Defining the Human Macula Transcriptome and Candidate Retinal Disease Genes Using *EyeSAGE*

Catherine Bowes Rickman,[1,2,3] *Jessica N. Ebright,*[1,3] *Zachary J. Zavodni,*[1] *Ling Yu,*[1] *Tianyuan Wang,*[4] *Stephen P. Daiger,*[5] *Graeme Wistow,*[6] *Kathy Boon,*[7] *and Michael A. Hauser*[4]

PURPOSE. To develop large-scale, high-throughput annotation of the human macula transcriptome and to identify and prioritize candidate genes for inherited retinal dystrophies, based on ocular-expression profiles using serial analysis of gene expression (SAGE).

METHODS. Two human retina and two retinal pigment epithelium (RPE)/choroid SAGE libraries made from *matched* macula or midperipheral retina and adjacent RPE/choroid of morphologically normal 28- to 66-year-old donors and a human central retina longSAGE library made from 41- to 66-year-old donors were generated. Their transcription profiles were entered into a relational database, *EyeSAGE*, including microarray expression profiles of retina and publicly available normal human tissue SAGE libraries. *EyeSAGE* was used to identify retina- and RPE-specific and -associated genes, and candidate genes for retina and RPE disease loci. Differential and/or cell-type specific expression was validated by quantitative and single-cell RT-PCR.

RESULTS. Cone photoreceptor-associated gene expression was elevated in the macula transcription profiles. Analysis of the longSAGE retina tags enhanced tag-to-gene mapping and revealed alternatively spliced genes. Analysis of candidate gene expression tables for the identified Bardet-Biedl syndrome disease gene (*BBS5*) in the BBS5 disease region table yielded *BBS5* as the top candidate. Compelling candidates for inherited retina diseases were identified.

CONCLUSIONS. The *EyeSAGE* database, combining three different gene-profiling platforms including the authors' multidonor-derived retina/RPE SAGE libraries and existing single-donor retina/RPE libraries, is a powerful resource for definition of the retina and RPE transcriptomes. It can be used to identify retina-specific genes, including alternatively spliced transcripts and to prioritize candidate genes within mapped retinal disease regions. (*Invest Ophthalmol Vis Sci.* 2006;47:2305–2316) DOI: 10.1167/iovs.05-1437

The macula is a highly specialized region of the primate retina that contains the highest densities of rod and cone photoreceptors, as well as the second-order neurons and ganglion cells that receive input from the photoreceptors. Because it encompasses the cone-enriched fovea, which mediates high-acuity central vision, diseases affecting the macula have devastating sight-related consequences. In fact, age-related macular degeneration (AMD) is the leading cause of untreatable new vision loss in the elderly, affecting more than 14 million patients in the United States alone.[1–3] Macular degenerations are characterized by dysfunction and, ultimately, the death of rod and cone photoreceptors and the adjacent retinal pigment epithelium (RPE). The mechanism of cone loss is poorly understood, but it appears that foveal cones, in particular, have a capacity for survival in a region of the retina with a propensity for degeneration.[4–7] The macula also contains the highest concentration of retinal ganglion cells (RGCs), whose death is responsible for the vision loss associated with glaucoma. We hypothesize that differential gene expression contributes to the establishment of the unique macular microenvironment and is crucial to maintaining the health of this region. Expression profiling of the macula should reveal genes that may contribute to disease susceptibility and others that enhance photoreceptor survival.

Gene expression profiling in the human macula is complicated by the technical difficulties inherent in isolating RNA from human donor tissue and is further confounded by the presence of pigmented RPE cells (for an excellent review of these technical hurdles, please see Chowers et al.[8]). Recent technical advances in large-scale, RNA-based technologies and analysis strategies (e.g., microarray technologies and serial analysis of gene expression, or SAGE[9]), have vastly increased the extent of the transcriptome that can be uncovered from small amounts of starting material.

SAGE is a powerful technique that provides quantitative and comprehensive gene expression profiling.[9] Conventional or shortSAGE uses short 14-bp tags of internal transcript signatures[9] to identify and quantify individual gene transcripts. We describe the construction of four new shortSAGE libraries representing topographic regions of human retina and RPE/choroid, as well as the first retinal longSAGE library. LongSAGE produces 21-bp transcript tags that, owing to its increased tag length, can be used for direct assignment to genome sequences and for identification of novel genes and alternative transcripts.[10,11] Analysis of the human retina and RPE transcriptomes has been approached using a wide array of large-scale expression profiling methodologies (please see Refs. 12,13 for reviews), including SAGE[14] and microarrays[8,15–17] used here and suppression subtraction hybridization.[12] These studies have been instrumental in developing knowledge of the retina and RPE transcriptomes. To enhance the utility of these data

FIGURE 1. Posterior eyecup of a human donor eye. Right eye obtained from a 64-year-old white male donor within 4 hours of death and stored in RNA preservative. The donor had no ocular history of disease. Macular (M) and midperipheral (P) regions, from which 4-mm diameter punches were taken.

we created *EyeSAGE,* a relational database for the analysis and presentation of human retina and RPE/choroid SAGE and microarray expression profiles, compared with transcript expression in the human body or brain. The unique strength of the *EyeSAGE* database is that by examining the tissue-specific patterns of gene expression, we are able to identify sets of transcripts that are characteristic of subpopulations of cells within the retina, including cell populations that would be difficult or impossible to isolate physically. By combining these cell-type–specific expression profiles with the results of genomic linkage studies, we have identified candidate genes for a variety of ocular genetic disorders. The *EyeSAGE* database is posted at the National Eye Institute's NEIBank Web site (http://neibank. nei.nih.gov/index.shtml/ provided in the public domain by the National Eye Institute, Bethesda, MD) and the candidate retinal disease gene expression tables are also available at our Web site (http://www.duke.edu/~bowes007/*EyeSAGE*.htm/ provided in the public domain by Duke University, Durham, NC), as well as through RetNet (http://www.sph.uth.tmc.edu/Ret-Net/ provided in the public domain by the University of Texas Houston Health Science Center, Houston, TX).

MATERIALS AND METHODS

Tissue Procurement, Processing, and RNA Extraction

Human donor eyes used in these studies were obtained from the North Carolina Eye Bank (Winston-Salem, NC) within 6 hours of death (with an average procurement time of less than 4 hours) and stored in an RNA stabilizer (RNA*later*; Ambion, Austin, TX). Donors had no history of ocular disease and no premorbid life support, which has been demonstrated to reduce the RNA quality.[18] The posterior eyecups were morphologically normal. Human retina SAGE libraries were constructed using RNA isolated from pooled 4-mm diameter punches of the macula (encompassing the fovea) and 4-mm diameter punches of pooled midperipheral retina from the *same* five donors designated, *4Mac* and *4Peri,* respectively (Fig. 1, Table 1).

TABLE 1. Human Donor Eyes Used for SAGE Libraries and Microarray Analysis

Donor	Age (y)	Gender	Death to Procurement (hr:min)	Cause of Death
4Mac and *4Peri* Retina library				
0904-01*	28	F	3:30	Brain death
0265-02	32	F	3:25	End-stage renal disease
0842-01*	39	F	2:00	Intracranial bleeding
0177-02	56	F	3:20	Breast cancer
0852-01*	66	F	1:30	Intracranial bleeding
4MacRPE and *4PeriRPE* RPE/ choroid library				
1318-03	40	M	3:28	Muscular dystrophy
1432-03	42	M	4:20	Intracranial bleeding
1236-03	44	F	5:46	Cerebrovascular accident
1327-03	47	F	3:03	Breast cancer
1401-03	59	M	4:30	Bronchogenic cancer
4cRet central retina long SAGE library				
0448-04	66	F	5:50	Subarachnoid hemorrage
0481-04	51	M	4:40	Acute renal failure
0246-03	63	M	5:55	Myocardial infarction
1413-04	51	M	3:30	Sepsis, small bowel obstruction
1416-04	41	M	4:50	Sepsis, D/T bacteremia UTI infection
1515-04	57	M	4:20	Liver cancer
1517-04	60	F	4:00	Renal failure
Summary	28–66		3:56	

4Mac, standard SAGE library derived from RNA isolated from pooled 4-mm diameter punches of the macula; *4Peri,* standard SAGE library derived from RNA isolated from pooled 4-mm diameter punches of mid-peripheral retina; *4MacRPE,* standard SAGE library derived from RNA isolated from pooled 4-mm diameter punches of the macula RPE/choroid; *4PeriRPE,* standard SAGE library derived from RNA isolated from pooled 4-mm diameter punches of the mid-peripheral RPE/choroid; *4cRet* longSAGE library derived from RNA isolated from pooled 4-mm diameter punches of the macula+mid-peripheral retina or central retina.

* Donor eyes used for SAGE and for microarray analysis, others used exclusively for SAGE.

TABLE 2. Human Eye SAGE Tag Counts

Library Name*	Total Tags†	Unique Tags‡	Unique Tags Minus Singles	Tags Not Reliably Matched to a UniGene Cluster§	Not Matched to Cluster Minus Singles	Unique to Library	Unique to Library Minus Singles
4MAC SAGE	101420	37556	10885	11408	1426	12894	632
4PERI SAGE	108383	39831	12260	10601	1402	13339	808
4MACRPE SAGE	77289	35351	8978	10356	1059	12394	569
4PERIRPE SAGE	114428	39527	17799	13017	3887	15826	4108
HPR1 SAGE (PeriB1)‖	59661	22812	6084	6116	681	6714	261
HPR2 SAGE (PeriB2)‖	105312	34066	10867	10051	1697	10539	705
HMAC2 SAGE‖	102417	33522	10296	10081	1597	10699	674
HRPE1 SAGE (RPEB1)‖	53666	38358	4877	18081	1844	26979	2782
Total human Retina/RPE short tags	722,576	160,723	61,878	68,394			
4cRET lSAGE (longSAGE)	98,408	58,707		47,381			

* See Table 1 for explanation of SAGE library names.
† Total number of tags counted, indicating the size of the library.
‡ Number of different tag sequences occurring in each library.
§ Number of tags in each library that could not be assigned a "Best Gene" match according to SAGE Genie.
‖ Sharon et al.[14]

Human RPE/choroid shortSAGE libraries were made from 4-mm diameter punches of macular RPE/choroid (*4MacRPE*) and *matching* 4-mm diameter punches of midperipheral RPE/choroid (*4PeriRPE*), taken from five donor eyes (Table 1). The longSAGE central retina library, *4cRET* was synthesized from RNA isolated from 4-mm diameter single punches of the macula and midperipheral retina from seven donors (Table 1).

Retina-derived RNA was isolated in Trizol (Invitrogen, Carlsbad, CA) plus glycogen and quantified (RiboGreen; Invitrogen, Eugene, OR).[19] RPE/choroid-derived RNA was isolated using the same methods and further treated to remove visible melanin contamination as previously described.[20] RPE-enriched RNA was prepared from RPE cells carefully brushed off the posterior cup and processed as previously described.[20] RNA quality was verified on a 0.8% agarose gel and by real-time quantitative RT-PCR (qRT-PCR) analysis of known tissue-specific genes.[19,21]

Synthesis and Analysis of SAGE Libraries

ShortSAGE libraries were constructed from 10 μg RNA, using *Nla*III as the anchoring enzyme and standard methodologies.[22] SAGE libraries were sequenced at Agencourt Bioscience (Beverly, MA) to a depth of approximately 100,000 tags per library (Table 2). The *4cRET* library was constructed with a longSAGE kit (I-SAGE; Invitrogen) and sequenced at Agencourt to a depth of 98,408 tags.

The SAGE 2000 software, version 4.12 (available at http://www.sagenet.org), was used to extract and tabulate SAGE tags. Mitochondrial and repetitive tags were removed and the best gene match for each reliable tag was assigned using resources available on the Cancer Genome Antamony Project (CGAP) SAGE Genie Web site (http://cgap.nci.nih.gov/SAGE).[23] Specifically, SAGE Genie's "best gene for the tag" table was used to match each long tag to its best UniGene cluster match. This is, in most cases, a nonredundant assignment. UniGene clusters were mapped to the human genomic assembly, as previously described.[24] Tag sequences, tag counts, and gene associations were stored in a relational database (Microsoft Access, Redmond, WA) for subsequent analysis and generation of the *EyeSAGE* database. Detailed SAGE library information and tag counts for the retina and RPE/choroid libraries are posted at CGAP's SAGE Genie Web site in conjunction with this study. Tag counts were normalized to 200,000 tags per library (according to the convention used to display tag counts at the SAGE Genie Web site).

cDNA Microarray-Based Expression Profiling

Total RNA was isolated from 4-mm trephine punches of pooled human maculas and 4-mm trephine punches of pooled midperipheral retina from the same donors used for the retina SAGE libraries (Table 1) and used to probe a human UniGene 1 LifeArray (UniGene: 8466 unique genes; Incyte Genomics, Palo Alto, CA). RNA (5 μg) from each retina region was submitted to Incyte for T7 amplification and array hybridization. To correct for variations in data, the average signal from all elements in the Cy3 channel (the Macula-derived probe) was divided by the average signal from all elements in the Cy5 (the peripheral retina-derived probe) channel, resulting in the balance coefficient. The Cy5 signal for each element was then multiplied by the balance coefficient, before calculating the balanced differential expression ratio. The balanced, differential expression ratio was calculated as Cy3/Cy5 if the Cy3 signal was greater, reported as a positive number, or Cy5/Cy3 if the Cy5 signal was greater, reported as a negative number. According to Incyte, a balanced, differential expression ratio greater than 1.7 (or less than −1.7) can be considered differentially expressed with 99% confidence. cDNA from the macula was labeled with Cy3 and cDNA from the peripheral retina, with Cy5, so that negative values indicate preferential expression in the peripheral retina and positive values in the macula. Array controls included sensitivity controls (ranging from 2 to 2000 pg); variable ratios of labeled cDNA to control for preferential labeling with dye, housekeeping genes, for which there were sufficient signal levels and no differential expression for ribosomal S9, tubulin, and 23-kDa *HBP*, and buffer-only array spots to control for background hybridization, all performed in quadruplicate.

Real-Time qRT-PCR

Total RNA extraction, cDNA synthesis and qRT-PCR (using intron-spanning primers) was performed as previously described.[19,21] Real-time quantitation of candidate mRNAs normalized to an endogenous reference(s) (i.e, β-actin [*ACTB*] in retina or glyceraldehyde-3-phosphate dehydrogenase [*GAPDH*], β-2-microglobulin [*B2M*] and ubiquitin C [*UBC*] in RPE samples[25]) was performed on a sequence-detection system (iCycler iQ; Bio-Rad) using SYBR-Green. The *x*-fold difference between candidate genes were normalized to a single endogenous control gene or to several that were then geometrically averaged[25] and calculated by the comparative threshold cycle (C_T) method ($2^{-\Delta\Delta Ct}$).[26] PCR primer sequences for each gene analyzed are available on request.

Photoreceptor Enrichment

The photoreceptor layer of human donor eyes was isolated using the sandwich method, as described by Nishizawa et al.[27] with the following modifications. A 7.5-mm diameter punch from over the macula or central peripheral retina through the sclera was removed and placed retina-side-down onto a PBS-soaked piece of Whatman filter paper. Starting from the edge, the sclera was grasped and pulled gently. The retina remained attached, ganglion cell-side-down, to the filter paper. A piece of 0.2 μm nitrocellulose membrane was then placed directly on top of the retina. The resulting sandwich was then inverted, and the filter paper was carefully peeled away, leaving the retina intact on the nitrocellulose paper. To split the retina, a piece of dry filter paper was placed directly onto the ganglion cell retina and firmly pressed, and then the two sides were pulled apart.

Using a sterile 6 or 4 mm trephine punch, a central punch was collected from both the nitrocellulose (photoreceptor layer attached) membrane and filter paper (inner retina attached) yielding four separate in situ cell samples (two per punch): macular photoreceptor layer and macular inner retina or peripheral photoreceptor layer and peripheral inner retina. For each pair of donor eyes, one eye was prepared using the sandwich method, and the other was prepared in the same manner as the tissue used for preparing the SAGE libraries in which whole 6- or 4-mm trephine punches of retina from over the macula and periphery were collected. Total RNA was isolated from the resultant six tissue samples, DNased, and cDNAs were synthesized with a cDNA synthesis kit (iScript; Bio-Rad, Hercules, CA). The efficacy of these mechanical cell separations was analyzed by qRT-PCR, which was used to compare quantitatively the expression levels of cone-, rod-, and inner retina–localized genes between the different tissue preparations.

Detection of Cone Transcripts with Single Photoreceptor Cell RT-PCR

Two-millimeter diameter punches of macular retina were isolated from human donor eyes stored at 4°C (RNAlater; Ambion) and placed in PBS. Each punch was gently triturated with a wide-bore pipette to float off individual cells. Single cone or rod photoreceptor cells were isolated on a micromanipulator (TransferMan NK2; Eppendorf, Westbury, NY) mounted on a microscope (Diaphot 200; Nikon, Tokyo, Japan) based on their visual rodlike phenotype.[28] Five microliters per well of cDNA mix described by McHeyzer-Williams et al.[29] was placed in a 72-well low-profile plate (Scientific; Robbins, Sunnyvale, CA). Captured single cells were ejected into the cDNA mix, one cell per well, with one in six wells receiving no cells and processed as negative controls. After cDNA synthesis at 37°C in an incubator for 90 minutes, the plates were stored at −80°C. Nested primer sets were designed to amplify mRNA transcripts without genomic amplification by spanning an intron. The cDNA reaction from each cell was split in half and placed in parallel first-round RT-PCR reactions in which primers for the cone-specific gene (*PDE6C*) were mixed with primers for a candidate cone gene in one tube, and primers that amplify the rod-specific *PDE6A* were combined with the candidate primers in the other. A second round of RT-PCR was performed with 1 μL of the first-round product as the template for the reaction with nested primer pairs for the genes amplified in the first round. Products were visualized on a 3.5% acryl agarose gel.

RESULTS

SAGE Libraries of Adult Human Retina and RPE/Choroid

ShortSAGE was performed to obtain comprehensive, genome-wide expression profiles from topographically specific (macula and midperiphery) 4-mm diameter regions of the human retina and adjacent RPE and choroid. Because these libraries were constructed from paired 4-mm punches from five donor eyes, true tissue-specific differences can be detected while minimiz-

ing donor-specific background. A combined macular and peripheral *4cRET* longSAGE library was generated to facilitate genome annotation of the shortSAGE libraries. Nearly half a million tags were obtained with an average of 100,000 tags per library. A summary of the tag numbers obtained from all five libraries is shown in Table 2, along with the tag counts of four previously published human posterior eye shortSAGE libraries.[14] Counting all eight shortSAGE libraries together, there are 160,723 unique shortSAGE tags in the retina/RPE transcriptome after removing mitochondrial and repetitive tags (Table 2). There are 61,878 tags present more than once in one shortSAGE library (two or more nonnormalized counts). In the total database, there are 37,868 unique UniGene clusters (23,959 expressed in the Sharon et al. [14] libraries and 26,303 expressed in ours). Only 12,394 of the clusters are expressed in libraries from both laboratories, which may reflect bias arising from the use of tissue from single donors as suggested by Blackshaw et al.[30] There are 68,394 short tags in the *EyeSAGE* database that lack reliable mapping to a UniGene cluster (i.e, tags may match to the sequence of a clone with an accession number or match sequence found in several UniGene clusters so they cannot be reliably assigned[23]) but may well represent alternative transcripts for a gene with a unique UniGene cluster number or novel eye transcripts.[31] These tags form the pool for gene discovery in the eye. The complete tag counts for all these libraries are posted on the CGAP SAGE Genie Web site, where they can be accessed for downloading or online analysis.[23]

Building the *EyeSAGE* Database

One goal of this study was to generate a comprehensive picture of gene expression in the human macula that is accurate, readily accessible, and can be used as a resource to identify and quantitate cell-type–specific or –associated genes. To this end we integrated large-scale expression data obtained from this tissue, by using different technologies: SAGE, longSAGE, and cDNA microarrays into a database that we named *EyeSAGE*. Starting with the short tag retina and RPE/choroid SAGE libraries summarized in Table 2, 160,723 unique tags were used as the first building block (column) for the *EyeSAGE* database. Each tag was analyzed and assigned a best gene match,[23] and UniGene cluster assignment (based on NCBI Build 182) if available. Genomic map positions (as nucleotide numbers along the chromosome) were assigned as previously described.[24] Columns of tag counts normalized to 200,000 for each tag in each posterior eye library were added. The *4cRET* longSAGE library was incorporated by matching the longSAGE tags with their reliable best gene matches, based on CGAP's SAGE Genie assignments, to the short tag, the sequence of which is the first 10 bases of the 17-bp tag. Next the tag counts for each tag in 39 additional normal tissue SAGE libraries (available at SAGE Genie) were added to incorporate expression information from a variety of tissue and cell types. Incyte cDNA microarray expression data of peripheral and macular retina were imported and linked to the SAGE data after using BLAT homology searches to assign a UniGene cluster number (Build 182) to each microarray probe. Using the convention at CGAP's SAGE Genie (http://cgap.nci.nih.gov/SAGE/AnatomicViewer) the SAGE libraries were normalized to 200,000 tags for pair-wise comparisons. The entire *EyeSAGE* database in Access was sorted by tag number and genes with expression (tags) totaling five or more (normalized to 200,000 therefore totaling approximately two or more raw tag counts/library), in the eight posterior-eye shortSAGE libraries combined, were exported into spreadsheet software (Excel; Microsoft; the entire Microsoft Access version of *EyeSAGE* is available on request). This step removes unique tags that occur as singletons in only

one retina or RPE/choroid library. This version of the *EyeSAGE* database was used for subsequent data mining (available at NEI-Bank, http://neibank.nei.nih.gov/index.shtml). The *EyeSAGE* database is an easily searchable, comprehensive expression dataset representing the posterior eye transcriptome. In its current form, *EyeSAGE* can be used to analyze tissue and cell-type expression of single genes or classes of genes, or to display ocular expression over user-defined genomic regions. It can also be mined to generate large-scale views of cell-type expression. Examples of specific queries follow.

Generating Cell-Type–Associated Gene Lists

We were particularly interested in examining cone-photoreceptor–associated gene expression. To derive a cone-associated profile using *EyeSAGE* we took advantage of the fact that cone photoreceptors are concentrated in the macula and that cone-specific transcripts should be present at higher levels in the retina than in other neural tissues. In contrast, inner retina neuron and glial cell-associated genes elevated in macula might also be found in the brain but not in non-neural tissues. These observations were translated into the following set of queries of the *EyeSAGE* database to generate a list of putative cone-enriched transcripts (*EyeSAGE* column heading is given in quotes and described in Table 3, legend):

1. Tag counts in the four retina shortSAGE libraries average three or more ("Retina Ave" $\geq$ 3).
2. Tag count is higher in macula-derived libraries than in peripheral retina libraries ("Mac/Peri" $\geq$ 1.1).
3. Total tag counts/genes in retina are higher than in brain libraries ("Retina/Neural" $>$ 1.2).

There are 270 transcripts in the *EyeSAGE* database that satisfy all these criteria and were therefore considered to represent candidate cone-photoreceptor–associated transcripts (Table 3 and Supplementary Table S1; all Supplementary Tables are online at http://www.iovs.org/cgi/content/full/47/6/2305/DC1). If one more condition is imposed to look for cone photoreceptor-enriched transcripts:

4. Average of tag count is significantly higher in the retina libraries than in the sum of tag counts in libraries from the rest of the body ("Retina Ave/Ave Body and Neural" $\geq$ 3)

the number of transcripts is reduced to 38 (see Table 3 for the top 20). Similar kinds of selection criteria can be used to generate lists of transcripts that are specific to or enriched in other cell types. In each case, the presence of known cell-specific genes was used as a reference to gauge the success of given parameters to return the desired cell-associated expression. A list of rod photoreceptor cell–associated transcripts was identified by selecting for genes with higher expression in peripheral retina than the macula and higher expression in the retina than the rest of the body (see Supplementary Table S2 online). RPE-enriched genes were identified by selecting for higher expression in the three RPE-derived SAGE libraries compared with retina and higher expression in the RPE than the rest of the body (Supplementary Table S3 online). Finally, a list of putative ganglion cell- and inner retina-associated transcripts were identified based on a query for tags with higher expression in the macula than in the peripheral retina and a higher macula-to-periphery tag count ratio in our 4-mm punch-derived libraries (*4Mac/4Peri*) than in the 6 mm punch-derived retina libraries (*HMac2/PeriB2*). This parameter was imposed because second-order neurons and ganglion cells are concentrated in the primate macula.[32] In addition, a requirement for higher tag counts in the other neural tissues but not in the rest of the body was used, because it is expected that inner retina–associated gene expression overlaps significantly with brain-

expression profiles (see Supplementary Table S4 online). A search for tags with counts totaling more than 15 combined in the retina libraries and average expression greater in retina than in all other represented tissues returned a list of more than 1000 tags for genes with highest expression in the retina (see Supplementary Table S5 online).

Validation of the *4Mac* and *4Peri* Retina-Derived SAGE Expression Profiles

The differential gene expression patterns revealed by SAGE analysis were validated with several approaches. Digital comparison of gene expression, as tag counts, of our shortSAGE retina libraries (*4Mac* and *4Peri*) to the longSAGE *4cRET* library and to the retina and RPE shortSAGE libraries generated by Sharon et al.[14] provided a qualitative assessment of differential expression of any genes found in all three of these groups. cDNA microarrays were probed with the same RNA used to generate the *4Mac* and *4Peri* retina SAGE libraries (see Table 1 for donors). The resultant microarray expression profiles (5836 genes out a total of 8466 on the array) were related by UniGene cluster number to the SAGE profiles in *EyeSAGE* and found to match fairly consistently the gene expression given by SAGE analysis. For example, 33 of the top 100 rod-associated genes were present on the array and, of these, 27 (82%) showed the expected higher expression in the peripheral retina (see Supplementary Table S2 online).

To validate more rigorously the differential expression detected by SAGE and (when available) microarray analysis, expression of several known retina-specific genes and candidate cell-associated genes in the macula and midperipheral retina was determined by qRT-PCR. Rod-specific, *PDE6A* (phosphodiesterase 6A, cGMP-specific, rod, alpha; GeneID: 5145), cone-specific, *PDE6C* (phosphodiesterase 6C, cGMP-specific, cone, alpha prime; GeneID: 5146). *GNAT2* (guanine nucleotide binding protein [G protein], α-transducing activity polypeptide 2; GeneID: 2780), and ganglion cell-associated *THY1* (Thy-1 cell surface antigen; GeneID: 7070) expression was compared with expression of selected candidate cone-associated genes as well as one candidate inner retina-associated gene, *UCHL1* (ubiquitin carboxyl-terminal esterase L1; GeneID: 7345; Fig. 2). In each case, this independent analysis verified the differential gene expression seen in the SAGE libraries. To further localize expression of selected cone-associated genes, qRT-RCR was performed on RNA isolated from the photoreceptor layer and inner retina of macula and peripheral retina (Fig. 3). The highest expression for the known cone photoreceptor gene, *GNAT2*, and candidate cone-associated genes, *HR* (hairless homologue [mouse]; GeneID: 55806) and *CPLX4* (complexin 4; GeneID: 225644) was detected in the macula photoreceptor layer-derived RNA containing the highest concentration of cone-derived transcripts. This provides strong evidence that the *HR* and *CPLX4* genes are transcribed in cones.

Single-cell PCR was used to confirm cone-associated expression for some relatively abundant candidate cone-associated genes. *SH3BGRL2* (SH3 domain binding glutamic acid-rich protein like 2; GeneID: 83699) was among the top 20 genes showing enrichment in cone photoreceptors (Table 3). *SH3BGRL2* maps within the 6q linkage region for several mapped but not yet identified retinal dystrophies including LCA5, MCDR1, and BCMAD (RetNet). Elevated expression of *SH3BGRL2* in the macula was validated by qRT-PCR (Fig. 2). Single photoreceptor cells were isolated with a micromanipulator from 2-mm diameter punches of human retina obtained over the macula (Fig. 4A). RT-PCR performed on RNAs isolated from single human photoreceptors showed coamplification of *SH3BGRL2* with the known cone-specific gene, *PDE6C*, but not in reactions with RNA isolated from photoreceptors in which rod-specific PDE6A was detected (Fig. 4B).

TABLE 3. Twenty Most Abundant Cone Photoreceptor-Associated Genes

Tag Sequence*	Long Tag†	LongSAGE plus Best Gene Match‡	Build 182 UniGene Number§	Symbol‖	Title¶	Location*
AATAAATTAG	AATAAATTAGTAAATGA	ZNF593	544758		Transcribed locus	
AGGCCGCTAG	AGGCCGCTAGCTCTGAG	ARR3	308	ARR3	Arrestin 3, retinal (X-arrestin)	Xcen-q21
CTTTCTCCTT	CTTTCTCCTTTTTCTAC	GNB3	534315	GNB3	Guanine nucleotide binding protein (G protein), beta polypeptide 3	12p13
AGCTCGCTCA	AGCTCGCTCAGTTTGGG	PDE6H	54471	PDE6H	Phosphodiesterase 6H, cGMP-specific, cone, gamma	12p13
TAATAAAATT	TAATAAAATTAAGCAAA	CPLX4	449884	CPLX4	Complexin 4	18q21.32
GCACATAGCA	GCACATAGCACTGTACT		180711		CDNA FLJ30906 fis, clone FEBRA2006055	
TGGACCAGGT	TGGACCAGGTTGGCCTC		541243		Transcribed locus, moderately similar to XP_516951.1	
TGTGTACCTA	TGTGTACCTATCTATCT		535996		Homo sapiens, clone IMAGE: 5276872, mRNA	
ATTTCCATTA	ATTTCCATTAAAGTCTG	HR	272367	HR	Hairless homologue (mouse)	8p21.2
GTCCCTCTCT	GTCCCTCTCTAACACCT		540818		Transcribed locus	
CTGTTGATTT	CTGTTGATTTCAGGTCT	GUCA1C	119693	PCDHB5	Protocadherin beta 5	5q31
TTGCAGGAAA	TTGCAGGAAAACACTCT	FAM19A3	439116	FAM19A3	Family with sequence similarity 19 (chemokine C-C motif-like), member A3	1p13.2
CAGGCTGGTG	CAGGCTGGTGGTGATGG	SIX6	194756	SIX6	Sine oculis homeobox homolog 6 (Drosophila)	14q22.3-q23
ACTGCAACTG	ACTGCAACTGTATTTCA	GNAT2	36973	GNAT2	Guanine nucleotide binding protein (G protein), alpha transducing activity polypeptide 2	1p13.1
TATGTATCCT	TATGTATCCTTAAAGAA	SH3BGRL2	302772	SH3BGRL2/FASH3B	SH3 domain binding glutamic acid-rich protein like 2	6q13-15
TTCCTTGAAG	TTCCTTGAAGCCCATCG	CABP2	278984	CABP2	Calcium binding protein 2	11q13.1
ACTTTTTTCA	ACTTTTTTCAAAAAAAA	CSF1	173894	CSF1	Colony-stimulating factor 1 (macrophage)	1p21-p13
TAAAAACTAC	TAAAAACTACACTTAGG	KIAA1345	292754	CCNT2	Cyclin T2	2q21.3
TGTAAACTTC	TGTAAACTTCACCCCCA	TRH	182231	TRH	Thyrotropin-releasing hormone	3q13.3-q21

* Tag Sequence, Sequence of the 10-bp short SAGE tag.

† Long Tag, Sequence of the 17-bp LongSAGE tag that matches the short tag in Tag Sequence, if there was a match in 4cRET Long.

‡ LongSAGE plus Best Gene Match, UniGene cluster number of the best gene match for the long SAGE tag in Long Tag.

§ Bld 182 UniGene Cluster Number, UniGene cluster number of the best gene match for the short SAGE tag in Tag Sequence.

‖ Symbol, Symbol of the best gene match for the short SAGE tag in Tag Sequence.

¶ Title, Expanded gene name describing the best gene match for the short SAGE tag in Tag Sequence.

Location, Cytogenetic map location.

** Chromosome, Chromosome Number.

†† Start, location of start of gene (in bases counting from p arm telomere) on the chromosome.

‡‡ End, location of end of gene (in bases counting from p arm telomere) on the chromosome.

§§ 4cRET Long, tag counts from LSAGE_Retina_central_normal_B_4cRetlong SAGE library prepared from pooled 4-mm punches human macula & central peri retina from 7 morphologically normal donor eyes.

‖‖ 4MAC, tag counts from SAGE_Retina_Macula_normal_B_4Mac; SAGE library prepared from 5 pooled 4-mm punches of human macula from morphologically normal donor eyes.

¶¶ 4PERI, tag counts from SAGE_Retina_Peripheral_normal_4Peri; SAGE library prepared from 5 pooled 4-mm punches of human peripheral retina from the same donor eyes as 4Mac.

Chromo-some**	Start††	End‡‡	4cRET Long§§	4MAC‖‖	4PERI¶¶	HMAC2##	PeriB2***	PeriB1†††	Eye Sum‡‡‡	Neural Sum§§§	Body Sum‖‖‖
6	2,402,858	2,403,280	10	55	41	94	74	34	382	0	0
X	69,271,175	69,284,103	17	101	85	49	25	44	442	0	0
12	6,824,223	6,828,231	40	164	155	111	70	91	888	33	14
12	15,022,276	15,026,066	28	112	78	21	8	7	360	5	0
18	55,113,633	55,136,861	19	73	35	45	57	10	271	5	85
11	59,326,777	59,329,926	10	43	30	12	13	13	183	5	44
19	52,025,683	52,026,054	2	8	9	16	6	0	44	0	0
X	1,393,488	1,395,116	1	30	2	33	13	3	96	32	13
8	22,028,528	22,045,326	3	41	11	23	6	0	112	38	12
17	41,283,570	41,283,839	1	14	11	12	11	13	96	14	0
5	140,494,983	140,497,888	4	30	13	12	9	13	97	6	27
1	112,981,489	112,981,887	1	18	4	10	6	10	60	10	0
14	60,045,690	60,048,278	2	12	4	12	6	7	54	1	6
1	109,857,755	109,867,747	1	6	2	12	6	0	32	0	0
6	80,397,718	80,470,088	5	63	15	61	17	13	213	56	158
11	67,042,993	67,047,443	1	4	6	10	4	0	32	0	0
1	110,165,498	110,184,403	4	10	0	55	36	0	104	31	96
2	135,509,534	135,548,145	13	28	7	14	9	3	78	17	46
			6	16	11	0	0	10	51	9	3

     ## HMAC2, Tag counts from SAGE_Retina_macula_normal_B_HMAC2; 6 mm macular retina punch from *one* 44-year-old male patient (http://cgap.nci.nih.gov/SAGE/SAGELibInfo?LID=164&ORG=Hs).

   *** PeriB2, Tag counts from SAGE_Retina_peripheral_normal_B_2; 6 mm peripheral retina punch from the same 44-year-old male patient as HMAC2.

     ††† PeriB1, Tag counts from SAGE_Retina_peripheral_normal_B_1; SAGE library prepared from a 6-mm peripheral retina punch from *one* 88-year-old female patient.

     ‡‡‡ Eye Sum, Summation of tag counts from all eight eye short SAGE libraries.

     §§§ Neural Sum, Summation of tag counts from nine publicly archived SAGE libraries prepared from the different regions of the brain [see EyeSAGE table or legend at NEIBAnk for description of each library].

     ‖‖‖ Body Sum, Tag count sum of 30 publicly archived SAGE libraries prepared from various non-neural tissues (see EyeSAGE table or legend at NEIBAnk for description of each library).
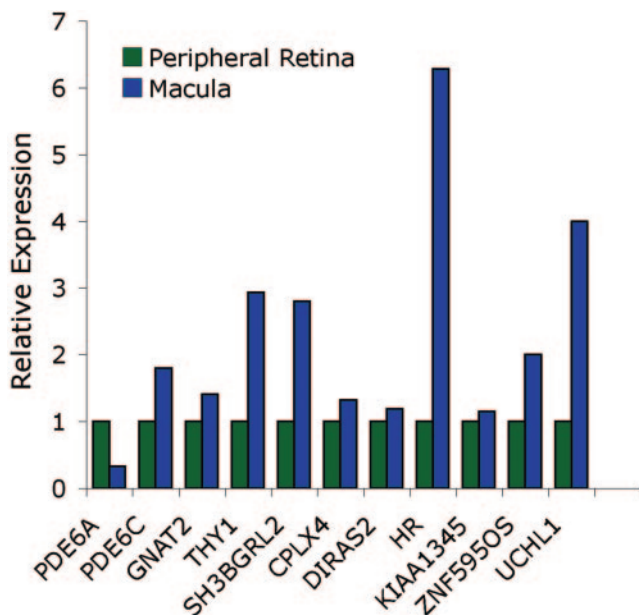
Figure 2. Real-time qRT-PCR summaries of known rod-specific (*PDE6A*), cone-specific (*PDE6C* and *GNAT2*), ganglion cell-associated (*THY-1*), and candidate cone photoreceptor–associated genes (*SH3BGRL2, CPLX4, DIRAS2, HR, KIAA1345, ZNF595OS,* and *UCHL1*) in the macula relative to expression in the peripheral retina.

Validation of the *4MacRPE* and *4PeriRPE* Retina-Derived SAGE Expression Profiles

Analysis of the SAGE profiles obtained from the *4MacRPE* and *4PeriRPE* libraries revealed an unexpectedly high number of retina-specific genes. This "contamination" may arise from tight adhesion of the RPE and photoreceptors, making it very difficult to cleanly separate RPE from retinal tissue. In addition, these libraries were derived from 4-mm trephine punches of RPE/choroid taken over the macula and adjacent central peripheral retina for regional comparisons. Each punch was made through the posterior eyecup with the retina still attached and without going completely through the sclera. The disc of retina was gently lifted off, followed by the disc of RPE/choroid. Retinal tissue could be carried over into the RPE along the
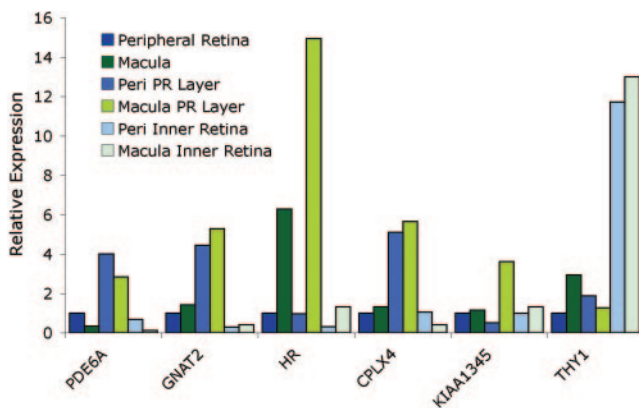
Figure 3. Real-time qRT-PCR summaries of cone-specific (*PDE6C* and *GNAT2*), candidate cone photoreceptor (*HR, CPLX4,* and *KIAA1345*) and ganglion cell-associated (*THY-1*) gene expression in regions of the human retina relative to expression in peripheral retina (calculated using the comparative C_T method). PR, photoreceptor layer; Peri, peripheral.
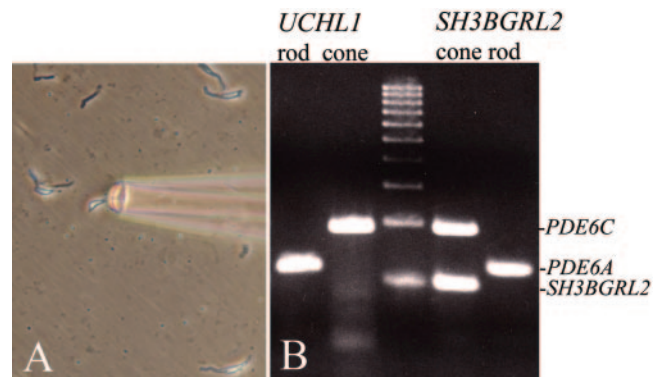

Figure 4. Single human photoreceptor (**A**) isolation and (**B**) RT-PCR. (**A**) Single photoreceptor cells were aspirated into micropipette. (**B**) Ethidium bromide–stained agarose gel of PCR products amplified in nested RT-PCR reactions of single photoreceptor cDNAs. Rod, rod-specific *PDE6A* gene primers (124-bp product) plus *UCHL1* primers (153-bp product) or plus *SH3BGRL2* primers (100-bp product); "cone" cone-specific *PDE6C* gene primers (195-bp product) plus *UCHL1* primers or *SH3BGRL2* primers. *UCHL1* is expressed in the inner retina and is therefore not detected in either photoreceptor derived cDNA pool. *Lanes 1, 2: UCHL1* reactions; *lane 3:* 100-bp ladder; *lanes 4, 5: SH3BGRL2* reactions.

perimeter of these punches. The youth of the donors, compared with the *RPEB1* library, was probably also a factor, because the adhesion of RPE to photoreceptors decreases with age. Postmortem time and the use of RNA preservative (RNA-*later*: Ambion) may also have contributed by increasing the adhesion of the tissue.

The *RPEB1* library generated by Sharon et al.[14] was derived from a freshly enucleated eye obtained from an 88-year-old patient in which the RPE was gently scraped off large fragments of the posterior eyecup after removal of the retina. This library may have a much lower number of retina-derived tags because the donor was older, no mechanical compression of the retina and RPE occurred, because no punches were taken, and/or because this was fresh tissue and RNA*later* was not used.[14] We used the tags present in the *RPEB1* library as one means to analyze the cell-source of the tags in the *4MacRPE* and *4PeriRPE* libraries. However, because our libraries surely contain legitimate RPE tags that are absent in the *RPEB1* library— because it was made from RNA from a single donor of advanced age and was only sequenced to half to two thirds the depth of the rest of the posterior eye libraries (~54,000 versus over 100,000; Table 2)—additional comparisons were undertaken. Comparison to the retina libraries from both laboratories is helpful, but these are all contaminated with RPE tags for the same reasons that the RPE libraries contain retina-derived tags described herein. Retinal contamination was even detected in RPE RNA derived from pools of RPE cells isolated by laser-capture microdissection.[33] Therefore, to test how well the parameters used to query the *EyeSAGE* database worked to generate a list of genes expressed in the RPE, qRT-PCR was performed on RNAs isolated from retina and RPE/choroid punches and compared to RNA isolated from pools of RPE cells mechanically purified from human donor eyes.[20] We determined that in the RPE/choroid the housekeeping gene *ACTB* fluctuated in parallel reactions but not in retina. We therefore tested other housekeeping genes, *GAPDH, B2M,* and *UBC,* for normalization of gene expression levels in qRT-PCR assays using RPE-derived RNAs.[25] Expression of *EMP3* (epithelial membrane protein 3; GeneID: 2014) and *MMP25* (matrix metallopeptidase 25 GeneID: 64386) was highest in the RPE-enriched samples like the known RPE-associated gene *RDH5*
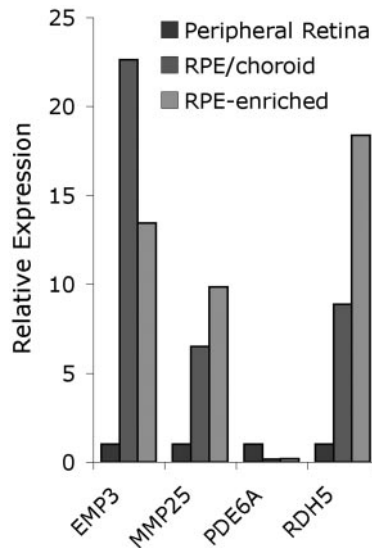
FIGURE 5. Real-time qRT-PCR summaries of rod-specific (*PDE6A*), RPE-specific (*RDH5*), and candidate RPE (*EMP3* and *MMP25*) gene expression in the peripheral retina, RPE/choroid, and RPE cells (RPE-enriched) relative to expression in the peripheral retina.

(retinol dehydrogenase 5 [11-cis and 9-cis]; GeneID: 5959; Fig. 5).

A longSAGE library generated from purified RPE cells could be used to annotate the *4MacRPE* and *4PeriRPE* libraries. This would improve tag-to-gene mapping and help identify macula-associated genes.

Detection of Splice Variants

The *4cRET* longSAGE provides a valuable means for detecting transcript variants of known retina genes. Alternative transcripts arising from tissue-specific splice forms or from alternative polyadenylation signals are known to occur at a high rate within the retina relative to the rest of the body.[34] One example of a transcript variant detected by longSAGE analysis is a short form of the transcript for *PDE6G* (phosphodiesterase 6G, cGMP-specific, rod, gamma; GeneID: 5148). The GenBank RefSeq accession number for PDE6G corresponds to a 1223-bp long mRNA (NM_002602), but this full-length transcript is not detected by short or longSAGE in any of the retina libraries. Instead, the most abundant tag in retina libraries maps to the *Nla*III site at position 847 corresponding to a shorter *PDE6G* transcript. This prediction from the SAGE data was tested by qRT-PCR using one forward primer at position 828 and two reverse primers at positions 949 and 1018. The shorter product amplified at a much higher rate (>9000 times higher relative expression than the longer transcript, data not shown), verifying the transcript length predicted by SAGE.

SAGE, particularly longSAGE, is also an excellent method for detecting antisense transcription, because the method itself is inherently directional. Detection of antisense transcription to date has primarily relied on the use of expressed sequence tag (EST) databases, where the directionality of the sequence has to be verified by the presence of canonical intron–exon splice junctions and/or poly(A) signals. A large number of ESTs lack these sequences and can therefore not be used for the analysis, leading to an underestimate of the presence of antisense transcription in the genome.[35] In contrast, a 21-bp longSAGE tag that maps to only one location in the genome, but is found on the opposite strand of a normally transcribed locus, provides enough information to implicate antisense transcription of that locus. We detected antisense transcripts of rhodop-

sin and *PDE6G* by the presence of high-count longSAGE tags matching the antisense or opposite transcript strand. In addition, the cone-associated candidate gene *ZNF593* was identified by its longSAGE expression, but the longSAGE tag did not match the RefSeq mRNA (accession number NM_015871). A BLAST search revealed that the tag matched to a location on 1p36 of the genomic DNA within the ZNF593 locus, but not within the mRNA sequence. An EST database search revealed the presence of the tag in several sequences, all derived from the eye. The UCSC Genome browser showed all these ESTs to be antisense transcripts of the ZNF593 locus, and the localization of the antisense transcripts is only in the eye whereas the localization of the sense transcription is ubiquitous (http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr1:26179935-26183592&hgsid=59876575&intronEst=pack). Primers were designed to the antisense transcript in a region that did not overlap with the sequence for the sense transcript. qRT-PCR verified a profile consistent with cone-associated expression as predicted by the SAGE profile (Fig. 2).

EyeSAGE Candidate Retinal Disease Gene Expression Tables

We found retinal diseases that have been mapped, but for which the disease gene has not yet been identified, using the RetNet table Summaries of Genes Causing Retinal Diseases, Table B (by Disease, http://www.sph.uth.tmc.edu/Retnet/sumdis.htm#B-diseases). The *EyeSAGE* database was queried for genes that map within these regions of linkage and that are also expressed in the retina and RPE. These lists of candidate disease genes for 50 mapped but not yet identified retina disease genes can be accessed through RetNet, listed by disease symbol, and also through NEIBank with links to the UCSC genome browser and Single-Nucleotide Polymorphism (SNP) database.

In the resultant tables, selecting candidates by eye expression levels served to reduce the number of candidate genes greatly for each locus. The *EyeSAGE* database correctly predicted the *BBS5* gene for the mapped-and-cloned retina disease region for the autosomal recessive Bardet-Biedl syndrome, BBS5.[36,37] The chromosomal markers flanking the disease locus were mapped by Beales et al.[36] to a 14-Mb region of chromosome 2 (bases 160851387–174882568). The query of our *EyeSAGE* database returned 131 genes with expression in the eye in this region (at least one tag in at least one eye library). The table was sorted from highest to lowest total tag counts in the eye ("Eye Sum" in all eight shortSAGE libraries combined). Thirty-seven candidate genes had a sum of at least 15 tags in the eye libraries (representing significantly high expression) and among these, the top candidate, with the highest sum of tag counts in the eye plus fairly ubiquitous expression (tags in the nonocular libraries), was the *BBS5* gene (see Supplementary Table S6 online at http://www.iovs.org/cgi/content/full/47/6/2305/DC1).[37] Of course, many disease-causing genes will not be expressed at such high levels. Thus, all expressed genes within linkage intervals should be considered candidates.

As another example, a table of candidates for the as yet unidentified CORD1 gene on chromosome 18 was generated (see Supplementary Table S7 online). The 18-Mb CORD1 disease region contains 223 genes or UniGene clusters, of which only 79 show a sum of 10 or more tag counts in the retina and RPE libraries. Of these, the top two retina-associated genes (based on little or no expression in any of the other libraries) are the *CPLX4* and *RAX* (retina and anterior neural fold homeobox, GeneID: 30062) genes.

The *EyeSAGE* database and cell-associated tables can also be used for prioritization of candidate genes for primary open-

angle glaucoma (POAG), a neurodegenerative disease characterized by death of the retinal ganglion cell (RGC). The *EyeSAGE* database was queried ("Mac/Peri" > 1; "6 mm/4 mm" <1; "Ret/Neural" <1; sorted highest to lowest by "RetAve/Body and Neural Ave"; see Supplementary Table S4 online) to produce a set of 2393 genes that are enriched in the inner retina—these genes would also be expected to be enriched for transcripts found in the RGC. When these transcripts are mapped back to the genomic assembly and compared with previously published regions of POAG linkage,[38–44] a set of 128 prioritized candidate genes emerges. This includes genes in pathways known or hypothesized to be involved in glaucomatous neurodegeneration, including apoptosis (*DAD, BBC3, BCL2L2*), axonal growth and regeneration (*RTN4, NAB1*), and calcium flux (*JPH4*). Moreover, this list of prioritized genes constitutes <6% of the almost 2400 UniGene clusters that map within regions of linkage. In this way, use of the cell-associated tables greatly reduces the number of genes that must be evaluated.

DISCUSSION

We are interested in identifying the patterns of gene expression that enable cells in the macula to cope with the increased metabolic demand and age-related changes unique to this region of the retina. Identifying such adaptations should help us understand why rods degenerate before cones in aging and early age-related macular degeneration (AMD).[1–3] *EyeSAGE* provides a new platform for addressing such questions. Four human regional retina or RPE/choroid-derived shortSAGE libraries and the first longSAGE library derived from human central retina were generated and analyzed. These were compiled into the *EyeSAGE* database where they were related to existing retina, brain, and other organ SAGE libraries obtained thorough SAGE Genie (http://cgap.nci.nih.gov/SAGE).[23] Profiles from cDNA microarrays probed with retina RNAs were incorporated into the *EyeSAGE* database. Because both expression profiles were obtained from the same donor pools of RNA (Table 1) and were generated using nonoverlapping techniques, the array data served to validate expression profiles of genes in the retina SAGE analysis. Plans to continue to improve the utility of *EyeSAGE* at NEIBank as a reference of the human retina transcriptome include: development of user-friendly web tools to query the database[45]; the addition of cell-type specific longSAGE libraries (e.g., purified human photoreceptors and mechanically isolated RPE cells); and the integration of publicly available expressed sequence tag information from human retina and RPE.[8, 46–52]

Previously, Sharon et al. found that the cone photoreceptor contribution in their two retina SAGE libraries was similar in the macula and the peripheral retina whereas the rod contribution was higher in the periphery.[14] This may seem counter intuitive because a 6-mm punch from the macula is enriched for cones (8:1 rod:cone) relative to a 6-mm peripheral retina punch (20:1 rod:cone). However, the macula also contains the highest concentration of ganglion cells (60%) and associated interneurons.[53] In addition, when comparing SAGE libraries generated from single donors, individual-to-individual variation can complicate the identification of real tissue or cell-associated gene expression differences.[30] Thus, we hypothesized that generating additional SAGE libraries from matched *pooled* donor sets could reduce the impact of individual variability and that additional human retina/RPE transcriptome profiles would validate gene expression differences identified previously if these were upheld in this pooled donor set. In fact, analysis of the resulting expression profiles in *EyeSAGE* bore this out (see Table 3 or supplementary cell-associated tables online for examples).

The *EyeSAGE* database, and particularly the first retina long-SAGE library (*4cRET*), provides a unique opportunity to use eye transcriptome data in novel ways. The 21-bp longSAGE tags can, in most cases, be mapped to a single physical location in the human genome. This specificity overcomes much of the redundancy and uncertainty that can occur with the 14-bp tag assignments, and allows for the use of NCBI BLAST searches to assign tags that are not assigned to a gene by current SAGE Genie and SAGEmap resources. These advantages allowed us to identify new retina and photoreceptor-specific candidate genes, and detect transcript variants and antisense transcripts of known genes. For example, unlike standard SAGE, longSAGE is able to uniquely identify the cone-associated gene *KIAA1345* (KIAA1345 protein; GeneID: 57545), which falls within the narrowest current mapping of the disease locus for MCDR2, an autosomal dominant inherited macular degeneration for which the disease gene has not yet been identified (RetNet, http://www.sph.uth.tmc.edu/RetNet/[54]).

LongSAGE analysis also facilitated identification of transcript variants and antisense transcription in the retina. For example, we identified a shorter transcript variant of PDE6G expressed at a much higher level than the reference sequence. LongSAGE also provided strong evidence for antisense transcription of PDE6G and rhodopsin, and allowed for identification of a retina-specific antisense transcript of the ubiquitously expressed gene *ZNF593* (zinc finger protein 593; GeneID: 51042). These findings are particularly intriguing in the context of the recent study by Alfano et al.[55] establishing the importance of natural antisense transcription in eye development.

Analysis of the longSAGE tags also enhanced annotation of known genes and of their corresponding short tags. There are often instances in which a short tag is designated as the best tag for more than one gene on SAGE Genie. It can be difficult to determine which gene the tag counts represent, because they can be representative of the expression of either gene or even the sum of both. The algorithms used by SAGE Genie to assign the best gene match to a tag favor genes that are more ubiquitously expressed, and retina-specific genes, which often have fewer archived cDNA sequences, are underrepresented by these methods.[23] Comparison to the longSAGE tag counts for the genes in the retina provides a means to estimate what percentage of the short tag counts correspond to expression of a given gene in the retina and thus which is the true tissue- or cell-associated gene. For example, the shortSAGE tag, CTGTT-GATTT, emerged in the cone-associated expression profile, but it was assigned to two different genes. LongSAGE analysis identified the correct cone-associated gene to be GUCA1C (Table 3), as has been previously reported.[56]

We also used the *EyeSAGE* database to produce candidate retina genes for mapped-but-not-yet-identified retina disease loci. We tested the validity of this approach by applying the data-mining paradigm for candidate retina disease genes to a retina disease, BBS5, for with the gene has been identified, *BBS5*,[37] and returned *BBS5* as the top candidate. Evaluation of the candidate disease genes for the autosomal dominant cone-rod dystrophy 1 (CORD1)[57] yielded the gene for complexin IV, *CPLX4*, among the top candidates (see Supplementary Table S7 online). Complexin IV has very recently been localized to murine photoreceptor ribbon synapses where it modulates transmitter release.[58] This is intriguing because *CPLX4* is similar to another gene, *RIMS1,* responsible for the cone–rod dystrophy, CORD7.[59] *RIMS1* codes for a presynaptic protein expressed in brain and photoreceptors that also localizes to ribbon synapses where it functions in glutamate neurotransmission. Based on *CPLX4*'s expression profile in *EyeSAGE* and by qRT-PCR where it is elevated in the macula photoreceptor layer relative to peripheral retina photoreceptor layer, *CPLX4*

behaves like a cone-associated gene (Table 3, Fig. 3). Future studies will determine whether *CPLX4* mutations are associated with CORD1.

In summary, these new SAGE libraries of the human retina and RPE/choroid and the relational transcriptome database *EyeSAGE* can be used to identify tissue and regional specificity of retinal gene expression, and global assessment of alternative transcription occurring in the human retina. These data can also be used to identify candidate retinal disease genes, either by using the candidate tables generated and now available on the RetNet and NEIBank Web sites, or by allowing researchers to query the database with new loci as they are identified. The new transcriptome information added by the work presented in this article and, more important, the database, which greatly expands the utilization of existing and these new expression profiles, will be an excellent resource for the vision research community as they explore questions related to gene expression in normal function and ocular disease.

References

1. Klein R, Klein BE, Linton KL. Prevalence of age-related maculopathy. The Beaver Dam Eye Study. *Ophthalmology.* 1992;99:933–943.
2. Mitchell P, Smith W, Attebo K, Wang JJ. Prevalence of age-related maculopathy in Australia. The Blue Mountains Eye Study. *Ophthalmology.* 1995;102:1450–1460.
3. Vingerling JR, Dielemans I, Bots ML, Hofman A, Grobbee DE, de Jong PT. Age-related macular degeneration is associated with atherosclerosis. The Rotterdam Study. *Am J Epidemiol.* 1995;142:404–409.
4. Curcio CA. Photoreceptor topography in ageing and age-related maculopathy. *Eye.* 2001;15:376–383.
5. Curcio CA, Medeiros NE, Millican CL. Photoreceptor loss in age-related macular degeneration. *Invest Ophthalmol Vis Sci.* 1996;37:1236–1249.
6. Curcio CA, Millican CL, Allen KA, Kalina RE. Aging of the human photoreceptor mosaic: evidence for selective vulnerability of rods in central retina. *Invest Ophthalmol Vis Sci.* 1993;34:3278–3296.
7. Jackson GR, Owsley C, Curcio CA. Photoreceptor degeneration and dysfunction in aging and age-related maculopathy. *Ageing Res Rev.* 2002;1:381–396.
8. Chowers I, Esumi N, Campochiaro P, Zack DJ. Studies on retinal and retinal pigment epithelial gene expression. *Novartis Found Symp.* 2004;255:131–145; discussion 2004;6:177–8.
9. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science.* 1995;270:484–487.
10. Saha S, Sparks AB, Rago C, et al. Using the transcriptome to annotate the genome. *Nat Biotechnol.* 2002;20:508–512.
11. Wahl MB, Heinzmann U, Imai K. LongSAGE analysis significantly improves genome annotation: identifications of novel genes and alternative transcripts in the mouse. *Bioinformatics.* 2005;21:1393–1400.
12. Schulz HL, Rahman FA, Fadl El Moula FM, Stojic J, Gehrig A, Weber BH. Identifying differentially expressed genes in the mammalian retina and the retinal pigment epithelium by suppression subtractive hybridization. *Cytogenet Genome Res.* 2004;106:74–81.
13. Swaroop A, Zack DJ. Transcriptome analysis of the retina. *Genome Biol.* 2002;3:REVIEWS1022.
14. Sharon D, Blackshaw S, Cepko CL, Dryja TP. 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.* 2002;99:315–320.
15. Chowers I, Gunatilaka TL, Farkas RH, et al. Identification of novel genes preferentially expressed in the retina using a custom human retina cDNA microarray. *Invest Ophthalmol Vis Sci.* 2003;44:3732–3741.
16. Chowers I, Liu D, Farkas RH, et al. Gene expression variation in the adult human retina. *Hum Mol Genet.* 2003;12:2881–2893.
17. Yoshida S, Yashar BM, Hiriyanna S, Swaroop A. Microarray analysis of gene expression in the aging human retina. *Invest Ophthalmol Vis Sci.* 2002;43:2554–2560.
18. Johnston NL, Cervenak J, Shore AD, Torrey EF, Yolken RH. Multivariate analysis of RNA levels from postmortem human brains as measured by three different methods of RT-PCR. Stanley Neuropathology Consortium. *J Neurosci Methods.* 1997;77:83–92.
19. Ebright JN, Bowes Rickman C. Rapid, reproducible real-time quantitative RT-PCR using the iCycler iQ real-time PCR detection system and iQ supermix. *Bioradiations.* 2003;110:32–35.
20. Yang P, Wiser JL, Peairs JJ, et al. Human RPE expression of cell survival factors. *Invest Ophthalmol Vis Sci.* 2005;46:1755–1764.
21. Thompson CL, Rickman CB, Shaw SJ, et al. Expression of the blue-light receptor cryptochrome in the human retina. *Invest Ophthalmol Vis Sci.* 2003;44:4515–521.
22. St Croix B, Rago C, Velculescu V, et al. Genes expressed in human tumor endothelium. *Science.* 2000;289:1197–1202.
23. Boon K, Osorio EC, Greenhut SF, et al. An anatomy of normal and malignant gene expression. *Proc Natl Acad Sci USA.* 2002;99:11287–11292.
24. Hauser MA, Li YJ, Takeuchi S, et al. Genomic convergence: identifying candidate genes for Parkinson's disease by combining serial analysis of gene expression and genetic linkage. *Hum Mol Genet.* 2003;12:671–677.
25. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;3:RESEARCH0034.
26. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* 2001;25:402–408.
27. Nishizawa Y, Komori N, Usukura J, Jackson KW, Tobin SL, Matsumoto H. Initiating ocular proteomics for cataloging bovine retinal proteins: microanalytical techniques permit the identification of proteins derived from a novel photoreceptor preparation. *Exp Eye Res.* 1999;69:195–212.
28. Yuodelis C, Hendrickson A. A qualitative and quantitative analysis of the human fovea during development. *Vision Res.* 1986;26:847–855.
29. McHeyzer-Williams LJ, Panus JF, Mikszta JA, McHeyzer-Williams MG. Evolution of antigen-specific T cell receptors in vivo: preimmune and antigen-driven selection of preferred complementarity-determining region 3 (CDR3) motifs. *J Exp Med.* 1999;189:1823–1838.
30. Blackshaw S, Kuo WP, Park PJ, et al. MicroSAGE is highly representative and reproducible but reveals major differences in gene expression among samples obtained from similar tissues. *Genome Biol.* 2003;4:R17.
31. Chen J, Sun M, Lee S, Zhou G, Rowley JD, Wang SM. Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proc Natl Acad Sci USA.* 2002;99:12257–12262.
32. Schein SJ. Anatomy of macaque fovea and spatial densities of neurons in foveal representation. *J Comp Neurol.* 1988;269:479–505.
33. Ishibashi K, Tian J, Handa JT. Similarity of mRNA phenotypes of morphologically normal macular and peripheral retinal pigment epithelial cells in older human eyes. *Invest Ophthalmol Vis Sci.* 2004;45:3291–3301.
34. Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* 2002;30:3754–3766.
35. Yelin R, Dahary D, Sorek R, et al. Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol.* 2003;21:379–386.
36. Beales PL, Katsanis N, Lewis RA, et al. Genetic and mutational analyses of a large multiethnic Bardet-Biedl cohort reveal a minor involvement of BBS6 and delineate the critical intervals of other loci. *Am J Hum Genet.* 2001;68:606–616.
37. Li JB, Gerdes JM, Haycraft CJ, et al. Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell.* 2004;117:541–552.
38. Allingham RR, Wiggs JL, Hauser ER, et al. Early adult-onset POAG linked to 15q11-13 using ordered subset analysis. *Invest Ophthalmol Vis Sci.* 2005;46:2002–2005.

39. Nemesure B, Jiao X, He Q, et al. A genome-wide scan for primary open-angle glaucoma (POAG): the Barbados Family Study of Open-Angle Glaucoma. *Hum Genet*. 2003;112:600–609.

40. Samples JR, Kitsos G, Economou-Petersen E, et al. Refining the primary open-angle glaucoma GLC1C region on chromosome 3 by haplotype analysis. *Clin Genet*. 2004;65:40–44.

41. Stoilova D, Child A, Trifan OC, Crick RP, Coakes RL, Sarfarazi M. Localization of a locus (GLC1B) for adult-onset primary open angle glaucoma to the 2cen-q13 region. *Genomics*. 1996;36:142–150.

42. Trifan OC, Traboulsi EI, Stoilova D, et al. A third locus (GLC1D) for adult-onset primary open-angle glaucoma maps to the 8q23 region. *Am J Ophthalmol*. 1998;126:17–28.

43. Wiggs JL, Allingham RR, Hossain A, et al. Genome-wide scan for adult onset primary open angle glaucoma. *Hum Mol Genet*. 2000; 9:1109–1117.

44. Wirtz MK, Samples JR, Rust K, et al. GLC1F, a new primary open-angle glaucoma locus, maps to 7q35–q36. *Arch Ophthalmol*. 1999;117:237–241.

45. Wistow G, Bernstein SL, Touchman JW, et al. Grouping and identification of sequence tags (GRIST): bioinformatics tools for the NEIBank database. *Mol Vis*. 2002;8:164–170.

46. Bernstein SL, Borst DE, Neuder ME, Wong P. Characterization of a human fovea cDNA library and regional differential gene expression in the human retina. *Genomics*. 1996;32:301–308.

47. Bernstein SL, Borst DE, Wong PW. Isolation of differentially expressed human fovea genes: candidates for macular disease. *Mol Vis*. 1995;1:4.

48. Malone K, Sohocki MM, Sullivan LS, Daiger SP. Identifying and mapping novel retinal-expressed ESTs from humans. *Mol Vis*. 1999;5:5.

49. Sinha S, Sharma A, Agarwal N, Swaroop A, Yang-Feng TL. Expression profile and chromosomal location of cDNA clones, identified from an enriched adult retina library. *Invest Ophthalmol Vis Sci*. 2000;41:24–28.

50. Sohocki MM, Malone KA, Sullivan LS, Daiger SP. Localization of retina/pineal-expressed sequences: identification of novel candidate genes for inherited retinal disorders. *Genomics*. 1999;58:29–33.

51. Wistow G, Bernstein SL, Wyatt MK, et al. Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: over 6000 non-redundant transcripts, novel genes and splice variants. *Mol Vis*. 2002;8:205–220.

52. Wistow G, Bernstein SL, Wyatt MK, et al. Expressed sequence tag analysis of human retina for the NEIBank Project: retbindin, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts. *Mol Vis*. 2002;8:196–204.

53. Rodieck RW. *The First Steps of Seeing*. Sunderland, MA: Sinauer; 1998.

54. Michaelides M, Johnson S, Poulson A, et al. An autosomal dominant bull's-eye macular dystrophy (MCDR2) that maps to the short arm of chromosome 4. *Invest Ophthalmol Vis Sci*. 2003;44:1657–1662.

55. Alfano G, Vitiello C, Caccioppoli C, et al. Natural antisense transcripts associated with genes involved in eye development. *Hum Mol Genet*. 2005;14:913–923.

56. Imanishi Y, Li N, Sokal I, et al. Characterization of retinal guanylate cyclase-activating protein 3 (GCAP3) from zebrafish to man. *Eur J Neurosci*. 2002;15:63–78.

57. Warburg M, Sjo O, Tranebjaerg L, Fledelius HC. Deletion mapping of a retinal cone-rod dystrophy: assignment to 18q211. *Am J Med Genet*. 1991;39:288–293.

58. Reim K, Wegmeyer H, Brandstatter JH, et al. Structurally and functionally unique complexins at retinal ribbon synapses. *J Cell Biol*. 2005;169:669–680.

59. Johnson S, Halford S, Morris AG, et al. Genomic organisation and alternative splicing of human RIM1, a gene implicated in autosomal dominant cone-rod dystrophy (CORD7). *Genomics*. 2003;81: 304–314.