Published in final edited form as:

Exp Eye Res. 2016 April ; 145: 347–351. doi:10.1016/j.exer.2016.01.010.

Non-coding RNA profiling of the developing murine lens

Shahid Y. Khan, Sean F. Hackett, and S. Amer Riazuddin*

The Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD, 21287, USA

Abstract

Non-coding RNAs (ncRNAs) are emerging as an important player in the regulation of genome integrity and gene expression, and they have been implicated in the pathogenesis of many diseases. The aim of the present study is to identify the repertoire of ncRNAs expressed in the developing mouse lens. We previously reported the mouse lens transcriptome, including mRNA and microRNA (miRNA) profiling at two embryonic (E15 and E18) and four postnatal (P0, P3, P6, and P9) time points. We analyzed the data from small RNA-Seq and mRNA-Seq libraries to investigate the ncRNA profile. Our analysis revealed expression of 12 different classes of ncRNA in the murine lens at six developmental time points. Annotation of small RNA data showed expression of 1,756 antisense ncRNA (asncRNA) in the mouse lens transcriptome. Likewise, we identified 82 P-element-induced wimpy testis (PIWI)-interacting RNA (piRNA), 345 transfer RNA (tRNA), 12 small nuclear RNA (snRNA), 167 small nucleolar RNA (snoRNA), 19 small Cajal body-specific RNA (scaRNA), six ribosomal RNA (rRNA), 18 tRNA-like structures, one MALAT1-associated small cytoplasmic RNA (mascRNA), one Vault RNA (vtRNA), and one Y RNA expressed in the developing mouse lens. In parallel, bioinformatic investigation of mRNA-Seq data identified expression of 1,952 long intergenic ncRNA (lincRNA) in the developing mouse lens. In conclusion, we report a comprehensive ncRNA profile in the murine lens at six developmental time points. To the best of our knowledge, this is first report investigating different classes of ncRNAs in the developing mouse lens and will be monumental in elucidating processes essential for the development of the ocular lens and the maintenance of its transparency.

Keywords

Non-coding RNA; RNA-Seq; Developing mouse lens; Transcriptome

1. Introduction

The mammalian genome transcribes a much larger repertoire of ncRNAs compared to protein-coding RNAs. It has been reported that nearly 70% of the human genome encodes for non-protein-coding RNAs while only 2% transcribes for protein-coding RNAs (Kapranov et al., 2007; Frith et al., 2005). Next-generation sequencing (NGS) based analysis,

*Corresponding author. The Wilmer Eye Institute, Johns Hopkins University School of Medicine, 600 N. Wolfe Street, Maumenee 840, Baltimore, MD, 21287, USA. riazuddin@jhmi.edu (S. Amer Riazuddin).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.exer.2016.01.010>.

RNA-Seq, and ChIP-Seq have significantly expanded our ability to identify functionally important novel ncRNAs. Among these, asncRNAs, lincRNA, miRNA, piRNA, rRNA, scaRNA, snRNA, snoRNA, and tRNA have been discovered in the last few years (Guil and Esteller, 2012; Morozova and Marra, 2008; Jung et al., 2010; Fasold et al., 2011; Hawkins et al., 2010).

Antisense ncRNA transcripts (transcription from the opposite strand of a protein coding or sense strand) play a significant role in the regulation of protein-coding genes (Villegas and Zaphiropoulos, 2015). They regulate target genes in cis or in trans. In cis, they interact with transcripts originating at the same DNA region, while in trans, they interact with genes located at distant loci (Villegas and Zaphiropoulos, 2015; Nishizawa et al., 2012). PIWI-interacting RNAs (piRNA), 24–31 nucleotides in length, are processed by the PIWI proteins (Siomi et al., 2011). The piRNA-PIWI complex, also known as piRNA-induced silencing complexes that repress transposons via transcriptional and/or post-transcriptional mechanisms (Siomi et al., 2011; Iwasaki et al., 2015). Both rRNA and tRNA are involved in mRNA translation while snRNAs are shown to regulate splicing of different RNA molecules (Watson, 1963; Frendewey and Keller, 1985; Grabowski et al., 1985). The snoRNA and scaRNA are named for their sub-cellular localization within the nucleoli, and interaction with Cajal bodies, respectively, and they are critical for rRNA, tRNA, snRNA, and mRNA sequence-specific modification and processing (Balakin et al., 1996; Bratkovic and Rogelj, 2014). LincRNAs, large ncRNA (>200 nucleotides in length), play an important role in gene regulation through either cis-acting or trans-acting mechanisms (Rinn and Chang, 2012).

We previously reported the developing mouse lens transcriptome, including mRNA and miRNA, using next-generation RNA sequencing (Khan et al., 2015). In this context, we investigated the small RNA and mRNA sequencing data to identify the ncRNA profile of the developing mouse lens transcriptome. These analyses identified the 12 different types of ncRNAs, including asncRNA, piRNA, rRNA, scaRNA, snRNA, snoRNA, tRNA, tRNA-like structures, mascRNA, vtRNA, Y RNA, and lincRNA.

2. Materials and methods

2.1. Small RNA bioinformatic and data analysis

Bioinformatic analysis of small RNA sequencing data (GEO series accession number GSE69221) was performed by Maverix small RNA-Seq pipeline (Maverix Biomix, San Mateo, CA). These analyses included the preprocessing of raw sequencing reads using FastQC to remove adapter sequences from both ends, followed by filtering low-quality reads based on quality scores using PRINSEQ and Trimmomatic (Schmieder and Edwards, 2011; Bolger et al., 2014). The preprocessed quality-passed reads were mapped against the *Mus musculus* (*GRCm38/mm10*) genome using Bowtie2 (Langmead and Salzberg, 2012). These mapped reads were used to generate counts for each ncRNA using custom scripts (Maverix Biomix). First, counts were normalized across samples using a counts per million (CPM) algorithm. Subsequently, differentially expressed ncRNAs were identified using DESeq and EdgeR packages, which are based on a model using the negative binomial distribution of raw reads (Anders and Huber, 2010; Robinson et al., 2010). The expression level at embryonic day 15 (E15) was used as a reference point.

2.2. mRNA library bioinformatic analysis and lincRNA identification

Bioinformatic analysis of mRNA sequencing data (GEO series accession number GSE69221) was performed to investigate lincRNA expression in the mouse lens transcriptome. Lasergene Genomics Suite (DNASTAR, Madison, WI, USA) was used to identify lincRNA in mRNA sequencing data. The paired-end reads were aligned into an assembly with SeqMan NGen version 12, using the default parameters. The *Mus musculus* lincRNA annotations and sequences were downloaded from Mouse GENCODE Version M8 (GRCm38.p4) and used as a reference for lincRNA identification in the mouse lens transcriptome. Mapped reads were subjected to quantification, normalization, and eventually assembled into transcripts with ArrayStar Version 12 (DNASTAR), using the default parameters. The expression data was normalized by calculating the reads per kilobase per million mapped reads (RPKM) values for each ncRNA. Finally, the Mouse Genome Database (MGD) was utilized to characterize the different ncRNA biotypes identified in this analysis (Eppig et al., 2015).

3. Results

We previously investigated the mouse lens transcriptome at two embryonic (E15 and E18) and four postnatal (P0, P3, P6 and P9) time points using next-generation RNA sequencing, which identified a total of 14,465 genes and 379 mature miRNAs in the developing mouse lens (Khan et al., 2015). We further investigated the small RNA and mRNA sequencing data to identify additional ncRNA species in the mouse lens. We used a commercially available small RNA-Seq analysis pipeline (Maverix Biomix Inc.) to investigate the different classes of ncRNA molecules in the small RNA data. The raw reads were processed for quality control check that resulted in the removal of low-quality reads (<15%). The reads that passed quality control were aligned using the Bowtie algorithm, which resulted in >80% of the total reads mapped to the *Mus musculus* genome for the six different developmental time points. The mapped reads suggested a wide range of ncRNA expression in the developing mouse lens (Fig. 1). Bioinformatic analysis identified the expression (>1.0 CPM) of 11 different classes of ncRNAs, including asncRNA, piRNA, rRNA, snoRNA, snRNA, scaRNA, tRNA, tRNA-like structures, mascRNA, vtRNA, and Y RNA. Among these ncRNAs, the majority of the reads mapped to asncRNA, piRNA, miRNA, tRNA, and rRNA (Fig. 1).

We identified a total of 1,756 asncRNAs in the developing mouse lens (Table 1 and Suppl. Table 1). Differential expression analysis identified 656 asncRNAs that were up-regulated and 396 asncRNAs that were down-regulated compared to their respective expression at E15 (Suppl. Fig. 1 and Suppl. Table 1), while 704 asncRNAs exhibited uniform expression at all developmental time points (Suppl. Table 1).

Likewise, our analysis revealed expression of 82 piRNAs in the developing mouse lens (Table 1 and Suppl. Table 2). Differential expression analyses of these 82 piRNAs suggested that 53 piRNAs exhibited an upward-expression trend, while the remaining 29 piRNAs demonstrated a downward-expression trend, compared with their respective expression levels at E15 (Suppl. Fig. 2 and Suppl. Table 2).

Approximately 7% of total reads were mapped to tRNA genes, and subsequent bioinformatic analysis identified a total of 345 tRNAs in the developing mouse lens (Table 1, Suppl. Fig. 3 and Suppl. Table 3). In addition to tRNA, a small fraction of reads (~0.1–0.2%) mapped to tRNA-like structures, identifying a total of 18 tRNA-like structures in the mouse lens transcriptome (Table 1 and Suppl. Table 3).

The snoRNAs are grouped based on two structures: first, a C/D box important for methylation and second, an H/ACA box important for pseudo-uridylation of rRNA, tRNA, and snRNA. We identified expression of 113 and 54 snoRNAs having conserved sequence motifs of C/D and H/ACA boxes, respectively, in the developing mouse lens (Table 1, Suppl. Fig 4 & 5, and Suppl. Table 4). Importantly, we observed an up-regulation of snoRNAs in the postnatal mouse lens transcriptome (Suppl. Table 4). Furthermore, we identified expression of 12 snRNAs and 19 scaRNAs in the mouse lens transcriptome (Table 1 and Suppl. Table 4).

In addition to major ncRNAs, we identified a small fraction of reads (1–3%) that mapped to six different rRNA transcripts in the mouse lens transcriptome (Table 1 and Suppl. Table 5). Last but not least, we identified one mascRNA, vtRNA, and Y RNA in the mouse lens transcriptome (Table 1 and Suppl. Table 5).

In parallel, we employed the Lasergene Genomics Suite (DNASTAR) to interrogate mRNA sequencing data to identify expression of lincRNA in the developing mouse lens. Some 394 million de-indexed paired-end reads were mapped to the *Mus musculus* lincRNA reference sequences (GENCODE Version M8; GRCm38.p4). Of these, 10.53 million paired-end reads (~2.5% of total reads) mapped to lincRNA reference sequences. A total of 3,495 lincRNA genes are expressed in the mouse genome (GENCODE Version M8; GRCm38.p4). Of these, we identified the expression (1.0 RPKM) of 1,952 lincRNA in the developing mouse lens transcriptome (Table 1 and Suppl. Table 6).

4. Discussion

We previously reported the mouse lens transcriptome, including mRNA and miRNA profiles, at multiple developmental stages through next-generation RNA sequencing (Khan et al., 2015). Here, we present a comprehensive ncRNA profile of the developing mouse lens. These analyses identified expression of 12 different classes of ncRNA species, including asncRNA, piRNA, rRNA, snoRNA, snRNA, scaRNA, tRNA, tRNA-like structures, mascRNA, vtRNA, YRNA, and lincRNA. To the best of our knowledge, this is the first comprehensive ncRNA profile of developing mouse lens investigated through next-generation RNA sequencing.

Recent studies have suggested a regulatory role for asncRNA transcripts through transcriptional initiation site blockage mediated by methylation of promoter site and/or chromatin modifications, and stability of sense mRNA (Pelechano and Steinmetz, 2013). We identified four different types of antisense RNA transcripts originating from repeat elements, ncRNAs, introns, and exons of the mouse genome. We determined the expression of 1,756 asncRNAs in the developing mouse lens. Aberrant expression of antisense transcripts has

also been reported in neurological disorders and multiple cancers (Pelechano and Steinmetz, 2013; Balbin et al., 2015).

We previously identified three Piwi-like protein-coding genes (*Piwi1*, *Piwi2*, and *Piwi4*) in the mouse lens transcriptome, (Khan et al., 2015) and identified expression of 82 piRNAs in the mouse lens transcriptome. Several studies reported the aberrant expression of PIWI-family proteins in somatic cancers in humans (Iwasaki et al., 2015; Suzuki et al., 2012). They also have been reported in diverse functions, including epigenetic regulation, genome rearrangement, and developmental regulation (Iwasaki et al., 2015).

Mature tRNAs are ~70–100 nucleotides long, fold into a ‘clover leaf’ secondary structure, and are essential adaptor molecules in the protein translation machinery (Torres et al., 2014). tRNA modifications have been associated with cancer, diabetes, neurological disorders, and mitochondrial-linked disorders (Torres et al., 2014; Abbott et al., 2014). We identified expression of 345 tRNAs in the developing mouse lens transcriptome. Our analysis further identified expression of 167 snoRNAs, 12 snRNAs, and 19 scaRNAs in the mouse lens. These ncRNAs are involved in posttranscriptional modification of rRNA, snRNA, and regulation of alternative splicing of pre-mRNA (Balakin et al., 1996; Bratkovic and Rogelj, 2014). Several studies have reported dysregulation or deletion of snoRNAs in multiple human cancers and neurodegenerative disorders (Stepanov et al., 2015; McMahon et al., 2015).

lincRNAs have been associated with numerous diseases, including diabetic retinopathy, ocular neovascularization, and central nervous system dysfunction (Jae and Dimmeler, 2015; Xu et al., 2014; Ng et al., 2013). Recently, Hong and colleagues reported expression of 254 lincRNAs in mouse lens epithelial and fiber cell populations (Hoang et al., 2014). Here, we identified expression of 1,952 lincRNAs, a comprehensive profile in the developing mouse lens transcriptome (Suppl. Table 6).

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

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are thankful to Maverix Biomics, San Mateo CA for the bioinformatic analysis of small RNA sequencing data.

Funding support

The work was supported by National Eye Institute Grant R01EY022714 (SAR).

References

- Abbott JA, Francklyn CS, Robey-Bond SM. Transfer RNA and human disease. *Front. Genet.* 2014; 5:158. [PubMed: 24917879]
- Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010; 11:R106. [PubMed: 20979621]
- Balakin AG, Smith L, Fournier MJ. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell.* 1996; 86:823–834. [PubMed: 8797828]
- Balbin OA, Malik R, Dhanasekaran SM, et al. The landscape of antisense gene expression in human cancers. *Genome Res.* 2015; 25:1068–1079. [PubMed: 26063736]
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics.* 2014; 30:2114–2120. [PubMed: 24695404]
- Bratkovic T, Rogelj B. The many faces of small nucleolar RNAs. *Biochim. Biophys. Acta.* 2014; 1839:438–443. [PubMed: 24735946]
- Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE. The mouse genome database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Res.* 2015; 43:D726–D736. [PubMed: 25348401]
- Fasold M, Langenberger D, Binder H, Stadler PF, Hoffmann S. DARIO: a ncRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res.* 2011; 39:W112–W117. [PubMed: 21622957]
- Fredericewey D, Keller W. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell.* 1985; 42:355–367. [PubMed: 3160483]
- Frith MC, Pheasant M, Mattick JS. The amazing complexity of the human transcriptome. *Eur. J. Hum. Genet.* 2005; 13:894–897. [PubMed: 15970949]
- Grabowski PJ, Seiler SR, Sharp PA. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell.* 1985; 42:345–353. [PubMed: 3160482]
- Guil S, Esteller M. Cis-acting noncoding RNAs: friends and foes. *Nat. Struct. Mol. Biol.* 2012; 19:1068–1075. [PubMed: 23132386]
- Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat. Rev. Genet.* 2010; 11:476–486. [PubMed: 20531367]
- 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. [PubMed: 25489224]
- Iwasaki YW, Siomi MC, Siomi H. PIWI-interacting RNA: its biogenesis and functions. *Annu. Rev. Biochem.* 2015; 84:405–433. [PubMed: 25747396]
- Jae N, Dimmeler S. Long noncoding RNAs in diabetic retinopathy. *Circ. Res.* 2015; 116:1104–1106. [PubMed: 25814678]
- Jung CH, Hansen MA, Makunin IV, Korbie DJ, Mattick JS. Identification of novel non-coding RNAs using profiles of short sequence reads from next generation sequencing data. *BMC Genomics.* 2010; 11:77. [PubMed: 20113528]
- Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* 2007; 8:413–423. [PubMed: 17486121]
- Khan SY, Hackett SF, Lee MC, Pourmand N, Talbot CC Jr, Riazuddin SA. Transcriptome profiling of developing murine lens through RNA sequencing. *Invest. Ophthalmol. Vis. Sci.* 2015; 56:4919–4926. [PubMed: 26225632]
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 2012; 9:357–359. [PubMed: 22388286]
- McMahon M, Contreras A, Ruggero D. Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. *Wiley Interdiscip. Rev. RNA.* 2015; 6:173–189. [PubMed: 25363811]
- Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics.* 2008; 92:255–264. [PubMed: 18703132]

- Ng SY, Lin L, Soh BS, Stanton LW. Long noncoding RNAs in development and disease of the central nervous system. *Trends Genet.* 2013; 29:461–468. [PubMed: 23562612]
- Nishizawa M, Okumura T, Ikeya Y, Kimura T. Regulation of inducible gene expression by natural antisense transcripts. *Front. Biosci. Landmark Ed.* 2012; 17:938–958. [PubMed: 22201782]
- Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. *Nat. Rev. Genet.* 2013; 14:880–893. [PubMed: 24217315]
- Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 2012; 81:145–166. [PubMed: 22663078]
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26:139–140. [PubMed: 19910308]
- Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 2011; 27:863–864. [PubMed: 21278185]
- Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* 2011; 12:246–258. [PubMed: 21427766]
- Stepanov GA, Filippova JA, Komissarov AB, Kuligina EV, Richter VA, Semenov DV. Regulatory role of small nucleolar RNAs in human diseases. *Biomed. Res. Int.* 2015; 2015:206849. [PubMed: 26060813]
- Suzuki R, Honda S, Kirino Y. PIWI expression and function in cancer. *Front. Genet.* 2012; 3:204. [PubMed: 23087701]
- Torres AG, Batlle E, Ribas de PL. Role of tRNA modifications in human diseases. *Trends Mol. Med.* 2014; 20:306–314. [PubMed: 24581449]
- Villegas VE, Zaphiropoulos PG. Neighboring gene regulation by antisense long non-coding RNAs. *Int. J. Mol. Sci.* 2015; 16:3251–3266. [PubMed: 25654223]
- Watson JD. Involvement of RNA in the synthesis of proteins. *Science.* 1963; 140:17–26. [PubMed: 13999211]
- Xu XD, Li KR, Li XM, Yao J, Qin J, Yan B. Long non-coding RNAs: new players in ocular neovascularization. *Mol. Biol. Rep.* 2014; 41:4493–4505. [PubMed: 24623407]

Fig. 1.

Pie chart diagrams illustrate the total number of reads mapped to each of the ncRNA species in the developing mouse lens transcriptome. Our data revealed that a high percentage of reads mapped to asncRNA, piRNA, miRNA, tRNA, and rRNA while a small percentage of reads were mapped to snoRNA, snRNA, scaRNA, and tRNA-like structures in the developing mouse ocular lens transcriptome. A small fraction of reads (1.0) were not mapped to any known ncRNA species Note: Pie chart diagrams represent ncRNA identified

in small RNA libraries at six developmental stages. **Note:** A: E15; B: E18; C: P0; D: P3; E: P6 and F: P9.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Non-coding RNAs identified in multiple developmental stages of the mouse lens transcriptome.

Non-coding RNAs	Expressed in at least one of the six developmental stages	Expressed in all six developmental stages
asncRNA	1,756	1,708
tRNA	345	343
piRNA	82	74
snoRNA	167	166
scaRNA	19	19
snRNA	12	12
tRNA-like	18	17
rRNA	6	6
mascRNA	1	1
vtRNA	1	1
Y-RNA	1	1
lincRNA	1,952	1,952

Note: antisense ncRNA: asncRNA; P-element-induced wimpy testis (PIWI)- interacting RNAs: piRNAs; transfer RNA: tRNA; small nuclear RNA: snRNA; small nucleolar RNA: snoRNA; small Cajal body-specific RNA: scaRNA; ribosomal RNA: rRNA; transfer RNA-like: tRNA-like; MALAT1-associated small cytoplasmic RNA: mascRNA; Vault RNA: vtRNA; and long intergenic ncRNA: lincRNA.