

Transcriptome Profiling of Developing Murine Lens Through RNA Sequencing

Shahid Y. Khan,¹ Sean F. Hackett,¹ Mei-Chong W. Lee,² Nader Pourmand,² C. Conover Talbot Jr,³ and S. Amer Riazuddin¹

¹The Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States

²Department of Biomolecular Engineering, University of California, Santa Cruz, California, United States

³Institute for Basic Biomedical Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States

Correspondence: S. Amer Riazuddin, The Wilmer Eye Institute, Johns Hopkins University School of Medicine, 600 N. Wolfe Street, Maumenee 840, Baltimore, MD 21287, USA; riazuddin@jhmi.edu.

Submitted: December 12, 2014

Accepted: June 4, 2015

Citation: Khan SY, Hackett SF, Lee M-CW, Pourmand N, Talbot CC Jr, Riazuddin SA. Transcriptome profiling of developing murine lens through RNA Sequencing. *Invest Ophthalmol Vis Sci.* 2015;56:4919-4926. DOI:10.1167/iovs.14-16253

PURPOSE. Transcriptome is the entire repertoire of transcripts present in a cell at any particular time. We undertook a next-generation whole transcriptome sequencing approach to gain insight into the transcriptional landscape of the developing mouse lens.

METHODS. We ascertained mouse lenses at six developmental time points including two embryonic (E15 and E18) and four postnatal stages (P0, P3, P6, and P9). The ocular tissue at each time point was maintained as two distinct pools serving as biological replicates for each developmental stage. The mRNA and small RNA libraries were paired-end sequenced on Illumina HiSeq 2000 and subsequently analyzed using bioinformatics tools.

RESULTS. Mapping of mRNA and small RNA libraries generated 187.56 and 154.22 million paired-end reads, respectively. We detected a total of 14,465 genes in the mouse ocular lens at the above-mentioned six developmental stages. Of these, 46 genes exhibited a 40-fold differential (higher or lower) expression at one the five developmental stages (E18, P0, P3, P6, and P9) compared with their expression level at E15. Likewise, small RNA profiling identified 379 microRNAs (miRNAs) expressed in mouse lens at six developmental time points. Of these, 49 miRNAs manifested an 8-fold differential (higher or lower) expression at one the five developmental stages, as mentioned above compared with their expression level at E15.

CONCLUSIONS. We report a comprehensive profile of developing murine lens transcriptome including both mRNA and miRNA through next-generation RNA sequencing. A complete repository of the lens transcriptome of six developmental time points will be monumental in elucidating processes essential for the development of the ocular lens and maintenance of its transparency.

Keywords: next-generation sequencing, transcriptome, lens development, mRNA, miRNA

The ocular lens is an excellent model for investigating both intricate details of development and pathophysiological mechanism of disease. The ocular lens is a transparent biconvex structure that helps focus light on the retina. The maintenance of lens transparency is essential for the normal visual function, and lack thereof leads to visual impairment by development of lens opacities. The lens arises from the head ectoderm that thickens to form the lens placode.^{1,2} The lens placode invaginates together with the optic vesicle to form the lens pit and the optic cup, respectively.^{1,2} Subsequently, the lens pit separates from the ectoderm to form the lens vesicle that gives rise to anterior and posterior single layers of cells.^{1,2} Cells of the former layer differentiate into the epithelium while cells forming the posterior layer differentiate into primary fiber cells forming the lens nucleus.^{1,2} The lens grows rapidly during late embryonic and early postnatal stages due to proliferation of epithelial cells that elongate and differentiate into secondary fiber cells continuously adding to the fiber mass.^{1,2}

Cataract, clouding of the lens, is the leading cause of blindness worldwide and accounts for one-third of the total cases of blindness in children.^{3,4} A significant fraction of cataractogenesis is familial with a total of 35 genes associated with congenital cataracts.^{5,6} Nonetheless, these genes constitute a small fraction

of the total genetic load, and the total number of genes associated with this debilitating disorder is presumed to be much higher.

Previously, serial analysis of gene expression (SAGE) and RNA microarrays have been used to investigate transcriptional activities of the mouse and human ocular lens.⁷⁻⁹ However, both of these techniques have inherent limitations including the extremely short length of the SAGE tags that may be shared by multiple genes, and the inbuilt limitation of microarrays that can detect only those transcripts included on the array itself. Next-generation sequencing (NGS) technology provides an efficient and cost-effective platform to investigate the transcriptional landscape accurately and with much greater sensitivity, which is essential to reliably detect and quantitate rare but physiologically relevant coding as well as noncoding RNA molecules. Recently, two studies reported to use NGS to investigate the mouse lens transcriptome.^{10,11} However, neither study examined the lens transcriptome through different developmental stages nor was microRNA (miRNA) investigated in the lens transcriptome.

miRNAs are important regulatory molecules that act as post transcriptional regulators and according to some estimates each miRNA can regulate the expression of 100 to 200 target genes.¹² Severe microphthalmia with no visible lens and a poorly stratified

corneal epithelium is observed in Dicer conditional knockout mice.¹³ Wolf and colleagues¹⁴ confirmed the indispensable role of miRNA-based regulation in differentiation and development of the lens tissue. This notion was further strengthened with the identification of the role of *Pax6* in the regulation of miR-204, where the latter acts as a negative modulator of gene expression during lens development in vertebrates.¹⁵

In this study, we employed NGS to create mouse lens expression profiles for both mRNA and miRNA at six developmental stages. Deep sequencing of mRNA libraries generated 187.56 million unique reads that identified expression of 14,465 genes in the mouse lens transcriptome. Likewise, sequencing of the small RNA libraries generated 154.22 million unique reads that retrieved 379 miRNAs expressed at different developmental stages.

MATERIALS AND METHODS

Samples Collection and RNA Isolation

The use of mice in this study was approved by Johns Hopkins Animal Care and Use committee (ACUC; Baltimore, MD, USA), and all experiments were performed in accordance with a protocol approved by Johns Hopkins ACUC. We acquired pregnant female C57BL/6 mice from Charles River laboratory (Wilmington, MA, USA). Lenses were obtained at six different developmental stages including embryonic day 15 and 18, and at postnatal day 0, 3, 6, and 9. To obtain the lenses, mice were first anesthetized by isoflurane and subsequently euthanized through cervical dislocation. The lenses were isolated using forceps under a microscope. Two biological replicates, each consisting of 8 to 9 pups, were used for each developmental stage. Lenses were dissolved immediately in Trizol reagent (Invitrogen, Carlsbad, CA, USA) and total RNA was extracted from each pool according to manufacturer's instructions. The quality and quantity of the total RNA was determined on a NanoDrop Lite spectrophotometer (Thermo Scientific, Inc., Grand Island, NY, USA).

Library Preparation for RNA Sequencing

We used Illumina TruSeq RNA Sample Preparation Kit v2 (Illumina, Inc., San Diego, CA, USA) to prepare mRNA sequencing libraries, according to manufacturer's instructions. Briefly, 4 µg of total RNA was used for polyA mRNA selection using streptavidin-coated magnetic beads followed by thermal fragmentation of selected mRNA. The fragmented mRNA was used as a template for cDNA synthesis by reverse transcriptase with random primers. This cDNA was further converted into double stranded DNA that was end-repaired to incorporate the specific index adaptors for multiplexing, followed by a purification step and amplification for 15 cycles. The quality and functionality of final amplified libraries were examined with a DNA high sensitivity chip on an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA) and quantitative (q)PCR according to manufacturer's instructions. Subsequently, RNA-Seq libraries with unique index sequences were pooled in equimolar ratio and the final size selection of 400 to 500 bp was performed on a Caliper LabChip XT DNA 750 chip (Caliper Life Sciences, Hopkinton, MA, USA).

Small RNA Library Preparation for miRNA Sequencing

We used Illumina TruSeq Small RNA Sample Prep Kit for small RNA library preparation according to manufacturer's instructions. Briefly, 1 to 2 µg of total RNA without selection for small

RNAs was processed for ligation to a 3' RNA adapter followed by ligation to 5' RNA adapters. The adapter-ligated RNA was subsequently converted to single-stranded cDNA using SuperScript II Reverse Transcriptase. The cDNA was amplified for 11 cycles using a common primer and primers with unique index sequences. The small RNA libraries were purified using AmPure XP beads (Beckman Coulter, Indianapolis, IN, USA) and examined on Agilent 2100 Bioanalyzer. Finally, the 130- to 170-bp size selection was performed on a 2% agarose gel.

Next-Generation RNA Sequencing

Initially, we sequenced two mRNA libraries (MLT1 and MLT2), two biological replicates representing the postnatal Day 3 on a MiSeq genome analyzer (Illumina, Inc.). Briefly, 20 pM of P3 library was used for cluster generation, followed by paired-end (2 × 150 bp) sequencing on the MiSeq genome analyzer. The total output reads were deindexed (separation of reads) based upon the unique bar-codes using Cassava software (v1.8; Illumina, Inc.).

Subsequently, all 12 mRNA and 12 small RNA bar-coded pooled libraries were clustered using TruSeq V3 (Illumina, Inc.) flow cells at 13 and 16 pM concentration, respectively. These libraries were paired-end sequenced using Illumina's TruSeq SBS Kit V3 (Illumina, Inc.) in two independent lanes on a HiSeq 2000 genome analyzer (2 × 100 bp; Illumina, Inc.) followed by deindexing based upon the unique barcodes using Cassava software (v1.8).

Bioinformatics and Data Analyses

The mRNA sequencing data were analyzed using two different pipelines: DNASTAR (DNASTAR, Madison, WI, USA), a commercially available tool, and Tophat (version 2.0.9; University of Maryland, College Park, MD, USA) with DESeq (version 1.15.3; EMBL, Heidelberg, Germany).^{16,17} Initially, we used DNASTAR for analysis of the sequencing data. The reads were aligned into an assembly with the SeqMan NGen version 11 using the default parameters. The *Mus musculus* (GRCm38.p2) reference genome was used for the mouse lens transcriptome assembly. The gene expression quantification, normalization, and statistical analysis were performed with ArrayStar Version 11 (DNASTAR) using default parameters. The expression data were normalized by calculating the reads per kilobase per million mapped reads (RPKM) values for each gene.¹⁸

In parallel, we employed Tophat/DESeq for an independent replication of the analysis performed on the DNASTAR platform. The analysis included preprocessing of FASTQ file format sequencing reads by removing the adapter sequences using SeqPrep (in the public domain, <https://github.com/jstjohn/SeqPrep>). The preprocessed sequencing reads were aligned against the *Mus musculus* (GRCm38/mm10) transcriptome using default parameters. Subsequently, duplicate reads caused by PCR amplification were removed using the rmdup option of SAMtools (version 0.1.18; Wellcome Trust Sanger Institute, Cambridge, UK).¹⁹ The proposed gene ontologies were identified using the Protein Analysis Through Evolutionary Relationships (PANTHER) classification system (in the public domain, <http://www.pantherdb.org/>).²⁰

The FASTQ files of sequencing reads for small RNA were subject to adapter sequence removal by SeqPrep (in the public domain, <https://github.com/jstjohn/SeqPrep>) followed by classification of reported and novel miRNAs using miRDeep2.²¹ Briefly, the sequencing short reads were aligned against the *Mus musculus* (GRCm38/mm10) reference genome with the mapper module of miRDeep2 that uses Bowtie (version 0.12.8; University of Maryland). Reported miRNAs were identified by miRDeep2 through reported miRNA hairpins and mature

miRNA sequences downloaded from the miRBase (Release 20) database (University of Manchester, Manchester, UK).²²

We used RPKM (mRNA) and RPM (miRNA) expression data from individual time points for principal component analysis (PCA) to perform direct comparisons between time points in developing ocular lens. Partek Genomics Studio (Partek, Inc., St. Louis, MO, USA) was used to create PCA plots where the expression values were in quantile normalized log₂ notation. In parallel, we created expression profile plots to examine mRNA and miRNA differential transcription using Spotfire DecisionSite with Functional Genomics (TIBCO Spotfire, Boston, MA, USA). The earliest developmental stage, E15, was set as the reference point, and the remaining five time points were separately compared to E15, thus yielding individual transcripts' fold change between E15 and increasingly greater development. For graphical purposes a virtual E15 to E15 comparison (0.0-fold change of course) was included.

RESULTS

To investigate the transcriptional landscape of ocular lens during development, we performed high-throughput RNA sequencing on mouse lens at two embryonic (E15 and E18) and four postnatal stages (P0, P3, P6, and P9), where E and P refer to embryonic and postnatal respectively, while the number corresponds to the day in that phase. The extracted lenses from each developmental stage were divided into two biological replicates for total RNA extraction. Twelve mRNA and 12 small RNA sequencing libraries (E15a, E15b, E18a, E18b, P0a, P0b, P3a, P3b, P6a, P6b, P9a, and P9b) were prepared using total RNA from each developmental stage as described in materials and methods.

Initially, we sequenced two mRNA libraries (MLT1 and MLT2), two biological replicates representing the postnatal Day 3, on a MiSeq genome analyzer. The deindexed data yielded a total of 10,213,826 and 13,547,058 reads for the two libraries, respectively, that were uploaded to SeqMan NGen version 11 for quality check and mapping to the *Mus musculus* genome. Quality control (QC) examination of these sequencing reads revealed that 99.5% of the reads were of approximately 150 bp. Of these, only 0.08% had ambiguous base calls with greater than 95% of the sequencing data yielding a PHRED score of 30 or above (a PHRED score of 30 represents 0.1% probability of a false calling). Approximately 60% of all sequencing reads demonstrated GC content of 40% to 60% while the remaining 40% reads represented a mixture of low (25%–40%) and high GC content (60%–70%). Our results suggest a homogenous distribution of each of the four nucleotide bases at each particular position of the read length ranging from 20 to 150 bp.

RNA sequencing data of both biological replicates were independently mapped against the *Mus musculus* (GRCm38.p2) reference genome, which resulted in 89.48% and 89.57% alignment for the two biological replicates, respectively. The expression data were normalized by calculating the reads per kilobase per million mapped reads (RPKM)¹⁸ values for each gene and the retest reliability conducted by examining the RPKM values of these biological replicates suggested a high correlation between both replicates with a Pearson coefficient value of 0.98 (Supplementary Fig. S1). The high reproducibility of these two replicates and successful outcome of the quality control parameters boosted our confidence. Subsequently, we pooled the 12 barcoded mouse lens mRNA and the 12 barcoded small RNA libraries and sequenced them in two independent lanes on a HiSeq 2000 genome analyzer (2 × 100 bp). All raw and processed sequencing data reported in this manuscript have been

deposited in NCBI's Gene Expression Omnibus,²³ and are accessible through GEO series accession number GSE69221.

The output reads were deindexed based upon the unique bar-codes using Cassava software, resulting in a total of 394.52 million raw reads. These deindexed raw reads were initially mapped to *Mus musculus* genome by SeqMan NGen version 11; of which 164.78 million paired-end reads passed the quality control filters and were uniquely aligned to the reference genome (data not shown). In parallel, the 394.52 million deindexed reads were also mapped to *Mus musculus* genome with Tophat algorithms. Of these, 187.56 million paired-end reads properly aligned to the mouse genome after removal of PCR duplicates by the rmdup option of SAMtools, which identified 58.63 million (19.0%) reads to be PCR duplicates (Table 1). DNASTAR and Tophat revealed 164.78 million and 187.56 million uniquely mapped paired-end reads representing nearly 275× and 313× sequence coverage, respectively, for a 60 Mb *Mus musculus* transcriptome.

The mapped reads (187.56 million) were assembled into transcripts, and the expression was measured and normalized using RPKM method. The analyses revealed expression (≥ 1.0 RPKM) of 14,465 genes present in at least one of six developmental stages (Table 2; Supplementary Table S1), which suggests that approximately 62.48% of the total mouse coding exome is expressed in lens transcriptome. We observed a higher number of genes expressed in the embryonic stages compared with number of genes expressed at postnatal stages (Table 2).

Subsequently, we investigated gene ontologies (GO) of the genes expressed in six developmental stages using the PANTHER classification system. The gene ontologies of mouse lens transcripts were characterized on the basis of the molecular function (MF), biological process (BP), and protein classification (PC). The analysis revealed a total 10 unique molecular function GO terms associated with mouse lens mRNAs that includes binding (34%), catalytic activity (34%), transcription regulator activity (10%), and (7%) structural molecule activity (Supplementary Fig. S2A–S2F). Similarly, a total of 14 biological process GO terms were identified in embryonic and postnatal mouse lens (Supplementary Fig. S3A–S3F). The mouse lens transcriptome revealed 29 distinctive protein classification GO terms including nucleic acid binding (16%), transcription factors (10%), cytoskeletal proteins (4.9%), chaperone (1.7%), and (0.7%) structural proteins (Supplementary Fig. S4A–S4F).

Deindexing of the raw data from the small RNA sequencing libraries yielded 239.25 million filter-passing reads (Table 1). The reads were subsequently processed with SeqPrep that removed 24.28 million (10.14%) adapter sequences (Table 1). The 214.97 million trimmed (ranging from 20–25 bp in size) reads were mapped to the *Mus musculus* reference genome. Of which, 154.22 million (71.73%) showed unique alignment while 19.26 million reads (8.96%) mapped to multiple targets (>5) in the mouse reference genome (Table 1). Mapping to the mouse genome generated 8 to 27 million uniquely aligned reads for 11 of the 12 small RNA libraries while one of the biological replicates of P3 produced only 2.3 million mapped reads (Table 1).

We searched for reported or novel miRNAs by examining the aligned data for miRNA hairpins and mature miRNA sequences. These analyses discovered expression (≥ 1.0 RPM) of 379 miRNAs present in at least one of the six developmental stages (Table 2; Supplementary Table S2), which constitutes approximately 32% of the mouse microRNAome. It is worthy of note that the total number of microRNA expressed at embryonic Day 15 (379) decreased to a total of 292 miRNAs by postnatal day 9 in mouse lens (Table 2). miR-92a-3p, miR-204-5p, miR-184-3p, miR-181a-5p, miR-99b-5p, miR-26a-5p, miR-92b-3p, and miR-127-3p were the overall most highly

TABLE 1. The Summary and Statistics of the Mouse Lens Next-Generation Whole Transcriptome Sequencing

Mouse Lens Transcriptome Sequencing (2 × 100 bp)								
D. Stage	Sample ID	TR, 10 ⁶	TR w/o Adapter, 10 ⁶	Total MR, 10 ⁶	MR w/o PCR Duplication, 10 ⁶	PPR w/o PCR Duplication, 10 ⁶	PPR w/o PCR Duplication, %	Sequenced Bases, Mb
E15	MLT-E15A	27.08	26.95	21.41	19.81	15.34	77.44	1534.61
	MLT-E15B	27.54	27.34	21.35	18.30	13.77	75.25	1377.30
E18	MLT-E18A	35.62	35.32	28.02	24.97	19.55	78.29	1955.12
	MLT-E18B	48.41	47.96	38.49	34.09	27.45	80.52	2745.63
P0	MLT-P0A	29.06	28.76	22.74	17.71	13.26	74.87	1326.25
	MLT-P0B	35.53	35.15	27.77	22.20	16.87	75.99	1687.54
P3	MLT-P3A	26.14	25.90	20.38	16.32	11.89	72.89	1189.77
	MLT-P3B	30.99	30.61	24.15	19.09	14.24	74.59	1424.47
P6	MLT-P6A	33.93	33.53	26.12	19.57	14.18	72.50	1418.92
	MLT-P6B	35.31	34.92	27.18	20.19	14.70	72.85	1470.82
P9	MLT-P9A	37.74	37.29	28.89	21.02	14.97	71.25	1497.85
	MLT-P9B	27.12	26.87	20.96	15.57	11.28	72.45	1128.19
Total		394.52	390.65	307.52	248.89	187.56	75.36	18756.54

Mouse Lens Small RNA Sequencing (2 × 100 bp)								
D. Stage	Sample ID	TR, 10 ⁶	TR w/o Adapter, 10 ⁶	UMR, 10 ⁶	UMR, %	MMR, >5, 10 ⁶	MMR, >5, %	Sequenced Bases, Mb
E15	MLM-E15A	19.11	17.41	10.82	62.16	2.95	16.96	1082.85
	MLM-E15B	35.06	30.19	16.23	53.75	4.85	16.09	1623.15
E18	MLM-E18A	28.27	25.71	18.23	70.90	3.21	12.50	1823.58
	MLM-E18B	37.69	35.92	27.42	76.35	2.67	7.44	2742.88
P0	MLM-P0A	12.13	10.68	8.26	77.29	5.59	5.23	826.23
	MLM-P0B	17.31	15.25	11.73	76.91	8.64	5.67	1173.17
P3	MLM-P3A	3.61	3.13	2.37	75.68	1.23	3.94	237.63
	MLM-P3B	14.66	13.06	10.49	80.33	5.60	4.29	1049.68
P6	MLM-P6A	18.09	16.08	12.42	77.23	8.87	5.52	1242.69
	MLM-P6B	15.62	13.86	10.66	76.92	7.93	5.72	1066.69
P9	MLM-P9A	20.92	18.63	13.80	74.08	9.13	4.90	1380.26
	MLM-P9B	16.71	14.98	11.73	78.31	8.63	5.76	1173.61
Total		239.25	214.97	154.22	71.73	19.26	8.96	15422.48

D. Stage, developmental stage; TR, total reads; MR, mapped reads; PPR, properly paired reads; UMR, uniquely mapped reads; MMR, multi mapped reads.

expressed miRNAs in mouse embryonic lens (Supplementary Table S2).

Principal component analysis can provide an overall understanding of the expression differential, in particular, cellular environments. It illustrates how samples at particular time points differ by using all the transcript values for all the samples and subsequently grouping these samples by similarity to represent many dimensions of the data (one per transcript) condensed into three dimensions. We employed PCA to provide global views of mRNA and miRNA expression in developing lens. These analyses reveal that there is indeed a time course running from E15 to P9, reflected in these time points chronological order across the first, most significant, dimension or component of the plots (Fig. 1A, 1B).

Time course analyses were run to depict how the transcription of mRNA and miRNA changes across time in the developing lens, to intuitively track the trajectory of each transcript across time. The expression profile plots show that a considerable proportion of the transcripts exhibit a change in their expression levels, either up or down, from their levels at E15 and, further, the degree of change increases as time passes (Fig. 2A, 2B). We identified a total of 46 genes that exhibited 40-fold expression differential compared with transcriptional levels at E15 (Supplementary Table S3). Likewise, we identified 49 miRNAs that manifested an 8-fold or higher expression

TABLE 2. The Total Number of Genes and miRNAs Identified in Multiple Stages of Mouse Lens Development

Developmental Stage	Genes ≥1.0 RPKM	Genes Expressed at Least in One of the Developmental Stages
E15	13,274	14,465
E18	13,900	
P0	12,560	
P3	12,940	
P6	12,130	
P9	12,229	

Developmental Stage	miRNAs ≥1.0 RPM	miRNAs Detected at Least in One of the Developmental Stages
E15	379	379
E18	331	
P0	316	
P3	316	
P6	303	
P9	292	

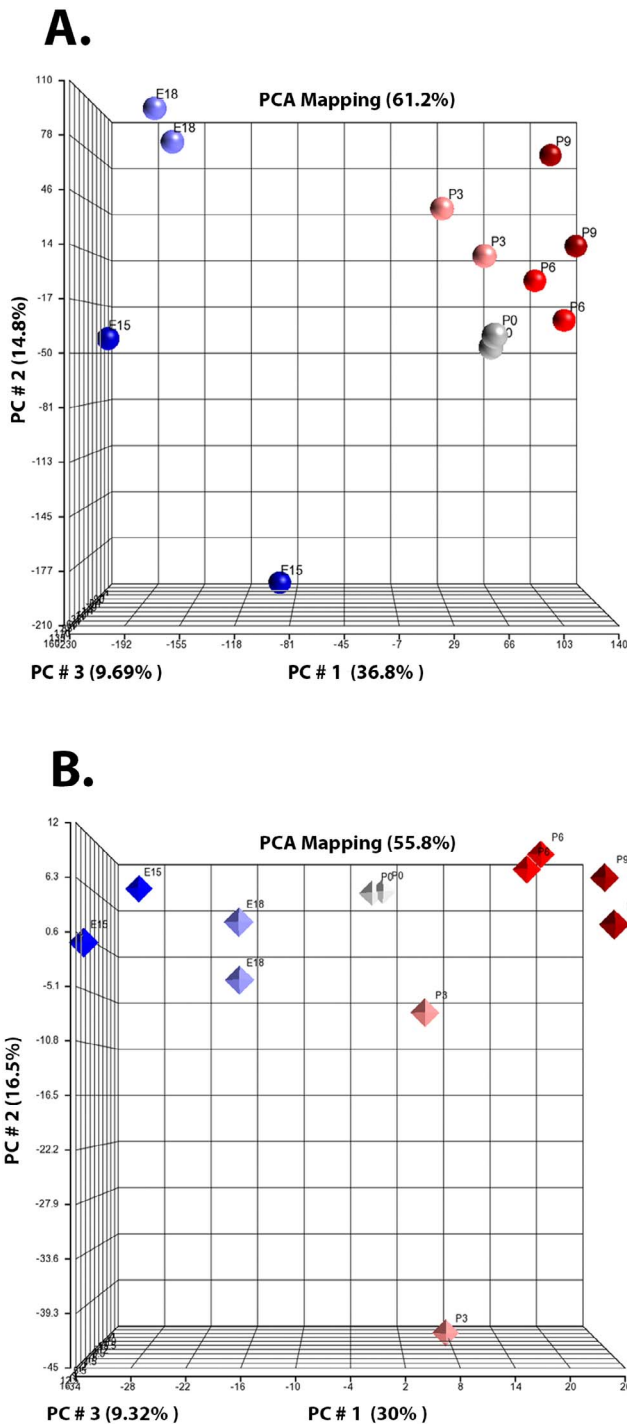


FIGURE 1. Principal component analysis illustrating direct comparisons among developmental time points with expression values from individual time points normalized to log₂ RPKM and RPM for (A) mRNA and (B) miRNA, respectively. The X, Y, and Z axes depict the three largest components of the total expression variation expressed in percent of the total, PC # 1, PC # 2, and PC # 3, respectively. These analyses reveal that there is a time course running from E15 to P9, which is reflected in these time points' chronological order across the first, most significant, dimension, or component of the plots. Note: the axis values are essentially arbitrary.

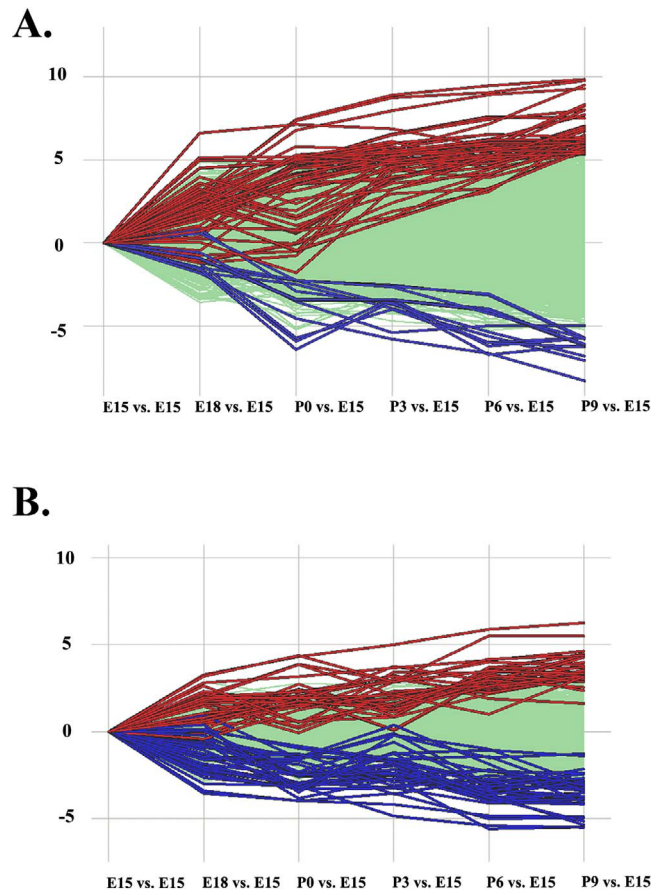


FIGURE 2. Time-course expression profile plot representing the developmental time-point comparisons for (A) mRNA and (B) miRNA. The earliest developmental stage, E15, was set as the starting point, and the other five-time point were separately compared with E15, thus yielding individual transcripts' fold changes between E15 and increasingly greater development. The Y axis depicts the log₂ fold change of the expression at that time point versus expression at E15. Transcripts that increased greatly (40-fold or greater for mRNA; 8-fold or greater for miRNA) are highlighted in *red*, and those that decreased greatly (minus 40-fold or less for mRNA; 8-fold or less for microRNA) are highlighted in *blue* and the identities provided in Supplementary Tables S3 and S4. Note: a virtual E15 to E15 comparison (0.0-fold change of course) was included for graphical illustration.

differential when compared with miRNA expression at E15. (Supplementary Table S4). It is worthy of note that mRNA expression profiles show a greater differential regulation from E15 levels compared with miRNAs (Fig. 2A, 2B; Supplementary Tables S3, S4).

DISCUSSION

Here, we report a comprehensive mRNA and miRNA transcriptome of mouse embryonic and postnatal lens tissue using next-generation RNA sequencing. The lenses were obtained from two embryonic and four postnatal stages, and two sequencing libraries were prepared for each of the six developmental time points; each library was then paired-end sequenced on HiSeq 2000 genome analyzer. Deep sequencing of mRNA libraries generated 187.56 million unique reads that identified expression of 14,465 genes in the mouse lens transcriptome. Likewise, sequencing of the small RNA libraries generated 154.22 million unique reads that retrieved 379 miRNAs expressed at different developmental stages. Taken

together, we present a complete repository of the lens transcriptome including mRNA and miRNA profiles at six developmental time points.

The earliest transcriptional details of the lens tissue come from expressed sequence tag (EST) analysis and microarray-based investigation of the human and mouse lens tissue, respectively. Initially, Wistow and colleagues⁷ explored the expression profile of the human lens through EST analysis and identified 1455 unique groups, of which, only two-thirds correspond to named genes in GenBank. Wride and colleagues⁸ investigated gene expression in mouse lenses using microarrays and identified 1668 genes in mouse lenses at levels significantly above background.

More recently, NGS-based transcriptome analyses of single developmental stages have been reported.^{10,11} Manthey and colleagues¹⁰ used high-throughput RNA sequencing that identified expression of more than 7700 genes at levels greater than 2 RPKM in E15.5 mouse lens. Of these, 7605 genes were present in our E15 mouse lens RNA-Seq dataset. Likewise, Hoang and colleagues¹¹ investigated expression profiles mouse lens epithelial and fiber cells and reported 13,732 and 10,850 genes expressed in lens epithelial and fiber cells, respectively. Although the total number of genes identified by Hoang and colleagues,¹¹ and that present in our developmental transcriptome is comparable (~14,000), surprisingly, only 75% (10,618 genes) of the transcriptome was found to be common in both datasets.

Cataracts are the leading cause of blindness worldwide with surgical removal of the cataractous lens providing the only treatment. The Cat-Map database includes 198 genes associated with congenital and age-related cataracts.⁵ All of them were expressed in at least one of the six mouse lens developmental time points examined in the current study. Among these, we identified 26 genes associated with autosomal dominant, and 11 genes associated with autosomal recessive congenital cataracts (Supplementary Table S5). We also identified 17 and 165 genes associated with age-related cataracts and cataractogenesis with other ocular and nonocular anomalies, respectively.

A majority of the genes in our dataset exhibited steady expression levels throughout the six developmental stages investigated in this study. Nonetheless, we identified genes that were significantly up- or downregulated in either embryonic or postnatal time points (Supplementary Table S3). Among these, we found crystallin beta B2 (*CryβB2*), collagen, type IV, alpha 3 and 4 (*Col4α3* and *Col4α4*) exhibited 866-, 253- and 630-fold higher expression, respectively, in P9 lenses compared with their levels at E15 (Supplementary Tables S1, S3) consistent with expression patterns reported previously.^{24,25} Likewise, we found that *Fgf1* displayed a 40-fold higher expression in P9 lenses compared with levels at E15 (Supplementary Tables S1, S3) consistent with its role reported previously in lens epithelial cell proliferation and fiber cell differentiation.^{26,27} Interestingly, *Fgf15/19* exhibited a higher expression during the embryonic stages that decreased substantially as the mice aged, leading to a 77-fold downregulation in P9 lenses compared with levels at E15 (Supplementary Tables S1, S3).

Sfrp2, a secreted frizzled-related protein is one of the earliest markers of lens induction that has shown to be expressed at E9.5 in mouse lens placode.²⁸ We identified two secreted frizzled-related protein genes, *Sfrp1* and *Sfrp2* in the mouse lens transcriptome. *Sfrp1* exhibited a steady expression pattern throughout the developmental stages investigated in this study while *Sfrp2* illustrated high expression at E15 that decreased almost 300-fold in P9 lenses (Supplementary Tables S1, S3). In contrast, lensin, periplakin, and desmoyokin are lens-specific proteins that are expressed dominantly in lens fiber cells. Lensin expresses dominantly in terminally differ-

entiating secondary lens fiber cells while periplakin and desmoyokin are expressed dominantly in mature fiber cells.^{29,30} Consistent with the previously reported expression pattern, all three genes exhibited a higher expression in P9 lenses compared with E15 (Supplementary Tables S1, S3).

We recently discovered multiple loss-of-function mutations in *FYCO1* responsible for autosomal recessive congenital cataract.³¹ *FYCO1* has been shown to promote plus-end directed microtubule transport.^{31,32} Taken together, the presence of components of autophagy in both lens epithelium and lens fiber cells suggests that autophagy is a critical process for maintenance of lens transparency. Matsui and colleagues³³ observed autophagy in embryonic lens cells and, more recently, Morishita and colleagues³⁴ reported that deletion of *Atg5* and *Pik3c3* in lens results in development of cataracts independent of programmed organelle degradation. To further explore the role of autophagy in developing ocular lens, we searched for autophagy-associated genes in the mouse lens transcriptome. Interestingly, we identified a total of 133 autophagy-associated genes including the 39 genes previously identified in human lens epithelium and fiber cells (Supplementary Table S6). A majority of these autophagy-associated genes (121) exhibit a gradual increase in expression from embryonic to postnatal mouse lens with only 12 autophagy-associated genes maintaining a steady expression from embryonic to postnatal lens (Supplementary Table S6). In a recent study, transcriptome analysis identified expression of 13,000 unique transcripts in the embryonic chicken lens (E13).³⁵ Interestingly, there is a relatively higher expression of autophagy-associated genes in both postnatal chicken and mouse lens transcriptomes consistent with the notion that autophagy plays a critical role in lens fiber cell differentiation.

miRNA have shown to be important regulators of protein-coding transcripts and miRNA profiling has been performed extensively in many mammalian tissues. Nevertheless, only a few microarray-based global miRNA expression studies have been reported for mouse and human ocular lens and human lens cell lines.³⁶⁻³⁹ In contrast, we used massively parallel deep sequencing technology to generate a comprehensive miRNA profile of embryonic and postnatal mouse lens. Our analyses identified miR-184 as the highly expressed miRNA in the mouse lens transcriptome while miR-204 exhibited the second highest expression. Both miR-184 and miR-204 exhibited a uniform expression through various stages of mouse lens development examined in this study. We detected mature miR-204 originating from both the 3p and 5p arms of the precursor miRNA with a dominant expression of miR-204-5p in the mouse lens. Previously, a heterozygous mutation in the seed region of miR-184 has been associated with familial autosomal dominant form of anterior segment dysgenesis.^{40,41} Likewise, miR-204 knock-down has been associated with abnormal Meis2-mediated regulation of the Pax6 transcriptional network that results in microphthalmia and abnormal lens formation.⁴²

Kubo and colleagues⁴³ investigated differentially expressed miRNAs by taking advantage of the Shumiyi cataract rat (SCR), a hereditary cataract model. The authors identified a significant downregulation of let-7c, miR-29a, miR-29c, and miR-126 in SCR lens. We detected expression of let-7c, miR-29a, and miR-29c exhibiting a consistent increase from embryonic to postnatal stages. Likewise, we identified both miR-126a and miR-126b with a steady expression over the six developmental stages investigated in this study (Supplementary Table S2). Moreover, we detected mature miR-126 originating from both arms of precursor miRNA with dominant expression of miR-126a and miR-126b from 5p and 3p, respectively, in mouse lens (Supplementary Table S2).

miRNA and FGF signaling regulate a wide range of cellular processes.¹⁴ Wolf and colleagues¹⁴ examined an *FGF2*-induced rat lens epithelial explant system that identified 131 FGF2-regulated miRNAs. They further demonstrated that miR-143, miR-155, and miR-301 downregulate expression of c-Maf, a lens-differentiation factor.¹⁴ We detected uniform expression of miR-143 and miR-155 in all six stages while miR-301 exhibited a higher expression in embryonic lenses that gradually decreased as the mice aged.

miRNA arm switching (3p and 5p) is a common mechanism that generates different mature miRNAs from the same pre-miRNA hairpin for regulation of targets genes in different tissues, or even within the same tissue at different developmental stages.^{44–48} We identified expression of 379 unique mature miRNAs in embryonic and postnatal mouse lens (Supplementary Table S2). Of these, 201 miRNAs exhibited expression from both 5p and 3p arms of the pre-miRNA hairpin while 52 and 70 miRNAs displayed expression from only the 5p arm or from only the 3p arm, respectively. Moreover, we identified 56 miRNAs in the mouse lens that did not exhibit any arm designation. Last but not least, we did not observe any arm switching between embryonic and postnatal stages in mouse lens, as previously observed in brain, ovary, testes, embryos, and newborns mice.⁴⁴

In conclusion, we present a complete repository of the lens transcriptome including mRNA and miRNA profiles at six developmental time points using next-generation RNA sequencing. Investigating the transcriptional landscape in model tissue that closely mimics human ocular lens will serve as a resource for lens research community and add to our understanding of the transcriptome's role in the maintenance of lens transparency and in disease manifestation.

Acknowledgments

The work was supported by National Eye Institute Grant R01EY022714 (SAR; Bethesda, MD, USA).

Disclosure: **S.Y. Khan**, None; **S.F. Hackett**, None; **M.-C.W. Lee**, None; **N. Pourmand**, None; **C.C. Talbot Jr**, None; **S.A. Riazuddin**, None

References

- McAvoy JW, Chamberlain CG, de Iongh RU, Hales AM, Lovicu FJ. Lens development. *Eye (Lond)*. 1999;13(pt 3b):425–437.
- Lovicu FJ, McAvoy JW. Growth factor regulation of lens development. *Dev Biol*. 2005;280:1–14.
- Robinson GC, Jan JE, Kinnis C. Congenital ocular blindness in children, 1945 to 1984 1502. *Am J Dis Child*. 1987;141:1321–1324.
- Hejtmancik JF, Smaoui N. Molecular genetics of cataract. *Dev Ophthalmol*. 2003;37:67–82.
- Shiels A, Bennett TM, Hejtmancik JF. Cat-Map: putting cataract on the map. *Mol Vis*. 2010;16:2007–2015.
- Shiels A, Hejtmancik JF. Genetics of human cataract. *Clin Genet*. 2013;84:120–127.
- Wistow G, Bernstein SL, Wyatt MK, et al. Expressed sequence tag analysis of adult human lens for the NEIBank Project: over 2000 non-redundant transcripts, novel genes and splice variants. *Mol Vis*. 2002;8:171–184.
- Wride MA, Mansergh FC, Adams S, et al. Expression profiling and gene discovery in the mouse lens. *Mol Vis*. 2003;9:360–396.
- Lachke SA, Ho JW, Kryukov GV, et al. iSyTE: integrated Systems Tool for Eye gene discovery. *Invest Ophthalmol Vis Sci*. 2012; 53:1617–1627.
- Manthey AL, Terrell AM, Lachke SA, Polson SW, Duncan MK. Development of novel filtering criteria to analyze RNA-sequencing data obtained from the murine ocular lens during embryogenesis. *Genom Data*. 2014;2:369–374.
- Hoang TV, Kumar PK, Sutharzan S, Tsonis PA, Liang C, Robinson ML. Comparative transcriptome analysis of epithelial and fiber cells in newborn mouse lenses with RNA sequencing. *Mol Vis*. 2014;20:1491–1517.
- Krek A, Grun D, Poy MN, et al. Combinatorial microRNA target predictions. *Nat Genet*. 2005;37:495–500.
- Li Y, Piatigorsky J. Targeted deletion of Dicer disrupts lens morphogenesis, corneal epithelium stratification, and whole eye development. *Dev Dyn*. 2009;238:2388–2400.
- Wolf L, Gao CS, Gueta K, et al. Identification and characterization of FGF2-dependent mRNA: microRNA networks during lens fiber cell differentiation. *G3 (Bethesda)*. 2013;3:2239–2255.
- Shaham O, Gueta K, Mor E, et al. Pax6 regulates gene expression in the vertebrate lens through miR-204. *PLoS Genet*. 2013;9:e1003357.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36.
- Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11:R106.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5:621–628.
- Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079.
- Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*. 2013;8:1551–1566.
- Friedlander MR, Chen W, Adamidi C, et al. Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol*. 2008;26:407–415.
- Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 2014;42:D68–D73.
- Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30:207–210.
- Van Leen RW, van Roozendaal KE, Lubsen NH, Schoenmakers JG. Differential expression of crystallin genes during development of the rat eye lens. *Dev Biol*. 1987;120:457–464.
- Kelley PB, Sado Y, Duncan MK. Collagen IV in the developing lens capsule. *Matrix Biol*. 2002;21:415–423.
- Chamberlain CG, McAvoy JW. Induction of lens fibre differentiation by acidic and basic fibroblast growth factor (FGF). *Growth Factors*. 1989;1:125–134.
- Robinson ML, Overbeek PA, Verran DJ, et al. Extracellular FGF-1 acts as a lens differentiation factor in transgenic mice. *Development*. 1995;121:505–514.
- Wawersik S, Purcell P, Rauchman M, Dudley AT, Robertson EJ, Maas R. BMP7 acts in murine lens placode development. *Dev Biol*. 1999;207:176–188.
- Wyatt K, Gao C, Tsai JY, Fariss RN, Ray S, Wistow G. A role for lengsin, a recruited enzyme, in terminal differentiation in the vertebrate lens. *J Biol Chem*. 2008;283:6607–6615.
- Straub BK, Boda J, Kuhn C, et al. A novel cell-cell junction system: the cortex adherens mosaic of lens fiber cells. *J Cell Sci*. 2003;116:4985–4995.
- Chen J, Ma Z, Jiao X, et al. Mutations in FYCO1 cause autosomal-recessive congenital cataracts. *Am J Hum Genet*. 2011;88:827–838.

32. Pankiv S, Alemu EA, Brech A, et al. FYCO1 is a Rab7 effector that binds to LC3 and PI3P to mediate microtubule plus end-directed vesicle transport. *J Cell Biol.* 2010;188:253-269.
33. Matsui M, Yamamoto A, Kuma A, Ohsumi Y, Mizushima N. Organelle degradation during the lens and erythroid differentiation is independent of autophagy. *Biochem Biophys Res Commun.* 2006;339:485-489.
34. Morishita H, Eguchi S, Kimura H, et al. Deletion of autophagy-related 5 (Atg5) and Pik3c3 in the lens causes cataract independent of programmed organelle degradation. *J Biol Chem.* 2013;288:11436-11437.
35. Chauss D, Basu S, Rajakaruna S, et al. Differentiation state-specific mitochondrial dynamic regulatory networks are revealed by global transcriptional analysis of the developing chicken lens. *G3 (Bethesda).* 2014;4:1515-1527.
36. Frederikse PH, Donnelly R, Partyka LM. miRNA and Dicer in the mammalian lens: expression of brain-specific miRNAs in the lens. *Histochem Cell Biol.* 2006;126:1-8.
37. Karali M, Peluso I, Gennarino VA, et al. miRNeye: a microRNA expression atlas of the mouse eye. *BMC Genomics.* 2010;11:715.
38. Tian L, Huang K, DuHadaway JB, Prendergast GC, Stambolian D. Genomic profiling of miRNAs in two human lens cell lines. *Curr Eye Res.* 2010;35:812-818.
39. Wu C, Lin H, Wang Q, et al. Discrepant expression of microRNAs in transparent and cataractous human lenses. *Invest Ophthalmol Vis Sci.* 2012;53:3906-3912.
40. Hughes AE, Bradley DT, Campbell M, et al. Mutation altering the miR-184 seed region causes familial keratoconus with cataract. *Am J Hum Genet.* 2011;89:628-633.
41. Iliff BW, Riazuddin SA, Gottsch JD. A single-base substitution in the seed region of miR-184 causes EDICT syndrome. *Invest Ophthalmol Vis Sci.* 2012;53:348-353.
42. Conte I, Carrella S, Avellino R, et al. miR-204 is required for lens and retinal development via Meis2 targeting. *Proc Natl Acad Sci U S A.* 2010;107:15491-15496.
43. Kubo E, Hasanova N, Sasaki H, Singh DP. Dynamic and differential regulation in the microRNA expression in the developing and mature cataractous rat lens. *J Cell Mol Med.* 2013;17:1146-1159.
44. Chiang HR, Schoenfeld LW, Ruby JG, et al. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.* 2010;24:992-1009.
45. Griffiths-Jones S, Hui JH, Marco A, Ronshaugen M. MicroRNA evolution by arm switching. *EMBO Rep.* 2011;12:172-177.
46. Ro S, Park C, Young D, Sanders KM, Yan W. Tissue-dependent paired expression of miRNAs. *Nucleic Acids Res.* 2007;35:5944-5953.
47. Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Res.* 2007;17:1850-1864.
48. de WE, Linsen SE, Cuppen E, Berezikov E. Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res.* 2009;19:2064-2074.