Generation of a foveomacular transcriptome

Alison Ziesel,[1] Steven Bernstein,[2] Paul W. Wong[1]

[1]Department of Ophthalmology, Emory University, Atlanta GA; [2]Department of Opthalmology and Visual Sciences, University of Maryland, Baltimore MD

Purpose: Organizing molecular biologic data is a growing challenge since the rate of data accumulation is steadily increasing. Information relevant to a particular biologic query can be difficult to extract from the comprehensive databases currently available. We present a data collection and organization model designed to ameliorate these problems and applied it to generate an expressed sequence tag (EST)–based foveomacular transcriptome.
Methods: Using Perl, MySQL, EST libraries, screening, and human foveomacular gene expression as a model system, we generated a foveomacular transcriptome database enriched for molecularly relevant data.
Results: Using foveomacula as a gene expression model tissue, we identified and organized 6,056 genes expressed in that tissue. Of those identified genes, 3,480 had not been previously described as expressed in the foveomacula. Internal experimental controls as well as comparison of our data set to published data sets suggest we do not yet have a complete description of the foveomacula transcriptome.
Conclusions: We present an organizational method designed to amplify the utility of data pertinent to a specific research interest. Our method is generic enough to be applicable to a variety of conditions yet focused enough to allow for specialized study.

Data management is a critical part of analyzing large data sets, and a sensible management system is needed to fully exploit their value. Methods sophisticated enough to cope with many data points and types are necessary to fully interrogate the results of modern gene profiling and/or expressed sequence tag (EST)–based studies. To best exploit all relevant data available from multiple data sources, it is necessary to design an organizational structure best suited for continuously adding data. In this manuscript, we present a platform developed to provide access to relevant foveomacular gene expression data, but which could be adapted to define the expression profile for any other tissue- or cell-specific phenomenon. Using the Perl scripting language and MySQL query language, we have developed a standardized system to extract relevant data from various public sources and compile the data in a form immediately useful to the end user.

Perl is well suited to our needs for several reasons: It is a well-established scripting language with a strong history in computational biology, and is particularly adept at text parsing, the function required to extract portions of data from online records. We chose MySQL as our relational database query language for several reasons, including availability, community support, and wide acceptance by the computational biology community. Relational databases are especially suited to this project, as they allow simple expansion of data types and sources.

Our interests focus on understanding the biology of the human fovea. The human fovea is a 1.5 mm diameter region at the center of the macula of the retina responsible for acute, detailed color vision [1]. The fovea and the macula (foveomacula, a 5 mm diameter region at the center of the retina) have morphology and function distinct from the peripheral retina, and are affected by a unique set of heritable and age-related disorders [2]. The major obstacle to studying the fovea is its relatively small size and the difficulty in obtaining human foveomacular tissue. Moreover, with the exception of analogous avian structures [3], primates are the only vertebrates that possess a foveal structure, meaning that no suitable non-primate mammal exists to model the human fovea. Although several genes that when mutated lead to macular degenerative states have been cloned and their functions determined [4], a significant number of disease states remain uncharacterized genetically, particularly those with a non-Mendelian inheritance pattern or sporadic development [5]. A systematic analysis of gene expression in healthy foveomacula is important for understanding normal and pathological states.

Several databases are relevant to the human foveomacula. They include RetNet, NCBI's Entrez, Retina Central's Retinome [6], EyeSAGE [7], and NEI's NEIBank [8]. The mandate of RetNet is to catalog clinical and genetic data regarding ocular disease in general. Entrez is intended to be a central repository of all genomic data and thus cannot be specific to

Correspondence to: Alison Ziesel, Department of Ophthalmology, Emory University, Atlanta GA, Phone: (404) 778 5531, FAX: (404) 778 2231: email: aziesel@emory.edu.

any one tissue or developmental state, although information centered on a single tissue can be potentially extracted. The Retinome project characterizes whole retina and RPE gene expression, and as expected there is some overlap in raw data used to build the Retinome and the foveome. EyeSAGE is a data source limited to Serial Analysis of Gene Expression (SAGE) tag data derived from the human retina, RPE, peripheral retina, and macula. SAGE analysis focuses on limited sequence runs of ten to 17 base pairs for identification of specific genes, which can be ambiguous on occasion [9-11]. NEIbank provides data regarding EST-derived cDNA libraries developed from ocular tissues from various organisms; however, it does not currently possess a sufficiently large or reliable foveal or foveomacular cDNA library.

We perceived the absence of a highly organized gene-centric data collection of foveomacular expressed genes. We performed a meta-analysis on preexisting macular ESTs (UI-E-CK1) [12] and fovea-derived macroarray data that have not been formally published [13-15]. Data sets that arise from the direct sequencing of unamplified human foveomacular cDNA clones or sequenced cDNA clones that have been screened with a mixed foveomacular cDNA probe are directly comparable. In both cases, the probability of detecting a specific foveomacularly expressed gene is a function of its relative expression level. Identifying a specific gene with screening arrayed clones with a representative, mixed cDNA probe depends on the sensitivity of a gene-specific probe and the presence of that target on the array. In most cases, identifying genes with this method is largely limited to high to middle abundant expressed genes.

METHODS

Human fovea tissue and RNA: Cadaver eyes were obtained from the National Disease Research Interchange (NDRI, Philadelphia, PA) and as such are exempt from IRB approval. Human eyes of Caucasian origin were obtained from the National Disease Research Interchange and from individual eye banks. Tissue samples were excluded if they were enucleated from donors with any reported ocular diseases or genetic abnormalities. To obtain human foveal tissue, human donor eyes were dissected on ice. The posterior pole of the eyes, containing the retina, was dissected free of the overlying vitreous. The retinal tissue surrounding the central foveola to a radius of 0.75 mm centered over the foveolar umbo was dissected essentially free of RPE and choroid, free of sclera, and flash frozen on dry ice. In all cases, dissected tissues were stored at −70 °C until RNA extraction. For RNA extraction, foveae tissue from ten pairs of donor eyes was pooled, total RNA was extracted, and poly(A+) RNA was prepared

using standard methodologies (Tel-test, Friendswood, TX; Oligotex, Qiagen, Chatsworth, CA).

Raw data analysis: Nylon membrane macroarrays bearing 248,832 human cDNA clones were purchased from Research Genetics (Huntsville, AL) and screened as recommended by Research Genetics. Clones on the Research Genetics arrays are double spotted on a set systematic grid pattern so that the location of hybridization signal for a pair of spots as defined by Research Genetics was used to elucidate gene identity. A second set of arrays, consisting of duplicate spotted 18,300 partially sequenced cDNA clones, all between 500 and 700 nucleotides in length, was purchased from Genome Systems (GDA version 1.3, St. Louis, MI) and screened as recommended by Genome Systems. Both sets of macroarrays were hybridized with a radioactively labeled mixed cDNA probe derived from pooled human fovea RNA. Mixed cDNA probes synthesized from poly(A+) RNA prepared from a pool of human fovea RNA were random-primed labeled using a combined mixture of ^{32}P nucleotides or ^{33}P nucleotides in a standard labeling reaction (Prime-It II, Stratagene, La Jolla, CA). Minimum activity used per hybridization was 5×10^6 CPM/ml of hybridization buffer (Hybrisol II, Oncor, Gaithersburg, MD). Hybridizations were performed using the protocol recommended by the manufacturer of the hybridization buffer. Blots were subjected to stringent hybridization and washes, followed by autoradiography at −80 °C [16] or scanned on a Molecular Dynamic scanner. The RNA used to make the cDNA probes for the two screens were derived from different sets of donors; therefore, each screen represents an independent experiment. The fovea cDNA probe used for screening the Research Genetics arrays were derived from a fovea tissue pool obtained from donors whose age ranged from 2 to 79 years; the mean donor age was 42.6 years. The fovea cDNA probe used for screening the Genome Systems arrays was derived from a separate fovea tissue pool obtained from donors spanning 12–80 years of age; the mean donor age was 50.3 years.

The resulting autoradiographs for the Research Genetics arrays were analyzed by four individuals according to the manufacturer's protocol. An autoradiograph signal was considered to represent a true fovea-expressed gene if all four individuals concurred. In the absence of consensus, a potential signal was considered negative, thus limiting false-positive results at the consequence of increasing false-negative results for this data set. Each confirmed positive hybridization signal address was retrieved from the Research Genetics EST database, which converted each identified signal into an IMAGE number. IMAGE IDs were then translated into GenBank accession numbers.
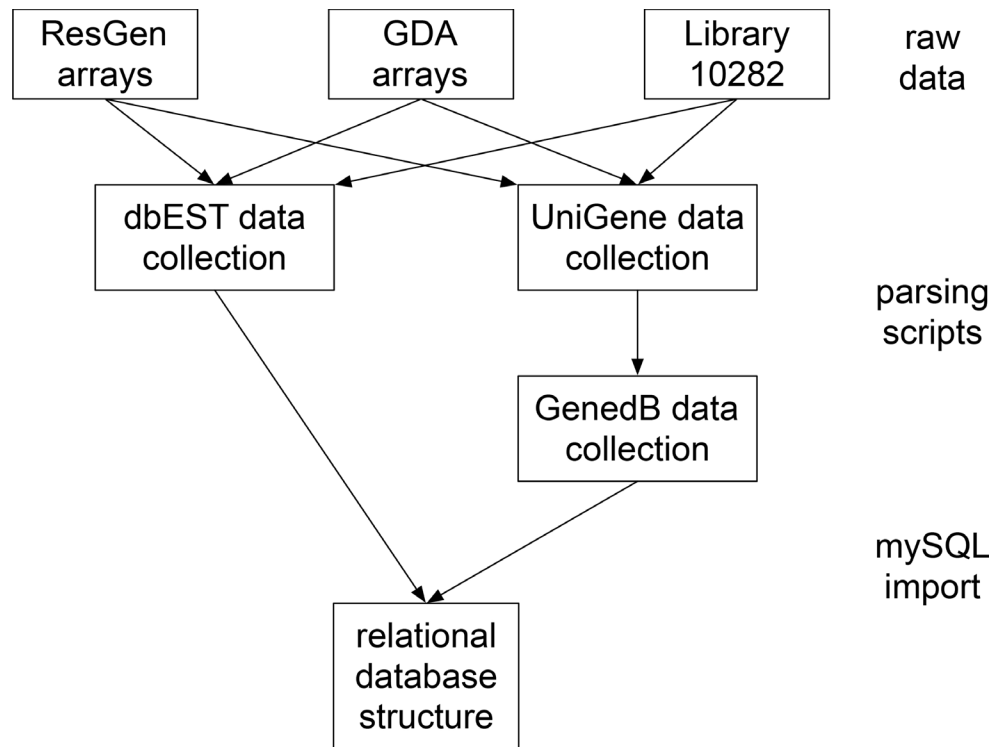
Figure 1. The flow of work from array interpretation to the generation of a usable database. Each of the three data sets was analyzed with a parsing script developed for each database, dbEST, UniGene, and Gene. The relational database design was dependent on the data we chose to harvest and is described further in Figure 2.

The molecular dynamic scans of screened GDA array filters were analyzed by Genome Systems using imaging software specialized for high-density array analysis (Array Vision; IMAGING Research, St. Catharines, Ontario, Canada). Accession numbers were collected for positive signals (the cut-off for a positive signal was twofold over the minimum value detected on the array) using the GDA software package. Accession numbers for ESTs derived from single pass sequencing of human macular clones from library 10,282 were obtained directly from the NCBI.

Data utilities: Perl scripts were written to search for each accession number in the DataBase of Expressed Sequence Tags (dbEST), UniGene, and Gene, and to harvest selected data from each database. Although GenBank accession numbers were our primary data type and are highly stable identifiers, Gene database identifiers are the central data type our database is modeled on, as this is a gene-centric effort. UniGene clusters were deemed too variable over time and were collected only to link ESTs identified to other ESTs thought to belong to the same transcript. Our Perl scripts formatted the retrieved results into a series of distinct files, corresponding to separate tables of our relational database. Figure 1 indicates the basic activities of each of our scripts. Figure 2 describes our MySQL relational database structure. A relational database structure is one in which related subtypes of data are organized into distinct tables, which are

linked to one another by sharing common elements. These common elements, unique identifiers for a specific data point in a set, are the connections between tables, and allow for variably complex queries of the data.

RESULTS

Identification of positive clones: From our Research Genetics macroarray screen, we identified 16,646 positive clones representing foveally expressed genes. Some IMAGE numbers have no associated accession numbers (indicating that they had not been sequenced), and lacked a dbEST entry. The 16,646 positive clones corresponded to 10,281 GenBank accession numbers that were used for further analysis. Our GDA arrays identified 3962 positive signals indicative of foveally expressed transcripts. At the time of most recent analysis (spring 2010), the foveomacular library 10,282 contained 6,279 ESTs, and data regarding this library is freely available through the NCBI. In total, 20,522 ESTs were examined (10,281 identified by Research Genetics macroarrays, 3,962 by GDA array, and 6,279 from library 10,282).

Data mining: All three data sets were subjected to the same subsequent analyses. For our initial data collection, we harvested data from three databases maintained by the National Center for Biotechnology Information (NCBI): dbEST (DataBase of Expressed Sequence Tags), UniGene,
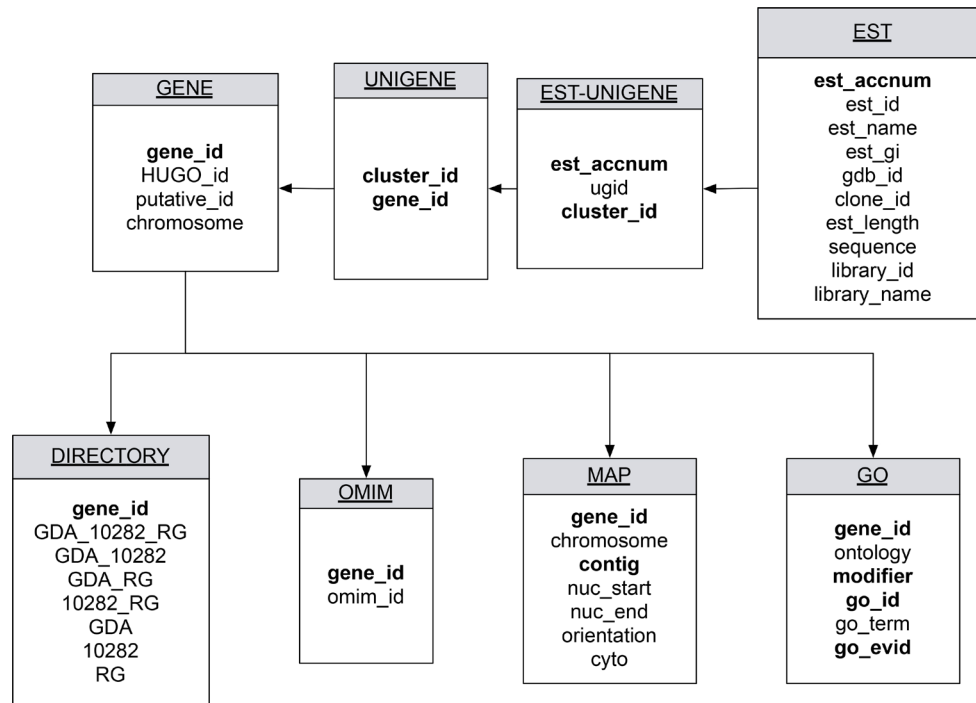
EST
est_accnum
est_id
est_name
est_gi
gdb_id
clone_id
est_length
sequence
library_id
library_name

GENE
gene_id
HUGO_id
putative_id
chromosome

UNIGENE
cluster_id
gene_id

EST-UNIGENE
est_accnum
ugid
cluster_id

DIRECTORY
gene_id
GDA_10282_RG
GDA_10282
GDA_RG
10282_RG
GDA
10282
RG

OMIM
gene_id
omim_id

MAP
gene_id
chromosome
contig
nuc_start
nuc_end
orientation
cyto

GO
gene_id
ontology
modifier
go_id
go_term
go_evid

Figure 2. The relational database schema. Each box describes a table, and each line describes a connection between tables. Bolded phrases within each table indicate key elements used to link between tables. Data for the table expressed sequence tag (EST) is derived from DataBase of Expressed Sequence Tags (dbEST); table EST-UNIGENE is derived from UniGene, and all remaining tables are derived from GenedB data. EST: Table describing data collected from dbEST. This includes GenBank accession number, EST ID, EST name, EST gene identifier (gi), GDB ID, IMAGE clone ID, length of the EST, sequence of the EST, and both ID number and name of the library from which the EST is derived. EST-UNIGENE: This table links the GenBank accession number to the UniGene clusters. The UGID is a novel identifier included for future expansion. UNIGENE: This table links the UniGene cluster to the Gene database identifiers. There is not necessarily a one-to-one relationship between UniGene and Gene. GENE: This describes data collected from the Gene database. This includes the gene symbol (HUGO ID), gene name, and chromosome of origin. GO: This table describes data relevant to Gene Ontology (GO) annotations. Each gene may have 0 or more GO annotations associated with it; of these annotations, they may have a term modifying their meaning (such as NOT or CONTRIBUTES_TO; "modifier"), a numerical identifier and text definition (GO ID and GO TERM) as well as all the evidence codes associated with that combination of modifier and GO annotation. MAP: This table describes map relevant data. Some genes may map to multiple chromosomes (for example, pseudoautosomal genes) so allowance is made for multiple map locations for a given gene. Chromosome, contig (useful for genes not fully mapped yet), nucleotide positions, orientation, and cytogenetic band (cyto) information is included. OMIM: This data describes OMIM data associated with a given Gene. DIRECTORY: This complex table tracks which experiment contributed genes to the entire set. Genes that are identified by a single experiment are listed under "GDA," "10282" or "RG" (Research Genetics); doubly identified genes by one of the three paired listings, and triply identified genes by the table GDA_10282_RG.

and Gene. These databases were chosen because they are well established, their interfaces lend themselves to automated searching, and all of the information contained within belongs to the public domain.

The data types we chose to collect and their organization are described in Figure 2. Since our three data sets are EST-based, dbEST was a natural starting point for information retrieval. dbEST provides several basic points of data for our genes of interest, namely, library of origin, available sequence, and other identifiers unique to the corresponding clone of interest (such as IMAGE number and GenBank gi (gene identifier)) that are useful for further data mining. The Gene database provides information regarding mapping, genomic organization, known disease states and phenotypes linked to that gene, and functional data in the form of Gene Ontology (GO) annotations.

In our examination of the UniGene database, we found that 17,437 (84%) of ESTs identified were associated with a UniGene cluster. Typically, multiple accession numbers are associated with a given UniGene cluster; when multiple occurrences of UniGene clusters were parsed out, the non-redundant list of positive clusters was limited to 6162, for an average EST-to-cluster ratio of approximately 3:1. In most cases, a single UniGene cluster is associated with a single Gene database entry, but occasionally multiple clusters associate with a single gene, or vice versa. In the end, 6,056 unique GenedB entries were identified. Appendix 1 contains additional database schema and database tables described in this publication.

Foveomacular transcriptome: The 6,056 identified human foveomacular genes varied, ranging from well-characterized genes to transcribed loci with little additional information.
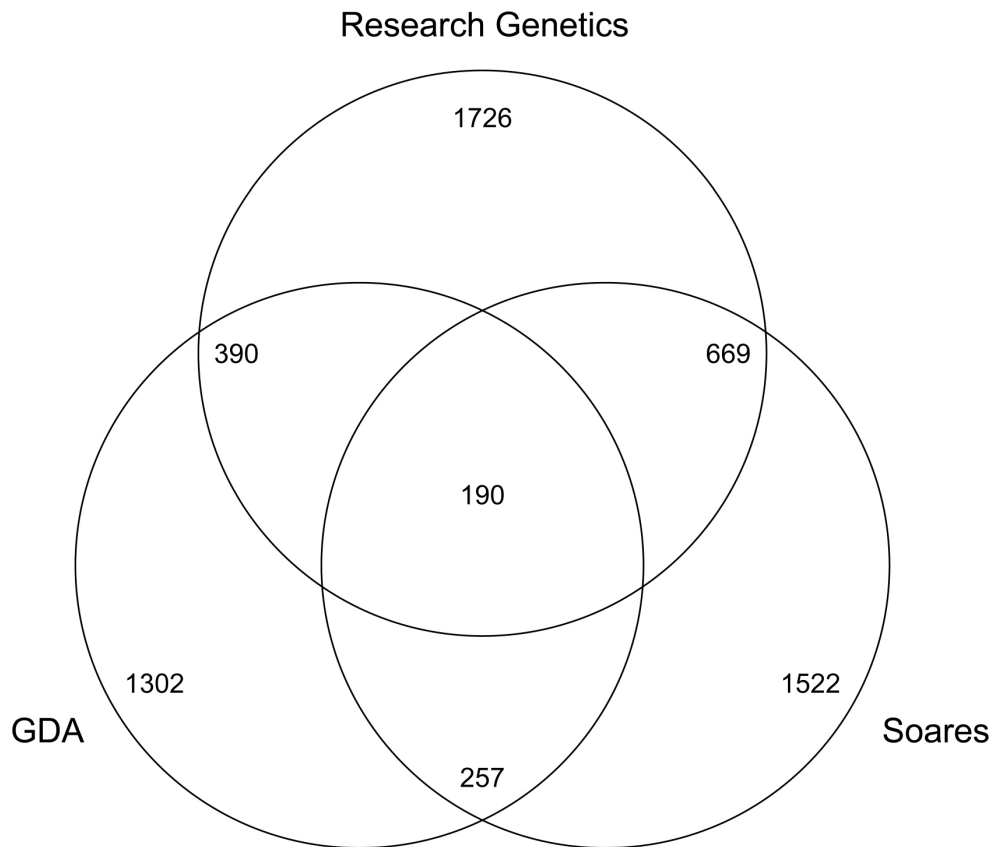
Research Genetics

Figure 3. A non-proportional Venn diagram describing overlap of genes identified by each of the three experiments. "Research Genetics" indicates the Research Genetics macroarrays, "GDA" indicates the GDA macroarrays, and "Soares" indicates library 10,282 produced by the laboratory of Bento Soares.

Identification of the same gene in independent data sources provides internal experimental confirmation of foveomacular gene expression status. There is a 3% overlap of genes identified (190 genes) in all three experimental data sources surveyed (Figure 3, Appendix 2). Twenty-two percent of the genes (1,316 genes) were identified in two of the three data sources surveyed.

In our collection, 5,979 genes (99%) had a defined nucleotide mapping position in the human genome. The distribution of these mapped genes by chromosome is shown in Figure 4A. A broad measure of normal distribution of foveomacular expressed genes can be obtained by comparing the gene collection of the current study against estimates of whole genome gene distribution as described by the human genome build used, 37.1. A chi-square test comparing the two sets indicated that there was a significant difference in the number of genes per chromosome for the fovea and whole genome sets (p>0.005). We calculated the ratio of foveomacular expressed genes to total genes per chromosome and compared this to the overall average (16.5% with a standard deviation of 2.2; Figure 4B). Regarding our collection of foveomacular expressed genes, there is a slight overrepresentation of these genes on chromosomes 12, 16, 17, 20, and

22, and an under representation on chromosome 15 and the sex chromosomes.

Functional annotation was available for 5,355 of our identified genes. Of these, 69% had ten or fewer GO terms annotated. The *transforming growth factor beta* gene (*TGFB*, Gene ID 7040) was the most annotated gene with 108 annotations. The distribution of GO annotation frequencies for our genes is provided in Figure 5, showing a steady decrease in annotation frequency. A minority of genes identified in this study are extremely well characterized, while the majority of genes are only lightly annotated if at all. To further organize genes regarding GO terms, we applied the clustering utility DAVID [17,18] to those 190 genes identified by all three of our data sources (Appendix 3). Clusters formed around common functions, including signaling, oxidative response, and metabolism. Our Perl scripts are included in Appendix 4 and are released under a Creative Commons Attribution-Non Commercial license (scripts previously described in [19,20]).

We previously demonstrated that there is a strong correlation of retinal disease genes to be expressed in the foveomacula as well as the peripheral retina [21]. We compiled a list of 125 genes implicated in retinopathic or maculopathic conditions from RetNet (Appendix 5). We reidentified 57
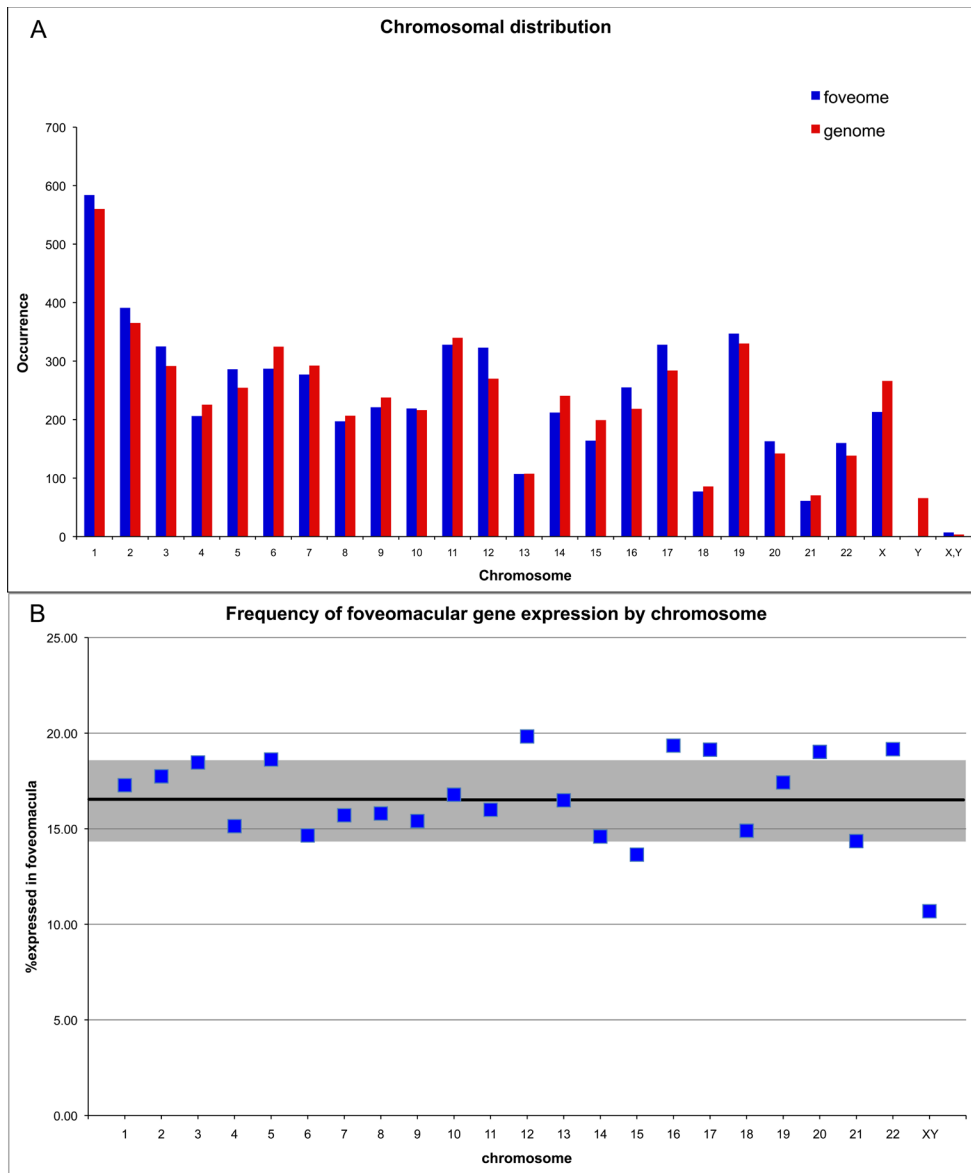
Figure 4. Distribution of the map location of our identified genes ("foveome"). **A**: All but one of our genes associated with the Y chromosome were also associated with the X chromosome ("X+Y"). Additionally, distributions for all known mapped genes are included for comparison ("genome"). **B**: Percentages of foveomacularly expressed genes per chromosomes (sex chromosomes treated as an autosomal pair). The bold black line indicates the mean frequency of expression, and the gray area represents one standard deviation above and below the mean.

(46%) of these retinopathic genes in our list of 6,056 foveomacularly expressed genes. We also compared our data to other published collections of foveomacularly expressed genes. Library 420 represents the first human fovea cDNA library made [16,22]; 100 ESTs defining 32 known human genes have been deposited in GenBank. Twenty of these genes were reidentified in our study. In an independent study, analysis of Incyte arrays identified 5,702 foveomacularly expressed genes [7]; 2,576 were reidentified in our data sets.

Comparing our EST-based foveomacular transcriptome with other published studies provides independent confirmation of foveomacular gene expression for 2,576 of the 6,056 genes collected in our current study; 3,480 genes were newly identified as foveomacularly expressed. Between the foveome and our two external comparisons with Library 420 and Incyte arrays, a total of 9,197 human foveomacular expressed genes can be defined (Appendix 2). Of these genes, 2,582 were found in at least one additional data set. Moreover, the published data sets (Incyte and Lib.420) identified 3,140 foveomacularly expressed genes that were not identified in our data set. The finding that there is incomplete overlap between experimental data sets indicates that none of the data sets considered define a full complement of human foveomacular expressed genes, and that more foveomacular genes remain to be discovered. Another possibility is that we have observed bias in gene expression due to individual genetics, age, and environmental backgrounds from donor
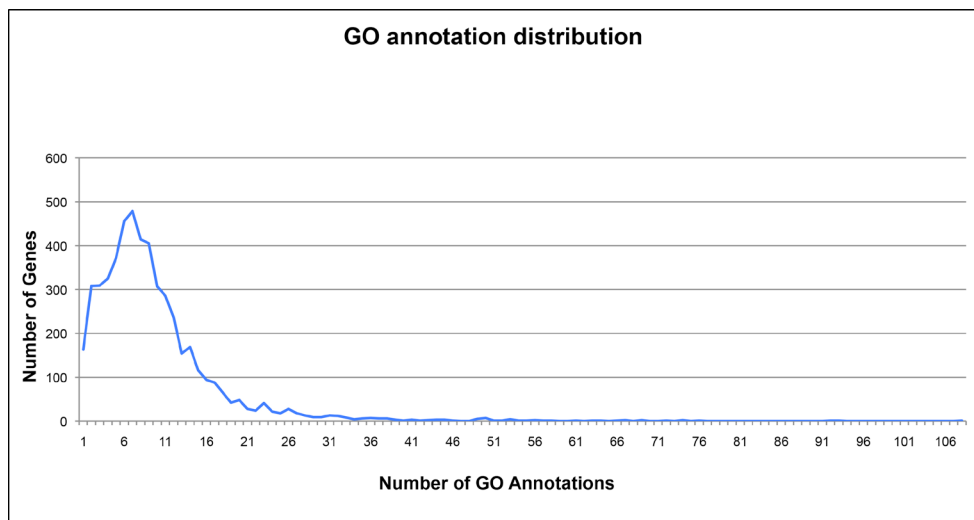
GO annotation distribution

Figure 5. Distribution of the map location of our identified genes ("foveome"). All but one of our genes associated with the Y chromosome were also associated with the X chromosome ("X+Y"). Description of the frequency of Gene Ontology annotations ascribed to our identified genes.

samples used to generate our cDNA probe. If this is the case, a great deal of variation may be detected between individuals, primarily as low expressed transcripts.

DISCUSSION

Although large all-encompassing databases, public and commercial, are excellent data resources for the molecular biologist or geneticist, their wealth of information can make navigation difficult and thus reduce utility. We established an EST-based foveomacular transcriptome consisting of 6,056 genes. A comparison of published data sets expanded this transcriptome to 9,197 genes. Our aim is to produce a boutique database: We deal only with the gene expression within a specific tissue, and are not necessarily interested in other genes. This notion can be applied to other tissues, developmental time points, or pathologies. Our EST-based foveomacular database is populated only with data relevant to foveomacular gene expression sources. Analysis of this database allowed us to gain some insight into foveomacular-expressed gene expression and gene function. Although the current relational database is qualitative and gene-centric, other data types can be incorporated into the database. Splice variants, quantitative expression levels, and risk factor allele data, for example, are welcome future additions to our collection.

Our foveomacular transcriptome data set collects only relevant data and organizes it into a static database that is periodically updated. We have chosen to do this to avoid the hazards of dynamic data collection; as data sources reorganize, connections are broken and potentially flawed data may be provided to users. A static database avoids those issues and encourages periodic reconsideration of the database structure.

Although our data collection described in this paper, will be useful to the foveomacular research community, our data management methods presented here will be useful to any group desiring focused data. Thoughtful organization of data greatly enhances the value of those data and makes possible insights that would otherwise be obscured.

APPENDIX 1.

Foveome MySQL tables and database schema. Working knowledge of MySQL is strongly recommended before using these files. To access the data, click or select the words "Appendix 1."

APPENDIX 2.

Genes identified by all three contributory data sets ('foveome') and re-identification in Incyte and Lib.420 studies. A '1' indicates identification in that set, a '0' indicates non-identification. To access the data, click or select the words "Appendix 2."

APPENDIX 3.

Results of DAVID clustering of 190 genes identified by all three experiments. Clusters with an enrichment score of 0.5 or greater are listed. To access the data, click or select the words "Appendix 3."

APPENDIX 4.

Archive containing Perl scripts used in this study. Included scripts collect dbEST, Gene, UniGene, and Gene Ontology data, and format it into tables appropriate for the included database schema. Also included is a README file detailing

script licensing. To access the data, click or select the words "Appendix 4."

APPENDIX 5. KNOWN RETINOPATHIC GENES RE-IDENTIFIED BY THIS STUDY.

Gene symbol, Gene ID, gene name and identification status are included. To access the data, click or select the words "Appendix 5."

REFERENCES

1. Hendrickson AE. Primate foveal development: A microcosm of current questions in neurobiology. Invest Ophthalmol Vis Sci 1994; 35:3129-33. [PMID: 8045707].

2. McDonald HR, Schatz H, Johnson RN, Madeira D. Acquired macular degeneration. In "Duane's Textbook of Ophthalmology", W. Tasman, editors. Vol 3. Philadelphia: Lippencott; p. 23.

3. Goldsmith TH. Optimization, constraint and history in the evolution of eyes. Q Rev Biol 1990; 65:281-322. [PMID: 2146698].

4. Ting AY, Lee TK, MacDonald IM. Genetics of age-related macular degeneration. Curr Opin Ophthalmol 2009; 20:369-76. [PMID: 19587596].

5. Mellough CB, Steel DH, Lako M. Genetic basis of inherited macular dystrophies and implications for stem cell therapy. Stem Cells 2009; 27:2833-45. [PMID: 19551904].

6. Schulz HL, Goetz T, Kaschkoetoe J, Weber BHF. The Retinome – Defining a reference transcriptome of the adult mammalian retina/retinal pigment epithelium. BMC Genomics 2004; 5:50-[PMID: 15283859].

7. Bowes Rickman C, Ebright JN, Zavodni ZJ, Yu L, Wang T, Daiger SP, Wistow G, Boon K, Hauser MA. Defining the human macula transcriptome and candidate retinal disease genes using EyeSAGE. Invest Ophthalmol Vis Sci 2006; 47:2305-16. [PMID: 16723438].

8. Wistow G, Bernstein SL, Touchman JW, Bouffard G, Wyatt MK, Peterson K, Behal A, Gao J, Buchoff P, Smith D. Grouping and identification of sequence tags (GRIST): Bioinformatics tools for the NEIBank database. Mol Vis 2002; 8:164-70. [PMID: 12107414].

9. Yamamoto M, Wakatsuki T, Hada A, Ryo A. Use of serial analysis of gene expression (SAGE) technology. J Immunol Methods 2001; 250:45-66. [PMID: 11251221].

10. Lee S, Clark T, Chen J, Zhou G, Scott LR, Rowley JD, Wang SM. Correct identification of genes from serial analysis of gene expression tag sequences. Genomics 2002; 79:598-602. [PMID: 11944993].

11. Ge X, Wang SM. Identifying nonspecific SAGE tags by context of gene expression. Methods Mol Biol 2008; 387:199-204. [PMID: 18287633].

12. Bonaldo MF, Lennon G, Soares MB. Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res 1996; 6:791-806. [PMID: 8889548].

13. Somiari RI, Wong P, Bernstein SL. A high-density differential screening strategy for identification of fovea genes with altered expression in the retina during aging. Biol Tech 1999; 13:577-81. .

14. ftp://ftp.ncbi.nih.gov//genbank/release.notes/gb90.release.notes

15. Ziesel A, Bernstein SL, Stepczynski J, Chrenek MA, Yong L, Carlson C, Clum R, Aggarwal A, Chambers ML, Wong PW. Defining fovea gene expression using existing expressed sequence tag (EST) databases. ARVO Annual Meeting; 1999 May 9-14; Fort Lauderdale (FL).

16. 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; 32:301-8. [PMID: 8838792].

17. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003; 4:3-[PMID: 12734009].

18. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nat Protoc 2009; 4:44-57. [PMID: 19131956].

19. Ziesel A, Bernstein S, Wong PW. Defining the human foveomacular transcriptome: compiling distinct data sets. ARVO Annual Meeting; 2004 April 25-29; Fort Lauderdale (FL).

20. Ziesel A, Bernstein S, Wong PW. Further definition of the human foveomacular transcriptome. ARVO Annual Meeting; 2005 May 1-5; Fort Lauderdale (FL).

21. Bernstein SL, Wong P. Regional expression of disease-related genes in human and monkey retina. Mol Vis 1998; 4:24-[PMID: 9815288].

22. Bernstein SL, Borst DE, Wong PW. Isolation of differentially expressed human fovea genes: candidates for macular disease. Mol Vis 1995; 1:4-[PMID: 9238082].