Understanding Inner Ear Development with Gene Expression Profiling

Zheng-Yi Chen,[1,2] David P. Corey[3,4]

[1] Neurology Service, Massachusetts General Hospital, WEL425, Boston, Massachusetts 02114

[2] Department of Neurology, Harvard Medical School, Boston, Massachusetts 02115

[3] Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115

[4] Howard Hughes Medical Institute, Chevy Chase, Maryland 20815

ABSTRACT: **Understanding the development of the inner ear requires knowing the spatial and temporal pattern of gene expression, and the functions of those gene products. In the last decade, hearing research has benefited tremendously from the progress of the human and mouse genome projects, as amply illustrated by the identification of many deafness genes in both human and mouse. However, the sheer amount of information generated from the genome project has far outpaced the rate at which it is utilized. Microarray technology offers a means to quantify the expression level of transcripts at a whole-genome scale. Cross-tissue comparisons will identify genes unique to the inner ear, which will expedite the identification of new deafness genes. Microdissection and subtraction after ablation of cell types can reveal genes expressed in certain cells, such as hair cells. Expression profiling of both inner ear and other tissues, under a variety of conditions (such as during development, with drug treatment or in knock-out animals), can be used for cluster analysis to group genes of similar expression. Coexpression can suggest functional pathways and interactions between known genes, and can identify new genes in a structure or pathway. In this review we give examples for both transcription factors and cochlear structures.** J Neurobiol 53: 276–285, 2002

Keywords: **inner ear development; gene expression profiling; human and mouse genome projects**

INTRODUCTION

Different strategies have been developed to identify genes expressed in the inner ear, including direct cDNA sequencing of inner ear libraries, macroarray analysis of a subtracted chick cochlea cDNA library, RT-PCR for candidate genes, and microarray profiling using tissues and cell lines derived from the inner ear (Gong et al., 1996; Heller et al., 1998; Skvorak et al., 1999) (for a detailed review of different techniques, see the article in this issue by Stefan Heller). Currently, there is no single method that is adequate for all gene discovery in the inner ear. Methods such as cDNA library sequencing, macroarray analysis, and RT-PCR can identify new genes or characterize known genes, but they can either be performed for a limited time (in case of cDNA sequencing and macroarray) or can only deal with a limited number of genes (such as RT-PCR). Microarray technology uses existing sequence information, and so is unable to identify completely novel genes. However, it can be used to study the expression pattern of tens of thousands of genes from the inner ear, under many conditions, with easy comparison across tissues and experiments (Lander, 1999). Because new genes are being discovered continuously through the human genome project and other cDNA sequencing projects, the number of genes not represented on microarrays will rapidly diminish. Thus microarray-based expression profiling shows great promise for address-

Correspondence to: David P. Corey (dcorey@hms.harvard.edu).

DOI 10.1002/neu.10125

ing a broad range of questions involving gene expression in inner ear development and function.

Microarray Technologies

The basic concept for microarray technology is simple: immobilize DNA targets that represent many thousands of known genes, unknown cDNAs, and ESTs, and then hybridize the targets with labeled RNAs from a cell sample. Typically, double-stranded cDNA is prepared from mRNA or from total RNA using a T7-oligo(dT) primer, and then amplified with RNA polymerase using biotinylated nucleotides. Fluorescently tagged streptavidin makes the hybridized sample visible in a scanner. By simultaneously measuring the hybridization signal for all targets one can determine the number of genes expressed in a tissue sample and the level of expression. Because many targets (ranging from 10,000 cDNAs to 400,000 oligonucleotides) can be immobilized on one surface, comprehensive expression patterns for a sample can be detected (Lander, 1999).

There are three main platforms of microarray technology: spotted cDNA microarrays, spotted oligonucleotide arrays, and oligonucleotide arrays prepared by combinatorial photosynthesis (Affymetrix GeneChip®). Spotted cDNA arrays use cDNA fragments prepared from PCR amplification of cDNA clones, which are then robotically spotted on a glass slide (Brown and Botstein, 1999). Over 10,000 independent cDNA clones can be spotted in an area of 1×1 cm. Arrays are hybridized with a double sample in which each sample is fluorescently labeled with a different dye. After hybridization, fluorescence signals are determined for each dye separately and the ratio of two fluorescent colors reflects the relative abundance of a specific gene in two samples. Because standardization is difficult, hybridization results from different studies cannot be easily compared, and so the expression ratio of experimental and control samples is the principal calibration. However, the spotted cDNA array is a very flexible platform and an average laboratory can produce its own collection of glass arrays. These arrays are particularly suited for the study of genes from species and tissues where amplified PCR fragments are not commercial available (such as chick and frog inner ears). Nevertheless, the production of many thousand high-quality PCR products for spotting can be a daunting task; as the numbers of clones climb into the thousands, production of spotted cDNA arrays can be better accomplished by a genome center or a commercial company.

Affymetrix uses a highly automated photolithography and solid-phase DNA synthesis to produce an oligonucleotide array (GeneChip) with very high density (400,000 polydeoxynucleotides on a 1.3×1.3-cm area) (Lipshutz et al., 1999). The small scale of GeneChip technology will make it possible to put all known genes and ESTs on a few chips, to produce a "whole-genome" chip. Hybridization is carried out using a single fluorescently labeled nucleic acid sample complementary to the oligonucleotides on the array. A laser scanner produces an image of hybridization intensity. Comparisons can be made between different samples on different chips because of manufacturing uniformity. On a GeneChip, each gene is represented by a matching set of 25-mer oligonucleotides (usually 16 per gene). For each 25-mer there are two probes that are synthesized adjacent to each other: perfect-match and mismatch oligonucleotides. The two oligonucleotides have exactly the same sequences except for one nucleotide in the center of the mismatch oligonucleotide, so that the signal derived from the mismatch oligonucleotide can be used as the control for nonspecific hybridization. The signal difference between the perfect-match and mismatch spots is used to determine the expression level of a gene. Because oligonucleotide sequences were derived from an evolving sequence database, including poorly sequenced ESTs, there can be errors in individual probes for a gene. We have found a surprisingly high number of errors in oligonucleotide sequences, when compared to established sequences for known genes. The redundancy of multiple probes per gene and the inclusion of mismatch controls may alleviate this problem, but all microarray results should be confirmed by more traditional methods.

The oligonucleotide spotted array is a platform in between the cDNA array and the GeneChip. It uses ink-jet technology to synthesize oligonucleotide *in situ* on the glass slides. This is a relatively new technology that has yet to be used by many laboratories.

Using Microarray Technology in the Inner Ear

Expression profiling in the inner ear is particularly challenging, due to the many cell types in this complex organ, the small number of cells for each type, and their physical location deep within the bony structure of the skull. The small number of cells requires a highly sensitive assay to detect the expression level, whereas the numerous intermixed cell types make it difficult to discern cell type-specific expression. We have used Affymetrix's GeneChips to study gene expression in the inner ear, because it covers the largest numbers of human and mouse genes available,

it is commercially available, and the results can be compared with studies from other laboratories using the same chips. For the examples presented here, we used the murine GeneChip Mu30K set (which consists of ~13,000 known genes and ~20,000 EST clusters) to profile the expression pattern in developing mouse inner ear and several other tissues (Chen and Corey, 2001). In other studies of human cochlea we have used the human GeneChip U95 set (covering ~10,000 known genes and ~~44,000 EST clusters).

A large number of genes was found to be expressed in the mouse cochlea at either P2 or P32. In total over ~50% of the genes and ESTs on the chip were detected according to the Affymetrix analysis software. This is not too surprising: single cell lines can express 10,000 to 20,000 genes as assessed by SAGE analysis, and there are many cell types within the cochlea. The variety of cell types and the limited amount of tissue in a mouse cochlea raises concerns about detection limits, however. In general, for GeneChips, about 50,000 cells of one type are needed for detection of mRNAs that occur at two to five copies/cell, and about two-thirds of all genes are expressed at <5 copies/cell (Zhang et al., 1997). To get 50,000 hair cells, for instance, we need ~15 mouse cochleas or ~30 mouse utricles. Less tissue can be used if the sample is further amplified, but this reduces confidence in the measurement of relative expression.

To test sensitivity, we looked at the expression of known hair-cell genes in a sample of 10 cochleas, estimated to contain ~30,000 hair cells. In the P2 and P32 mouse cochleas some hair-cell genes such as myosin-VIIa and calretinin were detected, yet others such as Math-1 were not. The lack of detection of some hair-cell genes is likely due to the small number of hair cells present in the complex sample. In a separate experiment using a mouse utricule sample that contained ~50,000 hair cells, most of the hair-cell genes were detected. Some hair-cell genes, such as PMCA2, were not detected either in the cochlea or the utricle samples. Their mRNA copy numbers are possibly below the sensitivity of detection by the GeneChip assay, or it may be that hybridization to the chosen oligonucleotides is inefficient. This study gave us confidence, however, that most mRNAs present at 5–10 copies/cell could be detected in a complex sample.

Many factors can influence the reproducibility of an expression study. Unlike tissues that are readily accessible, cochleas are difficult to dissect such that exactly the same tissue is recovered each time. Duplicate hybridizations were performed for P2 and P32 mouse cochlea to determine the degree and source of variability. For P2 mice, 20 cochleas were collected and total RNA was extracted, then the RNA was divided into halves and each half was used to synthesize cDNA and cRNA. Differences in measured expression between the two samples should reflect the variations inherent in cDNA and cRNA synthesis, variation between chips, and subsequent hybridization and scanning. For P32 mice, two batches of 10 cochleas were collected independently. Each batch then went through RNA extraction, cDNA and cRNA synthesis, and hybridization. Differences between these samples should include the additional variation brought about by the difference between sample collection and RNA extraction. Overall, there was good correlation between the two P2 samples, and between the two P32 samples. However, a larger degree of variation was observed between the two P32 than the two P2 samples, as demonstrated by the correlation coefficient (Fig. 1). For GeneChip experiments on inner ear it is better to dissect enough tissue to pool total RNA, then run duplicates with the total RNA. The GeneChip system itself is very reproducible, as shown both by others (Mahadevappa and Warrington, 1999) and by our own data: when highly homogenous samples were used the correlations of expression were over 99%. Nevertheless, most experiments require duplicate samples run in parallel to reduce the intrinsic noise of the system.

Identification of Cell-Specific Genes by Subtraction

The reproducibility of GeneChips allows identification of cell type-specific genes by subtraction between data sets. For instance, it would be difficult to produce a pure sample of 50,000 hair cells, and analysis of a smaller sample with further amplification may not give accurate representation of expression levels. However, hair cells can be removed from a sample in several ways. Dissected utricles can be treated with aminoglycoside antibiotics in culture to kill hair cells, before the tissue is processed for RNA. Alternatively, utricles or cochleas can be dissected from mice with a targeted deletion of the Math1 gene, which lack hair cells (Bermingham et al., 1999), and from their wild-type littermates. Many of the genes expressed in the normal samples but not in those lacking hair cells will be hair-cell specific.

Identification of Structural Components by Subtraction

Some mutant mice have hair cells, but show more subtle defects in morphology. Phenotypes affecting hair bundles are particularly striking; such mice in-

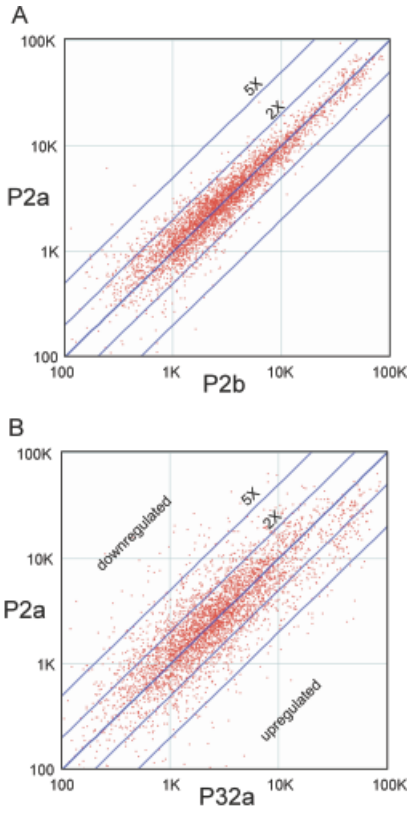

Figure 1 Scatter plots showing the correlation of expression levels. Average difference values (equivalent of signal intensity, determined by the analysis software) were used to quantify expression, and were plotted on a log scale. The blue lines demarcate a twofold (2×) or fivefold difference (5×) between samples in average difference. (A) Two mouse cochlea samples, both acquired at P2, compared to each other (P2a vs. P2b). Most (but not all) genes show expression levels within 2×. (B) Two mouse cochlea samples, acquired at P2 and at P32. The significantly lower correlation in (B) indicates changes in gene expression during development.

clude *shaker-1, shaker-2, pirouette, waltzer, Ames waltzer, Beethoven*, and mice with a deletion of the Brn-3.1 gene. In some cases, the mutations cause defects in particular structural proteins, such as myosins in *shaker-1* and *shaker-2*. It is possible, although not certain, that the dysfunction of one protein can cause changes in gene expression of functionally related proteins, and that a mutation in one stereocilia protein changes expression of another stereocilia protein. If so, a subtraction between mutant and wild type may reveal those interacting proteins. Interpretation could be difficult; for instance, synaptic vesicle proteins may be downregulated in a hair cell that lacks transduction because its bundle never formed properly. In other cases, the gene defect is in a transcription factor, such as Brn-3.1. The downstream targets of that factor, needed for formation of a hair bundle, will necessarily be either down- or upregulated in the mutant, depending on whether Brn-3.1 is an activator or repressor. These target genes—either direct or indirect—will include those that encode proteins of the hair bundle.

Identification of Downstream Targets of Developmental Transcription Factors

Many transcription factors have been shown to be important in the development of the inner ear, including Math1, Hes-1, and Brn-3.1 (Erkman et al., 1996; Xiang et al., 1997, 1998; Bermingham et al., 1999; Shailam et al., 1999; Zheng and Gao, 2000; Zheng et al., 2000). Few if any of the downstream targets for these transcription factors have been identified. In principle, subtractions of gene expression data from wild-type mice and those lacking a functional transcription factor can reveal these targets, and under certain circumstances (Brn-3.1) the interpretation can be relatively straightforward. In other cases (Math1), the absence of the transcription factor eliminates an entire class of cell, and so a subtraction will turn up all genes expressed in that cell type, whether or not directly controlled by the factor. In such cases, the negative assay provided by a mutant must be replaced or supplemented by a positive assay, to identify genes induced.

We have used two standard assays with analysis by GeneChips. In one, cell lines derived from the inner ear were transfected with a transcription factor. Cells were cultured for 12–48 h, and RNA extracted for GeneChip analysis. Gene expression was compared to cells transfected with an irrelevant marker. A difficulty with this approach is that only a fraction of cells is transfected. For instance, if the factor upregulates transcription by fivefold, but only 20% of cells are transfected, total mRNA for that gene is only doubled—barely at the detection limit for GeneChips. To address this problem, we transfect with a vector that contains GFP under a separate promoter, and use fluorescence-activated cell sorting to purify transfected cells. Another difficulty is that transfection itself may induce large changes in gene expression, and even random variation between cells transfected with the transcription factor or with marker alone can produce spurious positives. In a different assay, the transcription factor is placed under the control of a tetracycline-controlled element and a stable cell line is created (Lee et al., 1999). When tetracycline is removed from the culture medium, expression of the transcription factor is induced, and—with some delay—expression of genes controlled by that transcrip-

tion factor. Target genes are identified by subtraction between cultures with and without tetracycline. Although this assay produces more specific expression of target genes, it uses a cell type that is not a hair cell, and that may not contain cofactors needed for expression of hair-cell genes.

Thus, none of the three approaches is perfect for identifying targets of inner-ear transcription factors, but genes that turn up in all three assays warrant further study. Moreover, a whole-genome approach, as is possible with GeneChips, is likely to turn up a large number of candidates. If the genes are confirmed by specific assays, their promoter regions can be examined for common binding motifs (Tavazoie et al., 1999; Hughes et al., 2000). The identification of the binding motif will likely indicate direct target genes instead of secondary targets, and can be used to search for additional targets not found in the initial screen.

Statistical Analysis of Gene Expression Data

In these examples, genes were identified by simple subtraction between two conditions. Because of the intrinsic noise in the system, a large number of false positives is produced, even with duplicate experiments. Much more sophisticated analyses can be done if expression data from many different conditions are compared. "Different conditions" can be a developmental series, various mutant animals, or different tissues, but the analysis relies on the idea that genes with similar expression profiles may interact, either as a signal transduction pathway or in a cellular structure.

An early example is a comprehensive survey by two independent groups of genes that vary in expression during the yeast cell cycle (Cho et al., 1998; Spellman et al., 1998). Cultured yeast cells were synchronized using different methods, and the hybridizations were carried out using either GeneChips or DNA microarrays. By following the genes whose expression levels change along with cell cycle, the experiments identified 800 cell cycle-regulated genes out of the 6000-gene yeast genome. The 800 genes contain a majority of known cell-cycle genes and many new genes whose cell-cycle roles had not been recognized previously. When genes are coregulated it may also indicate that they likely share some common *cis*-regulatory elements in their promoters that are recognized by the same transcription factor. Recognition of such elements could reveal the transcription factors that bind them. For instance, the analysis of clustered galactose-response genes yielded the known binding site of Gal4, and six of nine genes known to be induced by galactose (Roth et al., 1998). Analysis of gene expression following heat shock revealed the cell-cycle activation motif, which is known to mediate cell-cycle dependent activation. Other examples include the dissection of the regulatory circuitry of yeast (Holstege et al., 1998), analysis of *Drosophila* development during metamorphosis (White et al., 1999), identification of genes involved in aging of mouse (Lee et al., 1999), cancer classification (Golub et al., 1999; Alizadeh et al., 2000), and dissection of signaling and circuitry of multiple MAPK pathways (Roberts et al., 2000).

Two general strategies are used for grouping genes by coexpression. One group includes the self-organizing map (SOM), principal component analysis, hierarchical analysis, and the k-mean method (Tamayo et al., 1999; Quackenbush, 2001). The expression pattern of each gene is assessed in each of the different experimental conditions. For instance, if the expression data from 11 conditions are analyzed using SOM, each gene is represented by a point in an 11-dimensional space. The algorithm then clusters genes that occur near each other in that space—that have similar patterns of expression among all conditions. Typically, genes are grouped into 20–40 clusters. For the inner ear, a single cluster might represent genes that are expressed in utricle, but only at P12 and not at E16, and that are not expressed significantly in whole cochlea or in macrophages. Cluster 14 in Figure 2 is an example. Although cluster analysis uses no preconceptions about candidate genes, one might expect to find within that cluster genes that prevent cell division by hair cells. This method is like the profiling used by advertisers seeking likely customers, by focusing on people who live in certain neighborhoods, subscribe to certain magazines, and have ordered from certain catalogs.

By virtue of coexpression, these genes are more likely to be involved in the same or related functional pathways (Brown and Botstein, 1999). For instance serum treatment alters gene expression in human fibroblasts. Using pairwise average-linkage cluster analysis, the genes that respond to serum treatment were clustered into functional groups, on the basis of their temporal patterns of expression (Eisen et al., 1998; Iyer et al., 1999). The immediate-early transcription factor families were activated at one time, while genes involved in angiogenesis were activated together but at a different time point. The genes with known roles in wound healing were also clustered.

A second strategy starts with a known gene, and asks which other genes show the most similar expression pattern by a correlation value. This can be done, for example, using the *t* test by specifying a *p*-value,

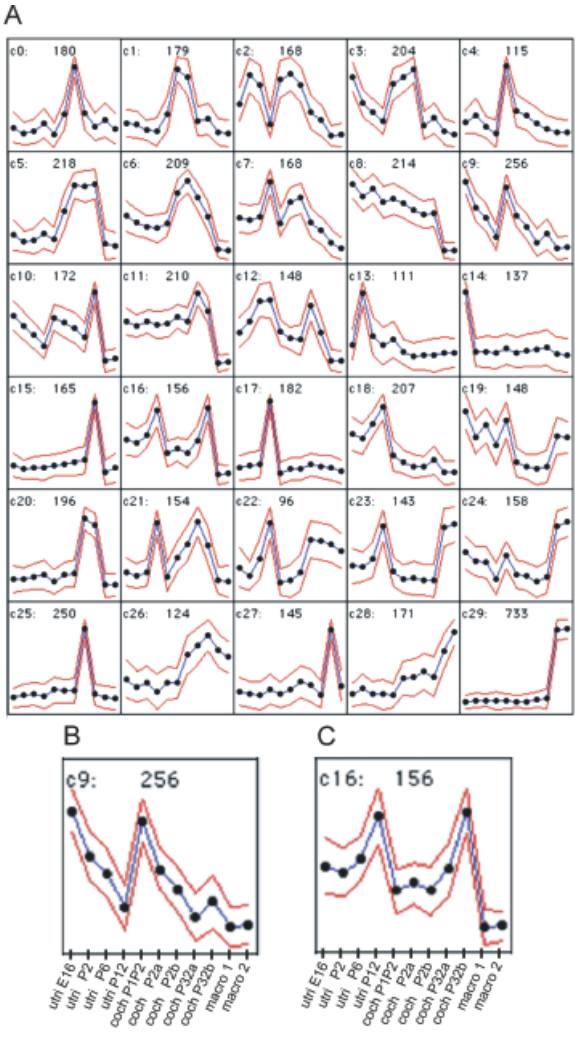

Figure 2 Cluster analysis with the self-organizing map (SOM) algorithm. (A) Thirty clusters of genes grouped based on differential expression profiles are shown. Data were obtained using 11 samples: four from utricular macula from embryonic day 16 to postnatal day 12, five from whole cochlea from postnatal day 1 to 32, and two from a macrophage cell line. The blue lines connect average values, whereas the red lines indicate variability. The number of genes in each cluster is indicated next to the cluster number (e.g., 156 genes in cluster 16). (B) Cluster 9 enlarged, with legend identifying samples. This cluster includes many genes downregulated during development in both utricular macula and cochlea. (C) Cluster 16, which includes many upregulated genes.

to identify genes with highly correlated expression patterns. In this case no attempt is made to group genes into clusters; one asks simply which genes have the most similar expression across all conditions. If the query gene is a structural protein, the coexpressed genes might be part of the same structure. This method is like that used to find members of a criminal network, where a single known individual is observed to identify his associates.

The two approaches are not exclusive, of course; one might use cluster analysis to group genes, and then ask which clusters contain known genes with a known function. Other members of the cluster could be involved in the same function. For each of these methods, new genes can be identified by probable function, or new functions identified for known genes. We will illustrate each approach with examples.

Identification of Functional Pathways by Cluster Analysis: Developmental Genes

The comparison between the P2 and P32 cochlea expression profiles revealed many genes differentially expressed. Among them certain genes involved in the structure of the inner ear—such as ototconin-95, α-tectorin, β-tectorin, and many collagens—were downregulated; whereas other genes—such as different types of Ca^{2+} binding proteins, channel genes, and ion transporters—were upregulated (Chen and Corey, 2001). During early development genes highly expressed are more likely to be involved in the maturation of the cochlea, whereas at later stages the fully developed cochlea is participating in a variety of functions including ion transport between the organ of Corti and stria vascularis, and transduction by hair cells.

More systematic clustering (Fig. 2) reveals groups of genes that may be involved in development. For example, cluster 9 contains utricle genes that are downregulated from E16 to P12, that are also downregulated from P2 to P32 in whole cochlea, but are not expressed in macrophages. This cluster includes many genes involved in DNA replication, RNA processing, and translation, and in cell-cycle control. Cluster 16 contains genes upregulated during the same period; included are genes encoding ion channels and transporters, cell adhesion molecules, and negative cell-cycle regulators.

This approach can be combined with experimental manipulations. For instance, embryonic cochleas cultured and treated with retinoic acid generate supernumerary hair cells (Kelley et al., 1993). Clustering GeneChip expression data from cultures with and without retinoic acid together with a data from a developmental series may provide more powerful grouping of cell-cycle genes.

Identification of Related Proteins by Correlation Value

Biochemical purification and genetic analyses have identified some components of inner ear structures,

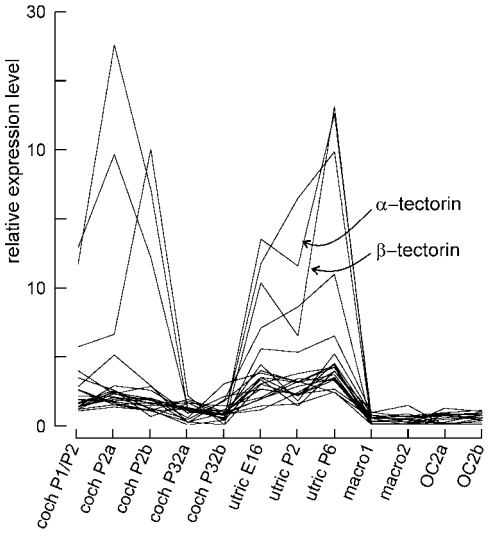

Figure 3 Correlation analysis used to identify genes with expression similar to α-tectorin, a principal component of the tectorial membrane. The Y-axis shows expression level after normalization. The highest correlation was found with β-tectorin. Other highly correlated genes include collagens IIα1, IXα3, and XV—all thought to be part of the tectorial membrane—and processing enzymes such as hepsin and a lysyl oxidase homolog precursor. The samples used for the correlation analysis include mouse cochlea (age P1/P2, P2, and P32), mouse utricle (E16, P2, and P6), a macrophage line, and a cell line derived from organ of Corti precursors.

such as α-tectorin and β-tectorin of the tectorial membrane, and otoconin of the otoconia (Killick et al., 1995; Wang et al., 1998; Verpy et al., 1999). However, those structures are likely to contain many more proteins, albeit at lower concentration. Also, because both the tectorial membrane and the otoconia are acellular structures, perhaps made by several cell types, identifying cell-type–specific genes may not identify other structural components. The comparison between the P2 and P32 mouse cochlea revealed that there was major downregulation of expression of genes encoding α-tectorin, β-tectorin, and otoconin, consistent with previous studies (Wang et al., 1998; Rau et al., 1999; Verpy et al., 1999).

With the data of Figure 2, but without cluster analysis, we asked which genes showed similar expression profiles as α-tectorin (Fig. 3). The group of genes identified included β-tectorin, collagen IIα1, and collagen IXα3, all of which are part of the tectorial membrane (Thalmann et al., 1986; Slepecky et al., 1992a, 1992b; Killick et al., 1995). Additional coexpressed genes include hepsin, lysyl oxidase homolog precursor, and collagen type XV. Hepsin is a transmembrane serine protease involved in digestion of extracellular matrix (Tsuji et al., 1991); lysyl oxidase crosslinks extracellular matrix substrates such as collagen and elastin by catalyzing the oxidative deamination of peptidyl lysine (Jourdan-Le Saux et al., 1999); and collagen type XV has been implicated in the tectorial membrane (Kalluri et al., 1998). Some or all these additional genes may play roles in the formation of the tectorial membrane. A similar approach could be applied to the study of the genes involved in the formation of otoconia.

In a separate study we asked which genes are coexpressed with the transcription factor SOX10, to identify its downstream targets. SOX10 is involved in the development of neural crest, and mutations in SOX10 have been associated with Hirschsprung disease (a hereditary cause of intestinal obstruction) and Shah-Waardenburg syndrome (WS4; a form of sensorineural deafness) (Pingault et al., 1998). SOX10 was detected in both P2 and P32 mouse cochleas by the GeneChip analysis. Among the genes highly correlated with SOX10 was ErbB-3, a receptor protein-tyrosine kinase precursor, which is a known target of SOX10 (Britsch et al., 2001).

Finally, we asked which genes are correlated with Math1, and found a variety of other transcription factors, including Hes1. Both Math1 and Hes1 are involved in hair-cell development (Zheng and Gao, 2000; Zheng et al., 2000).

Identification of Deafness Genes

This approach can also expedite the discovery of deafness genes. Over 20 nonsyndromic deafness genes have been cloned so far, out of over 70 mapped deafness loci (http://dnalab-www.uia.ac.be/dnalab/hhh). Many more loci will certainly be discovered in the future. Although traditional positional cloning or candidate gene approaches to isolate deafness genes have been very successful, identification of a deafness gene is a long, labor-intensive process. Most of the deafness families available for linkage are relatively small, making it difficult to narrow the region of a candidate gene. In many instances linkage can place a deafness locus within ~5 cM region, roughly 5 million bp of DNA. If we assume that the human genome encodes ~30,000 genes, there are ~50 candidate genes in that region that have to be analyzed to identify the causative gene. Using microarrays, it is possible to substantially reduce the number of candidates. For example, with SAGE analysis of human and mouse retina, two groups recently identified a set of retina-enriched genes; not surprisingly, most of the genes responsible for inherited blindness were within that set (Blackshaw et al., 2001; Sharon et al., 2002). A similar type of analysis can lead to cochlea-en-

riched genes, which we have found to have approximately 10-fold higher probability of being involved in nonsyndromic deafness. A gene also enriched in certain other tissues may be a good candidate for syndromic deafness. Norrie disease, a syndrome characterized by congenital blindness and progressive hearing loss, is caused by mutations in the gene encoding norrin (Chen et al., 1995). Cluster analysis showed that among a diverse set of tissues this gene is highly expressed in both the inner ear and the retina. Other such genes are candidates for forms of Usher syndrome.

Limitations

Although functional genomics by microarray analysis is a powerful tool, improvements will need to address some issues. First, a relatively large number of cells of one type (~50,000) is needed for hybridization. A few methods have been described to produce an adequate amount of aRNA for hybridization from as few as 100 cells (Luo et al., 1999; Wang et al., 2000; Baugh et al., 2001). As they all involve two rounds of amplification, the sensitivity of assay is further reduced. Second, even with a reasonable number of cells, low-copy-number mRNAs (e.g., one to two copies/cell) will not be detected by the GeneChip assay or any other microarray approaches, using the current protocol for amplification. Although it is possible to use a normalized inner ear cDNA library as the source for aRNA synthesis, that approach would not be quantitatively representative of the expression level. Third, microarray technology relies on using genes or sequences already discovered. Although the genome project has made tremendous progress and predicted genes should be incorporated on chips eventually, genes from unique tissues like inner ear may be underrepresented. Finally, the measurement of expression level is not absolutely quantitative. One way around the last two problems is to use SAGE (serial analysis of gene expression) to study the expression level, as it does not depend on previous knowledge of sequence information and is quantitative (Velculescu et al., 1995). However, because it is very laborious and costly, SAGE can only be used for a limited number of samples. The combination of microarray and SAGE analysis may help to identify most of the rare transcripts.

An On-Line Database

Broad dissemination of data generated from microarray studies is important both for maximizing utilization, and for understanding coexpression with cluster methods. We expect to have all the data from our studies presented in a Web-accessible public database. Towards that objective we have set up a public database with some data sets, which includes links to other public databases such as UniGene and MGI (http://www.mgh.harvard.edu/depts/coreylab/index.html). We hope this will enhance research in hearing and deafness.

REFERENCES

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM, et al. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503–511.

Baugh LR, Hill AA, Brown EL, Hunter CP. 2001. Quantitative analysis of mRNA amplification by in vitro transcription. Nucleic Acids Res 29:E29.

Bermingham NA, Hassan BA, Price SD, Vollrath MA, Ben-Arie N, Eatock RA, Bellen HJ, Lysakowski A, Zoghbi HY. 1999. Math1: an essential gene for the generation of inner ear hair cells. Science 284:1837–1841.

Blackshaw S, Fraioli RE, Furukawa T, Cepko CL. 2001. Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes. Cell 107:579–589.

Britsch S, Goerich DE, Riethmacher D, Peirano RI, Rossner M, Nave KA, Birchmeier C, Wegner M. 2001. The transcription factor Sox10 is a key regulator of peripheral glial development. Genes Dev 15:66–78.

Brown PO, Botstein D. 1999. Exploring the new world of the genome with DNA microarrays. Nat Genet 21:33–37.

Chen Z-Y, Corey DP. 2001. An inner ear gene expression database. J Assoc Res Otolaryngol 3:140–148.

Chen ZY, Denney RM, Breakefield XO. 1995. Norrie disease and MAO genes: nearest neighbors. Hum Mol Genet 4:1729–1737.

Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman D, Lockhart DJ, Davis RW. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2:65–73.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863–14868.

Erkman L, McEvilly RJ, Luo L, Ryan AK, Hooshmand F, O'Connell SM, Keithley EM, Rapaport DH, Ryan AF, Rosenfeld MG. 1996. Role of transcription factors Brn-3.1 and Brn-3.2 in auditory and visual system development. Nature 381:603–606.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. 1999. Molecular

classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286:531–537.

Gong TW, Hegeman AD, Shin JJ, Adler HJ, Raphael Y, Lomax MI. 1996. Identification of genes expressed after noise exposure in the chick basilar papilla. Hear Res 96:20–32.

Heller S, Sheane CA, Javed Z, Hudspeth AJ. 1998. Molecular markers for cell types of the inner ear and candidate genes for hearing disorders. Proc Natl Acad Sci USA 95:11400–11405.

Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. Cell 95:717–728.

Hughes JD, Estep PW, Tavazoie S, Church GM. 2000. Computational identification of *cis*-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. J Mol Biol 296:1205–1214.

Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson J Jr, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO. 1999. The transcriptional program in the response of human fibroblasts to serum [see comments]. Science 283:83–87.

Jourdan-Le Saux C, Tronecker H, Bogic L, Bryant-Greenwood GD, Boyd CD, Csiszar K. 1999. The LOXL2 gene encodes a new lysyl oxidase-like protein and is expressed at high levels in reproductive tissues. J Biol Chem 274: 12939–12944.

Kalluri R, Gattone VH II, Hudson BG. 1998. Identification and localization of type IV collagen chains in the inner ear cochlea. Connect Tissue Res 37:143–150.

Kelley MW, Xu XM, Wagner MA, Warchol ME, Corwin JT. 1993. The developing organ of Corti contains retinoic acid and forms supernumerary hair cells in response to exogenous retinoic acid in culture. Development 119: 1041–1053.

Killick R, Legan PK, Malenczak C, Richardson GP. 1995. Molecular cloning of chick β-tectorin, an extracellular matrix molecule of the inner ear. J Cell Biol 129:535–547.

Lander ES. 1999. Array of hope. Nat Genet 21:3–4.

Lee CK, Klopp RG, Weindruch R, Prolla TA. 1999. Gene expression profile of aging and its retardation by caloric restriction [see comments]. Science 285:1390–1393.

Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. 1999. High density synthetic oligonucleotide arrays. Nat Genet 21:20–24.

Luo L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR, Erlander MG. 1999. Gene expression profiles of laser-captured adjacent neuronal subtypes [published erratum appears in Nat Med 1999 Mar;5(3):355]. Nat Med 5:117–122.

Mahadevappa M, Warrington JA. 1999. A high-density probe array sample preparation method using 10- to 100-fold fewer cells. Nat Biotechnol 17:1134–1136.

Pandey A, Mann M. 2000. Proteomics to study genes and genomes. Nature 405:837–846.

Pingault V, Bondurand N, Kuhlbrodt K, Goerich DE, Prehu MO, Puliti A, Herbarth B, Hermans-Borgmeyer I, Legius E, Matthijs G, Amiel J, Lyonnet S, Ceccherini I, Romeo G, Smith JC, Read AP, Wegner M, Goossens M. 1998. SOX10 mutations in patients with Waardenburg-Hirschsprung disease. Nat Genet 18:171–173.

Quackenbush J. 2001. Computational analysis of microarray data. Nat Rev Genet 2:418–427.

Rau A, Legan PK, Richardson GP. 1999. Tectorin mRNA expression is spatially and temporally restricted during mouse inner ear development. J Comp Neurol 405:271–280.

Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, He YD, Dai H, Walker WL, Hughes TR, Tyers M, Boone C, Friend SH. 2000. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science 287: 873–880.

Roth FP, Hughes JD, Estep PW, Church GM. 1998. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat Biotechnol 16:939–945.

Shailam R, Lanford PJ, Dolinsky CM, Norton CR, Gridley T, Kelley MW. 1999. Expression of proneural and neurogenic genes in the embryonic mammalian vestibular system. J Neurocytol 28:809–819.

Sharon D, Blackshaw S, Cepko CL, Dryja TP. 2002. Profile of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium determined through serial analysis of gene expression (SAGE). Proc Natl Acad Sci USA 99:315–320.

Skvorak AB, Weng Z, Yee AJ, Robertson NG, Morton CC. 1999. Human cochlear expressed sequence tags provide insight into cochlear gene expression and identify candidate genes for deafness. Hum Mol Genet 8:439–452.

Slepecky NB, Cefaratti LK, Yoo TJ. 1992a. Type II and type IX collagen form heterotypic fibers in the tectorial membrane of the inner ear. Matrix 12:80–86.

Slepecky NB, Savage JE, Yoo TJ. 1992b. Localization of type II, IX and V collagen in the inner ear. Acta Otolaryngol 112:611–617.

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell 9:3273–3297.

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrousky E, Lander ES, Golub TR. 1999. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA 96:2907–2912.

Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. 1999. Systematic determination of genetic network architecture. Nat Genet 22:281–285.

Thalmann I, Thallinger G, Comegys TH, Thalmann R. 1986. Collagen—the predominant protein of the tectorial membrane. ORL J Otorhinolaryngol Relat Spec 48:107–115.

Tsuji A, Torres-Rosado A, Arai T, Le Beau MM, Lemons RS, Chou SH, Kurachi K. 1991. Hepsin, a cell membrane-associated protease. Characterization, tissue distribution, and gene localization. J Biol Chem 266:16948–16953.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. 1995. Serial analysis of gene expression. Science 270:484–487.

Verpy E, Leibovici M, Petit C. 1999. Characterization of otoconin-95, the major protein of murine otoconia, provides insights into the formation of these inner ear biominerals. Proc Natl Acad Sci USA 96:529–534.

Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM. 2000. High-fidelity mRNA amplification for gene profiling. Nat Biotechnol 18:457–459.

Wang Y, Kowalski PE, Thalmann I, Ornitz DM, Mager DL, Thalmann R. 1998. Otoconin-90, the mammalian otoconial matrix protein, contains two domains of homology to secretory phospholipase A2. Proc Natl Acad Sci USA 95:15345–15350.

White KP, Rifkin SA, Hurban P, Hogness DS. 1999. Microarray analysis of *Drosophila* development during metamorphosis. Science 286:2179–2184.

Xiang M, Gan L, Li D, Chen ZY, Zhou L, O'Malley BW Jr, Klein W, Nathans J. 1997. Essential role of POU-domain factor Brn-3c in auditory and vestibular hair cell development. Proc Natl Acad Sci USA 94:9445–9450.

Xiang M, Gao WQ, Hasson T, Shin JJ. 1998. Requirement for Brn-3c in maturation and survival, but not in fate determination of inner ear hair cells. Development 125: 3935–3946.

Zheng JL, Gao WQ. 2000. Overexpression of Math1 induces robust production of extra hair cells in postnatal rat inner ears. Nat Neurosci 3:580–586.

Zheng JL, Shou J, Guillemot F, Kageyama R, Gao WQ. 2000. Hes1 is a negative regulator of inner ear hair cell differentiation. Development 127:4551–4560.

Zheng L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW. 1997. Gene expression profiles in normal and cancer cells. Science 276:1268–1272.