta 1 integrin genes by rodent mast cells. J. Immunol. 149:1964–1972, 1992.

HEDRICK SM, COEN DI, NIELSEN EA, DAVIS MM. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature 308:149–153, 1984.

JONES DT, REED RR. Golf: an olfactory neuron specific G-protein involved in odorant signal transduction. Science 244:790–795, 1989.

MORTON NE. Genetic epidemiology of hearing impairment. Ann. N. Y. Acad. Sci. 630:16–31, 1991.

ROBERTSON NG, HELLER S, LIN JS, RESENDES BL, WEREMOWICZ S, DENIS CS, BELL AM, HUDSPETH AJ, MORTON CC. A novel conserved cochlear gene, OTOR: Identification, expression analysis, and chromosomal mapping. Genomics 66:242–248, 2000.

ROBERTSON NG, KHETARPHAL U, GUTIERREX–ESPELETA GA, BIEBER FR, MORTON CC. Isolation of novel and known genes from a human fetal cochlear cDNA library using subtractive hybridization and differential screening. Genomics 23:42–50, 1994.

ROBERTSON NG, LU L, HELLER S, MERCHANT SN, MCKENNA M, EAVY RD, NADOL JB, MYAMOTO RT, LUBIANCA NETO JF, HUDSPETH AJ ET AL. Mutations in a novel cochlear gene cause DFNA9, a human nonsyndromic deafness with vestibular dysfunction. Nat. Genet. 20:299–303, 1998.

ROBERTSON NG, SKVORAK AB, YIN Y, WEREMOWICZ S, JOHNSON KR, KOVATCH KA, BATTEY JF, BIEBER FR, MORTON CC. Mapping and characterization of a novel cochlear gene in human and in mouse: a positional candidate gene for a deafness disorder, DFNA9. Genomics 46:345–354, 1997.

SKVORAK AB, ROBERTSON NG, YIN Y, WEREMOWICZ S, HER H, BIEBER FR, BEISEL KW, LYNCH ED, BEIER DR, MORTON CC. An ancient conserved gene expressed in the human inner ear: identification, expression analysis, and chromosomal mapping of human and mouse antiquitin (*ATQI*). Genomics 46:191–199, 1997.

SKVORAK AB, WENG Z, YEE A, ROBERTSON NG, MORTON CC. Human cochlear expressed sequence tags provide insight into cochlear gene expression and identify candidate genes for deafness. Hum. Mol. Genet. 8:439–452, 1999.

STEEL KP, KROS CJ. A genetic approach to understanding auditory function. Nat. Genet. 27:143–149, 2001.

VAN CAMP G, SMITH RJH (2001) Hereditary Hearing Loss home page. http://www.uia.ac.be/dnalab/hhh.

VAN CAMP G, WILLEMS PJ, SMITH RJ. Nonsyndromic hearing impairment: unparalleled heterogeneity. Am. J. Hum. Genet. 60:758–764, 1997.